Supervised Learning of Two-Layer Perceptron under the Existence of External Noise — Learning Curve of Boolean Functions of Two Variables in Tree-Like Architecture —

Tatsuya UEZU[1] * and Shuji KIYOKAWA[2]

[1] *Graduate School of Humanities and Sciences, Nara Women's University, Nara 630-8506, Japan*

[1] *Faculty of Science, Nara Women's University, Nara 630-8506, Japan*

(Received March 28, 2016)

We investigate the supervised batch learning of Boolean functions expressed by a two-layer perceptron with a tree-like structure. We adopt continuous weights (spherical model) and the Gibbs algorithm. We study the Parity and And machines and two types of noise, input and output noise, together with the noiseless case. We assume that only the teacher suffers from noise. By using the replica method, we derive the saddle point equations for order parameters under the replica symmetric RS ansatz. We study the critical value $\alpha_{\rm C}$ of the loading rate α below which only the para phase exists for cases with and without noise. We find that $\alpha_{\rm C}$ is nonzero for the Parity machine, while it is zero for the And machine. We derive the exponent $\bar{\beta}$ of order parameters expressed as $(\alpha-\alpha_{\rm C})^{\bar{\beta}}$ when α is near to $\alpha_{\rm C}$. Furthermore, in the Parity machine, when noise exists, we find a spin glass solution, in which the overlap between the teacher and student vectors is zero but that between student vectors is nonzero. We perform Markov chain Monte Carlo simulations by simulated annealing and also by exchange Monte Carlo simulations in both machines. In the Parity machine, we study the de Almeida-Thouless stability, and by comparing theoretical and numerical results, we find that there exist parameter regions where the RS solution is unstable, and that the spin glass solution is metastable or unstable. We also study asymptotic learning behavior for large α and derive the exponents $\hat{\beta}$ of order parameters expressed as $\alpha^{-\hat{\beta}}$ when α is large in both machines. By simulated annealing simulations, we confirm these results and conclude that learning takes place for the input noise case with any noise amplitude and for the output noise case when the probability that the teacher's output is reversed is less than one-half.

1. Introduction

Previously, we studied the supervised learning of a simple perceptron with continuous weights (spherical model) and with discrete weights (Ising perceptron) by using the Gibbs algorithm under the existence of external noise.[1–6] In particular, we studied the learning curve for noiseless, output noise, and input noise cases, and found that learning behaviors depend on the type of noise. That is, for the spherical model, we found that there exists an optimal temperature at which the generalization error is a minimum only for the output noise model.[3,4] On the other hand, for the Ising perceptron, in which the synaptic weights are discrete and take values of ± 1, there exists perfect learning, such that a student vector completely coincides with a teacher vector at a finite number of examples, only for the output noise model.[5,6]

When the number of layers increases, what kinds of phenomena take place is an interesting theme to be studied. In addition, it is very important to study two-layer perceptrons. One of the reasons is that if the number of neurons in the hidden units is sufficiently large, they can solve any classification problem such as the exclusive OR (XOR) problem, which cannot be solved by a simple perceptron as is well known. Another reason is technological: multilayer perceptrons are applied to many practical problems such as pattern recognition,[7] combinatorial optimization,[8] and so forth. There have been many studies on the learning of multilayer networks.[9–15] Among other types of learning, we focus on learning of Boolean functions expressed by a two-layer perceptron with a tree-like architecture. This model has been studied by using the Gibbs and Bayes algorithms with and without noise by Schottky etal..[16,17] By defining α as p/N, where p and N are the numbers of examples and input synapses, respectively, it has been reported that the critical value $\alpha_{\rm C}$ above which learning takes place is nonzero for the Parity machine but is 0 for the And machine. In ref. 15, the authors call the phenomenon of the nonzero $\alpha_{\rm C}$ the "Aha effect", which is obtained with and without noise. On the other hand, $\alpha_{\rm C}$ is always 0 for learning in simple perceptrons.[1–6] The phenomenon of the nonzero $\alpha_{\rm C}$ has also been reported for the online learning of the Parity machine in the two-layer perceptron for output noise by Kabashima.[18] Furthermore, it has been reported that learning itself does not take place when the noise amplitude is large.[18]

In this paper, we assume that only the teacher suffers from noise for simplicity. We use the replica method and perform Markov chain Monte Carlo simulations (MCMCs) together with exchange Monte Carlo simulations (EXMCs). We assume the replica

*E-mail address: uezu@cc.nara-wu.ac.jp

symmetric (RS) ansatz. We study the Parity machine and the AND machine. In the Parity machine, firstly, we focus on whether $\alpha_{\rm C}$ is 0 or not and the behaviors of the learning curve when learning begins. Secondly, we study asymptotic learning behavior for large α. Thirdly, we study the de Almeida-Thouless (AT) stabilities[19] of the RS solutions in the cases that the temperature is low and the noise amplitude is large. In both cases, the spin glass solution exists theoretically, but numerically it does not appear. In the AND machine, we study learning behaviors for $\alpha \simeq 0$ and for large α, theoretically and numerically.

The structure of this paper is as follows. In sect. 2, we formulate the problem by the replica method and describe the free energy for general Boolean functions and for input and output noise together with the noiseless case. In sect. 3, we study the Parity machine. We derive the saddle point equations (SPEs) and expressions for the generalization error, and investigate the critical $\alpha_{\rm C}$ and the behaviors of the learning curve for α around $\alpha_{\rm C}$, and we study asymptotic learning behavior for large α. In sect. 4, we study the AT stability. In sect. 5, we show the results of numerical simulations and compare them with the theoretical results for the Parity machine. For the And machine, theoretical analysis is given in sect. 6, and numerical and theoretical results are compared in sect. 7. Section 8 contains a summary and discussion. In Appendix A, we describe the details of the derivation of the free energy of general Boolean functions for input and output noise together with the noiseless case. In Appendices B and C, we derive SPEs for the Parity and And machines, respectively.

2. Formulation

We study a two-layer perceptron with a tree-like structure. Input layers consist of K units, and each unit has M input points (see Fig. 1). We assume that both the teacher and student have the same architecture but only the teacher suffers from noise. We denote the teacher vector by $\boldsymbol{w}^t = (\boldsymbol{w}_1^t, \boldsymbol{w}_2^t, \cdots, \boldsymbol{w}_K^t)$ and the student vector by $\boldsymbol{w}^s = (\boldsymbol{w}_1^s, \boldsymbol{w}_2^s, \cdots, \boldsymbol{w}_K^s)$. Common examples are given to both the teacher and student. The μth example is denoted by $\boldsymbol{x}^\mu = (\boldsymbol{x}_1^\mu, \boldsymbol{x}_2^\mu, \cdots, \boldsymbol{x}_K^\mu)$. Vectors $\boldsymbol{w}_l^t, \boldsymbol{w}_l^s$, and $\boldsymbol{x}_l^\mu$ are all M-dimensional and their norms are set to $\sqrt{M}$.

$$\boldsymbol{w}_l^t = (w_{l1}^t, w_{l2}^t, \ldots, w_{lM}^t),\ |\boldsymbol{w}_l^t| = \sqrt{M},\ l = 1, 2, \ldots, K, \tag{1}$$

$$\boldsymbol{w}_l^s = (w_{l1}^s, w_{l2}^s, \ldots, w_{lM}^s),\ |\boldsymbol{w}_l^s| = \sqrt{M},\ l = 1, 2, \ldots, K, \tag{2}$$

$$\boldsymbol{x}_l^\mu = (x_{l1}^\mu, x_{l2}^\mu, \ldots, x_{lM}^\mu),\ |\boldsymbol{x}_l^\mu| = \sqrt{M},\ l = 1, 2, \ldots, K. \tag{3}$$

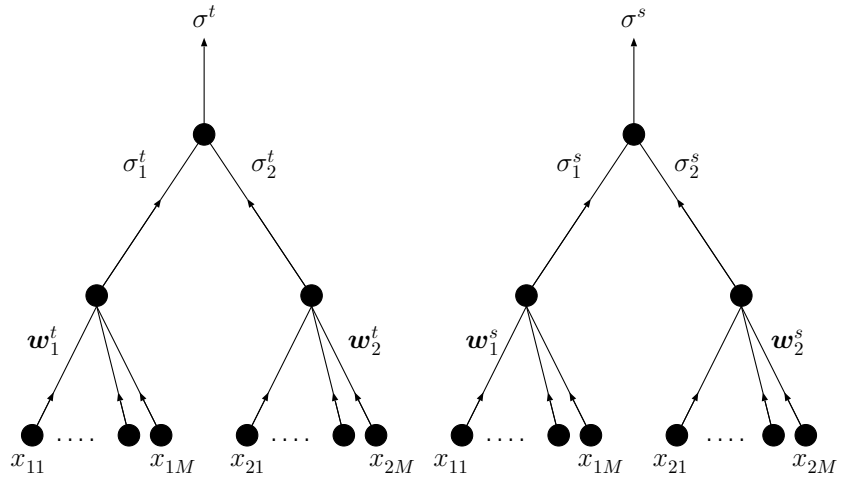

Fig. 1. Schematic figure of tree-like structure of two-layer perceptron.

The outputs to an example vector $\boldsymbol{x}^\mu = \{\boldsymbol{x}_l^\mu\}$ by the teacher and student are

$$\sigma_{l,\mu}^t = \mathrm{sgn}\Big(\frac{(\boldsymbol{x}_l^\mu, \boldsymbol{w}_l^t)}{\sqrt{M}}\Big), \tag{4}$$

$$\sigma_\mu^t = B_t(\sigma_{1,\mu}^t, \ldots, \sigma_{K,\mu}^t), \tag{5}$$

$$\sigma_{l,\mu}^s = \mathrm{sgn}\Big(\frac{(\boldsymbol{x}_l^\mu, \boldsymbol{w}_l^s)}{\sqrt{M}}\Big), \tag{6}$$

$$\sigma_\mu^s = B_s(\sigma_{1,\mu}^s, \ldots, \sigma_{K,\mu}^s). \tag{7}$$

Here, $(\boldsymbol{x}, \boldsymbol{w})$ denotes the inner product and $B_t(\sigma_1 \ldots, \sigma_K)$ and $B_s(\sigma_1, \ldots, \sigma_K)$ are Boolean functions, that is, mappings from $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_K)$ to σ, where $\sigma_i = \pm 1$ and $\sigma = \pm 1$. $\mathrm{sgn}(x)$ is defined as

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{for } x > 0, \\ -1 & \text{for } x \leq 0. \end{cases}$$

2.1 Learning algorithm

Let us define the training set ξ_p by the p-set of $(\boldsymbol{x}^\mu, \sigma_\mu^t)$,

$$\xi_p = \{(\boldsymbol{x}^\mu, \sigma_\mu^t), \mu = 1, \cdots, p\}. \tag{8}$$

When the training set ξ_p is given, we define the energy of a student $\boldsymbol{w}$ as the number of discrepancies between the output σ_μ^t by the teacher and the output σ_μ^s by the student,

$$E[\boldsymbol{w}, \xi_p] \quad = \quad \sum_{\mu=1}^{p} \Theta(-\sigma_\mu^t \sigma_\mu^s), \tag{9}$$

where $\Theta(u)$ is the Heaviside function, i.e., $\Theta(u) = \frac{1}{2}(\mathrm{sgn}(u) + 1)$. The learning strategy we adopt is the Gibbs algorithm, in which a synaptic weight $\boldsymbol{w}$ is selected with a probability proportional to $e^{-\beta E[\boldsymbol{w},\xi_p]}$, where β is the inverse temperature, $\beta = \frac{1}{T}$. The algorithm in the limit $T \to +0$ corresponds to the minimum-error algorithm, in which only the synaptic weights with minimum energies are selected. Therefore, the temperature represents the measure of tolerance in selecting synaptic weights. Let us define the overlap $\{R_l\}$ between the weights of a student and the teacher, and the overlap $\{q_l\}$ between the weights of students as

$$R_l = \frac{1}{M}\left(\boldsymbol{w}_l^t, \boldsymbol{w}_l^s\right),\ l = 1, \cdots, K, \tag{10}$$

$$q_l = \frac{1}{M}\left(\boldsymbol{w}_l^s, \boldsymbol{w}_l^s\right),\ l = 1, \cdots, K. \tag{11}$$

When $R_l = 1$, the teacher and student vectors in the lth unit coincide, and if it is 0, they are orthogonal. We define the loading rate α as

$$\alpha = \frac{p}{KM}. \tag{12}$$

The generalization error ϵ_g is defined as the probability that the outputs by the teacher and student to a new example are different. That is,

$$\epsilon_g \quad = \quad \langle \Theta(-\sigma^s \sigma^t) \rangle_{p,x}, \tag{13}$$

where $\langle \cdot \rangle_{p,x}$ implies the average over the teacher's output and examples. We study the learning curve $\epsilon_g(\alpha)$ and the α dependences of overlaps R and q, and study the temperature dependences of these quantities.

2.2 Replica analysis

We derive the SPEs by the replica method. The partition function Z is expressed as

$$Z = \int \Big[\prod_{l=1}^{K} d\boldsymbol{w}_l^s\Big]\Big[\prod_{l=1}^{K} \delta((\boldsymbol{w}_l^s)^2 - M)\Big] e^{-\beta E[\boldsymbol{w}^s,\xi_p]}. \tag{14}$$

Here, $d\boldsymbol{w}_l^s = dw_{l1}^s \ldots dw_{lM}^s$. We denote the probability that the teacher output is σ by $P(\sigma)$. We assume the self-averaging of the logarithm of the partition function, that is, $\ln Z$ is equal to the value averaged over teacher's output, examples, and teacher's weight vectors.

$$\ln Z = \langle \ln Z \rangle_{p,x,w^t}.$$

In order to calculate $\ln Z$, we use the replica method. We prepare n sets of students $\{\boldsymbol{w}_l^{s,\alpha}\}_{\alpha=1,n}$ with the same set of the training set ξ_p and the same teacher vector $\boldsymbol{w}^t$. Thus, from Eq. (14), we have

$$Z^n = \int \Big[\prod_{\alpha}\prod_{l=1}^{K} d\boldsymbol{w}_l^{s,\alpha}\Big]\Big[\prod_{\alpha}\prod_{l=1}^{K}\delta((\boldsymbol{w}_l^{s,\alpha})^2 - M)\Big] e^{-\sum_\alpha \beta E[\boldsymbol{w}^{s\alpha},\xi_p]}. \tag{15}$$

We define

$$R_l^\alpha = \frac{1}{M}(\boldsymbol{w}_l^{s\alpha}, \boldsymbol{w}_l^t), \quad l = 1, \cdots, K, \tag{16}$$

$$q_l^{\alpha\beta} = \frac{1}{M}(\boldsymbol{w}_l^{s\alpha}, \boldsymbol{w}_l^{s\beta}), \quad l = 1, \cdots, K. \tag{17}$$

By using the standard recipe, under the ansatz of the replica symmetry, i.e., $R_l = R_l^\alpha$ and $q_l = q_l^{\alpha\beta}$ $(l = 1, \cdots, K)$, $\langle \ln Z\rangle_{p,x,w^t}$ is expressed as

$$\frac{1}{N}\langle \ln Z\rangle_{p,x,w^t} = \frac{1}{n}\alpha G_1 + \frac{1}{n}G_2. \tag{18}$$

Therefore, the free energy per input unit F/N is given by

$$F/N = -T\langle \ln Z\rangle_{p,x,w^t}/N = -T\frac{1}{n}(\alpha G_1 + \frac{1}{n}G_2). \tag{19}$$

Irrespective of the type of Boolean function and noise, $\frac{G_2}{n}$ is expressed as

$$\frac{G_2}{n} = \frac{1}{K}\sum_{l=1}^{K}\frac{1}{2}\Big(\ln\frac{2\pi}{E_l + F_l} + E_l + \frac{F_l - \bar{R}_l^2}{E_l + F_l}\Big) + \frac{1}{K}(i\sum_l \bar{R}_l R_l + \sum_l \frac{F_l q_l}{2}), \tag{20}$$

where E_l is introduced to express the normalization of students' vectors, and F_l and $\bar{R}_l$ are conjugate variables to q_l and R_l, respectively. See Appendix A for the derivation. By eliminating conjugate variables using their saddle point conditions, the expression for G_2 is further rewritten as

$$\frac{G_2}{n} = \frac{1}{2K}\sum_{l=1}^{K}\Big[\ln\Big(2\pi(1-q_l)\Big) + 1 + \frac{q_l - R_l^2}{1-q_l}\Big]. \tag{21}$$

Below, for the input and output noise and noiseless cases, we give expressions for G_1 for general Boolean functions. The detail of the derivation of these expressions is also given in Appendix A.

2.2.1 Input noise

We assume that an independent noise $\boldsymbol{\zeta}_l$ enters each example $\boldsymbol{x}_l$ input to the teacher and that $(\zeta_l)_i$ is Gaussian noise with mean 0 and standard deviation τ_l. The probability $p(\sigma|\{h_l^t\})$ that the teacher's output is σ is the probability that $\sigma = B(\{\sigma_l\})$. Here,

$h_l^t = \frac{1}{\sqrt{M}}(\boldsymbol{w}_l^t, \boldsymbol{x}_l)$. This probability is expressed as

$$p(\sigma|\{h_l^t\}) = \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_{l=1}^{K} H\Big(- \sigma_l \frac{h_l^t}{\tau_l}\Big), \tag{22}$$

where we define

$$H(x) = \int_x^\infty \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = \int_x^\infty Dt, \tag{23}$$

$$Dt \equiv dt\ h(t),\ h(t) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \tag{24}$$

$$\triangle_t^\sigma(\{\sigma_l\}) = \begin{cases} 1 & \text{for } \sigma = B_t\big(\{\sigma_l\}\big), \\ 0 & \text{for } \sigma = -B_t\big(\{\sigma_l\}\big). \end{cases} \tag{25}$$

$\frac{G_1}{n}$ is expressed as

$$\frac{G_1}{n} = \Big[\int \prod_l Dt_l\Big] \sum_{\sigma=\pm 1} \mathrm{Tr}_{\sigma_l} \triangle_t^\sigma (\{\sigma_l\}) \Big[\prod_l H\Big(\sigma_l \tilde{X}_l\Big)\Big] \ln \Phi_{\beta,s}^\sigma(\{Y_l\}), \tag{26}$$

where

$$\Phi_{\beta,s}^\sigma(\{Y_l\}) = \mathrm{Tr}_{\{\sigma_l'\}}\Big(\tilde{\triangle}_{s,\beta}^\sigma(\{\sigma_l'\}) \prod_l H(\sigma_l' Y_l)\Big), \tag{27}$$

$$\tilde{X}_l = \tilde{\zeta}_l t_l,\ Y_l = \gamma_l t_l,\ \gamma_l = \sqrt{\frac{q_l}{1-q_l}},\ \tilde{\zeta}_l = \frac{R_l}{\sqrt{\eta_l q_l - R_l^2}},\ \eta_l = 1 + \tau_l^2, \tag{28}$$

$$\triangle_s^\sigma(\{\sigma_l\}) = \begin{cases} 1 & \text{for } \sigma = B_s\big(\{\sigma_l\}\big), \\ 0 & \text{for } \sigma = -B_s\big(\{\sigma_l\}\big), \end{cases} \tag{29}$$

$$\tilde{\triangle}_{s,\beta}^\sigma(\{\sigma_l^\alpha\}) \equiv e^{-\beta} + (1 - e^{-\beta})\ \triangle_s^\sigma\ \big(\{\sigma_l^\alpha\}\big). \tag{30}$$

Note that $\Phi_{\beta,s}^\sigma(\{Y_l\})$ is common for the noiseless and input and output noise cases but differs according to the Boolean function $B_s\big(\{\sigma_l\}\big)$, that is, it differs for the Parity and And machines.

2.2.2 Output noise

In the output noise case, the sign of the output σ is reversed with a nonzero probability. We assume that only the teacher suffers from output noise. Let λ be the probability that the teacher's output is reversed. Then, the probability that teacher's output is σ is

$$p(\sigma|\{h_l^t\}) = \lambda + (1-2\lambda)\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l \Theta(\sigma_l h_l^t). \tag{31}$$

$\frac{G_1}{n}$ is expressed as

$$\frac{G_1}{n} = \Big[\int \prod_l Dt_l\Big] \sum_{\sigma=\pm 1} \{\lambda + (1-2\lambda)\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H\Big(\sigma_l \tilde{X}_l\Big)\} \ln \Phi^\sigma_{\beta,s}(\{Y_l\}), \tag{32}$$

$$X_l = \zeta_l t_l,\ Y_l = \gamma_l t_l, \gamma_l = \sqrt{\frac{q_l}{1-q_l}}, \zeta_l = \frac{R_l}{\sqrt{q_l - R_l^2}}. \tag{33}$$

$\Phi^\sigma_{\beta,s}(\{Y_l\})$ is given by Eq. (27). Note that the expression for ζ_l is different from that for the input noise case.

2.3 Noiseless case

Quantities for the noiseless case are obtained by substituting $\tau_l = 0$ in the input noise model or by substituting $\lambda = 0$ in the output noise model. Thus, $\frac{G_1}{n}$ is

$$\frac{G_1}{n} = \sum_{\sigma=\pm 1} \{\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l \int Dt_l H(\sigma_l X_l)\} \ln \Phi^\sigma_{\beta,s}(\{Y_l\}), \tag{34}$$

$$X_l = \zeta_l t_l,\ Y_l = \gamma_l t_l,\ \gamma_l = \sqrt{\frac{q_l}{1-q_l}},\ \zeta_l = \frac{R_l}{\sqrt{q_l - R_l^2}}. \tag{35}$$

In this paper, we study the case that $B_s = B_t = B$; thus, $\Delta_s^\sigma = \Delta_t^\sigma$.
In the next section, we study the Parity machine.

3. Parity Machine

Hereafter, we consider the case of two units, that is $K = 2$. The output of the Parity machine is 1 if the outputs of the two units are the same and -1 otherwise, that is,

$$B(\sigma_1, \sigma_2) = \sigma_1 \sigma_2. \tag{36}$$

Below, we show the SPEs. Detailed derivations are given in Appendix B.

3.1 Input noise

We study the case of $\tau_l = \tau$. As shown later, our numerical results obtained by MCMCs show the relations $q_l = q$ and $_l = R$ for $l = 1, 2$. Thus, we assume these relations. We obtain

$$\frac{G_1}{n} = 2\int Dt_1 Dt_2 H_2(\tilde{X}_1, \tilde{X}_2) \ln \Phi_+(Y_1, Y_2), \tag{37}$$

where $H_2(\tilde{X}_1, \tilde{X}_2) = H(\tilde{X}_1)H(\tilde{X}_2) + H(-\tilde{X}_1)H(-\tilde{X}_2)$ and $\Phi_+(Y_1, Y_2) \equiv \Phi^{+1}_{\beta,s}(\{Y_l\}) = H_2(Y_1, Y_2) + e^{-\beta} H_2(Y_1, -Y_2)$. Then, the SPEs are given by

$$q^2 - R^2 = 2\alpha(1 - e^{-\beta})\sqrt{q(1-q)}\ I_1^{(\mathrm{i})}, \tag{38}$$

$$R(\eta q - R^2)^{3/2} = -2\alpha\eta q(1-q)\ I_2^{(\mathrm{i})}, \tag{39}$$

$$\tilde{X}_l = \tilde{\zeta} t_l,\ Y_l = \gamma t_l,\ \gamma = \sqrt{\frac{q}{1-q}},\ \tilde{\zeta} = \frac{R}{\sqrt{\eta q - R^2}},\ \eta = 1+\tau^2, \tag{40}$$

$$I_1^{(\mathrm{i})} = 2I_{1,1} = 2\int Dt_1 Dt_2 t_1 h(Y_1) H_a(Y_2) \frac{H_2(\tilde{X}_1, \tilde{X}_2)}{\Phi_+(Y_1, Y_2)}, \tag{41}$$

$$I_2^{(\mathrm{i})} = 2I_{2,1} = 2\int Dt_1 Dt_2 t_1 h(\tilde{X}_1) H_a(\tilde{X}_2) \ln \Phi_+(Y_1, Y_2), \tag{42}$$

where $H_a(\tilde{X}_2) = H(\tilde{X}_2) - H(-\tilde{X}_2)$.

3.2 Output noise

$\frac{G_1}{n}$ is

$$\frac{G_1}{n} = 2\int Dt_1 Dt_2 \big\{\lambda + (1-2\lambda) H_2(X_1, X_2)\big\} \ln \Phi_+(Y_1, Y_2). \tag{43}$$

The SPEs are

$$q^2 - R^2 = 2\alpha(1-e^{-\beta})\sqrt{q(1-q)}\ I_1^{(\mathrm{o})}, \tag{44}$$

$$R(q-R^2)^{3/2} = -2\alpha q(1-q)\ I_2^{(\mathrm{o})}, \tag{45}$$

$$I_1^{(\mathrm{o})} = 2\int Dt_1 \int Dt_2 \{\lambda + (1-2\lambda) H_2(X_1, X_2)\} t_1 \frac{h(Y_1) H_a(Y_2)}{\Phi_+(Y_1, Y_2)}, \tag{46}$$

$$I_2^{(\mathrm{o})} = 2(1-2\lambda)\int Dt_1 \int Dt_2 h(X_1) H_a(X_2) t_1 \ln \Phi_+(Y_1, Y_2), \tag{47}$$

$$X_l = \zeta t_l,\ \zeta = \frac{R}{\sqrt{q-R^2}},\ Y_l = \gamma t_l,\ \gamma = \sqrt{\frac{q}{1-q}}. \tag{48}$$

3.3 Noiseless case

In this case, the SPEs are given by substituting $\tau = 0$ in the expressions for the input noise model or substituting $\lambda = 0$ in the expressions for the output noise model.

$$q^2 - R^2 = 2\alpha(1-e^{-\beta})\sqrt{q(1-q)}\ I_1^{(\mathrm{n})}, \tag{49}$$

$$R(q-R^2)^{3/2} = -2\alpha q(1-q)\ I_2^{(\mathrm{n})}, \tag{50}$$

$$I_1^{(\mathrm{n})} = 2\int Dt_1 \int Dt_2 H_2(X_1, X_2) t_1 \frac{h(Y_1) H_a(Y_2)}{\Phi_+(Y_1, Y_2)}, \tag{51}$$

$$I_2^{(\mathrm{n})} = 2\int Dt_1 \int Dt_2 h(X_1) H_a(X_2) t_1 \ln \Phi_+(Y_1, Y_2), \tag{52}$$

$$X_l = \zeta t_l,\ \zeta = \frac{R}{\sqrt{q-R^2}},\ Y_l = \gamma t_l\ \gamma = \sqrt{\frac{q}{1-q}}. \tag{53}$$

3.4 Generalization error

Here, we give the expressions for the generalization error for the Parity machine.

(1) Noiseless case

$$\epsilon_g = \frac{1}{2} - \frac{2}{\pi^2}\sin^{-1} R_1 \sin^{-1} R_2. \tag{54}$$

(2) Input noise

$$\epsilon_g = \frac{1}{2} - \frac{2}{\pi^2}\sin^{-1}(\frac{R_1}{\sqrt{1+\tau^2}})\sin^{-1}(\frac{R_2}{\sqrt{1+\tau^2}}). \tag{55}$$

(3) Output noise

$$\epsilon_g = \frac{1}{2} - (1-2\lambda)\frac{2}{\pi^2}\sin^{-1} R_1 \sin^{-1} R_2. \tag{56}$$

3.5 Behavior for small α and critical value $\alpha_{\rm C}$

In the Parity machine, there exists nonzero critical values of α, $\alpha_{\rm C}$ and $\alpha_{\rm L}$, above which learning takes place when T is fixed. In order to obtain the formulae for $\alpha_{\rm C}$ and $\alpha_{\rm L}$, assuming $q \sim 0$ and $R \sim 0$, we expand I_1 and I_2 in terms of q and R. In the course of the study, we note that there are two cases of ζ and $\tilde{\zeta}$. In the case of $\zeta \sim \mathcal{O}(1)$ or $\tilde{\zeta} \sim \mathcal{O}(1)$, which occurs in the learning state, $\alpha_{\rm C}$ is defined. On the other hand, in the case of $\zeta = 0$ and $\tilde{\zeta} = 0$, which occurs in the spin glass state, $\alpha_{\rm L}$ is defined. Firstly, we study the former case.

