Statistical-mechanical approach for analog neural network model used in image restoration

H. Shouno
Dept. of Biophysical Engineering, School of Enggineering Science,
Osaka Univ., Machikaneyama 1-3, Toyonaka, Osaka, JAPAN
M. Okada
JST, ERATO, Kawato Dynamic Brain Project
2-2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto, JAPAN

Abstract ***The ability to restore an image from signals received through a noisy channel is an important concern. This issue is related to the physics theory of spin-glass. In the theory, the Ising spin system is usually used for image restoration; however, a lot of calculation time is needed to obtain precise solution. As a result many researchers substitute the Ising spin model with the analog neural network model. We analyzed the analog neural network ability applied to the image restoration problem using the mean field theory.***

1 Introduction

In this paper, we analyze the image restoration abilities of an analog neural network model. The ability to restore an image received through a noisy communication channel is an important concern. Addressing the image restoration problem using the mean field theory of the Ising spin model was discussed by Nishimori & Wong [1]. This problem is related to the spin-glass theory in physics.

In image restoration, the structure of the decoder/receiver model is usually assumed to be identical to the structure of the encoder/sender. However, in our research, the decoder/receiver does not have a structure identical to that of the encoder/sender, because the difference in the robustness of the noise from an identical structure poses a very interesting problem.

In general, combinatorial optimization should be used for the Bayesian approach of the image restoration [2]. The present work provides a good example for comparing the abilities of the analog neural network approach with those of the Ising spin system not only for image restoration, but also for combinatorial optimization problem.

The analog neural network model is faster than the stochastic Ising spin model in the sense of computation time; however, the quality of the solution obtained from the analog neural network model has not been discussed.

In a computer simulation, the analog neural network model is easy to implement, and a lot of researchers have applied it to optimization problems; however, few have discussed the goodness of the solution.

Recently, Nishimori & Wong analyzed the image restoration problem and the error-correcting problem theoretically using the Ising spin model. They further confirmed their theory with the simulated annealing method. For the image restoration problem, it is easy to define the goodness measure of solutions. Thus, we analyzed the application of the analog neural network model to the image restoration problem theoretically, and compared it with the results of the Nishimori & Wong's study.

2 Model and Analysis

Now let us define the problem of image restoration such as Figure 1. The original image, $\{\xi_i\}$, has a binary value of $\{-1, 1\}$. We assume the source image, $\{\xi_i\}$, has a prior probability of:

$$P_s(\{\xi\}) \propto \exp(\beta_s \sum_{i<j} \xi_i \xi_j). \qquad (1)$$

we can then consider the parity check code to be $\{\xi_i\xi_j\}$. The sender transmits the raw code $\{\xi_i\}$ and the parity check code $\{\xi_i\xi_j\}$ to the receiver through a noisy channel. The receiver receives the noise added code $\{\tau_i\}$ instead of the raw code $\{\xi_i\}$, and also receives $\{J_{ij}\}$ instead of the parity check code $\{\xi_i\xi_j\}$.

The problem of image restoration is the reconstruction of an image, $\{\sigma_i\}$, using only the de-

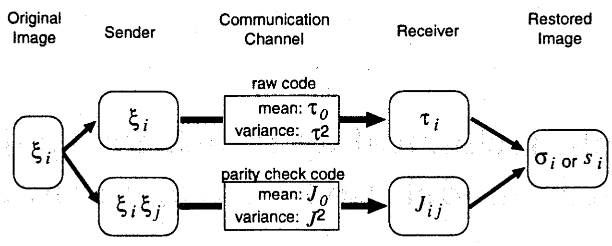

Figure 1: Schematic diagram of the image restoration problem

graded information $\{\tau_i\}$ and $\{J_{ij}\}$. When $\{J_{ij}\}$ and $\{\tau_i\}$ are received, Nishimori & Wong pointed out that the posterior probability of the restored sequence $\{\sigma_i\}$ can be written as follows [1].

$$P(\{\sigma_i\}|\{J_{ij}\},\{\tau_i\}) \propto \exp(\beta \sum_{i<j} J_{ij}\sigma_i\sigma_j + h\sum_i \tau_i\sigma_i) P_m(\{\sigma_i\}) \quad (2)$$

where, the image prior $P_m(\cdot)$ is

$$P_m(\{\sigma_i\}) \propto \exp(\beta_m \sum_{i<j} \sigma_i\sigma_j). \quad (3)$$

