Spin glasses and information

Hidetoshi Nishimori

Department of Physics, Tokyo Institute of Technology,
Oh-okayama, Meguro-ku, Tokyo 152-8551, Japan

Abstract

After a brief introduction to information theory, we review the close relationship between the theory of spin glasses and information processing, error-correcting codes in particular. An interesting equivalence of the solvability condition of the spin glass problem and the optimal inference condition in information theory is pointed out.

Key words: spin glass, information processing, error-correcting codes
PACS: 05.50.+q, 89.70.+c

1 Introduction

Spin glasses are magnetic materials with strong disorder in exchange interactions between spins. At low temperatures the spin degrees of freedom are randomly frozen in the sense that a spatially-random spin pattern, once realized, does not change with time. One of the principal objectives of the spin glass theory is to clarify the conditions under which such a strange state can exist [1,2].

The model system of common use for this purpose is the Edwards-Anderson (EA) model [3]. It has quenched (frozen) randomness in each exchange interaction, independently distributed from other interactions. As is always the case in any many-body systems, the starting point of statistical-mechanical analyses is the mean-field theory. The infinite-range version of the EA model, the Sherrington-Kirkpatrick (SK) model [4], is the basis of the mean-field theory of spin glasses. After a good amount of efforts by a number of researchers, it is now well established that the SK model has a stable spin glass phase at low temperatures. This spin glass phase has an extraordinary property that very many stable states exist, which are not related by symmetry operations and are therefore essentially different from each other.

One of the issues of current interest is to what extent the properties of the SK model are shared by realistic finite-dimensional systems. This is a very difficult problem due to the scarceness of mathematically well established facts. One of the rare examples of mathematically reliable results on finite-dimensional EA models is the exact solution for the internal energy on the so-called Nishimori line (NL) drawn in the phase diagram [2,5]. A surprising fact is that the theory concerning the NL turned out to have a very close relationship with a central issue of information theory, error-correcting codes. The condition for optimal inference of the original message out of the noisy output of transmission channel is now known to be equivalent to the NL condition [2,6].

In the present article I will review this relationship with physics audience in mind. The next section provides a brief introduction to information theory. Then the problem is formulated explicitly in terms of the spin glass theory in section 3. The last section is devoted to discussions.

2 Information theory: A brief introduction

Suppose that we wish to send a bit sequence, termed a message, '1011011' through a noisy transmission channel to somebody. The receiver should remove the noise and infer the original message out of the noisy output of the channel, which may be '1001011' for example. For this purpose it is known to be effective to introduce redundancy before transmission, a process called coding or encoding. For instance, the first bit '1' may be encoded as '111', a simple repetition, and is fed to the channel. If the noise rate of the channel is very small, as is the case in most modern transmission channels, at most a single bit out of the three bits '111' is affected by the noise and the output may be '110'. Then the receiver infers the first bit of the original message by the majority rule as '1'. This method is called the repetition code.

More sophisticated is the parity check code. Consider to add the parity check bit '1' to the original message '1011011' by counting the number of 1s. Since there exist odd 1s, we add '1' (otherwise '0' is added). The encoded message '10110111' is then fed into the channel. The output of the channel may then be '10010111' with the third bit flipped by the noise. The receiver checks the parity, finding an inconsistency. In this way, an error detection succeeds.

There are a number of ways to correct errors, not just to detect errors. A simple example is the block code depicted in Fig. 1. To send 9 bits '110101111' for example, one generates a 4×4 matrix by calculating parities for columns and rows as in Fig. 1(a). If a single bit out of the encoded 16 bits is affected by the noise in the channel, the output may be like Fig. 1(b). The receiver checks the parity and finds an inconsistency in the third column/row, concluding that

1	1	0	0
1	0	1	0
1	1	1	1
1	0	0	1

(a)

1	1	0	0
1	0	1	0
1	1	0	1
1	0	0	1

(b)

Fig. 1. Block code. The original 9 bits are encoded into 16 bits as in (a). The output may be like (b) with a single bit affected by the noise.

the '0' in the third column/row is erroneous and corrects it to '1'.