3.5.1 Learning state

(i) Noiseless case

Let us calculate the expansions of $I_1^{(\rm n)}$ [Eq. (51)] and $I_2^{(\rm n)}$ [Eq. (52)] with respect to γ. Using the expansions

$$\Phi_+(Y_1, Y_2) = \frac{1}{2}(1+e^{-\beta})\left[1 + \frac{2a}{\pi}\gamma^2 t_1 t_2\left(1 - \frac{\gamma^2}{6}(t_1^2+t_2^2) + \mathcal{O}(\gamma^4)\right)\right], \tag{57}$$

$$H_a(Y_2) \simeq -\frac{2}{\sqrt{2\pi}}\gamma t_2 + \frac{1}{3\sqrt{2\pi}}\gamma^3 t_2^3 + \mathcal{O}(\gamma^5), \tag{58}$$

we have

$$I_1^{(\rm n)} = -\frac{4\gamma}{\pi^2(1+e^{-\beta})}\left(\frac{\zeta}{\sqrt{1+\zeta^2}}\left(1 - \frac{\gamma^2}{6}\frac{3+2\zeta^2}{1+\zeta^2}\right)\frac{\zeta}{1+\gamma^2}\frac{1}{\sqrt{1+\gamma^2+\zeta^2}} - a\gamma^2(1+\gamma^2)^{-3/2} + \mathcal{O}(\gamma^4)\right) \tag{59}$$

$$I_2^{(\rm n)} = -\frac{4a\gamma^2}{\pi^2}\frac{\zeta}{(1+\zeta^2)^2}\left(1 - l_2(\zeta^2)\gamma^2 + \mathcal{O}(\gamma^4)\right), \tag{60}$$

$$l_2(\zeta^2) = \frac{3+\zeta^2}{3(1+\zeta^2)}, \tag{61}$$

where

$$a = \frac{e^{\beta}-1}{e^{\beta}+1}. \tag{62}$$

When ζ is not a small quantity, we obtain

$$I_1^{(\mathrm{n})} = -\frac{4\gamma}{\pi^2(1+e^{-\beta})}\frac{\zeta^2}{1+\zeta^2}\Big(1-l_1(\zeta^2)\gamma^2+\mathcal{O}(\gamma^4)\Big), \tag{63}$$

$$l_1(\zeta^2) = \frac{2(3+2\zeta^2)}{3(1+\zeta^2)} + (1+\frac{1}{\zeta^2})a. \tag{64}$$

We have the relations

$$\frac{q^2-R^2}{\gamma\sqrt{q(1-q)}} = \frac{\gamma^2-\zeta^2}{(1+\gamma^2)(1+\zeta^2)}, \tag{65}$$

$$\frac{R(q-R^2)^{3/2}}{q(1-q)\gamma^2} = \frac{\zeta}{(1+\zeta^2)^2}. \tag{66}$$

Substituting these relations into the SPEs (49) and (50) yields

$$\frac{x-y}{(1+x)(1+y)} = \frac{8a}{\pi^2}\alpha\frac{x}{1+x}\Big(1-l_1(x)y\Big)+\mathcal{O}(y^2), \tag{67}$$

$$1 = \frac{8a}{\pi^2}\alpha\Big(1-l_2(x)y\Big)+\mathcal{O}(y^2), \tag{68}$$

where we define $x \equiv \zeta^2$ and $y \equiv \gamma^2$. R and q are

$$R = \sqrt{\frac{xy}{(1+x)(1+y)}}, \tag{69}$$

$$q = \frac{y}{1+y}. \tag{70}$$

From Eq. (68), the solution y exists for $\alpha > \alpha_{\mathrm{C}}^{(\mathrm{n})}$. We obtain

$$\alpha_{\mathrm{C}}^{(\mathrm{n})} = \frac{\pi^2}{8a}. \tag{71}$$

Defining $\varepsilon = \frac{\alpha-\alpha_{\mathrm{C}}^{(\mathrm{n})}}{\alpha_{\mathrm{C}}^{(\mathrm{n})}}$, the SPEs (67) and (68) are rewritten as

$$\frac{x-y}{(1+x)(1+y)} = (1+\varepsilon)\frac{x}{1+x}\Big(1-l_1(x)y\Big)+\mathcal{O}(y^2), \tag{72}$$

$$1 = (1+\varepsilon)\Big(1-l_2(x)y\Big)+\mathcal{O}(y^2). \tag{73}$$

We expand x and y by ε.

$$x = x_0 + x_1\varepsilon + \cdots, \tag{74}$$

$$y = y_0 + y_1\varepsilon + \cdots. \tag{75}$$

From Eq. (73), we obtain

$$y = \frac{\varepsilon}{(1+\varepsilon)l_2(x)} + \mathcal{O}(y^2). \tag{76}$$

Thus,

$$y_0 = 0,\ y_1 = \frac{1}{l_2(x_0)}. \tag{77}$$

From Eq. (72), we obtain

$$y_1 = \frac{x_0}{x_0 l_1(x_0) - (1+x_0)}. \tag{78}$$

From this, we obtain

$$(x_0+1)(ax_0 - (1-a)) = 0. \tag{79}$$

Since $x = \zeta^2 \geq 0$, we have

$$x_0 = \frac{1-a}{a} = \frac{2}{e^\beta - 1}. \tag{80}$$

Thus, we obtain

$$R \simeq \sqrt{\frac{6(e^\beta-1)}{(3e^\beta-1)(e^\beta+1)}}\sqrt{\frac{\Delta\alpha}{\alpha_{\rm C}^{(\rm n)}}}, \tag{81}$$

$$q \simeq \frac{3(e^\beta+1)}{3e^\beta-1}\frac{\Delta\alpha}{\alpha_{\rm C}^{(\rm n)}}, \tag{82}$$

$$\epsilon_g \simeq \frac{1}{2} - \frac{2}{\pi^2}R^2 \simeq \frac{1}{2} - \frac{12}{\pi^2}\frac{e^\beta-1}{(3e^\beta-1)(e^\beta+1)}\frac{\Delta\alpha}{\alpha_{\rm C}^{(\rm n)}}, \tag{83}$$

where $\Delta\alpha = \alpha - \alpha_{\rm C}^{(\rm n)}$. Since R is positive, we call this a learning (L) solution. The exponents $\bar{\beta}$ of R and ϵ_g are defined by $R \propto (\Delta\alpha)^{\bar{\beta}_R}$, $q \propto (\Delta\alpha)^{\bar{\beta}_q}$, and $\Delta\epsilon \equiv \epsilon_g|_{R=0} - \epsilon_g \propto (\Delta\alpha)^{\bar{\beta}_\epsilon}$, respectively. Thus, we obtain $\bar{\beta}_R = 1/2$, $\bar{\beta}_q = 1$, and $\bar{\beta}_\epsilon = 1$.

(ii) Input noise

Expressions $I_1^{(\rm i)}$ and $I_2^{(\rm i)}$ are obtained from $I_1^{(\rm n)}$ and $I_2^{(\rm n)}$ by replacing ζ with $\tilde{\zeta}$. We have the relations

$$-\frac{q^2-R^2}{\gamma\sqrt{q(1-q)}} = \frac{\eta\tilde{x} + (\eta-1)\tilde{x}y - y}{(1+\tilde{x})(1+y)}, \tag{84}$$

$$\frac{R(\eta q - R^2)^{3/2}}{q(1-q)\gamma^2} = \frac{\eta^2\sqrt{\tilde{x}}}{(1+\tilde{x})^2}, \tag{85}$$

where $\tilde{x} = \tilde{\zeta}^2$ and $y = \gamma^2$. By similar analysis to the noiseless case, we obtain

$$\alpha_{\rm C}^{(\rm in)} = \eta\frac{\pi^2}{8a}, \tag{86}$$

$$y_0 = 0, \tag{87}$$

$$y_1 = \frac{\eta\tilde{x}_0}{\eta\tilde{x}_0 l_1(\tilde{x}_0) - (1+\tilde{x}_0)}, \tag{88}$$

$$\tilde{x}_0 = \frac{1-\eta a}{\eta(1+a)-1}. \tag{89}$$

The condition that $\tilde{x}_0$ is non-negative is

$$\beta < \ln\left(\frac{2+\tau^2}{\tau^2}\right). \tag{90}$$

In this case, we have an L solution as

$$R \simeq 2\sqrt{\eta}\sqrt{\frac{3}{9e^\beta+1}}\sqrt{\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (in)}}}, \tag{91}$$

$$q \simeq \frac{9(e^\beta+1)}{9e^\beta+1}\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (in)}}, \tag{92}$$

$$\epsilon_g \simeq \frac{1}{2} - \frac{2}{\pi^2}\frac{R^2}{\eta} \simeq \frac{1}{2} - \frac{24}{\pi^2}\frac{1}{9e^\beta+1}\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (in)}}. \tag{93}$$

Thus, we obtain $\bar{\beta}_R = 1/2$, $\bar{\beta}_q = 1$, and $\bar{\beta}_\epsilon = 1$.

(iii) Output noise

$I_1^{\rm (o)}$ and $I_2^{\rm (o)}$, respectively given by Eqs. (46) and (47), are expressed by using those for the noiseless case as

$$I_1^{\rm (o)} = \lambda J_1 + (1-2\lambda) I_1^{\rm (n)}, \tag{94}$$

$$I_2^{\rm (o)} = (1-2\lambda) I_2^{\rm (n)}, \tag{95}$$

$$J_1 = 2\int Dt_1 \int Dt_2 t_1 \frac{h(Y_1)H_a(Y_2)}{\Phi_+(Y_1,Y_2)}. \tag{96}$$

J_1 is estimated as

$$J_1 = -\gamma^3 \frac{8a}{\pi^2(1+e^{-\beta})}\frac{1}{(1+\gamma^2)^{3/2}} + \mathcal{O}(\gamma^5). \tag{97}$$

We set $x = \zeta^2$ and $y = \gamma^2$ and obtain

$$\alpha_{\rm C}^{\rm (o)} = \frac{\pi^2}{8a(1-2\lambda)}, \tag{98}$$

$$y_0 = 0, \tag{99}$$

$$y_1 = \frac{3(1-2\lambda)(e^\beta+1)}{(3-2\lambda)e^\beta - (1+2\lambda)}, \tag{100}$$

$$x_0 = \frac{2\Big(1-\lambda(1+e^\beta)\Big)}{e^\beta - 1}. \tag{101}$$

Note that the solution can be negative. The condition that the solution is non-negative is

$$\beta \leq \ln \frac{1-\lambda}{\lambda}. \tag{102}$$

In this case, we obtain an L solution as

$$R \simeq \sqrt{\frac{6(1-\lambda(e^\beta+1))}{(3-2\lambda)e^\beta-(1+2\lambda)}}\sqrt{\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (o)}}}, \tag{103}$$

$$q \simeq y \simeq \frac{3(1-2\lambda)(e^\beta+1)}{(3-2\lambda)e^\beta-(1+2\lambda)}\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (o)}}, \tag{104}$$

$$\epsilon_g \simeq \frac{1}{2}-(1-2\lambda)\frac{2}{\pi^2}R^2 \simeq \frac{1}{2}-\frac{1-2\lambda}{\pi^2}\frac{12(1-\lambda(e^\beta+1))}{(3-2\lambda)e^\beta-(1+2\lambda)}\frac{\Delta\alpha}{\alpha_{\rm C}^{\rm (o)}}. \tag{105}$$

Thus, we obtain $\bar{\beta}_R = 1/2$, $\bar{\beta}_q = 1$, and $\bar{\beta}_\epsilon = 1$. Now, let us consider the latter case.

3.5.2 Spin glass state

In this subsection, we study the case that the positivity condition for $\tilde{x}_0$ or x_0 is broken. We define $\beta_{\rm SG}$ as $\ln(\frac{2+\tau^2}{\tau^2})$ in the input noise case and $\ln(\frac{1-\lambda}{\lambda})$ in the output noise case. We define $A_{\rm noise}$ as

$$A_{\rm noise} = \begin{cases} \eta & \text{for input noise,} \\ \frac{1}{1-2\lambda} & \text{for output noise.} \end{cases} \tag{106}$$

$\beta_{\rm SG}$ is characterized by

$$e^{\beta_{\rm SG}} = \frac{A_{\rm noise}+1}{A_{\rm noise}-1}, \tag{107}$$

that is,

$$a(\beta_{\rm SG}) = \frac{e^{\beta_{\rm SG}}-1}{e^{\beta_{\rm SG}}+1} = \frac{1}{A_{\rm noise}}. \tag{108}$$

For $\beta > \beta_{\rm SG}$, we found two kinds of solutions, in one solution $R > 0$ and $q > 0$, and in the other solution $R = 0$ and $q > 0$. The former we call the L solution. Since $R = 0$ and $q \neq 0$, we call the latter a spin glass (SG) solution. The SG solution satisfies

$$q^{3/2} = 2\alpha(1-e^{-\beta})I_1^{\rm (n)}, \tag{109}$$

$$I_1^{\rm (n)} = \int Dt_1 \int Dt_2 t_1 \frac{h(Y_1)H_a(Y_2)}{\Phi_+(Y_1,Y_2)}. \tag{110}$$

q does not depend on $A_{\rm noise}$. The critical value of α, $\alpha_{\rm SG}$, above which the SG solution exists for any β is

$$\alpha_{\rm SG} = \frac{\pi^2}{8a^2}. \tag{111}$$

$\alpha_{\rm SG}$ also does not depend on $A_{\rm noise}$. The L solution bifurcates from the SG solution, and the critical capacity $\alpha_{\rm L}$ above which the L solution exists for $\beta > \beta_{\rm SG}$ is determined by

$$\frac{1}{1+\gamma^2} = 2\alpha_{\rm L}(1-e^{-\beta})\frac{1}{\gamma^3}\tilde{I}_1(\gamma), \tag{112}$$

$$A_{\rm noise} = \frac{4\alpha_{\rm L}}{\pi}\frac{1}{\gamma^2}\tilde{I}_2(\gamma), \tag{113}$$

$$\tilde{I}_1(\gamma) = 2\int Dt_1 \int Dt_2 t_1 \frac{h(\gamma t_1)H(\gamma t_2)}{\Phi_+(\gamma t_1, \gamma t_2)}, \tag{114}$$

$$\tilde{I}_2(\gamma) = \int Dt_1 \int Dt_2 t_1 t_2 \ln\Big(\Phi_+(\gamma t_1, \gamma t_2)\Big). \tag{115}$$

For noisy cases, $\alpha_{\rm c}$ is expressed as $\alpha_{\rm c} = aA_{\rm noise}\alpha_{\rm SG}$. From Eq. (108), $\alpha_{\rm C}(T_{\rm SG}, A_{\rm noise}) = \alpha_{\rm SG}(T_{\rm SG})$, where $T_{\rm SG} = 1/\beta_{\rm SG}$. Furthermore, it is easily proved that as q tends to 0, $\alpha_{\rm L}(T, A_{\rm noise})$ tends to $\alpha_{\rm C}(T, A_{\rm noise})$ with fixed $A_{\rm noise}$. This happens at $T = T_{\rm SG}$. That is, $\alpha_{\rm C}(T_{\rm SG}, A_{\rm noise}) = \alpha_{\rm L}(T_{\rm SG}, A_{\rm noise}) = \alpha_{\rm SG}(T_{\rm SG})$ holds.In Fig. 2, we display the T dependences of $\alpha_{\rm L}$, $\alpha_{\rm C}$, and $\alpha_{\rm SG}$ with $A_{\rm noise}$ fixed. P, L, and SG denote the para, learning, ans spin glass states, respectively. From the figure, there seems to be a parameter region in which only the SG solution exists. However, numerical results do not exhibit the SG state as shown in the next section. In order to find appropriate solutions, we study the AT stability of the solutions in sect. 4.

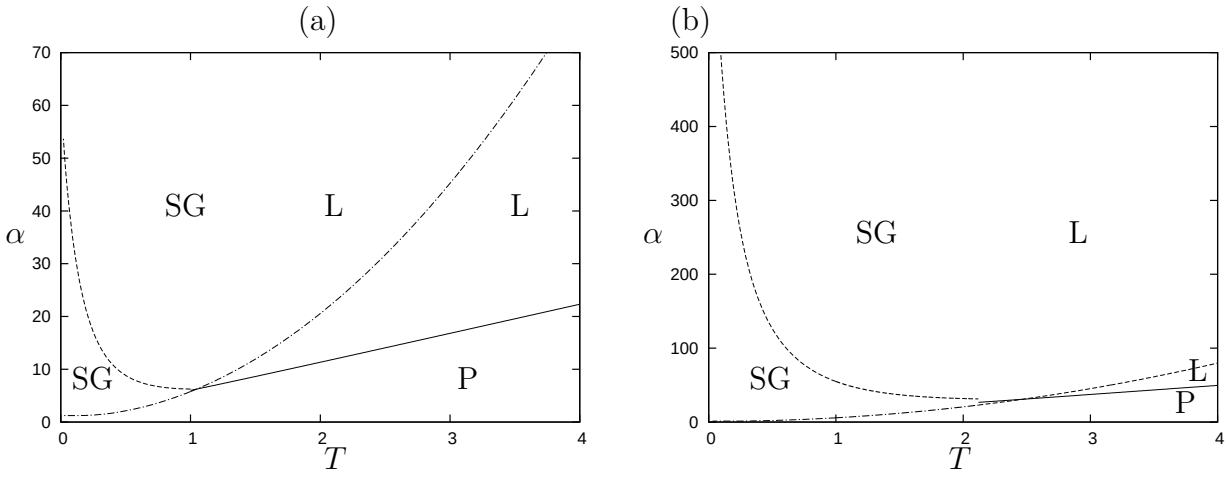

Fig. 2. T dependences of $\alpha_{\rm C}(T)$, $\alpha_{\rm L}(T)$, and $\alpha_{\rm SG}(T)$. Solid curve: $\alpha_{\rm C}$, dashed curve: $\alpha_{\rm L}$, dashed dotted curve: $\alpha_{\rm SG}$. P: para state, L: learning state, SG: spin glass state. (a) $A_{\rm noise} = 2.25, \tau = \frac{\sqrt{5}}{2}, \lambda = \frac{5}{18}$, (b) $A_{\rm noise} = 5, \tau = 2, \lambda = \frac{2}{5}$.

3.6 Learning behavior for large α

In this subsection, we study R and q when α tends to ∞. From numerical simulations, we observed that R and q tend to 1 as $\alpha \to \infty$. Thus, we define $\Delta R = 1 - R$ and $\Delta q = 1 - q$ and study the asymptotic forms of ΔR and Δq. The strategy used to derive the asymptotic behavior is to evaluate I_1 and I_2 by performing variable transformations. For the noiseless and output noise cases, we set $t_1 = u_1/\gamma$ for both I_1 and I_2. For the input noise case, for I_1 we set $t_1 = u_1/\gamma$, and for I_2 we set $t_1 = u_1/\gamma$ and $I_2 = u_2/\gamma$ and perform integration by parts.

3.6.1 Noiseless case

For $R \sim 1$ and $q \sim 1$, $\gamma \gg 1$ and $\zeta \gg 1$ follow. We define $\chi \equiv \frac{q-R^2}{1-q}$. We obtain consistent results when χ is assumed to tend to a constant as $\alpha \to \infty$. We estimate $I_1^{(\mathrm{n})}$ and $I_2^{(\mathrm{n})}$ as

$$I_1^{(\mathrm{n})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} g_1(\chi, \beta), \tag{116}$$

$$I_2^{(\mathrm{n})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} g_2(\chi, \beta), \tag{117}$$

$$g_1(\chi, \beta) = \int Duu \frac{H(-u/\chi)}{\hat{H}(u)}, \tag{118}$$

$$g_2(\chi, \beta) = \chi^2 \int Duu \ln \hat{H}(\chi u) = \chi^3 (1 - e^{-\beta}) \int Du \frac{h(\chi u)}{\hat{H}(\chi u)} > 0, \tag{119}$$

where $\hat{H}(u) = H(-u) + e^{-\beta} H(u)$. From the SPEs, we obtain

$$\Delta q \simeq q_0 \alpha^{-2}, \tag{120}$$

$$\Delta R \simeq R_0 \alpha^{-2}, \tag{121}$$

where $q_0 = \left(\sqrt{2\pi} \chi^3 / (4 g_2) \right)^2$ and $R_0 = q_0 - (1 - e^{-\beta}) 2 q_0^{3/2} g_1 / \sqrt{2\pi}$. χ is determined by

$$2R_0 = (1 + \chi^2) q_0. \tag{122}$$

To derive these results, $g_2 > 0$ is necessary, which is satisfied. We define the exponents $\hat{\beta}$ as $\Delta R \propto \alpha^{-\hat{\beta}_R}$, $\Delta q \propto \alpha^{-\hat{\beta}_q}$, and $\Delta\epsilon \equiv \epsilon_g - \epsilon_g|_{R=1} \propto \alpha^{-\hat{\beta}_\epsilon}$. Since $\Delta\epsilon \simeq \frac{2}{\pi}\sqrt{2}\sqrt{\Delta R}$, $\hat{\beta}_\epsilon = \hat{\beta}_R/2$ follows. Thus, we have $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\hat{\beta}_\epsilon = 1$.

3.6.2 Input noise

In this case, $\tilde{\zeta} = \sqrt{\frac{R}{\eta q - R^2}} \to \frac{1}{\sqrt{\eta - 1}}$ as $\alpha \to \infty$. We define $\tilde{\chi} = \sqrt{\frac{\eta q - R^2}{1-q}}$, which tends to ∞ as $\alpha \to \infty$. We obtain

$$I_1^{(\mathrm{i})} \simeq \frac{Q^2}{\sqrt{2\pi}} g_3(\beta), \tag{123}$$

$$I_2^{(\mathrm{i})} \simeq -\frac{\sqrt{\eta - 1}}{\eta \pi^2}(1 - e^{-\beta})\hat{g}(\beta), \tag{124}$$

$$g_3(\beta) = -\int Duu\frac{1}{\hat{H}(u)} = (1 - e^{-\beta})\int Du\frac{h(u)}{\left(\hat{H}(u)\right)^2} > 0, \tag{125}$$

$$\hat{g}(\beta) = \int Du_1 Du_2 \frac{2\Psi_+(u_1, u_2) - (1 - e^{-\beta})H_a(u_1)H_a(u_2)}{\left(\Psi_+(u_1, u_2)\right)^2} > 0. \tag{126}$$

Thus, we obtain

$$\Delta q \simeq q_0 \alpha^{-1}, \tag{127}$$

$$\Delta R \simeq R_0 \alpha^{-1/2}, \tag{128}$$

where $q_0 = (\eta - 1)\pi^2 / \left(2(1 - e^{-\beta})\hat{g}\right)$ and $R_0 = (1 - e^{-\beta})q_0^{3/2} g_3 / \sqrt{2\pi}$. To derive these results, $\hat{g} > 0$ and $g_3 > 0$ are necessary, which are satisfied. We note that $q_0 > 0$ for $\eta > 1$ and that learning occurs for any $\tau > 0$, irrespective of the noise amplitude. $\Delta\epsilon$ is evaluated as $\Delta\epsilon \simeq \frac{4}{\pi^2\sqrt{(\eta-1)}} \sin^{-1}\left(\frac{1}{\sqrt{\eta}}\Delta R\right)$. Thus, we obtain $\hat{\beta}_R = \hat{\beta}_\epsilon = 1/2$ and $\hat{\beta}_q = 1$.

3.6.3 Output noise

As in the noiseless case, we assume that χ tends to a constant as $\alpha \to \infty$. We estimate $I_1^{(\mathrm{o})}$ and $I_2^{(\mathrm{o})}$ as

$$I_1^{(\mathrm{o})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} \tilde{g}_1(\chi, \beta), \tag{129}$$

$$I_2^{(\mathrm{o})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} \tilde{g}_2(\chi, \beta), \tag{130}$$

$$\tilde{g}_1(\chi, \beta) = -\lambda g_3(\chi, \beta) + (1 - 2\lambda) g_1(\chi, \beta), \tag{131}$$

$$\tilde{g}_2(\chi, \beta) = (1 - 2\lambda) g_2(\chi, \beta). \tag{132}$$

Equations (129) and (130) are similar to Eqs. (116) and (117). $\tilde{g}_2$ is positive for $\lambda < 1/2$. Thus, we obtain the same asymptotic forms of Δq and ΔR as in the noiseless case, and q_0 and R_0 are obtained by replacing g_1 and g_2 with $\tilde{g}_1$ and $\tilde{g}_2$ in the expressions for q_0

and R_0 in the noiseless case.

$$\Delta q \simeq q_0 \alpha^{-2}, \tag{133}$$

$$\Delta R \simeq R_0 \alpha^{-2}. \tag{134}$$

$\Delta\epsilon$ is evaluated as $\Delta\epsilon \simeq (1-2\lambda)\frac{4}{\pi}\sqrt{2}\sqrt{\Delta R}$. Thus, we obtain $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\hat{\beta}_\epsilon = 1$.

From these results, we conclude that learning always occurs in the noiseless case, for $\tau > 0$ in the input noise case, and for $\lambda < 1/2$ in the output noise case for large α.

4. AT Stability

As shown later, there are discrepancies between the theoretical and numerical results in the input and output noise cases for small T with fixed $A_{\rm noise}$ or for large $A_{\rm noise}$ with fixed T. In these regions, the SG solution was expected to exist but it was not found numerically. In order to resolve the discrepancies, we study the AT stability. As usual, we denote the eigenvalues of the Hessian matrix as λ_i and λ_i', $i = 1, 2, 3$. $i = 3$ corresponds to the replicon mode. We assume that $R_l = R$ and $q_l = q$ for $l = 1, 2$. We found that $\lambda_1 = \lambda_2$ and $\lambda_1' = \lambda_2'$. By using the standard recipe, we calculate these eigenvalues for general Boolean functions. Below, we give the formulae for $R = R_l$ and $q = q_l (l = 1, \cdots, K)$ for simplicity.