We also assume that both images $\{\xi_i\}$ and $\{\sigma_i\}$ have a same prior probability distribution. Thus, the difference between $P_s(\cdot)$ and $P_m(\cdot)$ is only parameter β_s and β_m. The receiver cannot guess the source image's prior parameter β_s; therefore, the receiver's prior parameter is substituted for β_m. In the posterior probability equation (2), the Hamiltonian appears in exponential function:

$$\beta H = -\beta \sum_{i<j} J_{ij}\sigma_i\sigma_j - h\sum_i \tau_i\sigma_i. \quad (4)$$

Therefore, the minimization of the Hamiltonian under the prior probability $P_m(\cdot)$ stands for the maximization of the posterior probability. In the ground state, that is, $T_m = \beta_m^{-1} = 0$, this maximization corresponds to the MAP (Maximum A Posteriori) estimation. In contrast with MAP, decoding under the finite temperature, that is, $T_m > 0$, corresponds to the the free energy minimization in the finite temperature. The finite temperature decoding corresponds to the MPM (Maximization of Posterior Marginals) estimation.

In this study, we treat the noise of the channel as the additive Gaussian. Hence, for a given image $\{\xi_i\}$, the Gaussian channel is given as:

$$P_{\rm out}(\{J_{ij}\},\{\tau_i\}|\{\xi_i\}) \propto \exp\left(-\frac{N\sum_{i<j}\left(J_{ij} - \frac{J_0}{N}\xi_i\xi_j\right)^2}{2J^2} - \frac{\sum_i(\tau_i - \tau_0\xi_i)^2}{2\tau^2}\right), \quad (5)$$

where J_0 and τ_0 are the mean and J^2/N and τ^2 are the variance of the Gaussian noises.

The average of any quantity $f(\{\sigma_i\})$ is calculated as:

$$[\langle f\rangle] = \sum_{\xi}\prod\int dJ \prod\int d\tau \, P_s(\{\xi_i\}) P_{\rm out}(\{J_{ij}\},\{\tau_i\}|\{\xi_i\}) \frac{\sum_{\{\sigma\}} f(\{\sigma_i\}) e^{-\beta H} P_m(\{\sigma_i\})}{\sum_{\{\sigma_i\}} e^{-\beta H} P_m(\{\sigma_i\})}, \quad (6)$$

where H is the Hamiltonian given in equation (4). The outer brackets $[\cdot]$ in equation(6) denote the averages over $\{\xi_i\}$, $\{J_i\}$, and $\{\tau_i\}$ with the weight $P_s P_{\rm out}$.

2.1 Analog neuron model

Hopfield and Tank proposed to use an analog neural network method for finding the minimum of the Hamiltonian or free energy[4]. We can easily extend their approach to the present case, and obtain the dynamics of the analog neural network,

$$\frac{dx_i}{dt} = -x_i + \tanh(\beta \sum J_{ij}x_j + \beta_m \sum x_j + h\tau_i), \quad (7)$$

where x_i is the output of the i-th analog neuron.

We employed the standard approach proposed by Bray-Sompolinsky-Yu [5] to analyze the equilibrium properties of equation (7). Each site has M binary neurons, and the output of the site is defined as the population of firing neurons

$$s_i = \frac{1}{M}\sum_a^M \hat{\sigma}_{ia}, \quad (8)$$

where $\hat{\sigma}_{ia}$ has a binary state that is $\{-1,+1\}$. Then, as the parameter M becomes infinity, each site can be taken as the analog value $[-1,+1]$. Corresponding to the Hamiltonian appeared in equation (4), we introduce the following Hamiltonian:

$$-\frac{\beta H}{M} = \beta \sum_{i<j} J_{ij}s_is_j + \beta_m \sum_{i<j} s_is_j + h\sum_i \tau_i s_i \quad (9)$$

In the large M limit, it is easily shown that equilibrium equation of (9) is equivalent to the equilibrium of the equation (7). The prior probability $P_m(\cdot)$ is included in this Hamiltonian.

