Journal of the Physical Society of Japan
Vol. 62, No. 9, September, 1993, pp. 2973-2975

Optimum Decoding Temperature for Error-Correcting Codes

Hidetoshi NISHIMORI

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

(Received June 22, 1993)

The conjecture of Ruján on error-correcting codes is proven. Errors in decoding of signals transmitted through noisy channels assume the smallest values when signals are decoded at a particular finite temperature. This finite-temperature decoding is compared with the conventional maximum likelihood decoding which corresponds to the $T=0$ case. The method of gauge transformation in the spin glass theory is useful in this argument.

[coding theory, decoding, spin glass, gauge transformation, Nishimori line]

Minimization of errors in the transmission of information through noisy channels is quite important in our modern life. One first codes (encodes) information into an appropriate form to be fed into a channel. At the receiving end of the channel, the noisy signal is decoded and possible errors are detected and corrected before the final information is passed to the receiver. A central problem in coding theory[1,2] is, therefore, how to code and decode information (often represented as a series of binary numbers) in order to minimize errors due to transmission noise. A number of methods have been proposed for this purpose.[1,2] A particularly interesting technique from a statistical physics point of view is the maximum-likelihood decoding, which may be stated in term of the spin glass theory as follows.[3-5]

Suppose that somebody (X) wishes to teach the configuration of Ising spins $\{\varepsilon_i=\pm 1\}_{i=1,\cdots,N}$ to a receiver (Y) at the other end of a noisy channel. If X transmits the configuration itself into the channel, there is no way for Y to determine the correct configuration since Y does not know which of $\varepsilon_i(i=1,\cdots,N)$ has been flipped by noise in the channel. Thus X must introduce additional information, i.e., redundancy, before sending the information so that Y can correctly retrieve the original configuration from the noisy signal. A simple way is to form a set of interactions, $\{J_{ij}=\varepsilon_i\varepsilon_j\}$ for example, and send this set. Y receives the noisy set $\{\bar{J}_{ij}\}$ and looks for the ground state of the Ising model Hamiltonian $H=-\sum \bar{J}_{ij}\sigma_i\sigma_j$. If there are no errors, H represents the frustration-free Mattis model whose ground state is exactly the original configuration $\sigma_i=\varepsilon_i$ (or its perfectly inverted version). When some of the $\bar{J}_{ij}$ have been inverted from the correct value, $\bar{J}_{ij}=-J_{ij}$, it is still possible to retrieve the true configuration as long as errors are isolated from each other because isolated frustration does not change the ground-state configuration. Serious problems arise if transmission error is large and bonds with inverted signs $\bar{J}_{ij}=-J_{ij}$ are located adjacent to each other. The ground state of such a case is, in general, different from the original configuration.

Sourlas has pointed out[3-5] that use of the random energy model[6] gives an asymptotically vanishing error rate of decoding. In other words, the original spin configuration is retrieved without errors by looking for the ground state of the Hamiltonian of transmitted interactions of the random energy model. Since the random energy model has infinitely many-body interactions, realization of Sourlas' prescription faces practical difficulties even though the idea is quite interesting theoretically. Ruján recently observed that decoding at a particular finite temperature may yield a smaller error rate than searching for the ground state.[7] An appealing aspect of his idea is that it is applicable to an arbitrary type of interactions such as the usual two-body, three-body and so on, in contrast to the Sourlas method using the random energy model. I prove, in this letter, that Ruján's conjecture is indeed correct. The method of gauge transformation in the spin glass theory[8] is a useful tool for this argument.