4.1 Input noise case

In the input noise case, we obtain the following formulae:

$$\lambda_1\lambda_1' = \frac{1-3q}{1-q}\int [\prod_l Dt_l] \sum_{\sigma=\pm 1} S^\sigma(\{\tilde{X}_l\})\mathcal{A}^\sigma, \tag{135}$$

$$\lambda_3\lambda_3' = \int [\prod_l Dt_l] \sum_{\sigma=\pm 1} S^\sigma(\{\tilde{X}_l\})\mathcal{B}^\sigma, \tag{136}$$

$$S^\sigma(\tilde{X}_1, \tilde{X}_2) = {\rm Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(\sigma_l \tilde{X}_l), \tag{137}$$

$$\tilde{X}_l = \tilde{\zeta} t_l, \ \tilde{\zeta} = \frac{R}{\sqrt{\eta q - R^2}},$$

$$\mathcal{A}_l^\sigma = \left(\widehat{z_l^2} + (\widehat{z}_l)^2\right)\left(\widehat{z_l^2} + 3(\widehat{z}_l)^2\right), \tag{138}$$

$$\mathcal{B}_l^\sigma \equiv \left(\widehat{z_l^2} + (\widehat{z}_l)^2\right)^2, \tag{139}$$

$$\widehat{z}_l = \frac{h(Y_l)}{\Phi_{\beta,s}^\sigma(\{Y_k\})} {\rm Tr}_{\{\sigma_k\}} \tilde{\triangle}_{s,\beta}^\sigma(\{\sigma_k\})[\prod_{k\neq l} H(\sigma_k Y_k)]\sigma_l, \tag{140}$$

$$\widehat{z_l^2} = -\frac{1}{\Phi^\sigma_{\beta,s}(\{Y_k\})}\mathrm{Tr}_{\{\sigma_k\}}\tilde{\triangle}^\sigma_{s,\beta}(\{\sigma_k\})\prod_k H(\sigma_k Y_k)\sigma_l Y_l|\varphi(\sigma_l Y_l)|,$$

$$= \gamma\frac{K^\sigma_l(\{Y_k\})}{\Phi^\sigma_{\beta,s}(\{Y_k\})}, \tag{141}$$

$$Y_l = \gamma t_l,\ \gamma = \sqrt{\frac{q}{1-q}},\ \varphi(x) = \frac{H'(x)}{H(x)} = -\frac{h(x)}{H(x)}, \tag{142}$$

$$K^\sigma_l(\{Y_k\}) = \mathrm{Tr}_{\{\sigma_k\}}\tilde{\triangle}^\sigma_{s,\beta}(\{\sigma_k\})[\prod_k H(\sigma_k Y_k)]\sigma_l t_l\varphi(\sigma_l Y_l). \tag{143}$$

We define the following quantities to evaluate the AT stability:

$$\Lambda_1 \equiv 2\alpha\lambda_1\lambda_1' - 1, \tag{144}$$

$$\Lambda_3 \equiv 2\alpha\lambda_3\lambda_3' - 1. \tag{145}$$

The condition that a solution is stable is that both Λ_1 and Λ_3 are negative. For the Parity machine with $K = 2$, we explicitly obtain the following:

$$\mathrm{Tr}_{\{\sigma_k\}}\tilde{\triangle}^\sigma_{s,\beta}H(\sigma_m Y_m)\sigma_l = \mathrm{Tr}_{\{\sigma_k\}}\Big(e^{-\beta} + (1-e^{-\beta})\triangle^\sigma_{s,\beta}(\{\sigma_k\})\Big)H(\sigma_m Y_m)\sigma_l$$

$$= (1-e^{-\beta})\mathrm{Tr}_{\{\sigma_k\}}\triangle^\sigma_{s,\beta}(\{\sigma_k\})H(\sigma_m Y_m)\sigma_l$$

$$= (1-e^{-\beta})\sigma H_a(Y_m),\ m \neq l, \tag{146}$$

$$K^\sigma_l(\{Y_k\}) = -\frac{1}{\gamma}(1-e^{-\beta})h(Y_l)Y_l\sigma H_a(Y_m),\ m \neq l, \tag{147}$$

$$\widehat{z_l} = \frac{h(Y_l)}{\Phi^\sigma_{\beta,s}(\{Y_k\})}g_{6,l}(\{t_k\}_{k\neq l}), \tag{148}$$

$$\widehat{z_l^2} = -(1-e^{-\beta})\frac{h(Y_l)Y_l}{\Phi^\sigma_{\beta,s}(\{Y_k\})}\sigma H_a(Y_m),\ m \neq l, \tag{149}$$

$$\mathcal{A}^\sigma_l = \frac{h(Y_l)^2}{[\Phi^\sigma_{\beta,s}(\{Y_k\})]^4}(1-e^{-\beta})^2\Big(H_a(Y_m)\Big)^2$$
$$\times[Y_l\sigma\Phi^\sigma_{\beta,s}(\{Y_k\}) - (1-e^{-\beta})H_a(Y_m)h(Y_l)]$$
$$\times[Y_l\sigma\Phi^\sigma_{\beta,s}(\{Y_k\}) - 3(1-e^{-\beta})H_a(Y_m)h(Y_l)],\ m \neq l, \tag{150}$$

$$\mathcal{B}^\sigma_l = \frac{h(Y_l)^2}{[\Phi^\sigma_{\beta,s}(\{Y_k\})]^4}(1-e^{-\beta})^2\Big(H_a(Y_m)\Big)^2$$
$$\times[Y_l\sigma\Phi^\sigma_{\beta,s}(\{Y_k\}) - (1-e^{-\beta})H_a(Y_m)h(Y_l)]^2,\ m \neq l. \tag{151}$$

$$\Phi_\sigma(Y_1, Y_2) = \Phi^\sigma_{\beta,s}(\{Y_k\}) = e^{-\beta} + (1-e^{-\beta})H_2(Y_1, \sigma Y_2). \tag{152}$$

The condition that a solution is stable is that both Λ_1 and Λ_3 are negative.

4.2 Output noise case

In the output noise case, $\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(\sigma_l \tilde{X}_l)$ and $\tilde{X}_l$ should be replaced by $\lambda + (1-2\lambda)\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(\sigma_l X_l)$ and $X_l = \zeta t_l = \frac{R}{\sqrt{q-R^2}} t_l$, respectively. Thus, for $K = 2$, we obtain the results for the output noise case by replacing $S^\sigma(\tilde{X}_1, \tilde{X}_2)$ and $\tilde{X}_l$ with $S^\sigma(X_1, X_2) = \lambda + (1-2\lambda)H_2(X_1, \sigma X_2)$ and X_l, respectively.

5. Numerical Results for Parity Machine

To begin with, we study learning behaviors for small α. First, we study the α dependences of R_l for $T = 1$. We performed MCMCs using the simulated annealing method.

Here, we explain the numerical method of MCMCs for Figs. 1-9. The same method is used unless otherwise stated. Starting from $\beta = 0.1$, when the system reached its equilibrium, β was increased in steps of 0.1 up to 1. We adopted 100 samples. For each sample, after 100 Monte Carlo sweeps (MC sweeps), we started taking thermal averages. Here, one MC sweep corresponds to $M \times K$ updates. From 100 MC sweeps, the running average values of R_l were calculated every 10 MC sweeps. After 150 MC sweeps, we started checking the convergence. If the difference between the average of R_1 up to k MC sweeps and that up to $k + 10$ MC sweeps was less than 0.001, we considered that the system had reached its equilibrium.

We performed simulations for several values of M, e. g., $M = 10, 20$, and 30 for the noiseless, input noise, and output noise cases, and found that the results for $M = 20$ and 30 are almost the same. Furthermore, we found that the simulation results for R_1 and R_2 are almost the same. Therefore, below we display the results of R_1 for $M \geq 20$. We set parameters so that $\epsilon_{\min}[\equiv \epsilon_g(R = 1)]$ is 5/18 for noisy cases if not otherwise specified. That is, for the input noise $\tau = 1/\sqrt{3}$ and the output noise $\lambda = 5/18$. In Figs. 3-5, we display the α dependences of R_1 and the generalization error ϵ_g for the noiseless, input noise, and output noise cases, respectively.

First of all, we note that the error bars are larger for the output noise case than for the noiseless and input noise cases. This is because the "strength" of noise A_{noise} is 9/4 for the output noise, while it is 0 and 4/3 for the noiseless and output noise cases, respectively. We note that for large α, the theoretical and numerical results agree reasonably well. However, for small α, where learning does not take place, the simulation

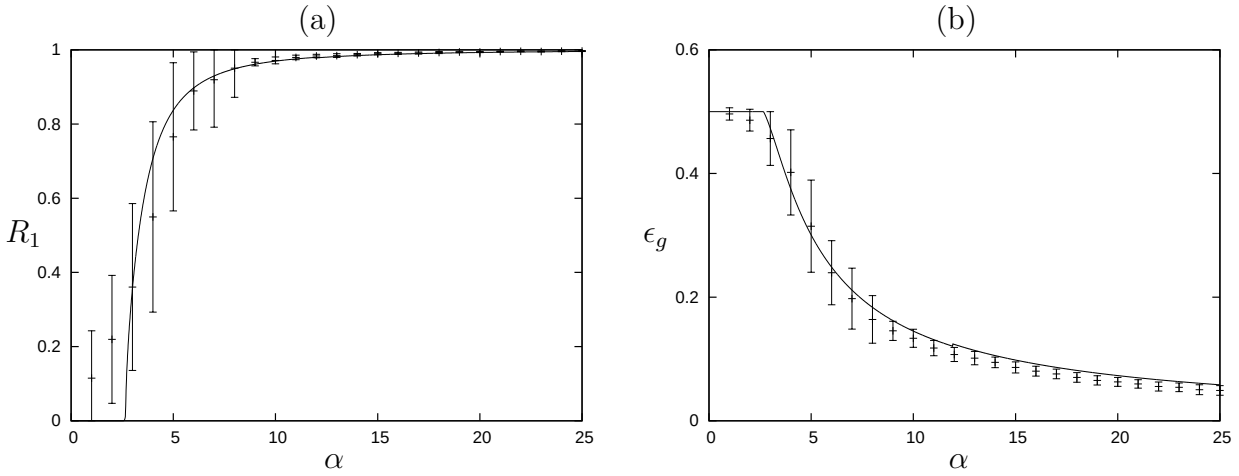

Fig. 3. α dependences of R_1 and the generalization error ϵ_g for noiseless case. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Averages are taken from 100 samples. Vertical lines are error bars. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. (a) R_1, (b) ϵ_g.

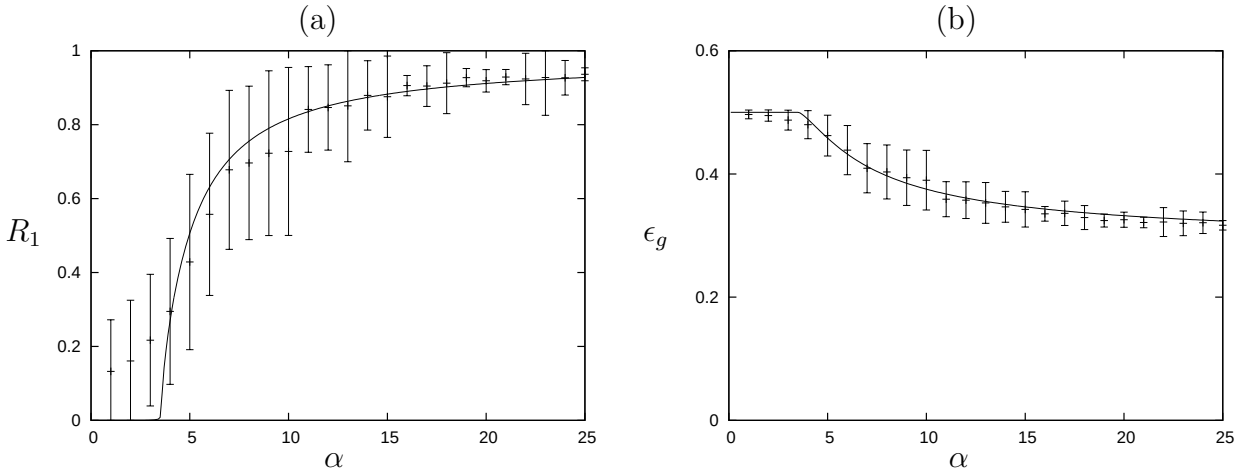

Fig. 4. α dependences of R_1 and the generalization error ϵ_g for case of input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Input: $\tau = 1/\sqrt{3}$. (a) R_1, (b) ϵ_g.

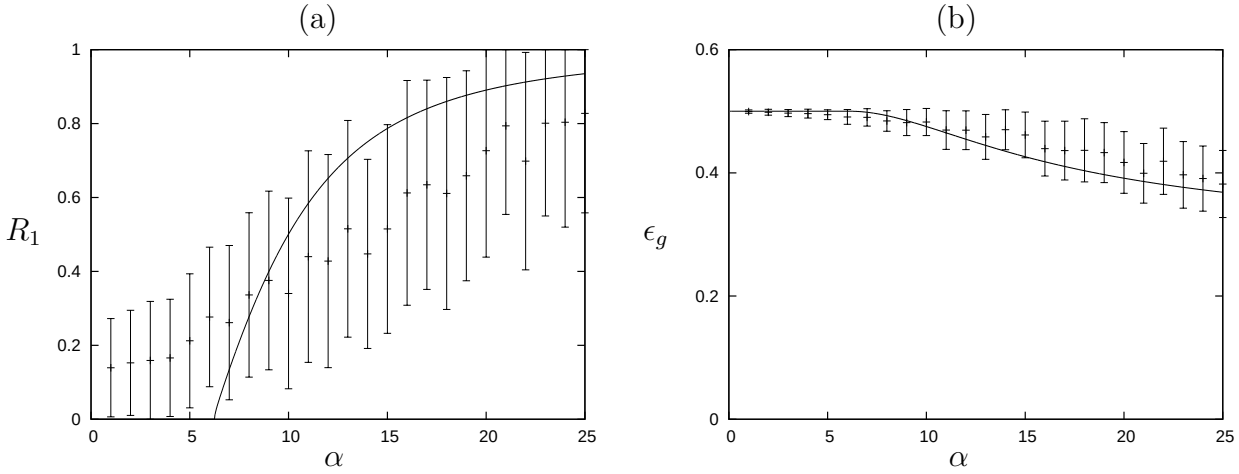

Fig. 5. α dependences of R_1 and the generalization error ϵ_g for case of output noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. $\lambda = 5/18$. (a) R_1, (b) ϵ_g.

results differ from theoretical values. The numerical data are all positive, whereas they are theoretically expected to be 0. . This is because we numerically calculated the absolute value of R_1, $|R_1|$, since the system has reflectional symmetry. That is, the output σ is the same for $\{\boldsymbol{w}_l\}$ and $\{-\boldsymbol{w}_l\}$. In order to investigate the value of $\alpha_{\rm C}$, we calculated the histograms of R_1 at small values of α (see Fig. 6). We observe that the value of R_1 that gives the maximum frequency, $R_{\rm max}$, changes from 0 to nonzero values as α increases. In Fig. 7, we display the α dependences of $R_{\rm max}$ and the theoretical results of $R_1(\alpha)$. This confirms the existence of the nonzero $\alpha_{\rm C}$, although there still exists a difference between the numerical and theoretical results. We consider that this is a finite size effect. In Figs. 8 and 9, we display the $\ln(\Delta\alpha)$ dependences of $\ln R_{\rm max}$ and $\ln\Delta\epsilon$ for $\alpha > \alpha_{\rm C}$, respectively. Here, $\Delta\alpha = \alpha - \alpha_{\rm C}$ and $\Delta\epsilon = \epsilon_g|_{R=0} - \epsilon_g$. Dotted lines denote the theoretical exponents $\bar{\beta}_R = 1/2$ and $\bar{\beta}_\epsilon = 1$. Numerical data are scattered but consistent with the theoretical data.

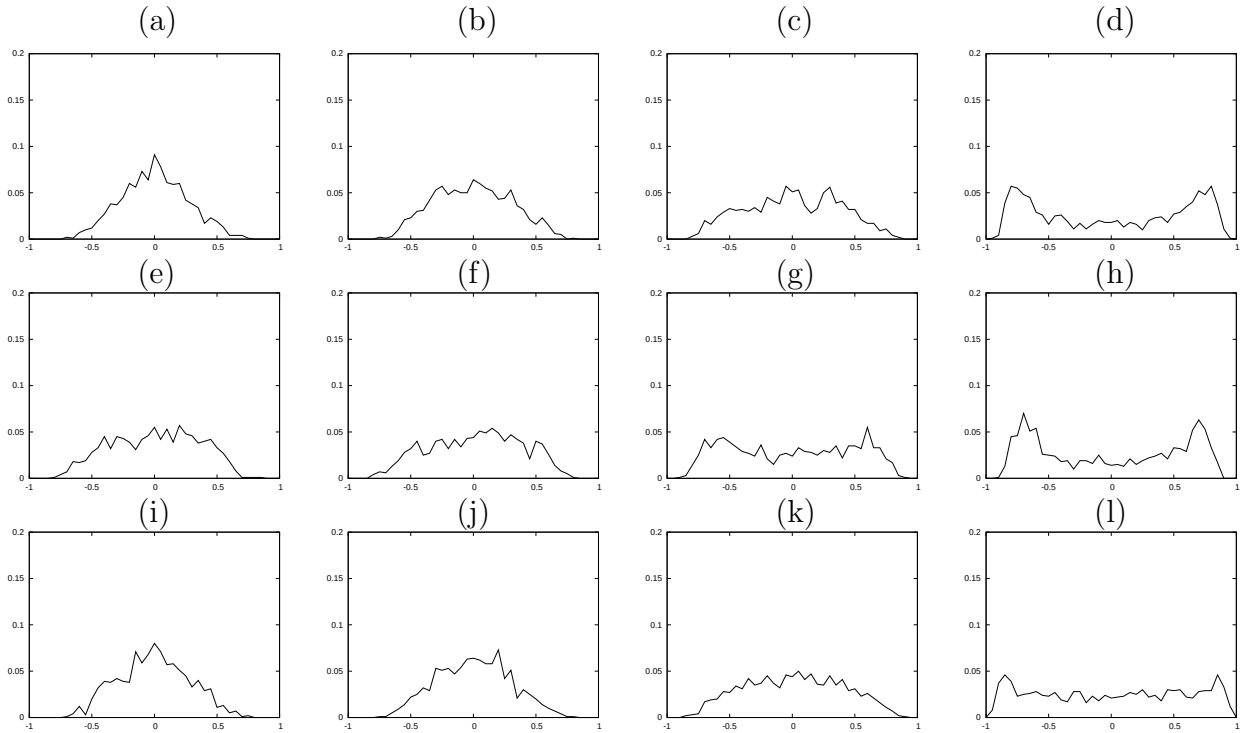

Fig. 6. Histograms of R_1 observed by MCMCs. $M = 20$, number of samples $= 1000, T = 1$. (a)-(d): Noiseless, $\alpha_{\rm C} = 2.66$. (a) $\alpha = 2$, (b) $\alpha = 2.5$, (c) $\alpha = 3$, (d) $\alpha = 4$. (e)-(h): Input noise, $\alpha_{\rm C} = 3.56, \tau = 1/\sqrt{3}$. (e) $\alpha = 3.5$, (f) $\alpha = 4$, (g) $\alpha = 5$, (h) $\alpha = 6$. (i)-(l): Output noise, $\alpha_{\rm C} = 6.01, \lambda = 5/18$. (i) $\alpha = 5$, (j) $\alpha = 6$, (k) $\alpha = 8$, (l) $\alpha = 14$.

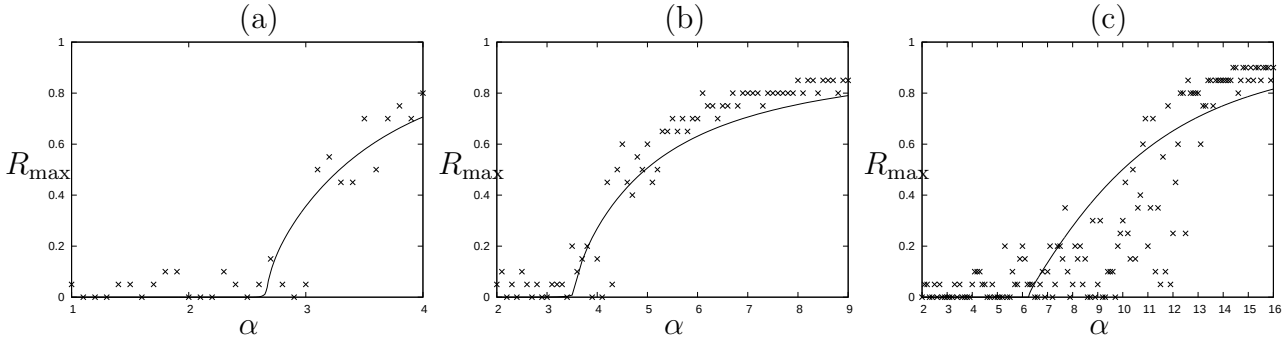

Fig. 7. α dependences of $R_{\max}$ for noiseless, input noise and output noise cases.
Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. $\epsilon_{\min} = 5/18$ for noisy case. (a) Noiseless, (b) input noise, $\tau = 1/\sqrt{3}$, (c) output noise, $\lambda = 5/18$.

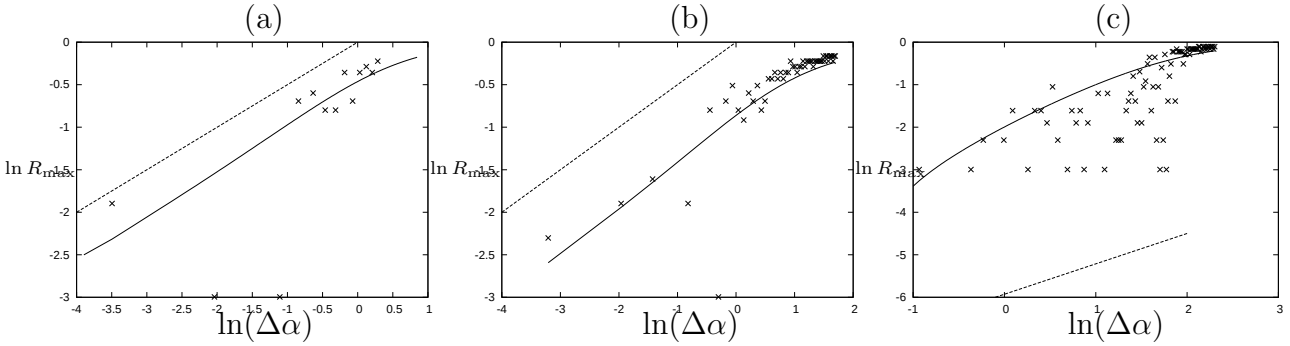

Fig. 8. $\ln \Delta\alpha$ dependences of $\ln R_{\max}$ for noiseless, input noise, and output noise cases. $\Delta\alpha = \alpha - \alpha_{\rm C}$.
Curves: RS solutions, symbols: MCMCs, dashed lines: theoretical value $\bar{\beta}_R = 0.5$. $M = 20$, $T = 1, \epsilon_{\min} = 5/18$. (a) Noiseless, (b) input noise, $\tau = 1/\sqrt{3}$, (c) output noise, $\lambda = 5/18$.

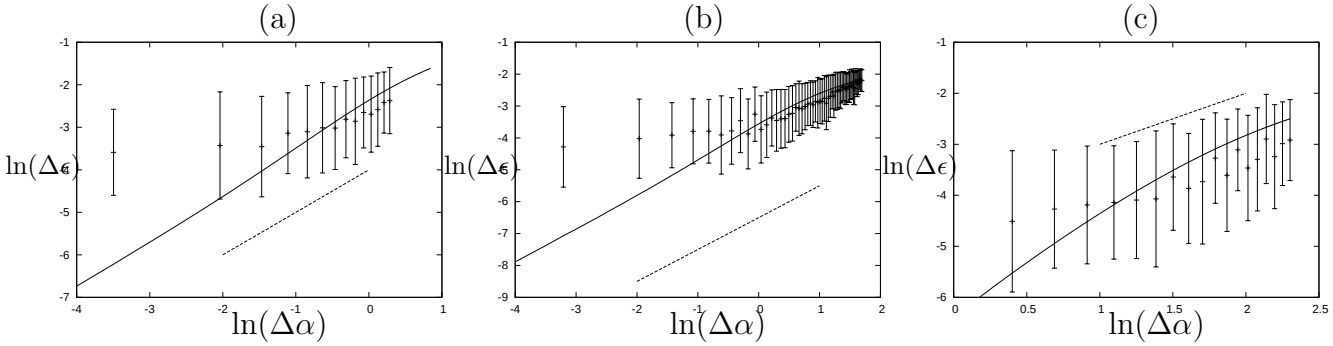

Fig. 9. $\ln \Delta\alpha$ dependences of $\ln \Delta\epsilon$ for noiseless, input noise, and output noise cases. $\Delta\alpha = \alpha - \alpha_c, \Delta\epsilon = \epsilon_g|_{R=0} - \epsilon_g$. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines: theoretical value $\bar{\beta}_\epsilon = 1$. (a) Noiseless, (b) input noise, $\tau = 1/\sqrt{3}$, (c) output noise, $\lambda = 5/18$.

Next, we study the temperature dependences of order parameters by performing EXMCs. In the EXMCs, for example, we prepared an odd number of temperatures, $T_1, T_2, \cdots, T_{2L+1}$. Every five MC sweeps, we exchanged temperatures T_1 and T_2, T_3 and $T_4, \cdots, T_{2L-1}$ and T_{2L}, or T_2 and T_3, T_4 and $T_5, \cdots, T_{2L}$ and T_{2L+1}. At each temperature, we calculated the averages of R_1 and q_1 after a transient of 10^4 MC sweeps. From 10^4 MC sweeps to 5×10^4 or 10^5 MC sweeps, we took data every 100 MC sweeps, and calculated the averages. We took the sample average for 100 samples. We use the same method in Fig. 10-13 unless otherwise stated. In Fig. 10, we display the theoretical and numerical results for the noiseless case for $\alpha = 10$. The numerical results confirm the validity of the theoretical results, at least for $T \leq 3.5$. We confirmed that both results agree for all temperatures when $\alpha = 30$ as long as learning takes place. However, there exists disagreement between the theoretical and numerical results in the paramagnetic region of $T \geq 4$ when $\alpha = 10$. That is, the numerical values of R_1 are positive at high temperatures despite the theoretical prediction of zero. The reason for this is that we took the average of the absolute value of R_1, $|R_1|$. In Fig. 11, we display the results for the input noise case with $\tau = \frac{\sqrt{5}}{2}$ for $\alpha = 10$ and those for the output noise case with $\lambda = \frac{5}{18}$ for $\alpha = 10$ and 30 in Figs. 12 and 13. The values of τ and λ correspond to $A_{\text{noise}} = 2.25$. As is seen in Figs. 11, 12 and 13, in addition to the paramagnetic region, there exists disagreement between theoretical and numerical results in the SG region in both the input and output noise cases. That is, we numerically obtained $R_1 > 0$ in the regions where the SG was expected theoretically for noisy cases. The discrepancy between numerical and theoretical results in the low-temperature region suggests that replica symmetry breaking (RSB) takes place at low temperatures as in the case of the learning of one-layer perceptrons with spherical weights.[3,4]

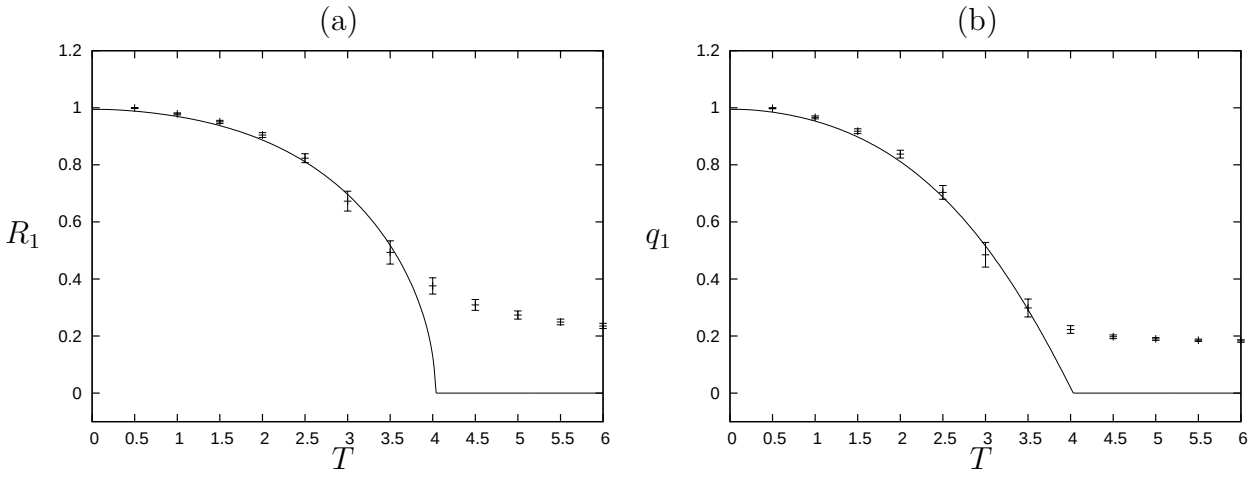

Fig. 10. T dependences of R_1 and q_1 for noiseless case. Curves: RS solutions, symbols: EXMCs. $M = 20$. Averages are taken from 100 samples. Vertical lines are error bars. (a) R, (b) q_1 for $\alpha = 10$.

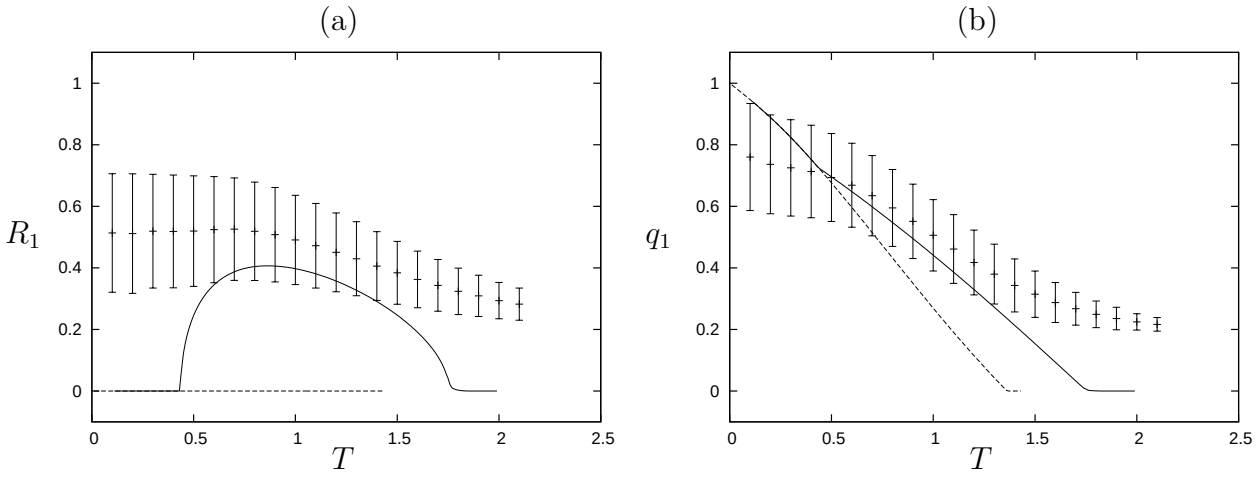

Fig. 11. T dependences of R_1 and q_1 for input noise. Curves: RS solutions, symbols: EXMCs. $M = 20$. $\tau = \sqrt{5}/2, A_{\rm noise} = 2.25$. (a) R_1, (b) q_1 for $\alpha = 10$.