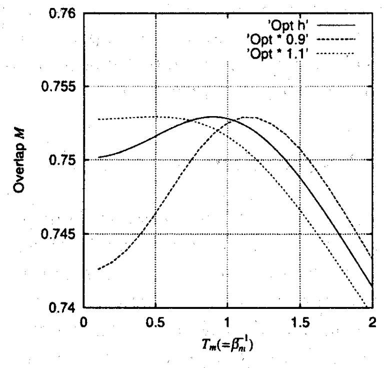

Figure 2: The overlap curve versus prior parameter T_m that is equal to β_m^{-1}. The random-field strength h is chosen to be $h = \tau_0\beta_m/\tau^2\beta_s$ (Opt h), $h = 0.9\mathrm{Opt}h$, or $h = 1.1\mathrm{Opt}h$. This figure is the same as the one that appeared in Nishimori and Wong study [1].

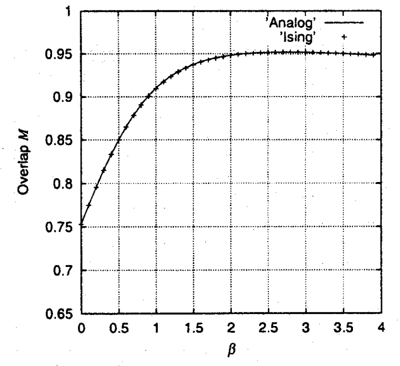

Figure 3: The overlap curve versus the parameter β. The variance of the Gaussian channel is small ($J = 0.60$). The performance of the analog neural network is almost the same as the Ising spin model.

In this way, the image restoration problem is replaced by the optimization problem to find the minimum state of equation (9) under the probability P_s.

2.2 Analysis

Since the equilibrium properties of equation (9) coincide with those of equation (7), we will analyze the system of equation (9), instead of the system of equation(7), through the following averaged replicated partition function. To find the minimum state of equation (9), we evaluate the following averaged replicated partition function,

$$[Z^n] = \operatorname*{Tr}_{\{\xi\},\{\hat{\sigma}\}} P_s(\{\xi_i\}) \sum_{\alpha} \sum_{a,b} P(\{\hat{\sigma}_{ia}^{\alpha}\}|\{J\}) \quad (10)$$

The standard replica calculation with the replica symmetric ansatz [6] leads to the expression of order parameters:

$$m_0 = \tanh(\beta_s m_0), \quad (11)$$

$$m = \frac{\mathrm{Tr}_\xi\ e^{\beta_s m_0 \xi} \int Dx\ \hat{F}(U(x))}{2\cosh(\beta_s m_0)}, \quad (12)$$

$$t = \frac{\mathrm{Tr}_\xi\ e^{\beta_s m_0 \xi} \xi \int Dx\ \hat{F}(U(x))}{2\cosh(\beta_s m_0)}, \quad (13)$$

$$q = \frac{\mathrm{Tr}_\xi\ e^{\beta_s m_0 \xi} \int Dx\ \hat{F}(U(x))^2}{2\cosh(\beta_s m_0)}, \quad (14)$$

$$\chi = \frac{\mathrm{Tr}_\xi\ e^{\beta_s m_0 \xi} \int Dx\ x\hat{F}(U(x))}{2\cosh(\beta_s m_0)\sqrt{h^2\tau^2 + \beta^2 J^2 q}}, \quad (15)$$

where the function $U(\cdot)$ is

$$U(x) = \sqrt{h^2\tau^2 + \beta^2 J^2 q}\,x + \beta_m m + (h\tau_0 + \beta J_0 t)\xi \quad (16)$$

and, the function $F(\cdot)$ is the solution of the following self-consistent equation:

$$F(x) = \tanh\left(U(x) + \beta^2 J^2 \chi F(x)\right) \quad (17)$$

These order parameter equations for the analog neural network decoder are almost the same as Nishimori & Wong's equations [1]. The only difference is the order parameter χ, which is called susceptibility, and its relation term

$$\beta^2 J^2 \chi F(x)$$

in the equation (17). This term is called the "Onsager reaction term" in physics. When we assume $\chi = 0$, the order parameter equations is identical to Nishimori & Wong's.