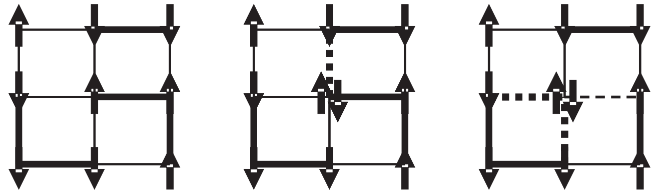
As can be guessed from these examples, it is necessary to introduce redundancy in the form of additional bits in the encoding process before we feed the message into the channel. When the original message has the length n and the encoded message has m bits, the ratio $R = n/m$ lies between 0 and 1 and is called the code rate or the transmission rate. For R close to 1, the transmission is efficient (because of a small number of additional bits) but the reliability is low (due to a small number of clues for the receiver to infer the original message). If R is small, on the other hand, we have less efficiency but high reliability. An important problem is therefore how large we can choose R while retaining reliability.

The answer to this question is given in the celebrated Channel Coding Theorem of Shannon. Suppose that the noise in the channel has the property to flip a bit to the other value (0 to 1 or 1 to 0) with probability p. The channel capacity is then defined as $C = 1 + p \log_2 p + (1-p)\log_2(1-p) (< 1)$. The Channel Coding Theorem states that there exists a code by which the asymptotically complete error correction is possible if $R < C$. Here the limit of infinite message length is called the asymptotic limit.

This is a highly non-trivial statement because one does not have to introduce infinite redundancy ($m \gg n$ or $R \to 0$) to achieve infinite reliability. A problem is that only the existence of such a useful code is shown and the explicit construction of reliable code is left untouched. Nevertheless the Channel Coding Theorem and related developments form the core of information theory and constitute the foundation of the modern information society.

3 Spin representation

It is instructive to rewrite the ideas of the previous section in terms of Ising spin systems. The value '1' of a bit can be mapped to the Ising spin value $-1 (= (-1)^1)$ and '0' to $1 (= (-1)^0)$. Then the sum of the bit values modulo



(a) (b) (c)

Fig. 2. The spin configuration in (a) is encoded as the set of interactions. The bold lines are for ferromagnetic interactions and the thin for antiferromagnetic interactions. The noise affects the interactions as shown dotted in (b). If the noise rate is larger, three interactions may be affected as in (c).

2 corresponds to the product of the Ising spin values ± 1, for instance $(-1)^1 \cdot (-1)^1 = (-1)^0$. The generation of a parity check bit from '1's and '0's is replaced by the generation of a product of Ising spins.

Let us consider the concrete example of Fig. 2. Suppose that one wants to send the Ising spin configuration of Fig. 2(a). A very simple idea corresponding to parity check is to generate products of spins at neighbouring sites i and j, $J_{ij}^0 \equiv S_i S_j (= \pm 1)$, which may be regarded as an interaction between S_i and S_j. This prescription corresponds to the Mattis model of spin glasses.

One then sends the set of all interactions $\{J_{ij}^0\}$ through the noisy channel. The receiver receives the noisy version of the interactions $\{J_{ij}\}$, in which some of the J_{ij}s are flipped from the original value ($J_{ij} = -J_{ij}^0$). This coding in Fig. 2 has the code rate $R = 9/12$ since the original configuration of 9 spins has been encoded in the terms of 12 interactions.

If it happens fortuitously that all the interactions have been transmitted without error ($J_{ij} = J_{ij}^0 \ \forall (ij)$), the receiver can retrieve the original spin configuration easily. Assume for simplicity that the sender and the receiver share the rule beforehand that the upper left spin is always fixed to 1 (up state) to break the overall inversion symmetry. Then the receiver, with only the information $\{J_{ij}^0\}$ at hand, determines the spin neighbouring the upper left one in Fig. 2(a) according to the sign of the interaction. It is not difficult to verify that each spin is assigned a unique configuration by repeating this procedure from site to site. In the jargon of the spin glass theory, the Mattis model involves no frustration and its ground state is uniquely determined.