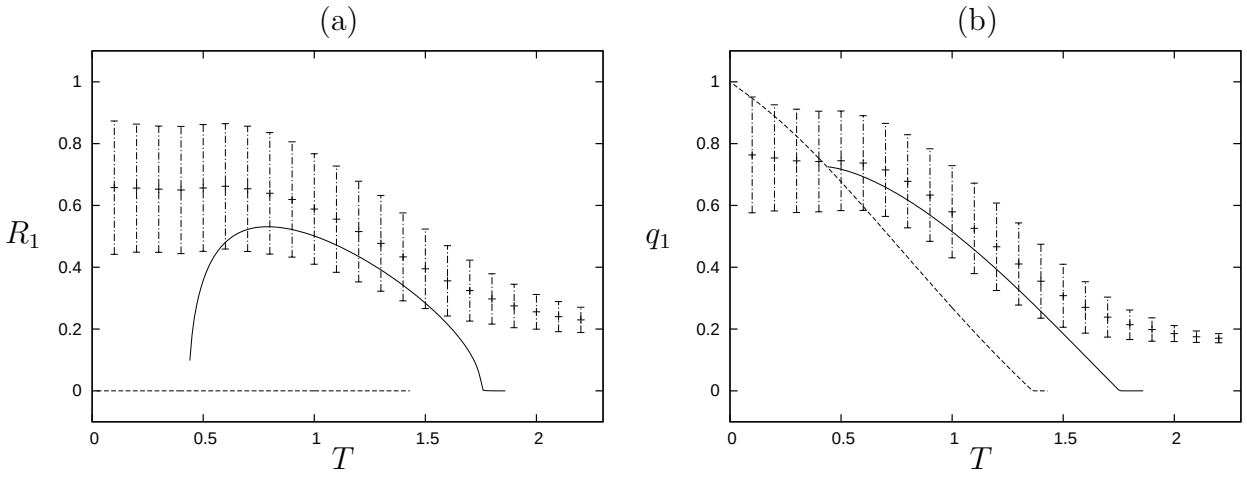

Fig. 12. T dependences of R_1 and q_1 for output noise. Curves: RS solutions, symbols: EXMCs. $\lambda = 5/18, A_{\rm noise} = 2.25$. (a) R, (b) q_1 for $\alpha = 10$, $M = 30$.

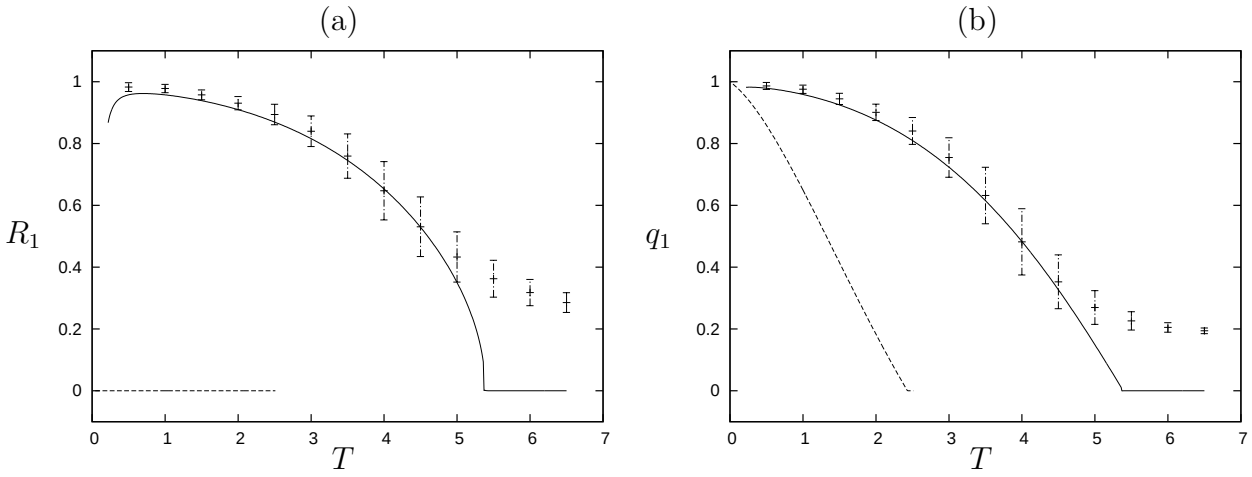

Fig. 13. T dependences of R_1 and q_1 for output noise. Curves: RS solutions, symbols: EXMCs. $\lambda = 5/18, A_{\rm noise} = 2.25$. (a) R, (b) q_1 for $\alpha = 30$, $M = 20$.

To resolve the discrepancy, we study the AT stability. We display Λ_1, Λ_3, and the free energy per input in Figs. 14-16 for the input noise case and in Figs. 17-22 for the output noise case. The free energy of the L solution is lower than that of the SG solution for all cases. That is, the SG solution is metastable even if it is AT stable. As the temperature is lowered, the L solution becomes AT unstable and then the SG solution becomes AT unstable in the input noise and output noise cases for $\alpha = 10$. In contrast, for $\alpha = 30$ in the case of output noise, the SG solution first becomes AT unstable and then the L solution becomes unstable as the temperature is lowered. Thus, in all the cases we studied, an RSB L solution is expected to appear. Therefore, in learning under the existence of noise, we conclude that learning takes place even in the limit of $T \to 0$.

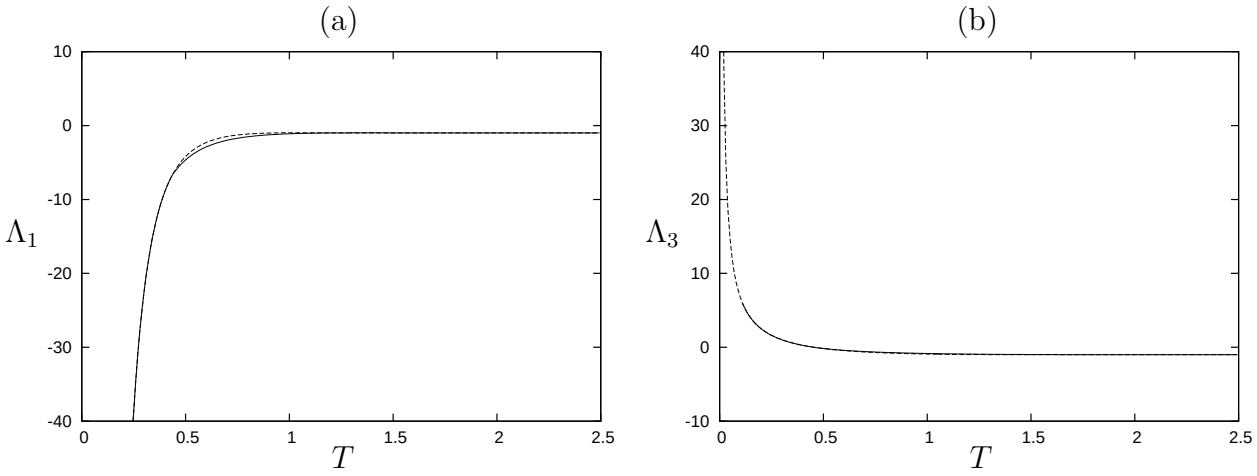

Fig. 14. Temperature dependences of Λ_1 and Λ_3 for case of input noise. $\tau = \sqrt{5}/2, \alpha = 10$. Solid curves: L solution, dashed curves : SG solution, dotted lines: zero line. (a) Λ_1, (b) Λ_3.

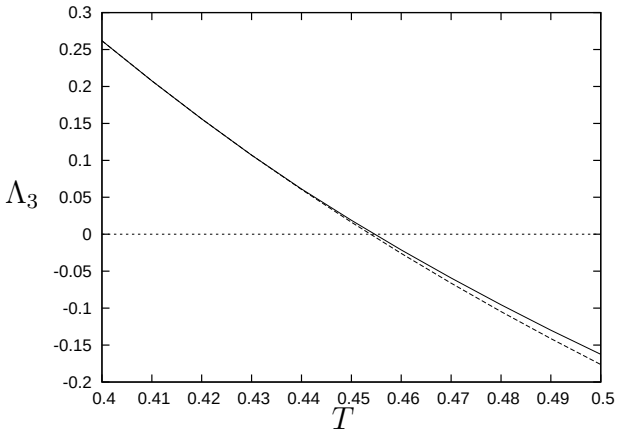

Fig. 15. Temperature dependences of Λ_3 [Enlargement of Fig. 13(b)] for case of input noise. $\tau = \sqrt{5}/2, \alpha = 10$.

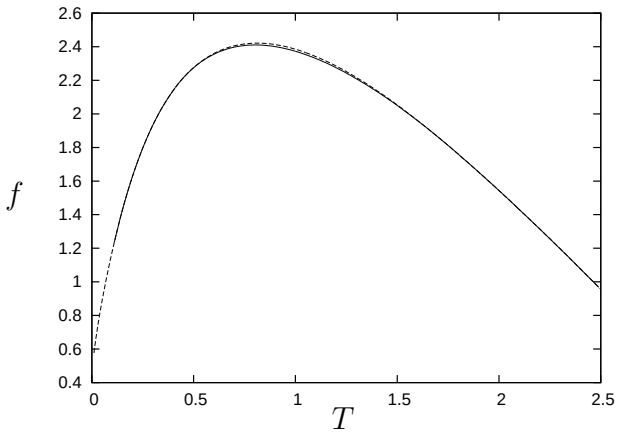

Fig. 16. Temperature dependences of free energy per input for case of input noise. $\tau = \sqrt{5}/2$. Solid curve: L solution, dashed curve: SG solution. $\alpha = 10$.

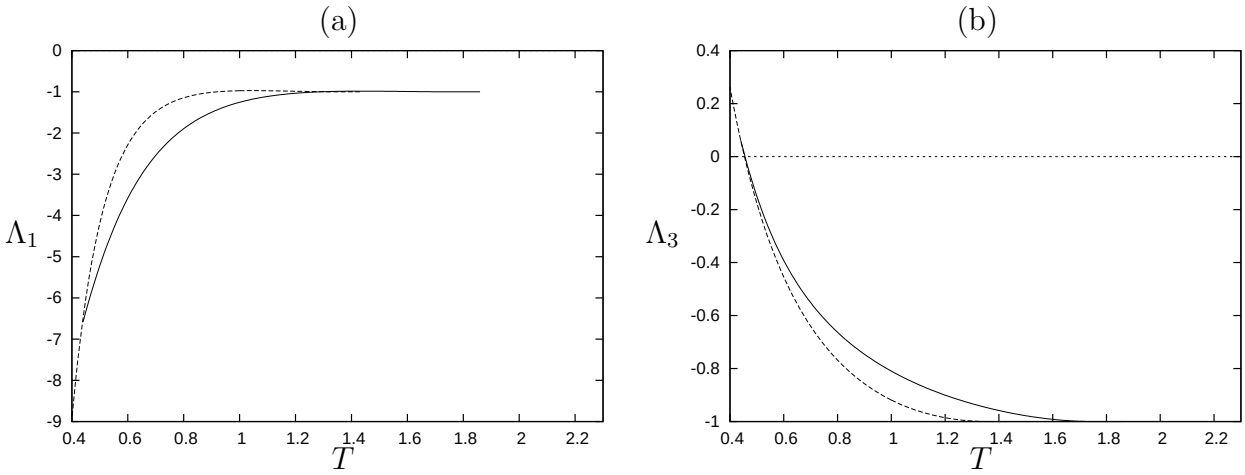

Fig. 17. Temperature dependences of Λ_1 and Λ_3 for case of output noise. $\lambda = 5/18, \alpha = 10$. Solid curves: L solution, dashed curves: SG solution, dotted lines: zero line. (a) Λ_1, (b) Λ_3.

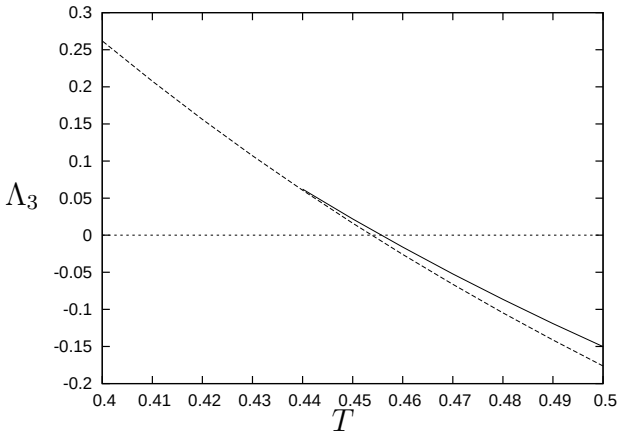

Fig. 18. Temperature dependences of Λ_3 for case of output noise [Enlargement of Fig. 16(b)]. $\lambda = 5/18, \alpha = 10$. Λ_3.

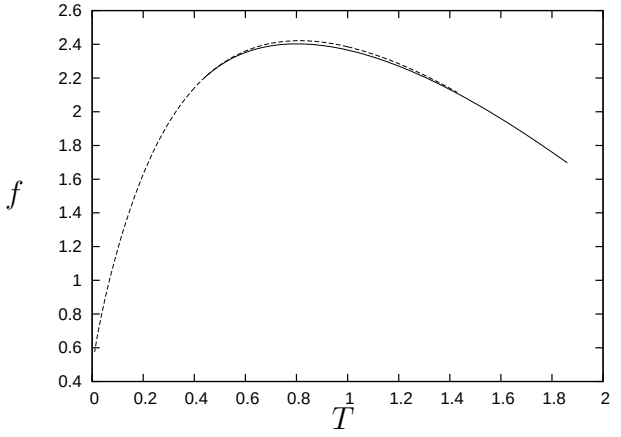

Fig. 19. Temperature dependences of free energy per input. $\lambda = 5/18, T = 1$. Solid curve: L solution, dashed curve: SG solution. $\alpha = 10$.

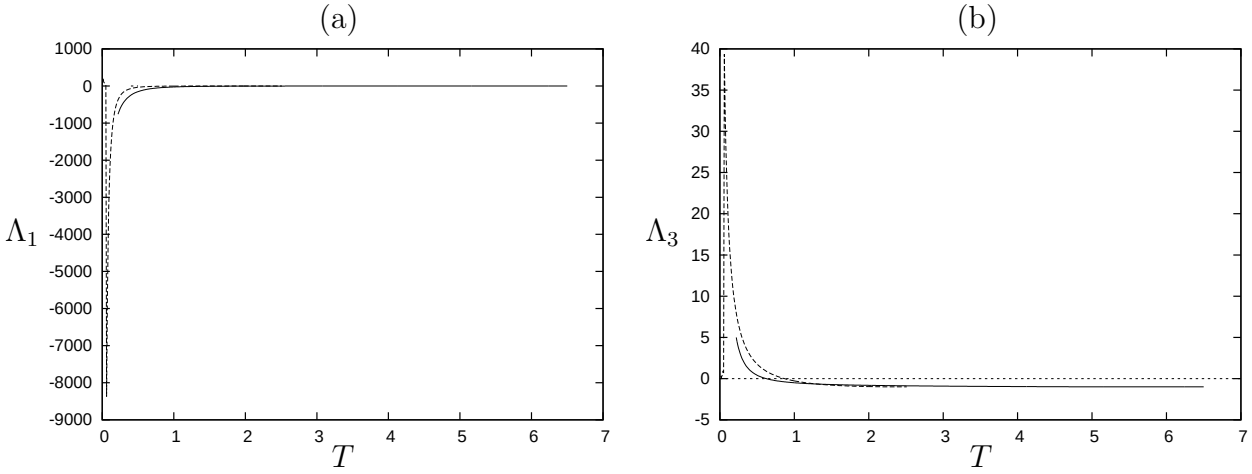

Fig. 20. Temperature dependences of Λ_1 and Λ_3 for case of output noise. $\lambda = 5/18, \alpha = 30$. Solid curves: L solution, dashed curves: SG solution, dotted lines: zero line. (a) Λ_1, (b) Λ_3.

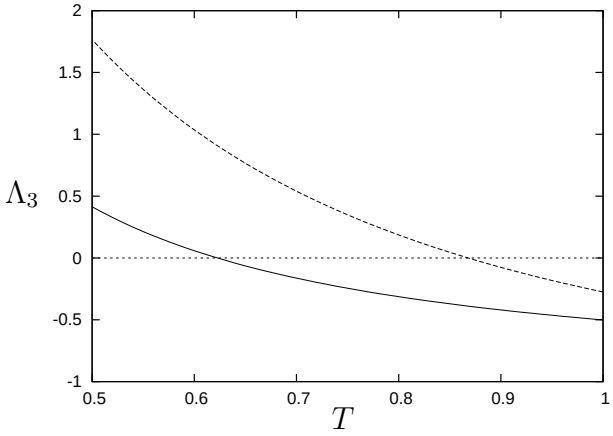

Fig. 21. Temperature dependences of Λ_3 for case of output noise [Enlargement of Fig. 16(b)]. $\lambda = 5/18, \alpha = 30$.
Λ_3.

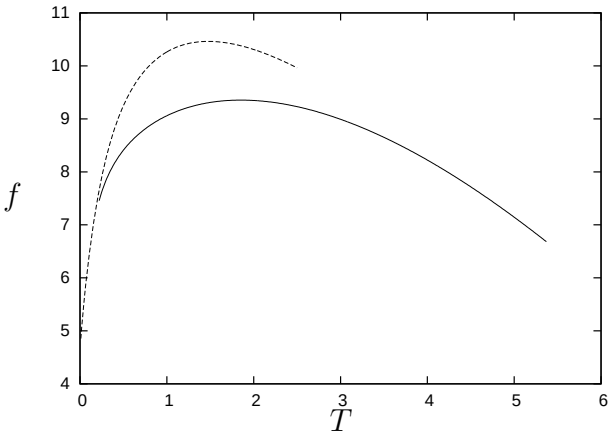

Fig. 22. Temperature dependences of free energy per input. $\lambda = 5/18, T = 1$. Solid curve: L solution, dashed curve: SG solution. $\alpha = 30$.

The value of $A_{\rm noise} = 2.25$ that we studied is rather small. Let us study the $A_{\rm noise}$ dependences of learning. In Fig. 23, we display the α dependences of R_1 and q_1 for a large value of noise, $\lambda = 0.4$, and in Fig. 24, the λ dependences of R_1 and q_1 for $\alpha = 35$ for several temperatures in the output noise case. From Fig. 23, it seems that as α increases, although the distributions of R_1 and q_1 become broad, learning does not take place for $\lambda = 0.4$. This is more clearly observed in Fig. 24. In these figures, we took the sample averages not only changing the initial conditions of the student but also the examples and teacher's outputs. If we fix the examples and the teacher's output, we obtain the results shown in Fig. 25.

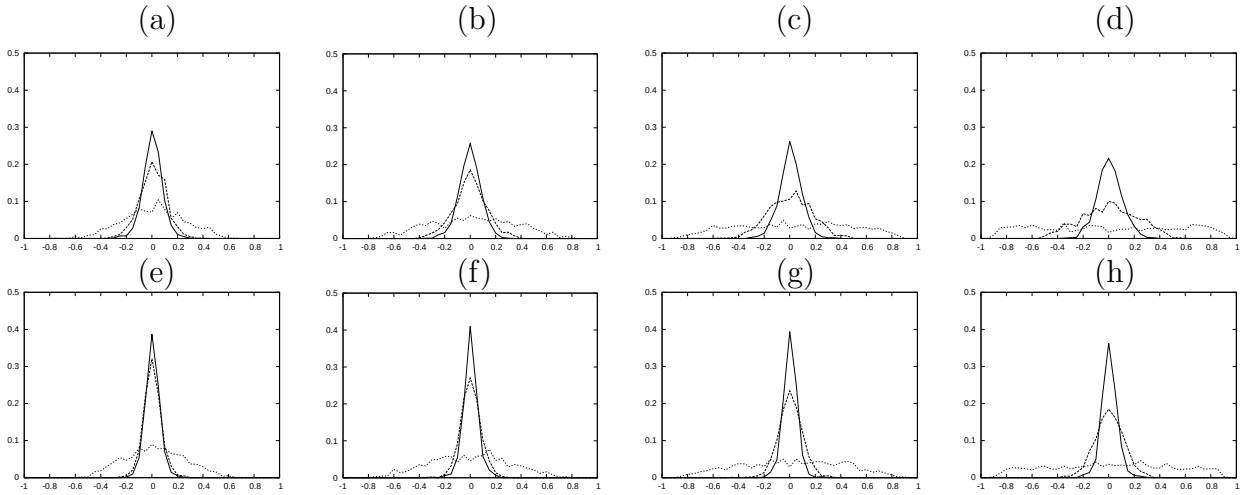

Fig. 23. α dependences of histograms of R_1 and q_1 obtained by MCMCs for case of output noise. $\lambda = 0.4, M = 30$, number of samples $= 1000$. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 0.5$. The convergence condition is $d_c = 10^{-5}$. Solid line: $T = 10$, dashed line: $T = 5$, dotted line: $T = 2$. Acceptance ratios at $T = 2 : 38\%$ $(\alpha = 25), 35\%$ $(\alpha = 35), 24\%$ $(\alpha = 45), 19\%$ $(\alpha = 55)$. (a)-(d): R_1, (e)-(h): q_1. (a), (e): $\alpha = 25$, (b), (f): $\alpha = 35$, (c), (g): $\alpha = 45$, (d), (h): $\alpha = 55$.

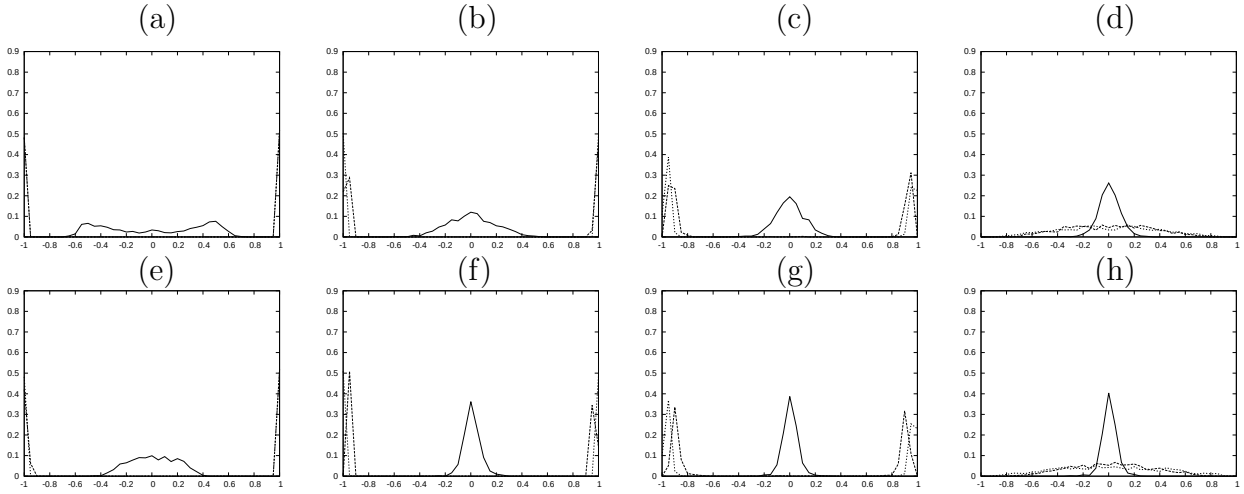

Fig. 24. λ dependences of histograms of R_1 and q_1 obtained by MCMCs for case of output noise. $\alpha = 35, M = 30$, number of samples = 1000. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. The convergence condition is $d_c = 10^{-5}$. Solid lines: $T = 10$, dashed lines: $T = 2$, dotted lines: $T = 1$. Acceptance ratios at $T = 1 : 5.6\%$ ($\lambda = 0.1$), 6.0% ($\lambda = 0.2$), 5.4% ($\lambda = 0.3$), 4.5% ($\lambda = 0.4$). (a)-(d): R_1, (e)-(h): q_1. (a), (e): $\lambda = 0.1$, (b), (f): $\lambda = 0.2$, (c), (g): $\lambda = 0.3$, (d), (h): $\lambda = 0.4$.

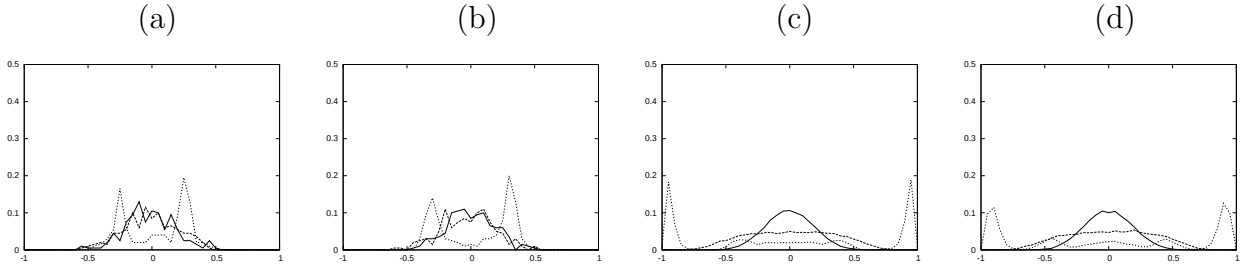

Fig. 25. Histograms of R_1, R_2, q_1 and q_2 at 100000 MCS for case of output noise. Number of initial conditions is 100x2. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. Solid lines: $T = 10$, dashed lines: $T = 2$, dotted lines: $T = 1$. $\lambda = 0.4, \alpha = 25, M = 30$. (a) R_1, (b) R_2, (c) q_1, (d) q_2.

We can observe the peak at $q_1 \simeq 1$ and $q_2 \simeq 1$ for $T = 1$. Next, we display the theoretical and numerical results in Fig. 26 for the output noise case. We found that there are three non-trivial solutions, the 1st L solution (L_1), the 2nd L solution (L_2), and the SG solution. We denote the L_1 solution as the solution that bifurcates from the SG solution and the L_2 solution as the solution that bifurcates from the para solution. We next study the AT stability and the free energy. The results are shown in Figs. 27 and 28, respectively. From these figures, we note that the L_1, L_2, and SG solutions are stable but the free energy of the L_2 solution is lower than that of the L_1 solution and the SG solution has the largest free energy. However, as is seen from Fig. 26, although

the L_2 solution agrees with the simulation results at higher λ, at lower λ the L_1 solution agrees with the simulation results. Therefore, here, there appears to be disagreement between the theoretical and numerical results. We consider the reason for this as follows. As written in the caption of Fig. 25, the acceptance ratio of the Metropolis method is only about 5 percent. This value is too low to reach the equilibrium and the system might be trapped to a local minimum of the free energy.

Aside from this disagreement, we conclude that learning takes place for the input noise case with any noise amplitude and for the output noise case when the probability that the teacher's output is reversed is less than one-half.

Now, let us study the asymptotic behavior of learning as $\alpha \to \infty$. In Figs. 29-34, we display the results obtained by MCMCs for the noiseless, input noise, and output noise cases. The method of MCMCs is the same as before. We display the α dependences of R_1 and q_1 and the $\ln(\alpha)$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for $\alpha \in [25, 50]$ in the noiseless case and for $\alpha \in [100, 200]$ in the noisy cases. Here, $\Delta R_1 = 1 - R_1$ and $\Delta q_1 = 1 - q_1$. The method used is simulated annealing with a mall α.The numerical results are consistent with the theoretically obtained exponents $\hat{\beta}_R$ and $\hat{\beta}_q$, as is seen from Figs. 26, 28, and 30.