3 Results

In this study, we evaluated the goodness by overlap:

$$M = [\xi_i \mathrm{sgn}(\langle\sigma_i\rangle)] = \frac{\mathrm{Tr}_\xi\ e^{\beta_s m_0 \xi} \int Dx\ \mathrm{sgn}(U(x))}{2\cosh(\beta_s m_0)}. \quad (18)$$

In conventional image restoration, the two-body exchange term; $\beta = 0$ does not exist. Under this condition, the results of the order parameter equations (11)~(15) are identical to the results of the Ising model analysis [1]. Nishimori & Wong pointed out that MAP estimation does not derive the best solution in a meaning of overlap. Their assertion is correct in the case of the

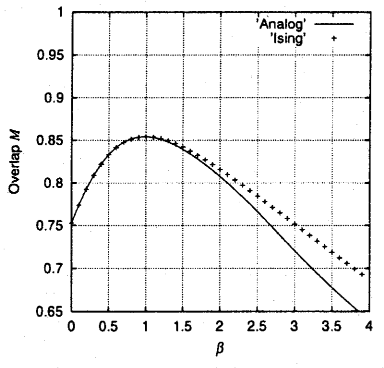

Figure 4: The overlap curve versus the parameter β. The variance of the Gaussian channel is large ($J = 1.00$). The performance of the analog neural network is worse in a large β.

analog neural network. Figure 2 shows the dependence of overlap M on the model prior parameter $T_m \equiv \beta_m^{-1}$. The original image parameters are $T_s \equiv \beta_s^{-1} = 0.9$, and the channel noise parameter $\tau_0 = \tau = 1$. The usual practice in image restoration is to use a Hamiltonian with a fixed ratio of h/β_m, using β_m as an adjustable parameter. We kept the ratio and examined three patterns of h. Nishimori & Wong derived the optimal value of h, that is, $h = \tau_0\beta_m/\tau^2\beta_s$. The ground-state limit; $T_m \to 0$ gives the MAP restoration. The maximum overlap is around $T_m = 0.9$ for the optimal h.

The difference between using the analog neural network and the Ising spin model appears in the existence of two-body exchange term β. When the mean control parameter of the Gaussian channel is set as; $J_0 = 1.0$. The performance of the analog neural network model is almost the same as the Ising spin model when the variance of the Gaussian channel is not large. Figure 3 shows the overlap for the two-body exchange parameter β. The remarkable difference does not exists in the condition that $J = 0.60$, which is the standard deviation of the Gaussian channel. However, the larger the parameter J becomes, the more the difference becomes apparent. When J is set larger, the performance of image restoration worsens in the large β area. Figure 4 shows the results where $J = 1.00$.

These results shows that the analog neural network does not improve the quality of image. However, the performance of image restoration is almost the same when the variance of the Gaussian channel is small. Even a large variance of the Gaussian channel is not a serious problem, because the receiver controls the effect of the parity check code by choosing the proper β.

4 Discussion

In this research, we evaluated the image restoration abilities of the analog neural network. With the conventional image restoration method, sending only image $\{\xi_i\}$, the estimated overlap with the analog neural network model is equivalent to that of the Ising spin network model. The difference occurs in sending parity check codes $\{\xi_i\xi_j\}$. Unfortunately, in this case the ability of the analog neural network does not improve over the Ising spin network model. However, in the case of a small noise variance, the performance of the analog neural network is as good as the Ising spin model. This difference comes from the Onsager reaction term.

Moreover, the calculation cost is much smaller than the Ising spin model. To use the Ising spin model, we must calculate states of neurons with a stochastic process. The analog neural network can perform this calculation with a deterministic process. This characteristic of the analog method is very beneficial.

References

[1] H. Nishimori and K. Y. M. Wong. *Statistical mechanics of image restoration and error-correcting codes. Physical Review E.*, 60(1):132–144, 1999.

[2] S. Geman and D. Geman. *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[3] S. Kirkpatrick, C. Gelatt, M. Vecchi. *Optimization by simulated annealing. Science*, 220:671–680, 1983.

[4] J. J. Hopfield and D. W. Tank. *Computing with Neural Circuits: A Model. Science.*, 233:625–633, 1986.

[5] A. J. Bray, H. Sompolinsky, and C. Yu. *On the 'naive' mean-field equations for spin glasses. J.Phys.*, 19:6389–6406, 1986.

[6] J. Hertz, A. Krough, and R. G. Palmer. *Introduction to the theory of neural computation. Addison Wesley.* , 1993.