Now suppose that one of the interactions is affected by noise and has been flipped to the other sign as illustrated in Fig. 2(b). Then the successive assignment of the spin orientation as before encounters an inconsistency at the central site: An up spin satisfies the constraint of three out of four interactions with the neighbouring spins but violates the remaining one. On the other hand, the assignment of a down spin at the same central site satisfies one of the four interactions whereas violates the other three. Stated otherwise, the up spin configuration gives a lower energy (actually the ground-state energy) than the case of the down spin. Therefore, in the present case of weak noise (with only a single interaction flipped) the search for the ground-state configuration for the given output $\{J_{ij}\}$ leads to a successful inference of the original spin configuration. In this way one can retrieve the correct information from the noisy output of the channel.

What happens if we have a larger noise rate? When three interactions are affected by the noise as illustrated in Fig. 2(c), the ground state spin configuration (i.e. the smallest number of violated interactions) has a down spin at the central site. The correct original configuration, however, has an up spin there. This example suggests that we may better look at finite-temperature states (i.e. excited states) than the ground-state configuration to correctly infer the original configuration when the noise rate is not very small. We shall show in the next section that this is indeed the case.

4 Equivalence to the spin glass theory

To formulate the inference procedure described in the preceding section more precisely, it is useful to express the noise rate in terms of a conditional probability. The probability that the output for an interaction is J_{ij}, given the input J_{ij}^0, is denoted as $P(J_{ij}|J_{ij}^0)$. The definition of the noise rate is that J_{ij} has the opposite sign from $J_{ij}^0 (= S_i S_j)$ with probability p:

$$P(J_{ij} = -S_i S_j | J_{ij}^0 = S_i S_j) = p. \tag{1}$$

For later convenience we define a variable β_p by the formula

$$\frac{e^{-\beta_p}}{e^{\beta_p} + e^{-\beta_p}} = p. \tag{2}$$

For small noise ($p \approx 0$), β_p is large, and for large noise ($p \approx 1/2$) β_p is close to 0. In this sense $T_p \equiv 1/\beta_p$ may be called the noise temperature: Low noise temperature $T_p \approx 0$ corresponds to small p, and high noise temperature to large p. It follows automatically from Eqs. (1) and (2) that the conditional probability for the noise-free case ($J_{ij} = J_{ij}^0$) is expressed similarly,

$$P(J_{ij} = S_i S_j | S_i S_j) = 1 - p = \frac{e^{\beta_p}}{e^{\beta_p} + e^{-\beta_p}}. \tag{3}$$

These formulas (1)-(3) are summarized in a compact form

$$P(J_{ij}|S_iS_j) = \frac{e^{\beta_p J_{ij} S_i S_j}}{e^{\beta_p} + e^{-\beta_p}}. \tag{4}$$

This equation applies to all interacting pairs of spins independently. Thus the conditional probability for the total transmission process is given by the product,

$$P(\mathbf{J}|\mathbf{J}^0) = \prod_{(ij)} P(J_{ij}|S_iS_j) \propto \exp\left(\beta_p \sum_{(ij)} J_{ij} S_i S_j\right), \tag{5}$$

where $\mathbf{J} = \{J_{ij}\}$ and $\mathbf{J}^0 = \{J_{ij}^0\}$. The task of the receiver is to infer the original spin configuration $\{S_i\}$ from the given noisy output $\mathbf{J}$. Since $\{S_i\}$ is uniquely determined by the noiseless $\mathbf{J}^0$, the receiver makes use of the conditional probability of $\mathbf{J}^0$ given $\mathbf{J}$, $P(\mathbf{J}^0|\mathbf{J})$. This expression has the entries exchanged from Eq. (5). The Bayes formula allows us to exchange the entries, giving $P(\mathbf{J}^0|\mathbf{J}) \propto P(\mathbf{J}|\mathbf{J}^0)$ in the present case as explained in Appendix A. Hence the problem can be solved by manipulating Eq. (5), which is nothing but the Boltzmann factor of a spin glass system, the so-called $\pm J$ Ising model, with the effective temperature T_p.