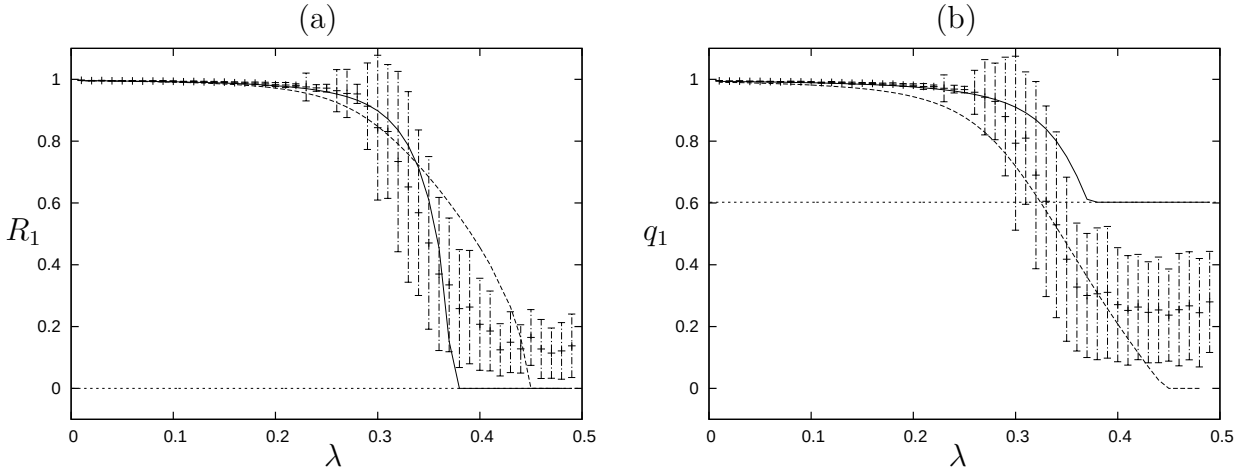

Fig. 26. λ dependences of R_1 and q_1 for case of output noise. Curves: RS solutions, solid curve: L_1 solution, dashed curve: L_2 solution, dotted curve: SG solution. Symbols: MCMCs, $M = 40$, $T = 1, \alpha = 25$. Averages are taken from 100 samples. Vertical lines are error bars. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. (a) R_1, (b) q_1.

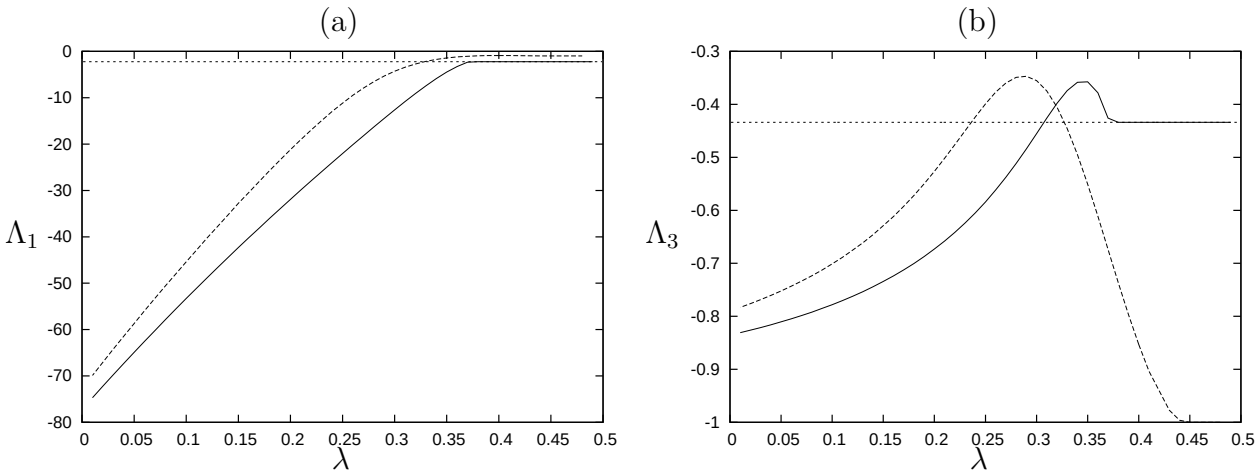

Fig. 27. λ dependences of Λ_1 and Λ_3 for case of output noise. $\alpha = 25$. Curves: RS solutions, solid curve: L_1 solution, dashed curve: L_2 solution, dotted curve: SG solution. (a) Λ_1, (b) Λ_3.

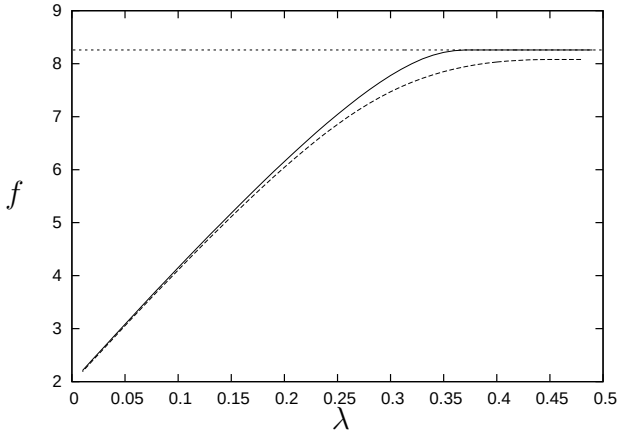

Fig. 28. λ dependence of free energy per input for case of output noise. $\alpha = 25$. Curves: RS solution, solid curve: L_1 solution, dashed curve: L_2 solution, dotted curve: SG solution.

Next, we study

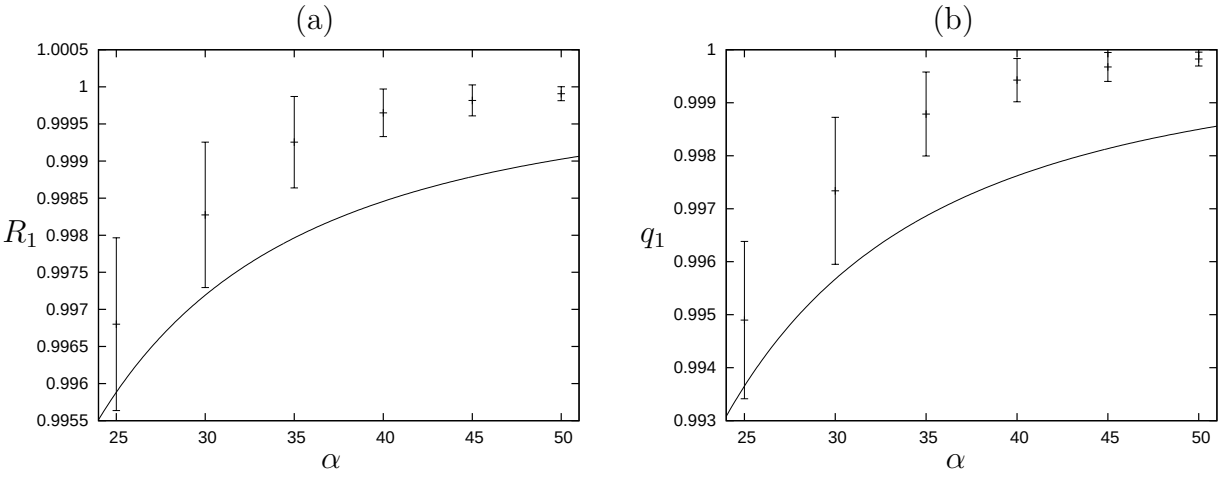

Fig. 29. α dependences of R_1 and q_1 for noiseless case. Curves: RS solutions, symbols: MCMCs. $M = 20, T = 1$. Averages are taken from 100 samples. Vertical lines are error bars. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. (a) R_1, (b) q_1.

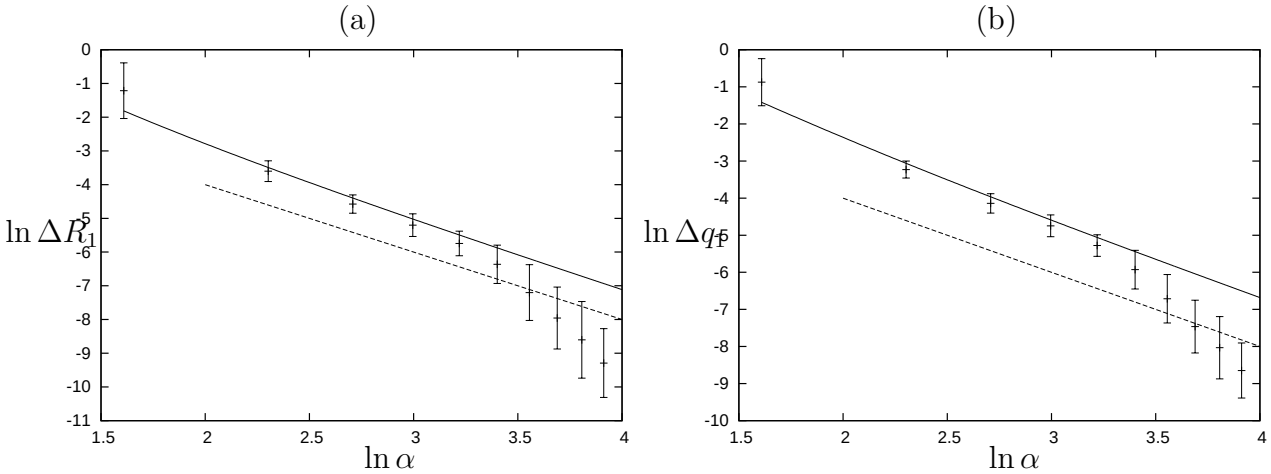

Fig. 30. $\ln \alpha$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for noiseless case. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R = 2, \hat{\beta}_q = 2$. (a) $\ln \Delta R_1$, (b) $\ln \Delta q_1$.

6. And Machine

Next, we study the And machine. As in the case of the Parity machine, we consider two units, that is $K = 2$. The output by the And machine is 1 if outputs by two units are 1, and -1 otherwise. In the below, we show the SPEs. The derivations are given in Appendix C. As in the Parity machine, we assume $R_l = R$ and $q_l = q$ for $l = 1, 2$. $\tilde{X}_l, X_l, Y_l, \gamma, \tilde{\zeta}$ and ζ are the same as those for the Parity machine.

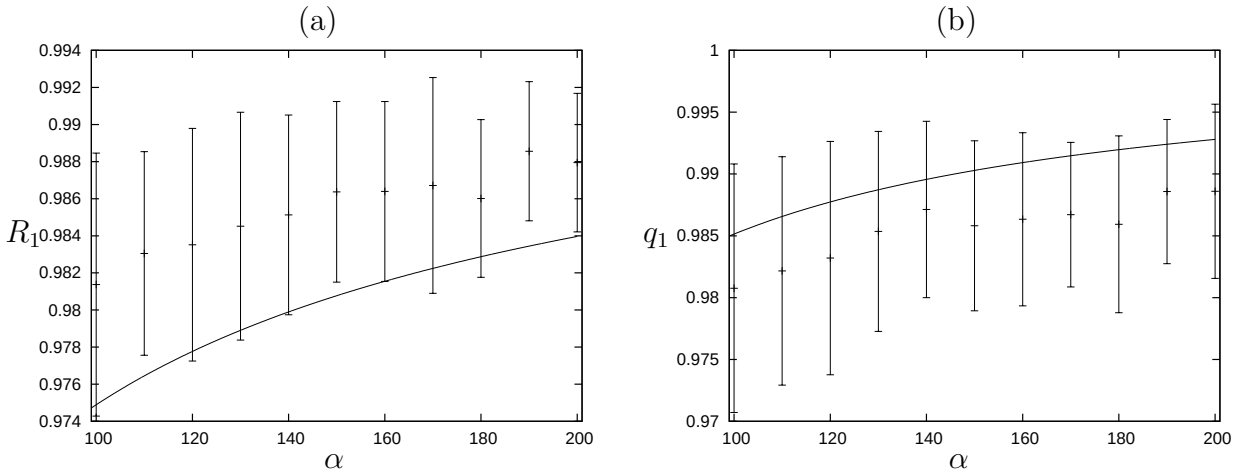

Fig. 31. α dependences of R_1 and q_1 for case of input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1, \tau = 1/\sqrt{3}$. (a) R_1, (b) q_1.

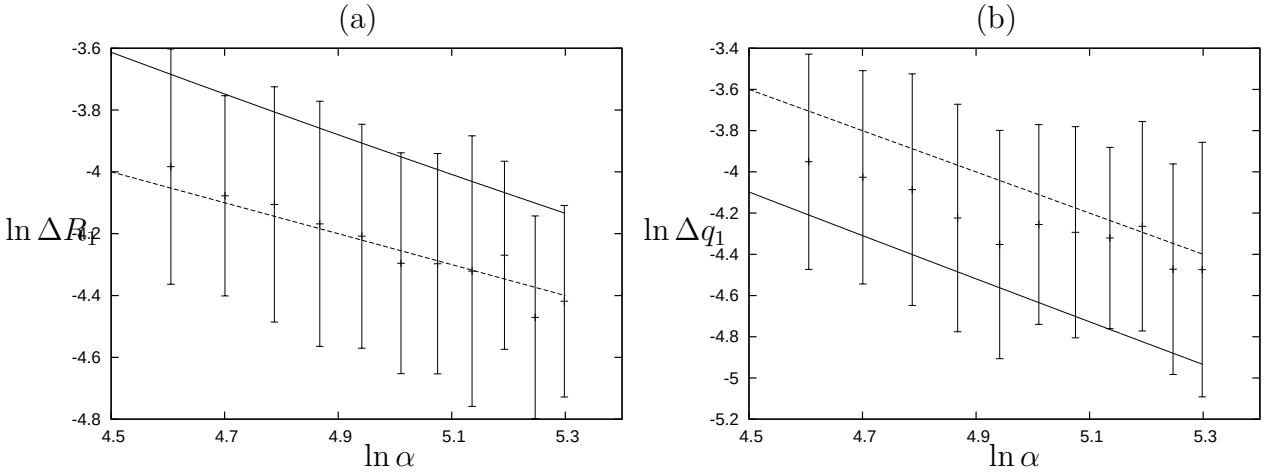

Fig. 32. $\ln \alpha$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for case of input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R = 1/2, \hat{\beta}_q = 1$. (a) $\ln \Delta R_1$, (b) $\ln \Delta q_1$.

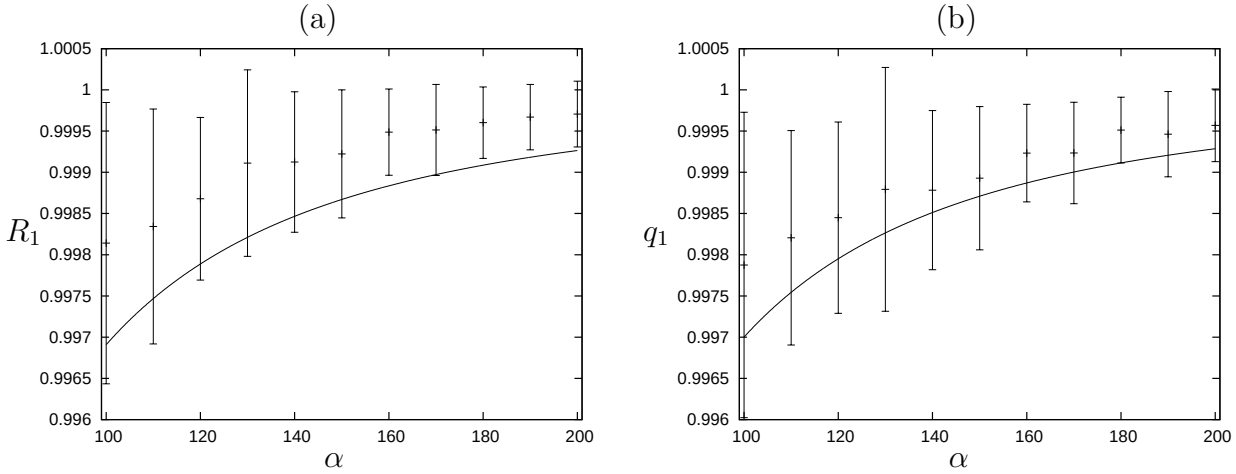

Fig. 33. α dependences of R_1 and q_1 for case of output noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1, \lambda = 5/18$. (a) R_1, (b) q_1.

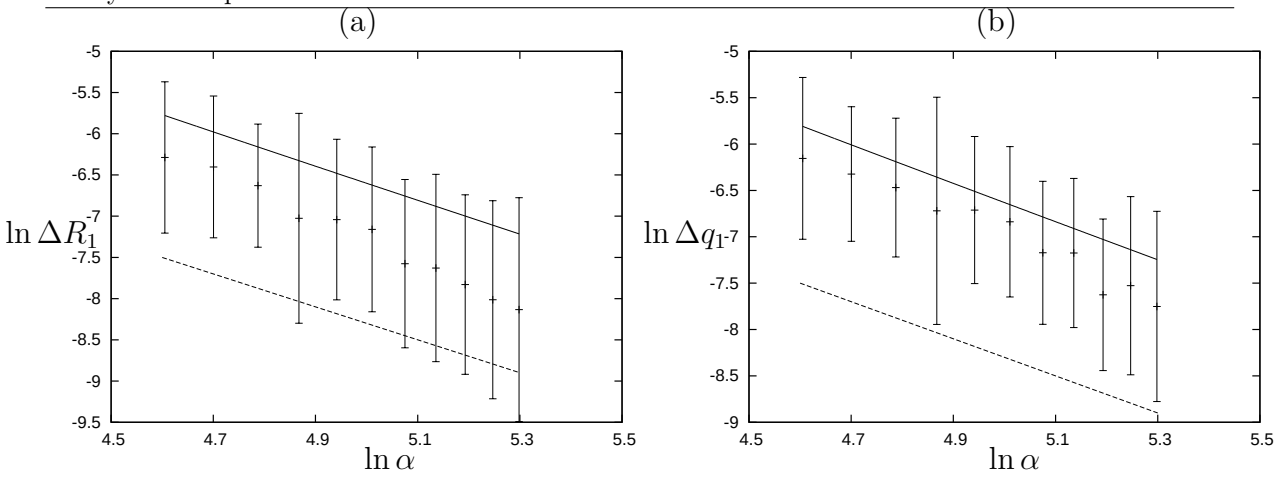

Fig. 34. $\ln\alpha$ dependences of $\ln\Delta R_1$ and $\ln\Delta q_1$ for case of output noise. Curves: RS solutions, symbols: MCMCs. $M=20, T=1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R=2, \hat{\beta}_q=2$. (a) $\ln\Delta R_1$, (b) $\ln\Delta q_1$.

6.1 Input noise

$\frac{G_1}{n}$ is

$$\frac{G_1}{n}=\int Dt_1Dt_2\{H(\tilde{X}_1)H(\tilde{X}_2)\ln\Phi_+(Y_1,Y_2)+\hat{H}(\tilde{X}_1,\tilde{X}_2)\ln\Phi_-(Y_1,Y_2)\}, \tag{153}$$

where $\Phi_+(Y_1,Y_2) = H_2(Y_1,Y_2) + e^{-\beta}\hat{H}(Y_1,Y_2) = e^{-\beta} + (1-e^{-\beta})H(Y_1)H(Y_2)$, $\Phi_-(Y_1,Y_2)=1+e^{-\beta}-\Phi_+(Y_1,Y_2)$, and $\hat{H}(Y_1,Y_2)=1-H(Y_1)H(Y_2)$.

$\frac{G_2}{n}$ is

$$\frac{G_2}{n}=\frac{1}{2}\Big(\ln(2\pi(1-q))+\frac{1-R^2}{1-q}\Big). \tag{154}$$

The SPEs are given by

$$q^2-R^2 = \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{i})}, \tag{155}$$

$$R(\eta q-R^2)^{\frac{3}{2}} = -\alpha\eta q(1-q)J_2^{(\mathrm{i})} \tag{156}$$

$$\begin{aligned} J_1^{(\mathrm{i})} &= -2\int Dt_1Dt_2t_1h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\ &\quad\times\{\Phi_+(Y_1,Y_2)-(1+e^{-\beta})H(\tilde{X}_1)H(\tilde{X}_2)\}, \end{aligned} \tag{157}$$

$$J_2^{(\mathrm{i})} = -2\int Dt_1Dt_2h(\tilde{X}_1)H(\tilde{X}_2)t_1\ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). \tag{158}$$

6.2 Output noise

$\frac{G_1}{n}$ is

$$\begin{aligned} \frac{G_1}{n} = \int Dt_1Dt_2\Big[&\big\{\lambda+(1-2\lambda)H(X_1)H(X_2)\big\}\ln\Phi_+(Y_1,Y_2) \\ &+\big\{\lambda+(1-2\lambda)\hat{H}(X_1,X_2)\big\}\ln\Phi_-(Y_1,Y_2)\Big]. \end{aligned} \tag{159}$$

The SPEs are

$$
\begin{aligned}
q^2 - R^2 &= \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{o})}, && (160)\\
R(q-R^2)^{\frac{3}{2}} &= -\alpha q(1-q)J_2^{(\mathrm{o})}, && (161)\\
J_1^{(\mathrm{o})} &= -2\int Dt_1Dt_2t_1h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\
&\quad \times[\Phi_+(Y_1,Y_2)-(1+e^{-\beta})\{\lambda+(1-2\lambda)H(X_1)H(X_2)], && (162)\\
J_2^{(\mathrm{o})} &= -2(1-2\lambda)\int Dt_1Dt_2h(X_1)H(X_2)t_1\ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). && (163)
\end{aligned}
$$

6.3 Noiseless case

The SPEs are

$$
\begin{aligned}
q^2 - R^2 &= \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{n})}, && (164)\\
R(q-R^2)^{\frac{3}{2}} &= -\alpha q(1-q)J_2^{(\mathrm{n})}, && (165)\\
J_1^{(\mathrm{n})} &= -2\int Dt_1Dt_2t_1h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\
&\quad \times[\Phi_+(Y_1,Y_2)-(1+e^{-\beta})H(X_1)H(X_2)], && (166)\\
J_2^{(\mathrm{n})} &= -2\int Dt_1Dt_2h(X_1)H(X_2)t_1\ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). && (167)
\end{aligned}
$$

6.4 Generalization error

Here, we give the expressions for the generalization error for the And machine.

(1) Noiseless

$$
\epsilon_g = \frac{3}{8}-\frac{1}{4\pi}\left(\sin^{-1}R_1+\sin^{-1}R_2\right)-\frac{1}{2\pi^2}\sin^{-1}R_1\sin^{-1}R_2. \qquad (168)
$$

(2) Input noise

$$
\epsilon_g = \frac{3}{8}-\frac{1}{4\pi}\left(\sin^{-1}\left(\frac{R_1}{\sqrt{\eta}}\right)+\sin^{-1}\left(\frac{R_2}{\sqrt{\eta}}\right)\right)-\frac{1}{2\pi^2}\sin^{-1}\left(\frac{R_1}{\sqrt{\eta}}\right)\sin^{-1}\left(\frac{R_2}{\sqrt{\eta}}\right) \qquad (169)
$$

(3) Output noise

$$
\begin{aligned}
\epsilon_g &= \frac{3}{8}+\frac{\lambda}{4}-\frac{1-2\lambda}{4\pi}(\sin^{-1}R_1+\sin^{-1}R_2) \\
&\quad -\frac{1-2\lambda}{2\pi^2}\sin^{-1}R_1\sin^{-1}R_2. && (170)
\end{aligned}
$$

6.5 Behavior for $\alpha \ll 1$

We obtain consistent results for $\zeta = \mathcal{O}(1)$ and $\tilde{\zeta} = \mathcal{O}(1)$.

6.5.1 Input noise

Let us estimate $K_1^{(\mathrm{i})}$. We estimate Φ_+ and Φ_- as follows:

$$\Phi_+(Y_1,Y_2) = e^{-\beta}+(1-e^{-\beta})H(\gamma t_1)H(\gamma t_2) = b[1-\frac{(1-e^{-\beta})}{2\sqrt{2\pi}b}(\gamma t_1+\gamma t_2)]+\mathcal{O}(\gamma^3), \quad (171)$$

$$b = \frac{1}{4}(1+3e^{-\beta}), \quad (172)$$

$$\Phi_-(Y_1,Y_2) = 1+e^{-\beta}-\Phi_+(\gamma t_1,\gamma t_2) = c[1+\frac{(1-e^{-\beta})}{2\sqrt{2\pi}c}(\gamma t_1+\gamma t_2)]+\mathcal{O}(\gamma^3), \quad (173)$$

$$c = \frac{1}{4}(3+e^{-\beta}), \quad (174)$$

$$\ln\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)} = A+B(\gamma t_1+\gamma t_2)+\mathcal{O}(\gamma^3), \quad (175)$$

$$A = \ln\frac{c}{b}, \quad B = \frac{(1-e^{-\beta})}{2\sqrt{2\pi}}(\frac{1}{b}+\frac{1}{c}). \quad (176)$$

Therefore, we have

$$J_2^{(\mathrm{i})} = -\frac{B\gamma}{\sqrt{2\pi}}\frac{1}{(1+\tilde{\zeta}^2)^{\frac{3}{2}}}+\mathcal{O}(\gamma^3). \quad (177)$$

Substituting this into Eq. (156), we have

$$R \simeq \frac{B}{\sqrt{2\pi\eta}}\alpha = R_0\alpha, \quad (178)$$

where $R_0 = \frac{B}{\sqrt{2\pi\eta}}$. From $\tilde{\zeta} = \frac{R}{\sqrt{\eta q-R^2}}$ and defining $q_0 = \frac{1+\tilde{\zeta}^2}{\tilde{\zeta}^2\eta}R_0^2$, we have

$$q \simeq q_0\alpha^2. \quad (179)$$

The exponents $\bar{\beta}$ are defined as $R \propto \alpha^{\bar{\beta}_R}$, $q \propto \alpha^{\bar{\beta}_q}$, and $\Delta\epsilon \equiv \epsilon_g|_{R=0}-\epsilon_g \propto \alpha^{\bar{\beta}_\epsilon}$ for small α. Since $\Delta\epsilon$ is expressed as $\Delta\epsilon \simeq \frac{R}{2\pi\sqrt{\eta}}$, we obtain $\bar{\beta}_R = \bar{\beta}_\epsilon$. Thus, we obtain $\bar{\beta}_R = \bar{\beta}_\epsilon = 1$ and $\bar{\beta}_q = 2$.

6.5.2 Output noise

Using the estimate of $J_2^{(\mathrm{i})}$, in Eq. (177), we have

$$J_2^{(\mathrm{o})} = -\frac{B\gamma(1-2\lambda)}{\sqrt{2\pi}}\frac{1}{(1+\zeta^2)^{\frac{3}{2}}}+\mathcal{O}(\gamma^3). \quad (180)$$

Substituting this into Eq. (161) yields

$$R \simeq \frac{B(1-2\lambda)}{\sqrt{2\pi}}\alpha = R_0\alpha, \quad (181)$$

where $R_0 = \frac{B(1-2\lambda)}{\sqrt{2\pi}}$. From $\zeta = \frac{R}{\sqrt{q-R^2}}$, we have

$$q \simeq q_0\alpha^2, \quad (182)$$

where $q_0 = \frac{1+\zeta^2}{\zeta^2} R_0^2$. $\Delta\epsilon$ is expressed as $\Delta\epsilon \simeq \frac{1-2\lambda}{2\pi} R$, and we obtain $\bar{\beta}_R = \bar{\beta}_\epsilon = 1$ and $\bar{\beta}_q = 2$.

6.5.3 Noiseless case

In this case, setting $\eta = 1$ in Eq. (178) or $\lambda = 0$ in Eq. (181), we have

$$R \simeq \frac{B}{\sqrt{2\pi}}\alpha, \tag{183}$$

$$q \simeq q_0 \alpha^2. \tag{184}$$

$\Delta\epsilon \simeq \frac{1}{2\pi} R$, and we have $\bar{\beta}_R = \bar{\beta}_\epsilon = 1$ and $\bar{\beta}_q = 2$.

6.6 Asymptotic behavior for $\alpha \gg 1$.

Now, let us study the asymptotic learning behavior. Using the same strategy as for the Parity machine, we derive the asymptotic forms for $\Delta q = 1 - q$ and $\Delta R = 1 - R$. Note that the SPEs are obtained by replacing $I_l^{(\mathrm{a})}$ with $J_l^{(\mathrm{a})}$ for $l = 1, 2$, and replacing 2α with α in the SPEs for the Parity machine.