Let us fix the notations. I explain the idea in terms of the two-body coding $\{J_{ij}=\varepsilon_i\varepsilon_j\}$ although this is not essential. Three-body coding $\{J_{ijl}=\varepsilon_i\varepsilon_j\varepsilon_l\}$ for instance can be treated in the same manner. The choice of interacting pairs $\langle ij\rangle$ is also arbitrary. The interactions are flipped, $\bar{J}_{ij}=-J_{ij}$, by transmission noise with probability p. This represents a binary symmetric channel.[1,2] One tries to retrieve the original spin configuration in some way from the Hamiltonian $H=-\Sigma\,\bar{J}_{ij}\sigma_i\sigma_j$. The retrieved configuration is denoted as $\{S_i=\pm 1\}$. The maximum-likelihood decoding[1,2] corresponds to the choice $S_i=\sigma_i$ where σ_i is the ground state of H. The bit error rate of decoding[1,2] p_e is related to the Mattis magnetization $m=(1/N)\,\Sigma_i\,S_i\varepsilon_i$ by the relation $p_e=(1-m)/2$. A perfect retrieval has $m=1$ or $p_e=0$. Our aim is to find a method to make p_e as small as possible or to enhance m.

Ruján[7] proposed to assign the decoded signal by the rule

$$S_i(T_N)=\frac{\langle\sigma_i\rangle_{T_N}}{|\langle\sigma_i\rangle_{T_N}|},$$

where the angular brackets denote the thermal average by the Hamiltonian $H=-\Sigma\,\bar{J}_{ij}\sigma_i\sigma_j$ at the temperature

$$T_N=\left(\frac{1}{2}\ln\frac{1-p}{p}\right)^{-1},$$

which is the temperature on the Nishimori line in the phase diagram of the $\pm J$ model.[8] The main result of the present analysis is

$$\bar{m}(T,p)\equiv[\varepsilon_i S_i(T)]_p\le\bar{m}(T_N,p),\qquad (1)$$

where the square brackets denote the average over errors due to transmission noise.

The proof of (1) consists of straightforward applications of the method of gauge transformation.[8] It is useful to note first that the gauge transformation $\sigma_i\to\sigma_i\varepsilon_i$ ($\forall i$) has no effect on thermal averages since the summation over $\sigma_i=\pm 1$ is either unchanged ($\varepsilon_i=+1$) or simply reversed in order ($\varepsilon_i=-1$). The interactions acquire additional signs through this gauge transformation $\bar{J}_{ij}\to\bar{J}_{ij}\varepsilon_i\varepsilon_j$. Since $\bar{J}_{ij}=-J_{ij}=-\varepsilon_i\varepsilon_j$ with probability p, the interaction after gauge transformation $\bar{J}_{ij}\to\bar{J}_{ij}\varepsilon_i\varepsilon_j$ is -1 with probability p and $+1$ with probability $1-p$. I denote the new interaction as τ_{ij} ($=\pm 1$). Thus $\tau_{ij}=-1$ (probability p) or $+1$ (probability $1-p$). The quantity $\bar{m}(T,p)$ is written as $[\langle\sigma_i\rangle_T/|\langle\sigma_i\rangle_T|]_p$ after the gauge transformation. In this "ferromagnetic gauge",[4-7] we have the following relation:

$$\bar{m}(T_N,p)=\left[\frac{\langle\sigma_i\rangle_{T_N}}{|\langle\sigma_i\rangle_{T_N}|}\right]_p=\left[\frac{\langle\sigma_i\rangle^2_{T_N}}{|\langle\sigma_i\rangle_{T_N}|}\right]_p$$
$$=[|\langle\sigma_i\rangle_{T_N}|]_p.\qquad (2)$$

I have applied the technique of gauge transformation[8] in deriving the third expression of (2). In short, the gauge-covariant factor $\langle\sigma_i\rangle_{T_N}$ in the second expression yields an extra factor $\langle\sigma_i\rangle_{T_N}$ after gauge transformation.