To actually infer the original spin configuration at site k for example, one calculates the probability of S_k from Eq. (5) by erasing all the other spin variables,

$$P(S_k) \propto \sum_{\backslash S_k} \exp\left(\beta_p \sum_{(ij)} J_{ij} S_i S_j\right), \tag{6}$$

where the first sum runs over all spin variables except for S_k. One then chooses the larger of $P(S_k = 1)$ and $P(S_k = -1)$ to determine S_k. This is equivalent to choosing the sign of the local magnetization $\langle S_k \rangle$ since

$$\langle S_k \rangle = P(S_k = 1) - P(S_k = -1). \tag{7}$$

In this sense the inference problem of the original message (spin configuration) is equivalent to the statistical mechanics of the spin glass at finite temperature T_p determined by the noise rate.

It is remarkable that the condition of $T = T_p$ turns out to be equivalent to the solvability condition of the spin glass problem: The exact value of the internal energy of the $\pm J$ Ising model of spin glasses is known to be given by the simple expression

$$E = -N_B \tanh \beta, \tag{8}$$

where N_B is the number of bonds (interactions) in the system and β is the inverse temperature $1/T$. The only condition for this result to apply is $T = T_p$, which defines the NL on the phase diagram, as shown in Appendix B. At present it is not fully understood why such a coincidence between completely different problems exists for error-correcting codes and the spin glass theory.

5 Summary

We have shown the relationship between the theory of spin glasses and information processing, error-correcting codes in particular. After a brief introduction to information theory, we have elucidated an Ising spin representation of error-correcting codes. Then the latter problem was shown to be expressible in terms of the spin glass theory. An interesting coincidence of the solvability condition of spin glass and the most probable decoding condition was pointed out.

In the present article we have illustrated the basic ideas using simple examples. There are more formal and general formulations of the topics treated here. The reader is referred to Ref. [2]. Also there exist many other related developments. For instance, the problem of image restoration (i.e. to remove noise from a given digital image) is known to be represented by the ferromagnetic spin system under random fields through the Bayes formula. In this case again, the optimal restoration performance is achieved under the condition $T = T_p$, the thermal temperature being equal to the noise temperature.

This is a very fascinating, expanding area of research activities, and we expect more surprising findings to come in the near future.

A Bayes formula

Let us consider the joint probability of two probabilistic events X and Y, $P(X,Y)$. This is the probability that both X and Y occur. The conditional probability $P(X|Y)$ to give the probability of X, given that Y has occurred, is defined by

$$P(X,Y) = P(X|Y)P(Y). \tag{A.1}$$

Similarly $P(Y|X)$ is defined by

$$P(X,Y) = P(Y|X)P(X). \tag{A.2}$$

From these equations we have

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}. \tag{A.3}$$

This is the Bayes formula. If $P(Y)$ is uniform (namely, independent of Y) and we are interested only in the Y-dependence of $P(Y|X)$, then the Bayes formula suggests

$$P(Y|X) \propto P(X|Y), \tag{A.4}$$

which allows us to exchange the entries of X and Y in the conditional probability.

In the situation described in Sec. 4, we assign $Y = \mathbf{J}^0$ and $X = \mathbf{J}$. In practice, one usually compresses the message before encoding in order to express the information most efficiently, namely with the minimum number of bits (a process called source coding). After the source coding, the message is represented by an apparently almost uniform sequence of '1's and '0's. Under such a circumstance $P(\mathbf{J}^0)$ is uniform because all spin configurations appear with almost the same probability. Hence we can use Eq. (A.4).