6.6.1 Noiseless case

We assume that χ tends to a finite value as for the Parity machine. We obtain the following results:

$$J_1^{(\mathrm{n})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} l_1(\chi, \beta), \tag{185}$$

$$J_2^{(\mathrm{n})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} l_2(\chi, \beta), \tag{186}$$

$$l_1(\chi, \beta) = -\frac{1}{2}\left(\int Duu \frac{1}{\hat{H}(-u)} + (1 + e^{\beta}) \int Duu \frac{H(u/\chi)}{\hat{H}(u)\hat{H}(-u)} \right), \tag{187}$$

$$l_2(\chi, \beta) = g_2(\chi, \beta). \tag{188}$$

$\Delta\epsilon$ is expressed as $\Delta\epsilon \simeq \frac{\sqrt{2}}{\pi}\sqrt{\Delta R}$. Since $l_2 > 0$, we obtain the same results as for the Parity machine, $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\hat{\beta}_\epsilon = 1$.

6.6.2 Input noise

We obtain the following results:

$$J_1^{(\mathrm{i})} \simeq \frac{Q^2}{\sqrt{2\pi}} l_3(\beta), \tag{189}$$

$$J_2^{(\mathrm{i})} \simeq -\frac{c(\eta)(\eta - 1)}{\pi\eta} 2(1 - e^{-\beta}) \hat{l}(\beta), \tag{190}$$

$$l_3(\beta) = g_3(\beta), \tag{191}$$

$$\hat{l}(\beta) = \int Du \frac{1}{\hat{H}(u)} (> 0), \tag{192}$$

$$c(\eta) = \frac{1}{2} - \frac{1}{2\pi} \cos^{-1}(\frac{1}{\sqrt{\eta}}). \tag{193}$$

$\Delta\epsilon$ is expressed as $\Delta\epsilon \simeq \frac{1}{2\pi\sqrt{\eta-1}}(1 + \frac{2}{\pi}\sin^{-1}\frac{1}{\sqrt{\eta}})\Delta R$. Since $l_3 > 0$ and $\hat{l} > 0$, we obtain the same results as for the Parity machine, $\hat{\beta}_R = \hat{\beta}_\epsilon = 1/2$ and $\hat{\beta}_q = 1$.

6.6.3 Output noise

We obtain the following results:

$$J_1^{(\mathrm{o})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} \tilde{l}_1(\chi, \beta), \tag{194}$$

$$J_2^{(\mathrm{o})} \simeq -\frac{2Q^2}{\sqrt{2\pi}} \tilde{l}_2(\chi, \beta), \tag{195}$$

$$\tilde{l}_1(\chi, \beta) = l_1(\chi, \beta) + (1 + e^\beta)\lambda \int Duu \frac{H(u/\chi)}{\hat{H}(u)\hat{H}(-u)}, \tag{196}$$

$$\tilde{l}_2(\chi, \beta) = (1 - 2\lambda) g_2(\chi, \beta). \tag{197}$$

$\Delta\epsilon$ is expressed as $\Delta\epsilon \simeq (1 - 2\lambda)\frac{\sqrt{2}}{\pi}\sqrt{\Delta R}$. If $\lambda < 1/2$, $\tilde{l}_2 > 0$ follows. Thus, we obtain the same results as for the Parity machine, $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\hat{\beta}_\epsilon = 1$.

7. Numerical results for And Machine

For both the input and output noise models, we set $\epsilon_{\min} = 13/72$. Thus, $\tau = 1/\sqrt{3}$ and $\lambda = 13/72$. The methods used for the numerical simulations are similar to those in the case of the Parity machine. First, we study the α dependences of R_l for $T = 1$. The results for $M = 20$ and 30 are almost the same. Thus, we show the results for $M = 20$ below. Furthermore, we found that the simulation results for R_1 and R_2 are almost the same. Therefore, we display results for only R_1. In Figs. 35 - 37, we display the α dependences of R_1 and the generalization error ϵ_g for the noiseless, input noise, and output noise cases, respectively. We note that for the entire range of α, theoretical and numerical results agree reasonably well. In order to study the exponents $\bar{\beta}$ for R_1 and ϵ_g, we display the $\ln\alpha$ dependences of $\ln R_1$ and $\ln\Delta\epsilon$ in Figs. 38 and 39, respectively. Here, $\Delta\epsilon = \epsilon_g|_{R=0} - \epsilon_g$. We also display the theoretical values of $\bar{\beta}_R$ and $\bar{\beta}_\epsilon$ by dashed lines. The numerical and theoretical results agree reasonably well. Next, we study the temperature dependences of the order parameters. We performed EXMCs by the same procedure as for the Parity machine. In Figs. 40 -42, we display numerical and theoretical results for the noiseless, input noise, and output noise cases. We take the sample average for 100

samples. For the noiseless case, as is seen from Fig. 36, the numerical results confirm the theoretical ones at all temperatures we studied. For the input noise case shown in Fig. 37, we also find that the numerical results and theoretical ones agree reasonably well at all temperatures we studied except for very low temperatures. On the other hand, as seen from Fig. 38, for the output noise case, the numerical results do not agree very well at low temperatures, although they agree reasonably well at high temperatures. From these results, as for the Parity machine, it seems that RSB takes place at low temperatures.

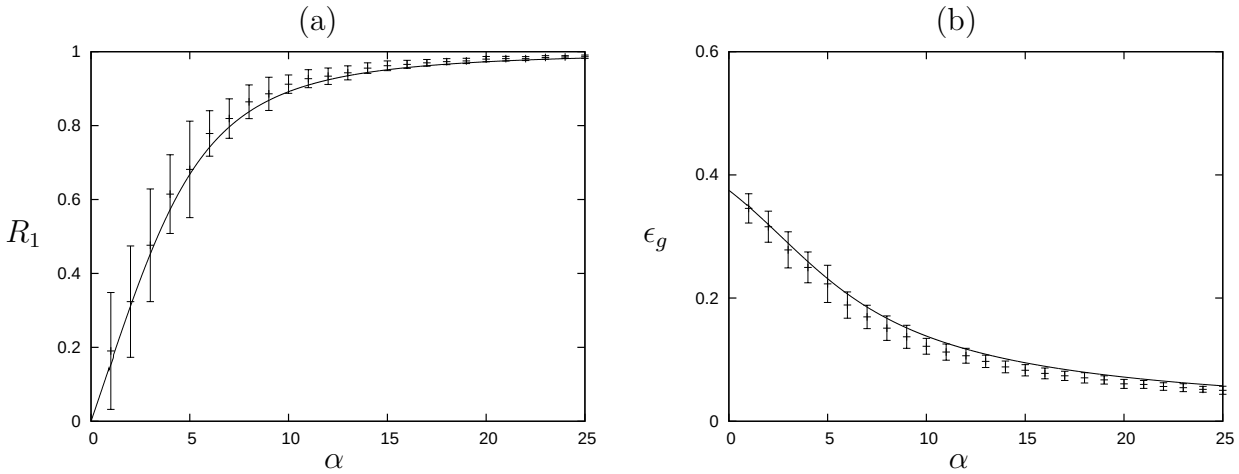

Fig. 35. α dependences of R_1 and generalization error ϵ_g for noiseless case. $M = 20$, $T = 1$. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. (a) R_1, (b) ϵ_g.

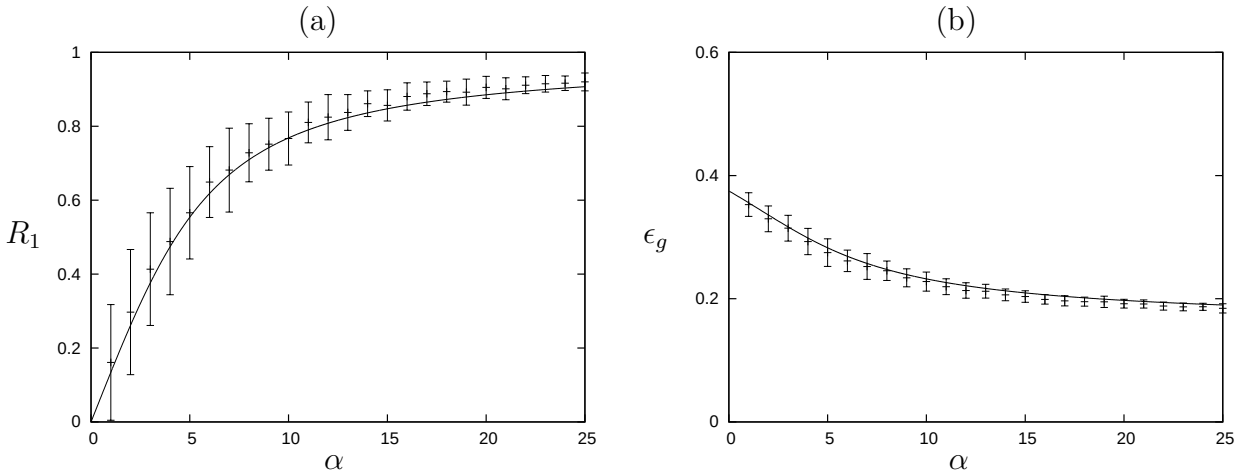

Fig. 36. α dependences of R_1 and generalization error ϵ_g for case of input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. $\tau = 1/\sqrt{3}$ (a) R_1, (b) ϵ_g.

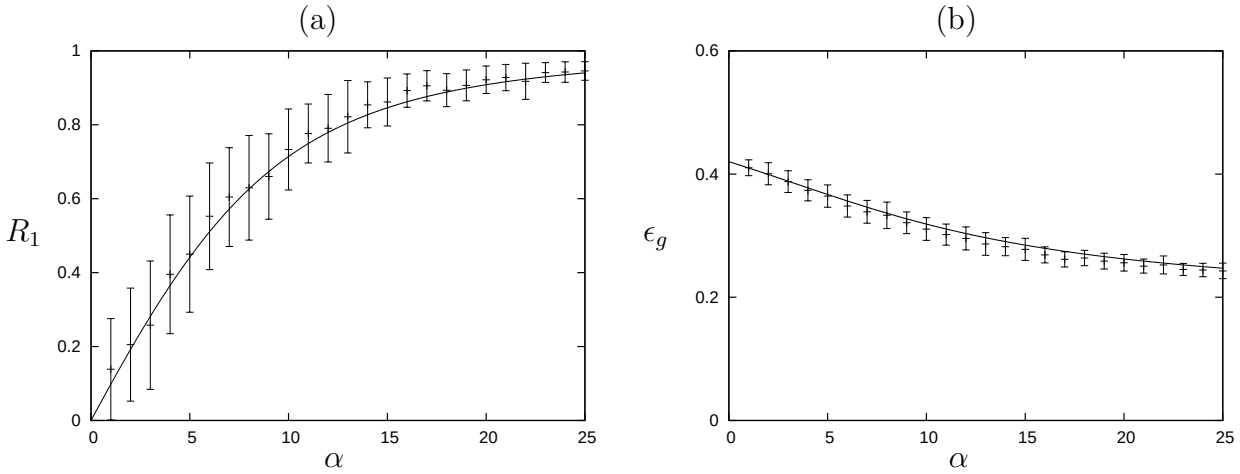

Fig. 37. α dependences of R_1 and generalization error ϵ_g for case of output noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. $\lambda = 13/72$. (a) R, (b) ϵ_g.

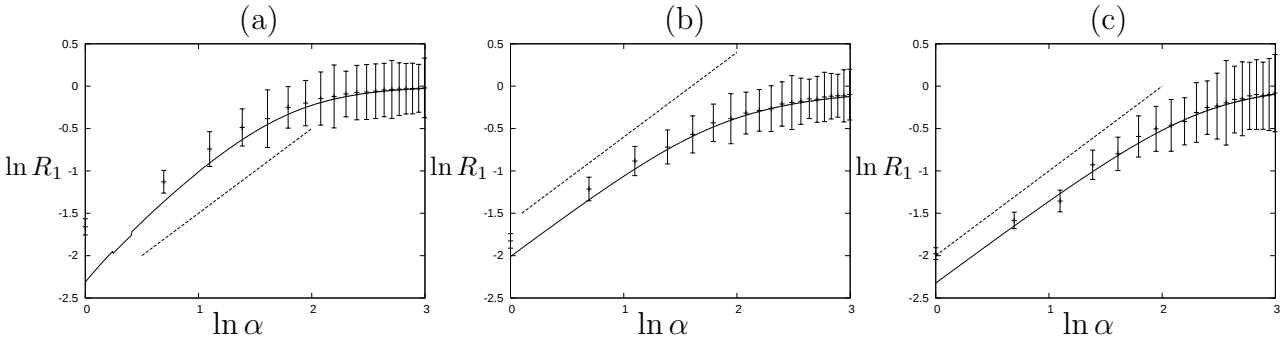

Fig. 38. $\ln \alpha$ dependences of $\ln R_1$ for noiseless, input noise, and output noise cases. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines: theoretical value $\bar{\beta}_R = 1$. (a) Noiseless, (b) input noise, $\tau = 1/\sqrt{3}$, (c) output noise, $\lambda = 13/72$.

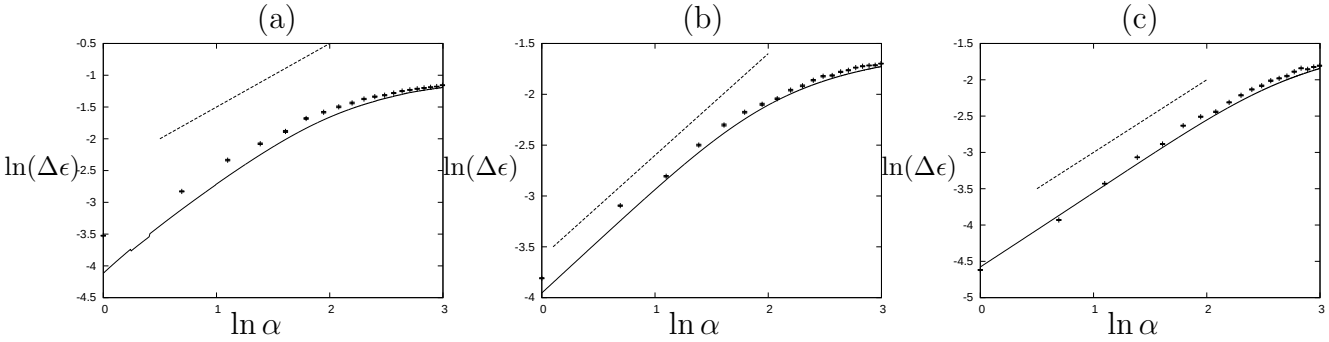

Fig. 39. $\ln \alpha$ dependences of $\ln(\Delta\epsilon)$ for noiseless, input noise, and output noise cases. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines: theoretical value $\bar{\beta}_\epsilon = 1$. (a) Noiseless, (b) input noise, $\tau = 1/\sqrt{3}$, (c) output noise, $\lambda = 13/72$.

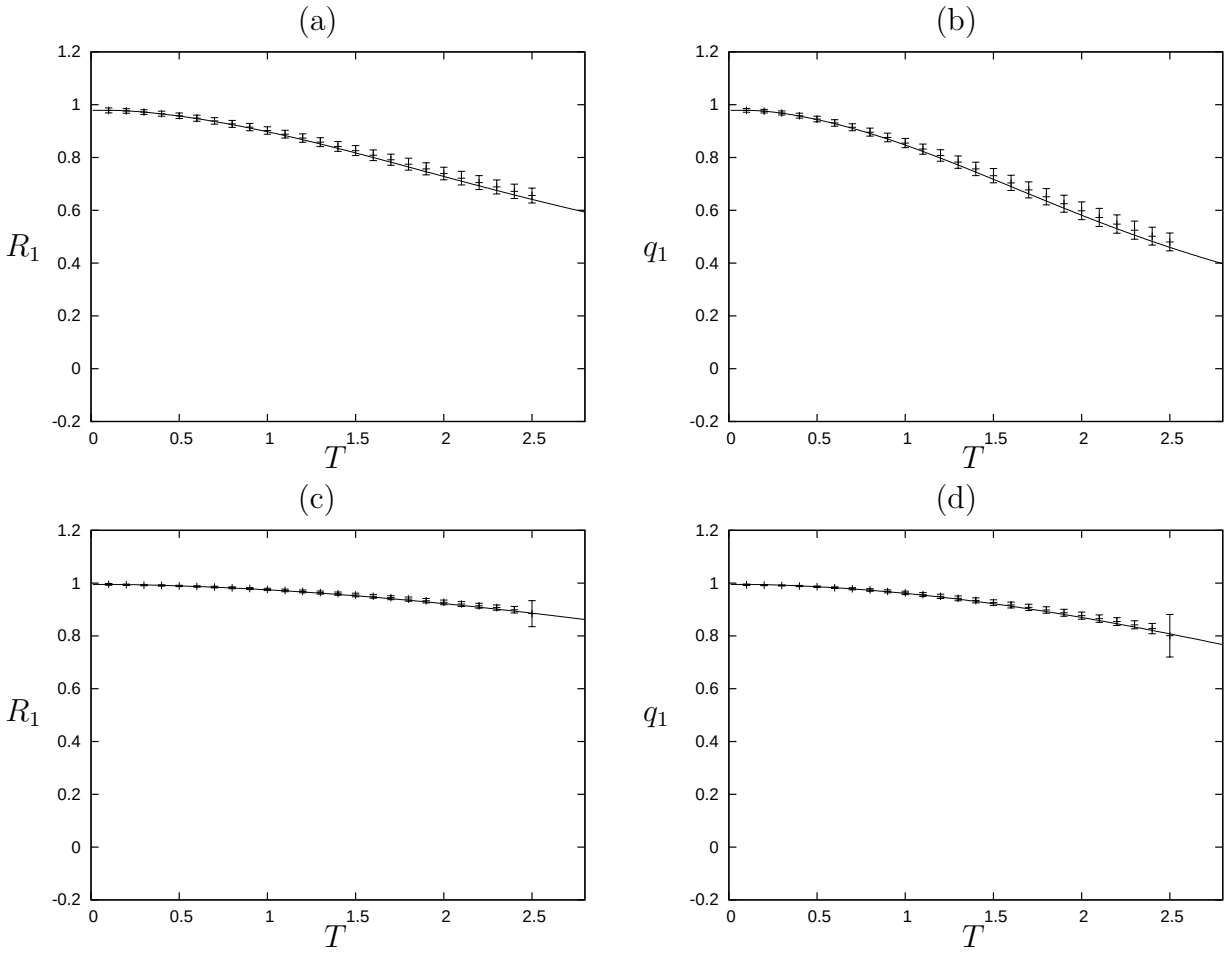

Fig. 40. T dependences of R_1 and q_1 for noiseless case. Curves: RS solutions, symbols: EXMCs. $M = 20$. (a) R_1, (b) q_1 for $\alpha = 10$, (c) R_1, (d) q_1 for $\alpha = 20$.

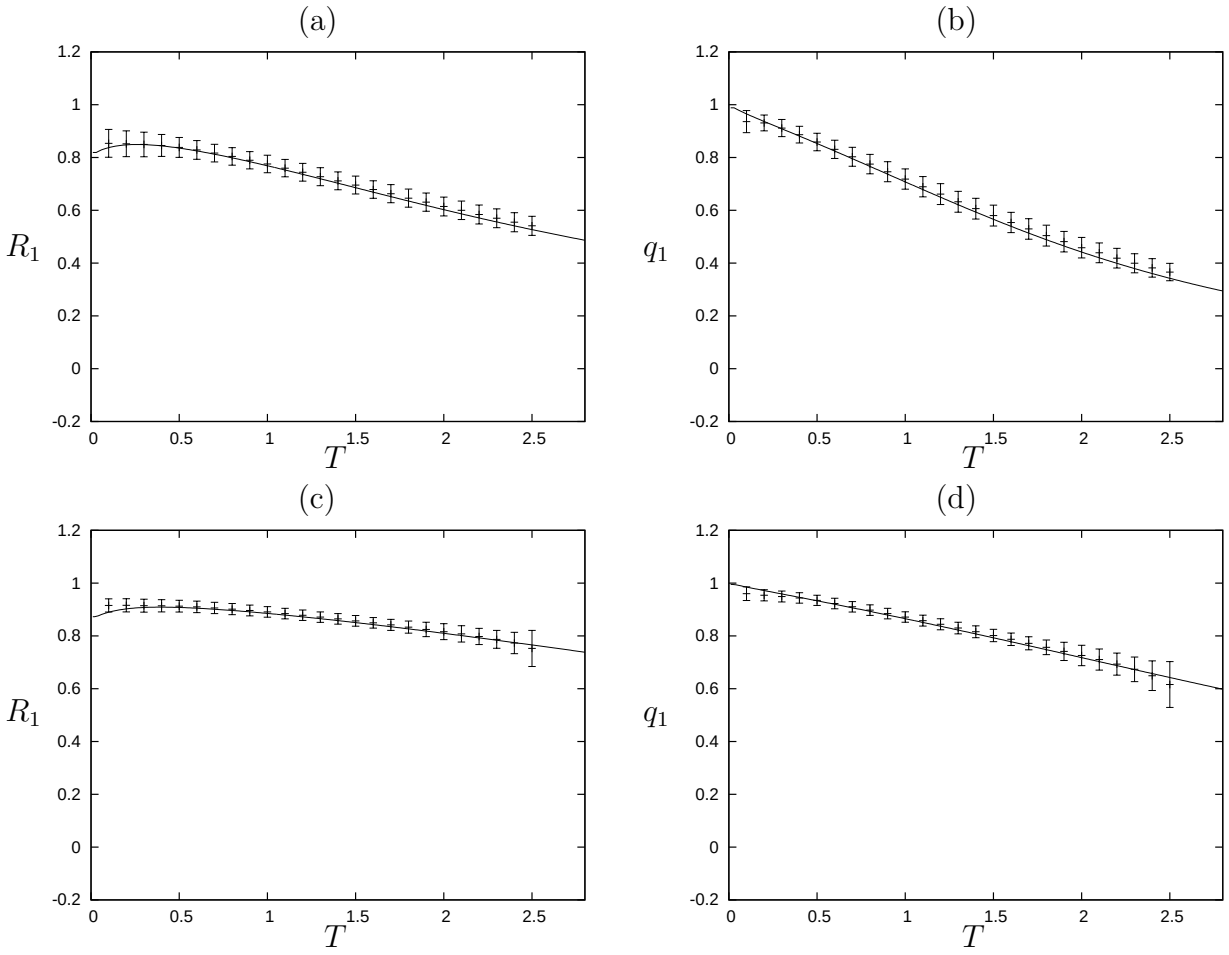

Fig. 41. T dependences of R_1 and q_1 for input noise. Curves: RS solutions, symbols: EXMCs. $M = 20$. Input: $\tau = 1/\sqrt{3}$. (a) R_1, (b) q_1 for $\alpha = 10$, (c) R_1, (d) q_1 for $\alpha = 20$.

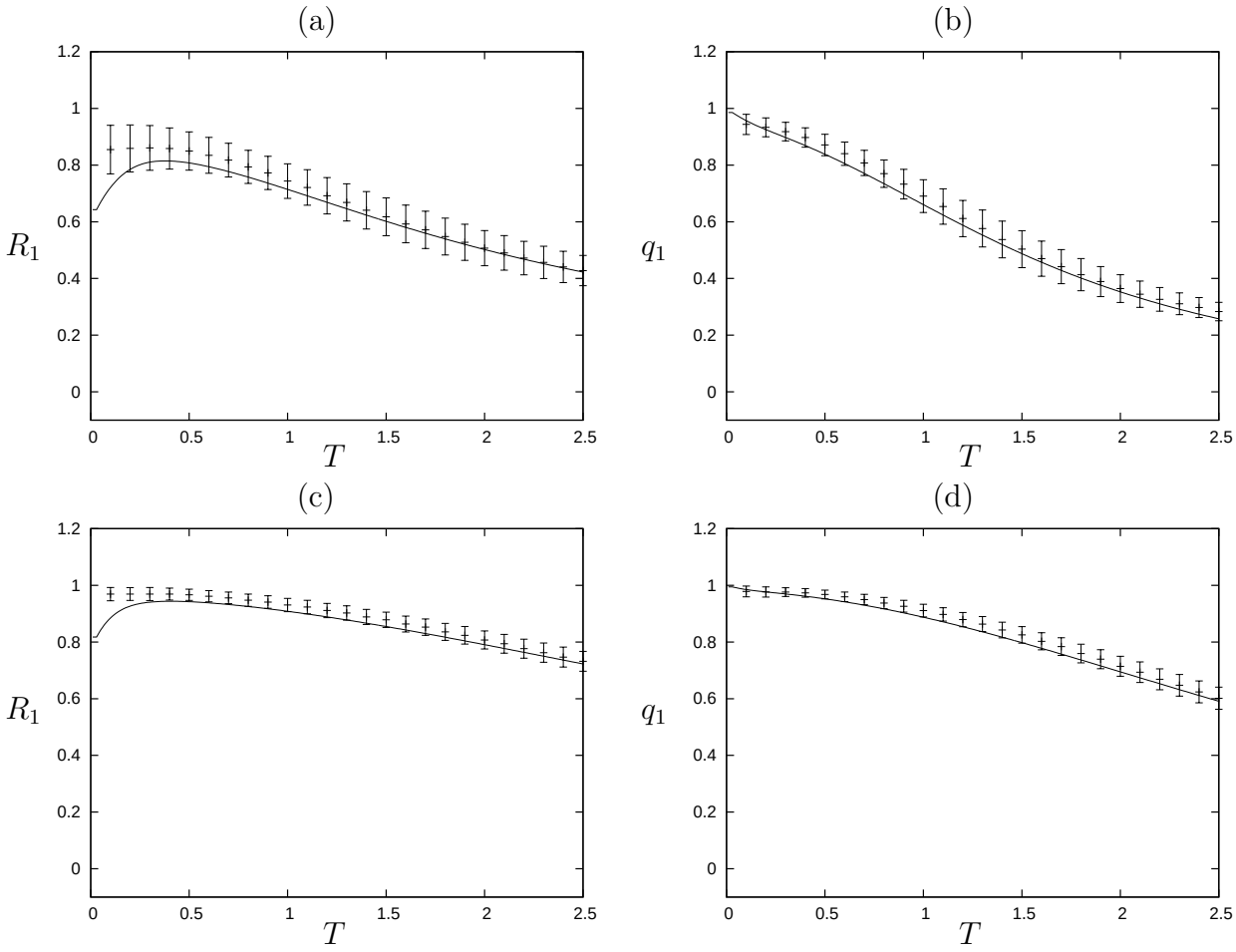

Fig. 42. T dependences of R_1 and q_1 for output noise. Curves: RS solutions, symbols: EXMCs. $M = 20$. Output: $\lambda = 13/72$. (a) R_1, (b) q_1 for $\alpha = 10$. (c) R_1, (d) q_1 for $\alpha = 20$.

Finally, we study the asymptotic behavior of learning as $\alpha \to \infty$. In Figs. 39-44, we display the results obtained by MCMCs for the noiseless, input noise, and output noise cases. We display the α dependences of R_1 and q_1 and the $\ln(\alpha)$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for $\alpha \in [25, 100]$ in the noiseless case and for $\alpha \in [25, 200]$ in the noisy cases. The method used is simulated annealing and is the same as that in the Parity machine. The numerical results are consistent with the theoretically obtained exponents $\hat{\beta}_R$ and $\hat{\beta}_q$, as is seen from Figs. 40, 42, and 44.

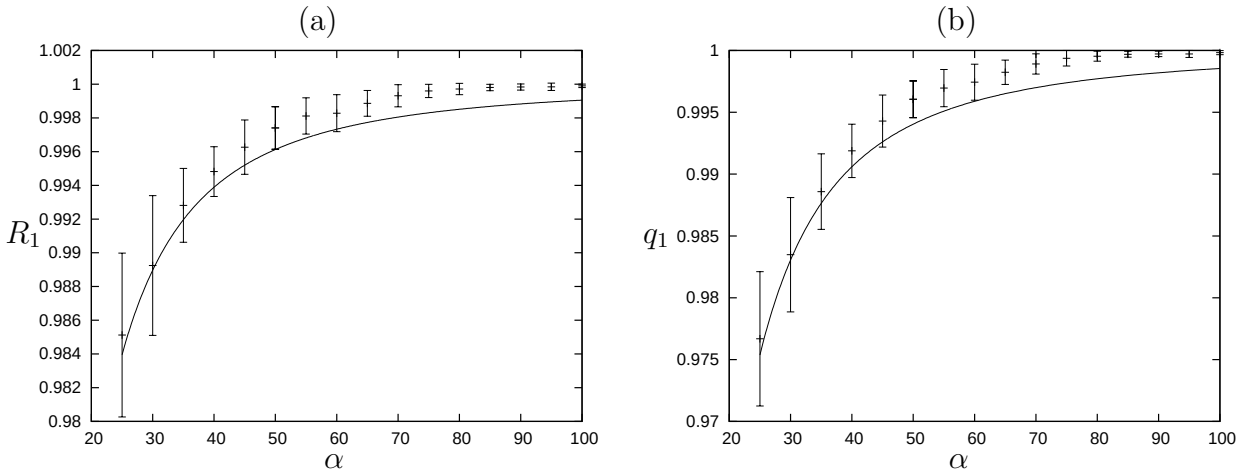

Fig. 43. α dependences of R_1 and q_1 for noiseless case. Curves: RS solutions, symbols: MCMCs. $M = 20, T = 1$. Averages are taken from 100 samples. Vertical lines are error bars. Annealing schedule: $\beta = 0.1, 0.2, \cdots, 1.0$. (a) R_1, (b) q_1.