Another useful relation is

$$\bar{m}(T,p)=\left[\frac{\langle\sigma_i\rangle_T}{|\langle\sigma_i\rangle_T|}\right]_p\le\left|\left[\frac{\langle\sigma_i\rangle_T}{|\langle\sigma_i\rangle_T|}\right]_p\right|$$
$$=\left|\left[\frac{\langle\sigma_i\rangle_{T_N}\langle\sigma_i\rangle_T}{|\langle\sigma_i\rangle_T|}\right]_p\right|.$$

Again a gauge transformation yielded the factor $\langle\sigma_i\rangle_{T_N}$. The last expression is bounded from above if one moves the absolute-value operation into the configurational average,

$$\bar{m}(T,p)\le[|\langle\sigma_i\rangle_{T_N}|]_p=\bar{m}(T_N,p),$$

where I have used (2) in the second equality. This completes the proof.

A remark on the implications of inequality (1) on spin glass physics is in order. The quantity $\bar{m}(T,p)$ is not the conventional thermodynamic magnetization $[\langle\sigma_i\rangle_T]_p$. Inequality (1) does not imply that the thermodynamic magnetization assumes the largest value on the Nishimori line. The present result rather means that the spin alignment, *neglecting the average spin reduction due to thermal fluctuations*, becomes largest on the Nishimori line. At a temperature lower than T_N, each spin may have a larger expectation value $|\langle\sigma_i\rangle_T|$ than at T_N. However, a spin is more misaligned, on average, with other spins at $T<T_N$ or at $T>T_N$ than at $T=T_N$.

The following comment may help the reader understand why the present finite-temperature decoding works well.[7)] For a finite value of p, the ground state of the Hamiltonian $H = -\sum \bar{J}_{ij}\sigma_i\sigma_j$ is, in general, different from the original configuration $\{\varepsilon_i\}$. This means that the original configuration has a higher energy. By replacing σ_i with ε_i and $\bar{J}_{ij}$ with $\tau_{ij}\varepsilon_i\varepsilon_j$ ($\tau_{ij} = -1$ with probability p), one gets $H = -\sum \tau_{ij} = -(1-2p)N_B$ where N_B is the number of interacting pairs. This energy of the original configuration is equal to the energy expectation value of the system on the Nishimori line.[8)] Therefore the search for spin configurations at T_N may give a better retrieval of the original configuration than that at $T=0$. This finding lead Ruján[7)] to the proposal of the finite-temperature decoding.

It is straightforward to apply the same argument to a Gaussian channel in which $\bar{J}_{ij}$ deviates from $J_{ij} = v\varepsilon_i\varepsilon_j$ according to the Gaussian distribution, where v is a real-valued signal amplitude.[1,2,7)] The final inequality (1) has the same form. The temperature T_N is defined by w^2/v, where w^2 is the variance of noise. The restriction to a two-body interaction is also not essential for binary, as well as Gaussian, channels. The whole argument applies to an arbitrary interaction type. The result remains unchanged. It is also possible to generalize the theory for non-Ising systems such as $Z(q)$ or $U(1)$ models. The method for gauge transformation established for these general models[9)] is applicable to the present argument.

I thank Professors N. Sourlas and P. Ruján for useful correspondence.

References

1) R. J. Eliece: *The Theory of Information and Coding* (Addison-Wesley, Reading, 1977) p. 1.
2) H. Imai: *Coding Theory* (The Institute of Electronics, Information and Communication Engineers, Tokyo 1990) p. 1 [in Japanese].
3) N. Sourlas: Nature **339** (1989) 693.
4) N. Sourlas: in *Statistical Mechanics of Neural Networks*, Lecture Notes in Physics No. 368 (Springer, Berlin, 1990) p. 317.
5) N. Sourlas: preprint (Ecole Normale Supérieure).
6) B. Derrida: Phys. Rev. B**24** (1981) 2613.
7) P. Ruján: Phys. Rev. Lett. **70** (1993) 2968.
8) H. Nishimori: Prog. Theor. Phys. **66** (1981) 1169.
9) Y. Ozeki and H. Nishimori: J. Phys. A**26** (1993) (to be published).