B Exact internal energy of the $\pm J$ model

Let us derive the exact internal energy (8) under the NL condition $T = T_p$ for the $\pm J$ Ising model [2,5]. The internal energy is defined as the thermal expectation value of the Hamiltonian, averaged over the configurations of the exchange interactions,

$$-E = \sum_{\{\tau_{ij}=\pm 1\}} \prod_{(ij)} \frac{e^{\beta_p \tau_{ij}}}{e^{\beta_p} + e^{-\beta_p}} \cdot \frac{\sum_S \left(\sum_{(ij)} \tau_{ij} S_i S_j\right) e^{\beta \sum \tau_{ij} S_i S_j}}{\sum_S e^{\beta \sum \tau_{ij} S_i S_j}}. \tag{B.1}$$

Note that the probability p is assigned to an antiferromagnetic interaction at each pair (ij) and $1-p$ to a ferromagnetic interaction, see Eqs. (2) and (3). It is useful to change the signs of running variables according to the rule

$$\tau_{ij} \to \tau_{ij}\sigma_i\sigma_j, \ S_i \to S_i\sigma_i, \tag{B.2}$$

where σ_i is an Ising variable arbitrarily fixed to 1 or -1 at each site. This is called the gauge transformation, and it does not affect the value of the internal energy since the gauge transformation is just to rename the running variables. The combination $\tau_{ij}S_iS_j$ is also gauge invariant:

$$\tau_{ij}S_iS_j \to \tau_{ij}\sigma_i\sigma_j \cdot S_i\sigma_i \cdot S_j\sigma_j = \tau_{ij}S_iS_j. \tag{B.3}$$

We then have

$$-E = \frac{1}{(2\cosh\beta_p)^{N_B}} \sum_{\tau} e^{\beta_p \sum \tau_{ij}\sigma_i\sigma_j} \cdot \frac{\sum_S \left(\sum_{(ij)} \tau_{ij}S_iS_j\right) e^{\beta\sum \tau_{ij}S_iS_j}}{\sum_S e^{\beta\sum\tau_{ij}S_iS_j}}, \tag{B.4}$$

where N_B is the number of interacting pairs. Since the right-hand side is independent of the assignment of the values to σ_i's, we can sum the right-hand side over all possible values of σ_i's and divide the result by 2^N (the number of such assignments) to have the same value of E,

$$-E = \frac{1}{2^N(2\cosh\beta_p)^{N_B}} \sum_{\tau}\sum_{\sigma} e^{\beta_p \sum \tau_{ij}\sigma_i\sigma_j} \cdot \frac{\sum_S \left(\sum_{(ij)} \tau_{ij}S_iS_j\right) e^{\beta\sum \tau_{ij}S_iS_j}}{\sum_S e^{\beta\sum\tau_{ij}S_iS_j}}. \tag{B.5}$$

If $\beta = \beta_p$ $(T = T_p)$, the sum over σ and the denominator cancel out and a simpler expression emerges,

$$-E = \frac{1}{2^N(2\cosh\beta)^{N_B}} \sum_{\tau} \frac{\partial}{\partial\beta} \sum_S e^{\beta\sum\tau_{ij}S_iS_j}. \tag{B.6}$$

The right-hand side can be easily evaluated by taking the sum over τ_{ij} first,

$$-E = \frac{1}{2^N(2\cosh\beta)^{N_B}} \sum_S \frac{\partial}{\partial\beta}(2\cosh\beta)^{N_B} = N_B\tanh\beta. \tag{B.7}$$

This is Eq. (8).

References

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

[2] H. Nishimori, *Statistical Physics of Spin Glasses and Information: An Introduction*, Oxford Univ. Press (Oxford, 2001).

[3] S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).

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

[5] H. Nishimori, Prog. Theor. Phys. **66**, 1169 (1981).

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