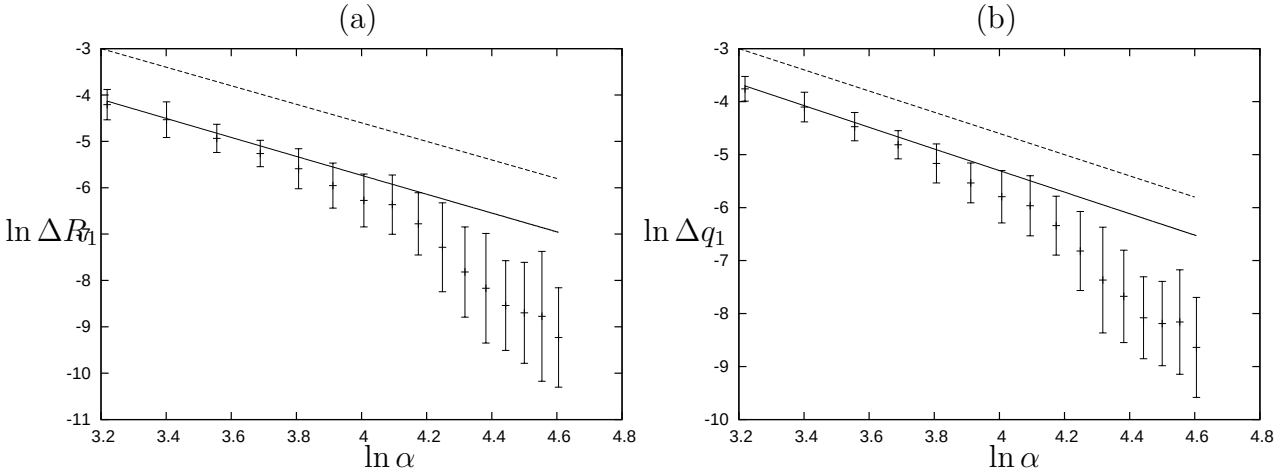

Fig. 44. $\ln \alpha$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for noiseless case. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R = 2, \hat{\beta}_q = 2$. (a) $\ln \Delta R_1$, (b) $\ln \Delta q_1$.

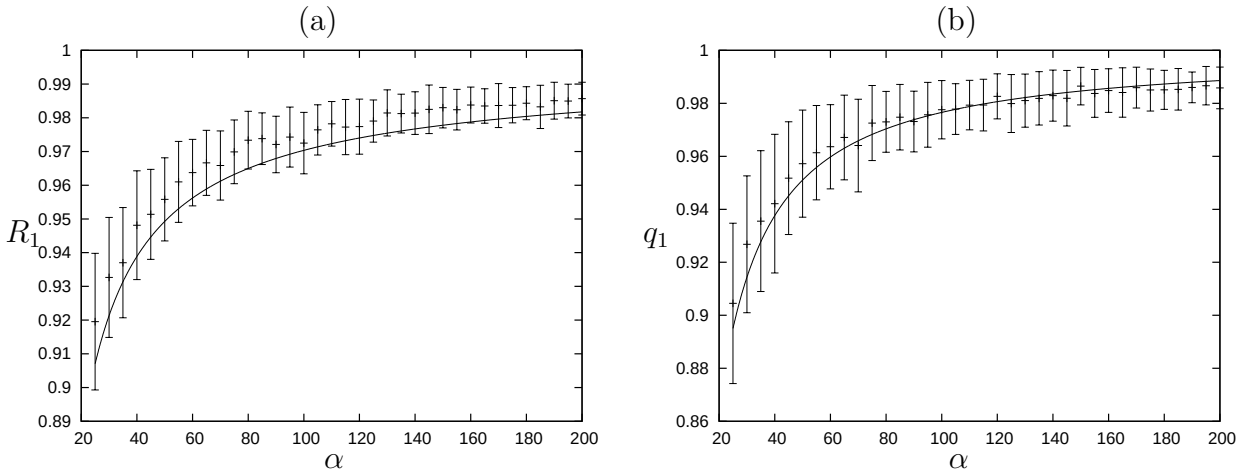

Fig. 45. α dependences of R_1 and q_1 for input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1, \tau = 1/\sqrt{3}$. (a) R_1, (b) q_1.

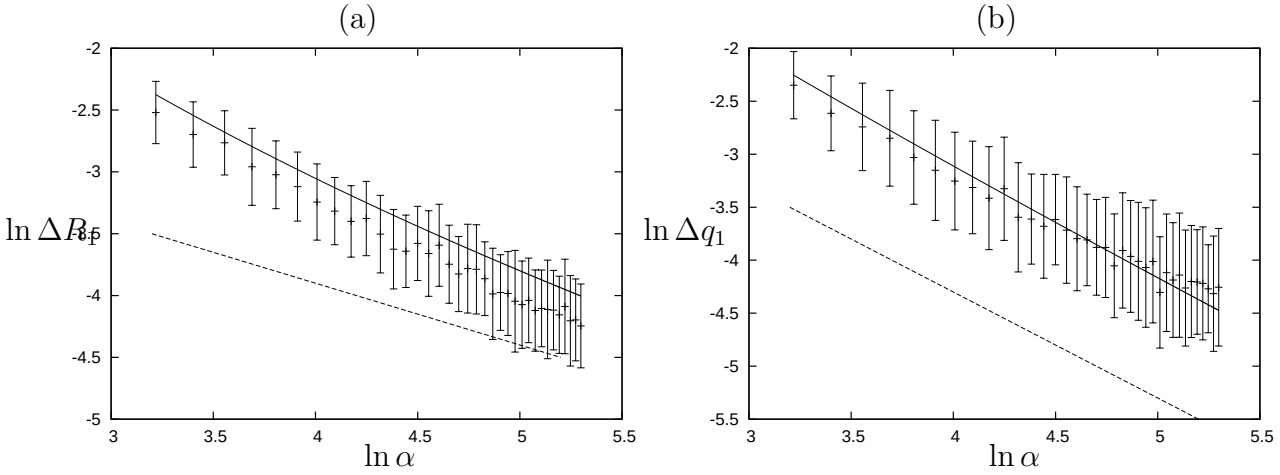

Fig. 46. $\ln \alpha$ dependences of $\ln \Delta R_1$ and $\ln \Delta q_1$ for input noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R = 1/2, \hat{\beta}_q = 1$. (a) $\ln \Delta R_1$, (b) $\ln \Delta q_1$.

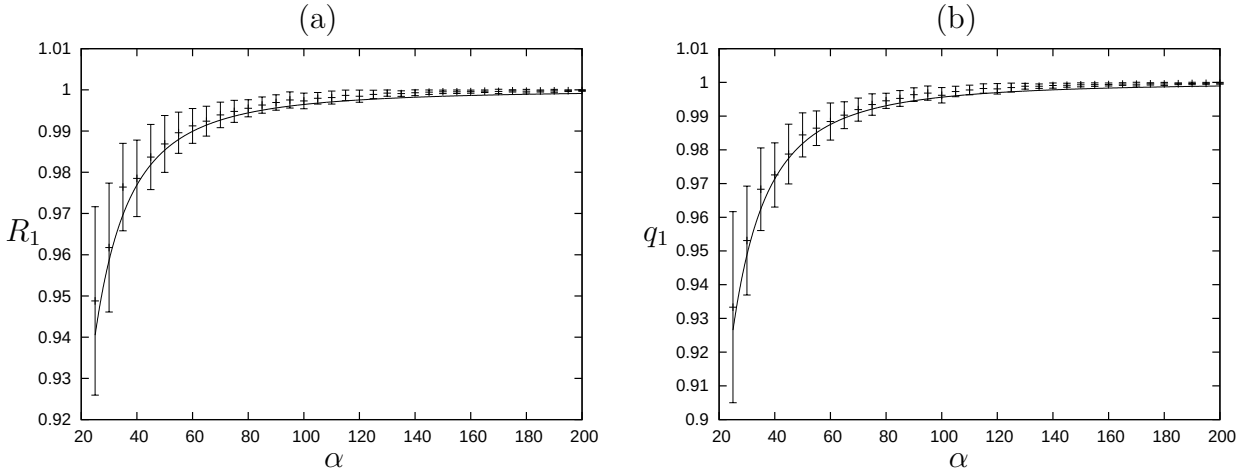

Fig. 47. α dependences of R_1 and q_1 for output noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1, \lambda = 5/18$. (a) R_1, (b) q_1.

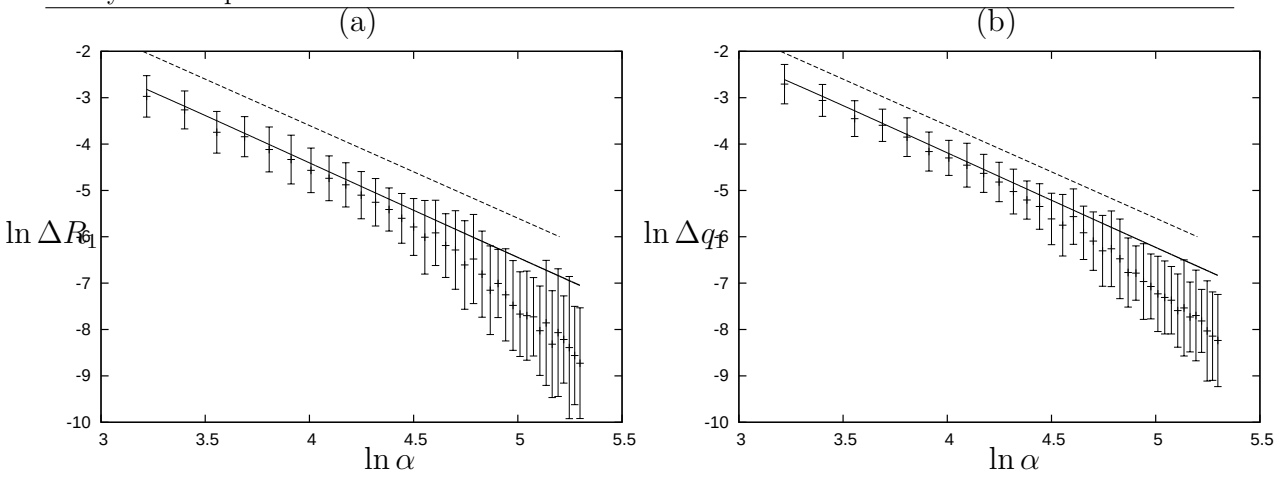

Fig. 48. $\ln\alpha$ dependences of $\ln\Delta R_1$ and $\ln\Delta q_1$ for output noise. Curves: RS solutions, symbols: MCMCs. $M = 20$, $T = 1$. Dashed lines denote theoretical values of exponents, $\hat{\beta}_R = 2, \hat{\beta}_q = 2$. (a) $\ln\Delta R_1$, (b) $\ln\Delta q_1$.

8. Summary and Discussion

In this paper, we studied the learning of Boolean functions expressed by a two-layer perceptron with and without external noise for small and large loading rates α. In order to make the analysis easier, we studied a tree-like architecture composed of two units. As the learning algorithm, we adopted the Gibbs algorithm. We assumed that both the teacher and students are expressed by two M-dimensional vectors with norm $\sqrt{M}$. By using the replica method, we derived the saddle point equations for order parameters R_l and q_l $(l = 1, 2)$, where R_l is the overlap between the teacher and student and q_l is the overlap between students for unit l. We studied the input and output noise cases. In the input noise case, an independent Gaussian noise with mean 0 and standard deviation τ was entered in each example input to the teacher. In the output noise case, the sign of the teacher's output was reversed with probability λ.

8.1 Parity machine

Firstly, we summarize the theoretical results within the RS ansatz.

We assumed $R_1 = R_2$ and $q_1 = q_2$, which were confirmed by numerical simulations. We theoretically found three phases, para (P), learning (L), and spin glass (SG) phases.

(1) P phase. $R_l = q_l = 0$, that is, learning does not take place.

(2) L phase. $R_l > 0$ and $q_l > 0$, that is, learning takes place.

(3) SG phase. $R_l = 0$ and $q_l > 0$. Learning does not take place but the average of each component of student vectors is not zero but random. This phase appears only for noisy cases.

For the noiseless case, we defined the critical loading rate $\alpha_{\rm C}^{(\rm n)}$ above which the L phase appears from the P phase when the temperature is fixed. For the noisy cases, we defined the noise amplitude $A_{\rm noise}$ as $A_{\rm noise} = \frac{1}{1-2\lambda}$ for the output noise case and $A_{\rm noise} = 1+\tau^2$ for the input noise case. Furthermore, we defined three critical capacities, $\alpha_{\rm L}, \alpha_{\rm C}$, and $\alpha_{\rm SG}$, and one critical temperature, $T_{\rm SG} = 1/\beta_{\rm SG}$. $T_{\rm SG}$ is a function of $A_{\rm noise}$. $\alpha_{\rm C}$ is defined for $T > T_{\rm SG}$, above which the L phase exists. $\alpha_{\rm SG}$ is defined for any temperature, and above which the SG phase exists. $\alpha_{\rm L}$ is defined for $T < T_{\rm SG}$, above which the L phase exists. $\alpha_{\rm L}$ and $\alpha_{\rm C}$ are functions of T and $A_{\rm noise}$, whereas $\alpha_{\rm SG}$ depends only on T. We sometimes denoted $\alpha_{\rm L}$ and $\alpha_{\rm C}$ by $\alpha_{\rm L}(T, A_{\rm noise})$ and $\alpha_{\rm C}(T, A_{\rm noise})$, respectively, in order to explicitly express the dependences on T and $A_{\rm noise}$. At $T = T_{\rm SG}$, $\alpha_{\rm L}, \alpha_{\rm C}$ and $\alpha_{\rm SG}$ coincide. We define $\alpha_0(A_{\rm noise}) = \alpha_{\rm L} = \alpha_{\rm C} = \alpha_{\rm SG}$. Since $\alpha_{\rm L}, \alpha_{\rm C}$, and $\alpha_{\rm SG}$ are increasing functions as long as they are defined with fixed $A_{\rm noise}$, we define their inverse functions $T_{\rm L}(\alpha), T_{\rm C}(\alpha)$, and $T_1(\alpha)$, respectively. We summarize the learning behavior when T is changed with α and $A_{\rm noise}$ is fixed for the cases of input and output noise.

(1) $\alpha_{\rm SG}(T = 0, A_{\rm noise}) > \alpha$. For every temperature, only the P phase exists and learning never takes place.

(2) $\alpha_{\rm SG}(T = 0, A_{\rm noise}) < \alpha < \alpha_0(A_{\rm noise})$. For $T_1(\alpha, A_{\rm noise}) < T$, the P phase exists and for $0 \leq T < T_1(\alpha, A_{\rm noise})$, the SG phase exists. That is, learning never takes place.

(3) $\alpha_0(A_{\rm noise}) < \alpha$. For $T_{\rm C}(\alpha, A_{\rm noise}) < T$, the P phase exists, for $T_1(\alpha, A_{\rm noise}) < T < T_{\rm C}(\alpha, A_{\rm noise})$, the L phase exists, for $T_{\rm L}(\alpha, A_{\rm noise}) < T < T_1(\alpha, A_{\rm noise})$, the SG phase and L phase exist, and for $0 \leq T < T_{\rm L}(\alpha, A_{\rm noise})$, the SG phase exists.

We defined the exponents $\bar{\beta}$ of R_l and q_l and the generalization error ϵ_g immediately after learning takes place. We found that $\bar{\beta}_R = 1/2$, $\bar{\beta}_q = 1$, and $\bar{\beta}_\epsilon = 1$ for the noiseless, input noise, and output noise cases.

Next, we studied asymptotic learning behavior for large α in order to clarify whether learning occurs under the existence of external noise. We defined the exponents $\hat{\beta}$ of $1 - R_l, 1 - q_l$ and $\epsilon_g - \epsilon_g|_{R=1}$ for large α. We found that for any temperature, when α becomes large, learning occurs for the noiseless case, as well as for the output noise case with $\lambda < 1/2$ and for the input noise case with any positive τ. We obtained $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\bar{\beta}_\epsilon = 1$ for the noiseless and output noise cases and $\hat{\beta}_R = \hat{\beta}_\epsilon = 1/2$ and $\bar{\beta}_q = 1$ for the input noise case.

Now, let us summarize the numerical results and compare them with the theoretical results. We performed Markov chain Monte Carlo simulations (MCMCs) and

exchange Monte Carlo simulations (EXMCs). For small α, by numerically estimating the maximum value $R_{\max}$ of the histogram of R_1, we compared numerical results of the α dependence of $R_{\max}$ with theoretical results of $R(\alpha)$ and obtained consistent results. For exponents $\bar{\beta}_R$ and $\bar{\beta}_\epsilon$, we also obtained consistent results between the theory and simulations. In order to study the temperature dependences of R_1 and q_1, we performed EXMCs. For the noiseless case, the numerical results confirmed the theoretical results. For the input and output noise cases, we could not observe the SG phase numerically. That is, when $\alpha_{\rm SG}(T = 0, A_{\rm noise}) < \alpha$ and for $T < T_1(\alpha, A_{\rm noise})$, the SG phase was expected theoretically, but numerically we found the L phase. In order to resolve this disagreement, we studied the AT stabilities and free energies of the RS solutions. We found that the free energy of the L solution is smaller than that of the SG solution for the input noise and output noise cases. As for the AT stability, we found two cases. As the temperature is lowered, the L solution becomes AT unstable and then the SG solution becomes AT unstable for the input and output noise cases for $\alpha = 10$. In contrast, for $\alpha = 30$ for the output noise case, the SG solution first becomes AT unstable and then the L solution becomes AT unstable as the temperature is lowered. Therefore, in all the cases we studied, an RSB L solution is expected to appear. Next, we studied the noise amplitude λ dependence of learning for the output noise case. We found three solutions, i.e., two learning solutions, the $\rm L_1$ and $\rm L_2$ solutions, and the SG solution. We numerically found that when λ is small, the $\rm L_1$ solution appears and when λ is large, the $\rm L_2$ solution appears, and the SG solution does not appear. We also studied the AT stabilities and free energies of these solutions. We found that these solutions are AT stable. The free energy of the $\rm L_1$ solution is smallest, that of the $\rm L_2$ solution is middle, and that of the SG is largest. The disagreement between the theory and simulations is considered that the acceptance ratio of the Metropolis method is so low that the system might be trapped to a local minimum of the free energy.
Finally, to investigate the asymptotic learning behavior, we studied the α dependence of R_1 and q_1 for large α and obtained consistent results between the theory and simulations for the exponents $\hat{\beta}_R$ and $\hat{\beta}_q$.

Thus, except for the low-temperature region and small noise amplitude region, the theoretical and numerical results agree reasonably well. From these results, we conclude the following.

1. Learning behavior

(1) $\alpha < \alpha_{\rm SG}(T = 0, A_{\rm noise})$. For every temperature, only the P phase exists.

(2) $\alpha_{\rm SG}(T = 0, A_{\rm noise}) < \alpha < \alpha_0(A_{\rm noise})$. For $T_1(\alpha, T_{\rm noise}) < T$, the P phase exists and for $T < T_1(\alpha, T_{\rm noise})$, the L phase exists.

(3) $\alpha_0(A_{\rm noise}) < \alpha$. For $T_{\rm C}(\alpha, T_{\rm noise}) < T$, the P phase exists and for $T < T_{\rm C}(\alpha, T_{\rm noise})$, the L phase exists.

2. Effects of noise

   Learning takes place even if input noise ($\tau > 0$) or output noise ($0 < \lambda < 1/2$) exists.

3. AT stability

   The replica symmetry is broken at low temperatures.

8.2 And machine

Firstly, we summarize the theoretical result within the RS ansatz.
In the And machine, contrary to the Parity machine, we found that the critical loading rate $\alpha_{\rm C}$ is 0 and the SG state does not exist. That is, learning takes place for any positive α and for any temperature. Thus, in this case, we defined the exponents $\bar{\beta}$ of R_l and q_l and the generalization error ϵ_g at $\alpha = 0$. We found that $\bar{\beta}_R = 1$, $\bar{\beta}_q = 2$, and $\bar{\beta}_\epsilon = 1$ for the noiseless, input noise, and output noise cases. The Exponents for R and q are different from those for the Parity machine. Regarding the asymptotic behavior, we obtained $\hat{\beta}_R = \hat{\beta}_q = 2$ and $\bar{\beta}_\epsilon = 1$ for the noiseless and output noise cases, and $\hat{\beta}_R = \hat{\beta}_\epsilon = 1/2$ and $\bar{\beta}_q = 1$ for the input noise case, as for the Parity machine.
Next, we summarize the numerical results and compare them with the theoretical results. We performed MCMCs and EXMCs and found that learning takes place for any positive α and for any temperature. For exponents $\bar{\beta}_R$ and $\bar{\beta}_\epsilon$, and also for exponents $\hat{\beta}_R$ and $\hat{\beta}_q$, we obtained consistent results between the theory and simulations. In order to study the temperature dependences of R_1 and q_1, we performed EXMCs. For the noiseless, input noise and output noise cases, the theoretical and numerical results agreed reasonably well except at low temperatures. We consider that replica symmetry breaking takes place at low temperatures as in the case of the Parity machine. From these results, we conclude the following.

1. Learning behavior

   For any $\alpha > 0$ and any temperature, learning takes place.

2. Effects of noise

Learning takes place even if input noise ($\tau > 0$) or output noise ($0 < \lambda < 1/2$) exists.

3. Stability

   The replica symmetry seems to be broken at low temperatures.

Now, let us compare our results with those obtained by previous studies. In Ref. 18, Kabashima studied the online learning of a two-layer perceptron of the Parity machine with non-overlapping receptive fields of $K = 2$. The amplitude of his example vector was set to $\sqrt{M}$, the learning algorithm was the least-action algorithm (LAA), and the output noise was studied. He found that if the noise rate λ exceeds the critical value λ_c, students cannot acquire any generalization ability even in the limit $\alpha \to \infty$. In Ref. 17, Schottky and Krey studied the influence of noise on multilayer perceptrons with non-overlapping receptive fields, i.e., the tree-like architecture. They treated the Committee machine, the Parity machine, and the And machine for input and output noise cases adopting the Gibbs and Bayes algorithms, and studied the behaviors of learning in the $q \to 1$ and $q \to 0$ limits. They found the "Aha effect", which occurs in the case that $\alpha_{\rm C}$ is positive. Contrary to the result of Kabashima, they did not obtain a critical noise level. In our model, a critical noise level does not exist. The reason for this difference seems to be the difference in the algorithms. That is, Kabashima used the LAA algorithm and online learning, whereas Schottky and Krey and we used the Gibbs algorithm and batch learning.

Acknowledgment

We are grateful to Nao Yasuda for performing several numerical simulations in the present work.

References

1) T. Uezu, Y. Kabashima, K. Nokura, and N. Nakamura, J. Phys. Soc. Jpn. **65**, 3797 (1996).

2) T. Uezu and Y. Kabashima, J. Phys. A **29** L55 (1996).

3) T. Uezu, J. Phys. A **30**, L777 (1997).

4) T. Uezu and S. Kiyokawa, J. Phys. Soc. Jpn., **80**, 044007 (2011).

5) T. Uezu and Y. Kabashima, J. Phys. A **29**, L439 (1996).

6) T. Uezu, J. Phys. Soc. Jpn. **71**, 1882 (2002).

7) C. M. Bishop: *Pattern Recognition and Machine Learning* (Springer, New York, 2006).

8) S. Haykin: *Neural Networks* (Prentice Hall International, Inc., Upper Saddle River, 1999).

9) T. L. H. Watkin, A. Rau, and M. Biehl, Rev. Mod. Phys. **65**, 499 (1993).

10) H. Schwarze and J. Hertz, Europhys. Lett. **20**, 375 (1992).

11) H. Schwarze and J. Hertz, J. Phys. A **26**, 4919 (1993).

12) H. Schwarze, J. Phys. A **26**, 5781 (1993).

13) K. Kang, C. Kwon, and Y. Park, Phys. Rev. E **48**, 4805 (1993).

14) M. Opper, Phys. Rev. Lett. **72** (1994) 2113.

15) G. Mato and N. Parga J. Phys. A **25**, 5047 (1992).

16) B. Schottky, J. Phys. A **28**, 4515 (1995).

17) B. Schottky and U. Krey, J. Phys. A **30**, 8541 (1997).

18) Y. Kabashima, J. Phys. A **27**, 1917 (1994).

19) J. R. L. de Almeida and D. J. Thouless, J. Phys. A: Math. Gen. **11**, 983 (1978).

9. Appendix A. Derivation of Free Energy f, $-\beta f = \frac{1}{N}\langle \ln Z \rangle_{p,x,w^t} = \alpha G_1 + G_2$, of General Boolean Functions by the Replica Method

In this appendix, we calculate the partition function by the replica method and derive the SPEs. Let us start by calculating $e^{-\beta E[\boldsymbol{w},\xi^p]}$,

$$e^{-\beta E[\boldsymbol{w},\xi^p]} = \prod_{\mu=1}^{K} e^{-\beta\Theta(-\sigma_\mu^t\sigma_\mu^s)}, \tag{198}$$

where $\Theta(x)$ is the Heaviside function,

$$\Theta(x) = \begin{cases} 1 & \text{for } x > 0, \\ 0 & \text{for } x \leq 0. \end{cases} \tag{199}$$

This is rewritten as

$$e^{-\beta E[\boldsymbol{w},\xi^p]} = \prod_{\mu=1}^{K} \{e^{-\beta} + (1-e^{-\beta})\Theta(\sigma_\mu^t \sigma_\mu^s)\}. \tag{200}$$

Defining the following quantities:

$$\sigma_\mu^{s\alpha} = B_s(\{\mathrm{sgn}(h_{l\mu}^{s\alpha})\}),\ h_{l\mu}^{s\alpha} = \frac{\boldsymbol{x}_l^\mu \cdot \boldsymbol{w}_l^{s\alpha}}{\sqrt{M}}, \tag{201}$$

$$d\mu(\boldsymbol{w}_l^s) \equiv d\boldsymbol{w}_l^s\ \delta((\boldsymbol{w}_l^s)^2 - M),\ d\mu(\boldsymbol{w}^s) \equiv \prod_{l=1}^{K} d\mu(\boldsymbol{w}_l^s), \tag{202}$$

we have

$$Z^n = \int \Big[\prod_{\alpha=1}^{n} d\mu(\boldsymbol{w}^{s\alpha})\Big] \prod_\mu \prod_\alpha \Big\{e^{-\beta} + (1-e^{-\beta})\Theta\Big(\sigma_\mu^t B_s(\{\mathrm{sgn}(h_{l\mu}^{s\alpha})\})\Big)\Big\}. \tag{203}$$

Let $p(\sigma|\{h_{l\mu}^t\})$ be the probability that the teacher takes a value of σ. Here, $h_{l\mu}^t$ is defined as

$$h_{l\mu}^t = \frac{\boldsymbol{x}_l^\mu \cdot \boldsymbol{w}_l^t}{\sqrt{M}}. \tag{204}$$

The average of Z^n with the probability distribution $p(\sigma|\{h_{l\mu}^t\})$, $\langle Z^n \rangle_p$, is expressed as

$$\begin{aligned} \langle Z^n \rangle_p = &\int \Big[\prod_\alpha d\mu(\boldsymbol{w}^{s\alpha})\Big] \prod_\mu \Big[\sum_{\sigma=\pm 1} p(\sigma|\{h_{l\mu}^t\})\Big] \\ &\times \prod_\alpha \Big[\mathrm{Tr}_{\{\sigma_l^\alpha\}}\Big(\tilde{\triangle}_{s,\beta}^{\sigma}(\{\sigma_l^\alpha\}) \prod_l \Theta\big(\sigma_l^\alpha h_{l\mu}^{s\alpha}\big)\Big)\Big], \end{aligned} \tag{205}$$

where we define

$$\triangle_s^{\sigma_s}\big(\{\sigma_l\}\big) = \begin{cases} 1 & \text{for } \sigma_s = B_s(\{\sigma_l\}), \\ 0 & \text{for } \sigma_s = -B_s(\{\sigma_l\}), \end{cases} \tag{206}$$

$$\tilde{\triangle}_{s,\beta}^{\sigma}(\{\sigma_l\}) \equiv e^{-\beta} + (1-e^{-\beta})\ \triangle_s^\sigma\ \big(\{\sigma_l\}\big). \tag{207}$$

Next, let us average over $\{\boldsymbol{x}_l^\mu\}$. Since $(\boldsymbol{x}_l^\mu)^2 = M$, the average over $\{\boldsymbol{x}_l^\mu\}$ for a function $g(\{\boldsymbol{x}_l^\mu\})$, $\langle g \rangle_x$, is

$$\langle g \rangle_x = \Big[\prod_{l,\mu} \frac{1}{S_M} \int d\boldsymbol{x}_l^\mu \delta((\boldsymbol{x}_l^\mu)^2 - M)\Big] g \tag{208}$$

$$= \Big[\prod_{l,\mu}\frac{1}{S_M}\int d\boldsymbol{x}_l^\mu \int_{-i\infty}^{i\infty}\frac{dK_{\mu l}}{2\pi i}e^{-K_{\mu l}\{\sum_j((x_{lj}^\mu)^2-M)\}}\Big]g, \tag{209}$$

where we define

$$S_M = \int d\boldsymbol{x}_l^\mu \delta((\boldsymbol{w}_l^{s\alpha})^2 - M). \tag{210}$$

$\langle g\rangle_x$ is rewritten as

$$\begin{aligned}\langle g\rangle_x &= \Big[\prod_{l,\mu}\frac{1}{S_M}\int d\boldsymbol{x}_l^\mu \int_{-i\infty}^{i\infty}\frac{dK_{\mu l}}{2\pi i}e^{-K_{\mu l}\{\sum_j((x_{lj}^\mu)^2-M)\}}\int dh_{l\mu}^t\delta(h_{l\mu}^t - \frac{\boldsymbol{w}_l^t\cdot\boldsymbol{x}_l^\mu}{\sqrt{M}}) \\ &\times\Big[\prod_\alpha dh_{l\mu}^{s\alpha}\delta(h_{l\mu}^{s\alpha}-\frac{\boldsymbol{w}_l^{s\alpha}\cdot\boldsymbol{x}_l^\mu}{\sqrt{M}})\Big]\Big]g \\ &= \Big[\prod_{l,\mu}\frac{1}{S_M}\int d\boldsymbol{x}_l^\mu \int_{-i\infty}^{i\infty}\frac{dK_{\mu l}}{2\pi i}e^{-K_{\mu l}\{\sum_j((x_{lj}^\mu)^2-M)\}} \\ &\times\int dh_{l\mu}^t\frac{dz_{l\mu}^t}{2\pi}\Big[\prod_\alpha dh_{l\mu}^{s\alpha}\frac{dz_{l\mu}^{s\alpha}}{2\pi}\Big]e^{iz_{l\mu}^t(h_{l\mu}^t-\frac{\boldsymbol{w}_l^t\cdot\boldsymbol{x}_l^\mu}{\sqrt{M}})+i\sum_\alpha z_{l\mu}^{s\alpha}(h_{l\mu}^{s\alpha}-\frac{\boldsymbol{w}_l^{s\alpha}\cdot\boldsymbol{x}_l^\mu}{\sqrt{M}})}\Big]g.\end{aligned} \tag{211}$$

Let us define $\boldsymbol{v}_{l\mu}$ as

$$(\boldsymbol{v}_{l\mu})_j = z_{l\mu}^t w_{lj}^t + \sum_\alpha z_{l\mu}^{s\alpha}w_{lj}^{s\alpha}. \tag{212}$$

By performing integrations over $\{x_{lj}^\mu\}(j=1,\cdots,l)$, $\langle Z^n\rangle_{p,x}\equiv\langle\langle Z^n\rangle_p\rangle_x$ is given by

$$\langle Z^n\rangle_{p,x} = A\times B, \tag{213}$$

$$\begin{aligned}A &= \int\Big[\prod_{l\alpha}d\boldsymbol{w}_l^{s\alpha}\delta\big((\boldsymbol{w}_l^{s\alpha})^2-M\big)\Big] \\ &\times\prod_\mu\Big[\sum_{\sigma\pm1}p(\sigma|\{h_{l\mu}^t\})\prod_\alpha\big(\mathrm{Tr}_{\{\sigma_l^\alpha\}}\tilde{\triangle}_{s,\beta}^\sigma(\{\sigma_l^\alpha\})\big)\prod_l\Theta\big(\sigma_l^\alpha h_{l\mu}^{s\alpha}\big)\Big],\end{aligned} \tag{214}$$

$$\begin{aligned}B &= \int\prod_{l,\mu}dh_{l\mu}^t\frac{dz_{l\mu}^t}{2\pi}\prod_{l,\mu,\alpha}dh_{l\mu}^{s\alpha}\frac{dz_{l\mu}^{s\alpha}}{2\pi}\prod_{\mu l}\Big[\frac{1}{S_M}\int\frac{dK_{\mu l}}{2\pi i}e^{MK_{\mu l}}\Big]\prod_{\mu l}\big(\frac{\pi}{K_{\mu l}}\big)^{\frac{M}{2}} \\ &\times\exp\Big[i\sum_{l,\mu}z_{l\mu}^t h_{l\mu}^t + i\sum_\alpha\sum_{l,\mu}z_{l\mu}^{s\alpha}h_{l\mu}^{s\alpha} - \sum_{l,\mu}\frac{(\boldsymbol{v}_{l\mu})^2}{4K_{\mu l}M}\Big].\end{aligned} \tag{215}$$

Firstly, we perform the following integration:

$$\begin{aligned}\int d\boldsymbol{x}_l^\mu\delta\big((\boldsymbol{x}_l^\mu)^2-M\big) &= \int_{-\infty}^{\infty}d\boldsymbol{x}_l^\mu\int_{-i\infty}^{i\infty}\frac{dK_{\mu l}}{2\pi i}e^{-K_{\mu l}\left((\boldsymbol{x}_l^\mu)^2-M\right)} \\ &= \int_{-i\infty}^{i\infty}\frac{dK_{\mu l}}{2\pi i}\left(\sqrt{\frac{\pi}{K_{\mu l}}}\right)^M e^{MK_{\mu l}}.\end{aligned} \tag{216}$$

The integral is evaluated at the saddle point $K_{\mu l} = 1/2$, and we obtain

$$\left(\sqrt{\frac{\pi}{K_{\mu l}}}\right)^M e^{MK_{\mu l}} = \sqrt{2\pi}^M e^{\frac{M}{2}} = (\sqrt{2\pi e})^M = S_M. \tag{217}$$

Hereafter, we set $K_{\mu l} = \frac{1}{2}$. Thus,

$$\begin{aligned} B &= \Big[\prod_{l\mu}\frac{1}{S_M}(2\pi)^{\frac{M}{2}}e^{\frac{M}{2}}\Big]\Big[\prod_{l,\mu}\int dh_{l\mu}^t\frac{dz_{l\mu}^t}{2\pi}e^{iz_{l\mu}^t h_{l\mu}^t}\Big] \\ &\quad\times\Big[\prod_{l,\mu,\alpha}\int dh_{l\mu}^{s\alpha}\frac{dz_{l\mu}^{s\alpha}}{2\pi}e^{iz_{l\mu}^{s\alpha}h_{l\mu}^{s\alpha}}\Big]e^{-\frac{1}{2M}\sum_{l,\mu}(\boldsymbol{v}_{l\mu})^2}. \end{aligned} \tag{218}$$

We define

$$R_l^\alpha = \frac{1}{M}(\boldsymbol{w}_l^{s\alpha}, \boldsymbol{w}_l^t), \tag{219}$$

$$q_l^{\alpha\beta} = \frac{1}{M}(\boldsymbol{w}_l^{s\alpha}, \boldsymbol{w}_l^{s\beta}). \tag{220}$$

Taking the average over $\boldsymbol{w}_l^t, (l = 1, \cdots, K)$ we have

$$\langle Z^n\rangle_{p,x,w^t} = \langle A\ B\rangle_{w^t} = \int\Big[\prod_{l,\alpha}dR_l^\alpha\frac{M}{2\pi}d\bar{R}_l^\alpha\Big]\Big[\prod_{l,\alpha<\beta}dq_l^{\alpha\beta}\frac{M}{2\pi i}F_l^{\alpha\beta}\Big]e^{\tilde{G}_1+\tilde{G}_2+\tilde{G}_3}, \tag{221}$$

where $\bar{R}_l^\alpha$ and $F_l^{\alpha\beta}$ are conjugate variables to R_l^α and $q_l^{\alpha\beta}$, respectively. Here, $e^{\tilde{G}_1}, e^{\tilde{G}_2}$ and $e^{\tilde{G}_3}$ are expressed as

$$\begin{aligned} e^{\tilde{G}_1} &= \Big[\prod_{l,\mu}\int dh_{l\mu}^t\frac{dz_{l\mu}^t}{2\pi}\Big]\Big[\prod_{l,\mu,\alpha}\int dh_{l\mu}^{s\alpha}\frac{dz_{l\mu}^{s\alpha}}{2\pi}\Big] \\ &\quad\times\exp\Big[\sum_{l,\mu}\{iz_{l\mu}^t h_{l\mu}^t + i\sum_\alpha z_{l\mu}^{s\alpha}h_{l\mu}^{s\alpha}\} \\ &\quad-\frac{1}{2}\{(z_{l\mu}^t)^2 + 2z_{l\mu}^t\sum_\alpha z_{l\mu}^{s\alpha}R_l^\alpha + \sum_\alpha(z_{l\mu}^{s\alpha})^2 + 2\sum_{\alpha<\beta}z_{l\mu}^{s\alpha}z_{l\mu}^{s\beta}q_l^{\alpha\beta}\}\Big] \\ &\quad\times\prod_\mu\Big[\sum_{\sigma\pm1}p(\sigma|\{h_{l\mu}^t\})\prod_\alpha\Big(\mathrm{Tr}_{\{\sigma_l^\alpha\}}\Big(\tilde{\triangle}_s^{\ \sigma}(\{\sigma_l^\alpha\})\prod_l\Theta(\sigma_l^\alpha h_{l\mu}^{s\alpha})\Big)\Big)\Big], \end{aligned} \tag{222}$$

$$\begin{aligned} e^{\tilde{G}_2} &= \int\prod_{l\alpha}d\boldsymbol{w}_l^{s\alpha}\delta\big((\boldsymbol{w}_l^{s\alpha})^2 - M\big)\prod_l d\boldsymbol{w}_l^t\delta\big((\boldsymbol{w}_l^t)^2 - M\big)\frac{1}{S_M} \\ &\quad\times e^{-i\sum_{\alpha l}\bar{R}_l^\alpha\boldsymbol{w}_l^{s\alpha}\cdot\boldsymbol{w}_l^t + \sum_{l,\alpha<\beta}F_l^{\alpha\beta}\boldsymbol{w}_l^{s\alpha}\cdot\boldsymbol{w}_l^{s\beta}}, \end{aligned} \tag{223}$$

$$\tilde{G}_3 = iM\sum_{l\alpha}\bar{R}_l^\alpha R_l^\alpha - M\sum_{l,\alpha<\beta}F_l^{\alpha\beta}q_l^{\alpha\beta}. \tag{224}$$

Now, let us assume the replica symmetry:

$$R_l^\alpha = R_l\ ,\ q_l^{\alpha\beta} = q_l,\ l = 1, \cdots, K. \tag{225}$$

Performing the integration over $\{z_{l\mu}^{s\alpha}\}$, $\{z_{l\mu}^{t}\}$, and $\{h_{l\mu}^{s\alpha}\}$, we have

$$e^{\tilde{G}_1} = \left\{ \prod_l \Big[dh_{l\mu}^t DS_{l\mu} \sqrt{\frac{1-q_l}{2\pi\{1-q_l+n(q_l-R_l^2)\}}} \Big] e^{\sum_l \{-\frac{1}{2Q_l}(h_{l\mu}^t)^2 + \frac{1-q_l}{2} n b_{l\mu}^2\}} \right.$$
$$\left. \times \Big[\sum_{\sigma\pm 1} p(\sigma|\{h_{l\mu}^t\}) \{\Phi_{\beta s}^{\sigma}(\{-\sqrt{1-q_l} b_{l\mu}\})\}^n \Big] \right\}^p, \tag{226}$$

where $Q_l = \frac{1-q_l+n(q_l-R_l^2)}{1-q_l+nq_l}$, $P_l = \frac{\sqrt{q_l-R_l^2}}{\sqrt{[1-q_l+n(q_l-R_l^2)](1-q_l)}}$, $b_{l\mu} = \frac{R_l}{1-q_l+n(1-q_l)} h_{l\mu}^t + P_l S_{l\mu}$, and $\Phi_{\beta s}^{\sigma}(\{\sqrt{1-q_l} b_{l\mu}\})$ is defined as

$$\Phi_{\beta s}^{\sigma}(\{\sqrt{1-q_l} b_{l\mu}\}) \equiv \mathrm{Tr}_{\{\sigma_l^\alpha\}} \left(\tilde{\triangle}_s^{\sigma}(\{\sigma_l^\alpha\}) \prod_l H(\sqrt{1-q_l} b_{l\mu} \sigma_l^\alpha) \right). \tag{227}$$

Hereafter, we omit the subscript μ and write the variables as

$$h_{l\mu}^t \to h_l^t \quad , \quad S_{l\mu} \to S_l \quad , \quad b_{l\mu} \to b_l.$$

We change the integration variable from S_l to b_l, given as

$$b_l = \frac{R_l}{1-q_l+n(q_l-R_l^2)} h_l^t + P_l S_l = \alpha_l S_l + \beta_l h_l^t, \tag{228}$$

where

$$\alpha_l = P_l \quad , \quad \beta_l = \frac{R_l}{1-q_l+n(q_l-R_l^2)}. \tag{229}$$

S_l is expressed as

$$S_l = \frac{1}{\alpha_l}(b_l - \beta_l h_l^t). \tag{230}$$

Thus, $e^{\tilde{G}_1}$ is rewritten as

$$e^{\tilde{G}_1} = \left\{ \prod_l \Big[\int \frac{dh_l^t}{\sqrt{2\pi}} \frac{db_l}{\alpha_l} e^{-\frac{1}{2}\frac{b_l^2}{\alpha_l^2} + L_l + n\frac{1-q_l}{2} b_l^2} \sqrt{\frac{1-q_l}{2\pi\{1-q_l+n(q_l-R_l^2)\}}} \right.$$
$$\left. \times \sum_{\sigma\pm 1} p(\sigma|\{h_l^t\}) \{\Phi_{\beta s}^{\sigma}(\{-\sqrt{1-q_l} b_l\})\}^n \Big] \right\}^p, \tag{231}$$

where

$$L_l = -\frac{1}{2}\frac{q_l}{q_l-R_l^2}(h_l^t)^2 + \frac{\beta_l}{\alpha_l^2} b_l h_l^t = -\frac{1}{2}\frac{q_l}{q_l-R_l^2}(h_l^t)^2 + \frac{1-q_l}{q_l-R_l^2} R_l b_l h_l^t. \tag{232}$$

Setting $\tilde{G}_1 = pG_1$, we finally obtain

$$e^{G_1} = \prod_l \Big[\int \frac{dh_l^t}{\sqrt{2\pi}} \frac{db_l}{\alpha_l} e^{-\frac{1}{2}\frac{b_l^2}{\alpha_l^2} + L_l + n\frac{1-q_l}{2} b_l^2} \sqrt{\frac{1-q_l}{2\pi\{1-q_l+n(q_l-R_l^2)\}}}$$
$$\times \sum_{\sigma=\pm 1} p(\sigma|\{h_l^t\}) \{\Phi_{\beta s}^{\sigma}(\{-\sqrt{1-q_l} b_l\})\}^n \Big]. \tag{233}$$

9.1 Input noise

An independent noise $\boldsymbol{\zeta}_l$ is entered in the input $\boldsymbol{x}_l$ to the teacher. $(\zeta_l)_i$ is assumed to be Gaussian with mean 0 and standard deviation τ_l. From Eq. (204), we have

$$\tilde{h}_{l\mu}^t \equiv \frac{\boldsymbol{w}_l^t \cdot (\boldsymbol{x}_l^\mu + \boldsymbol{\zeta}_l)}{\sqrt{M}} = h_{l\mu}^t + \frac{\boldsymbol{w}_l^t \cdot \boldsymbol{\zeta}_l}{\sqrt{M}}. \tag{234}$$

Let us define $y_l = \dfrac{\boldsymbol{w}_l^t \cdot \boldsymbol{\zeta}_l}{\sqrt{M}}$. The averages of y_l and y_l^2 are

$$\langle y_l \rangle = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} w_{l,i} \langle \zeta_{l,i} \rangle = 0, \tag{235}$$

$$\langle y_l^2 \rangle = \frac{1}{M} \sum_{i,j} w_{l,i} w_{l,j} \langle \zeta_{l,i}\ \zeta_{l,j} \rangle = \frac{1}{M} \sum_{i} (w_{l,i})^2 \tau_l^2 = \tau_l^2. \tag{236}$$

Hereafter, we assume $\tau_l = \tau$. Thus, y_l is a sum of many independent random variables, and by the central limit theorem, it is a Gaussian random variable with the mean 0 and the standard deviation τ. Its probability density function $p(y_l)$ is

$$p(y_l) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{y_l^2}{2\tau^2}}. \tag{237}$$

The probability $p(\sigma|\{h_l^t\})$ that the teacher's output is σ is the probability that $B(\sigma_1, \sigma_2) = \sigma$. Thus, by defining $\triangle_t^\sigma(\{\sigma_l\})$ as

$$\triangle_t^\sigma(\{\sigma_l\}) = \begin{cases} 1 & \text{for } \sigma = B_t(\{\sigma_l\}), \\ 0 & \text{for } \sigma = -B_t(\{\sigma_l\}), \end{cases} \tag{238}$$

we obtain

$$\begin{aligned} p(\sigma|\{h_l^t\}) &= \int dy_1 dy_2\ p(y_1)p(y_2)\ \triangle_t^\sigma\left(\{\mathrm{sgn}(h_l^t + y_l)\}\right) \\ &= \mathrm{Tr}_{\{\sigma_l\}}\ \triangle_t^\sigma\ (\{\sigma_l\}) \prod_l \int dy_l\ p(y_l) \Theta\{\sigma_l(h_l^t + y_l)\}. \end{aligned} \tag{239}$$

Let us define x_l by $y_l = \tau x_l$ for $l = 1, 2$. Then, the integration with respect to x_l yields

$$p(\sigma|\{h_l^t\}) = \mathrm{Tr}_{\{\sigma_l\}}\ \triangle_l^\sigma\ (\{\sigma_l\}) \prod_l H\left(-\sigma_l \frac{h_l^t}{\tau} \right). \tag{240}$$

Therefore, we obtain the following result:

$$\int dh_l^t\ e^{L_l} H\left(-\sigma_l \frac{h_l^t}{\tau} \right) = \sqrt{\frac{2\pi(q_l - R_l^2)}{q_l}} e^{A_l} H\left(\frac{-(1-q_l)R_l b_l \sigma_l}{\sqrt{(q_l\tau^2 + q_l - R_l^2)q_l}} \right), \tag{241}$$

where we define

$$A_l = \frac{(1-q_l)^2 R_l^2 b_l^2}{2q_l(q_l - R_l^2)}. \tag{242}$$

Thus, we have

$$[\prod_l \int dh_l^t e^{L_l}]p(\sigma|\{h_l^t\}) = \mathrm{Tr}_{\{\sigma_l\}} \triangle_l^\sigma (\{\sigma_l\}) \Big[\prod_l \sqrt{\frac{2\pi(q_l - R_l^2)}{q_l}} e^{A_l} H\left(\frac{-(1-q_l)R_l b_l \sigma_l}{\sqrt{(q_l\tau^2 + q_l - R_l^2)q_l}}\right)\Big] \quad (243)$$

Therefore, we obtain

$$e^{G_1} = \sum_{\sigma=\pm 1} \mathrm{Tr}_{\{\sigma_l\}} \triangle_l^\sigma (\{\sigma_l\}) \prod_l \left[\frac{1-q_l}{\sqrt{q_l}} \int \frac{1}{\sqrt{2\pi}} db_l e^{-\frac{b_l^2}{2}\frac{(1-q_l)^2}{q_l}} \right.$$
$$\left. \times H\left(\frac{-(1-q_l)R_l b_l \sigma_l}{\sqrt{(q_l\tau^2 + q_l - R_l^2)q_l}}\right)\right] \{\Phi_{\beta s}^\sigma(\{-\sqrt{1-q_l} b_l\})\}^n. \quad (244)$$

Making a variable transformation from b_l to $t_l = \frac{1-q_l}{\sqrt{q_l}} b_l$ yields

$$e^{G_1} = \sum_{\sigma=\pm 1} \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \Big[\prod_l \int Dt_l\Big] \Big[\prod_l H\Big(- \frac{R_l \sigma_l t_l}{\sqrt{q_l\tau^2 + q_l - R_l^2}}\Big)\Big] \Big(\Phi_{\beta,s}^\sigma(\{-\gamma_l t_l\})\Big)^n \quad (245)$$

From Eq. (227), we have

$$\Phi_{\beta s}^\sigma(\{\zeta_l t_l\}) = \mathrm{Tr}_{\{\sigma_l^\alpha\}} \tilde{\triangle}_{s,\beta}^\sigma(\{\sigma_l^\alpha\}) \prod_l H\Big(\gamma_l t_l \sigma_l^\alpha\Big), \quad (246)$$

$$\gamma_l = \sqrt{\frac{q_l}{1-q_l}}. \quad (247)$$

By taking the limit $n \to 0$, Eq. (245) becomes

$$e^{G_1} \simeq 1 + n \sum_{\sigma=\pm 1} \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \Big[\prod_l \int Dt_l\Big] \Big[\prod_l H\Big(- \frac{R_l \sigma_l t_l}{\sqrt{q_l\tau^2 + q_l - R_l^2}}\Big)\Big] \ln \Phi_{\beta,s}^\sigma(\{-\gamma_l t_l\}). \quad (248)$$

Therefore, from Eq. (221), $\langle Z^n \rangle_{p,x,w^t}$ is obtained as

$$\langle Z^n \rangle_{p,x,w^t} = \int \Big[\prod_{l,\mu} dR_l^\alpha \frac{M}{2\pi} d\bar{R}_l^\alpha\Big] \Big[\prod_{l,\alpha<\beta} dq_l^{\alpha\beta} \frac{M}{2\pi i} d\bar{F}_l^{\alpha\beta}\Big] e^{\tilde{G}_1 + \tilde{G}_2 + \tilde{G}_3}. \quad (249)$$

We evaluate the integration of Eq. (249) at the saddle point,

$$\langle Z^n \rangle \sim e^{\tilde{G}_1 + \tilde{G}_2 + \tilde{G}_3}, \quad (250)$$

$$\langle \ln Z \rangle = \frac{1}{n}(\tilde{G}_1 + \tilde{G}_2 + \tilde{G}_3). \quad (251)$$

Since $\tilde{G}_1 = pG_1$, from Eq. (248), by setting $t \to -t$, $\frac{G_1}{n}$ is obtained as

$$\frac{G_1}{n} = \sum_{\sigma=\pm 1} \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \Big[\prod_l \int Dt_l\Big] \Big[\prod_l H\Big(\frac{R_l \sigma_l t_l}{\sqrt{(1+\tau^2)q_l - R_l^2}}\Big)\Big] \ln \Phi_{\beta,s}^\sigma(\{\gamma_l t_l\}) \quad (252)$$

Next, we calculate $e^{\tilde{G}_2}$. Similar to the case of $K_{\mu l}$, we can set $E_l^t = 1$. We obtain

$$\delta\big((\boldsymbol{w}_l^{s\alpha})^2 - M\big) = \int_{-i\infty}^{i\infty} \frac{dE_l^\alpha}{4\pi i} e^{-\frac{E_l^\alpha}{2}\{(\boldsymbol{w}_l^{s\alpha})^2 - M\}}. \quad (253)$$

Therefore, from Eq. (223), we have

$$
\begin{aligned}
e^{\tilde{G}_2} &= \int [\prod_{l\alpha} d\boldsymbol{w}_l^{s\alpha} \frac{dE_l^\alpha}{4\pi i}][\prod_l d\boldsymbol{w}_l^t \frac{1}{4\pi i}\frac{1}{S_M}] \\
&\quad \times \exp[-\frac{1}{2}\sum_l (\boldsymbol{w}_l^t)^2 + \frac{M}{2}K - \frac{1}{2}\sum_{l,\alpha} E_l^\alpha \{(\boldsymbol{w}_l^{s\alpha})^2 - M\} \\
&\quad -i\sum_{l\alpha} \bar{R}_l^\alpha \boldsymbol{w}_l^{s\alpha}\cdot \boldsymbol{w}_l^t + \sum_{l,\alpha<\beta} F_l^{\alpha\beta}\boldsymbol{w}_l^{s\alpha}\cdot \boldsymbol{w}_l^{s\beta}]. \qquad (254)
\end{aligned}
$$

Integrating over $w_{l,j}^t$, we obtain

$$
\begin{aligned}
e^{\tilde{G}_2} &= e^{\sum_{lj} G_{2,lj}}, \qquad (255) \\
e^{\tilde{G}_{2,lj}} &\equiv \int [\prod_\alpha dw_{lj}^{s\alpha} \frac{dE_l^\alpha}{4\pi i}\frac{\sqrt{2\pi}}{4\pi i} e^{\mathcal{L}} e^{\sum_\alpha \frac{1}{2}E_l^\alpha} e^{\frac{1}{2}} \frac{1}{(s_M)^{1/M}}, \qquad (256) \\
\mathcal{L} &= -\sum_\alpha \frac{E_l^\alpha}{2}(w_{lj}^{s\alpha})^2 + \sum_{\alpha<\beta} F_l^{\alpha\beta} w_{lj}^{s\alpha} w_{lj}^{s\beta} - \frac{1}{2}(\sum_\alpha \bar{R}_l^\alpha w_{lj}^{s\alpha})^2. \qquad (257)
\end{aligned}
$$

The RS conditions are

$$
E_l^\alpha = E_l \quad , \quad F_l^{\alpha\beta} = F_l \quad , \quad \bar{R}_l^\alpha = \bar{R}_l, \quad l = 1, \cdots, K. \qquad (258)
$$

Thus, we have

$$
\mathcal{L} = -\frac{E_l}{2}\sum_\alpha (w_{lj}^{s\alpha})^2 + F_l \sum_{\alpha<\beta} w_{lj}^{s\alpha} w_{lj}^{s\beta} - \frac{\bar{R}_l^2}{2}(\sum_\alpha w_{lj}^{s\alpha})^2. \qquad (259)
$$

By integrating over $w^{s\alpha}$, we have

$$
e^{\tilde{G}_{2,lj}} = \Big[\int \prod_\alpha \frac{dE_l^\alpha}{4\pi i}\Big]\frac{\sqrt{2\pi}}{4\pi i}\Big(\frac{2\pi}{E_l+F_l}\Big)^{\frac{n}{2}} \sqrt{\frac{E_L+F_l}{E_l+\bar{R}^2+(1-n)(F_l-\bar{R}_l^2)}} e^{\frac{1}{2}} \frac{1}{(s_M)^{1/M}}. \qquad (260)
$$

Since the right-hand side of Eq. (260) has no j dependences, we have

$$
\tilde{G}_2 = \sum_{l,j} \tilde{G}_{2,lj} = M\sum_l \tilde{G}_{2,lj} \equiv M\sum_l \hat{G}_{2,l}. \qquad (261)
$$

Thus, we obtain

$$
e^{\hat{G}_{2,l}} = \Big[\int \prod_\alpha \frac{dE_l^\alpha}{4\pi i}\Big]\frac{\sqrt{2\pi}}{4\pi i}\Big(\frac{2\pi}{E_l+F_l}\Big)^{\frac{n}{2}} \sqrt{\frac{E_L+F_l}{E_l+\bar{R}^2+(1-n)(F_l-\bar{R}_l^2)}} \times e^{n\frac{E_l}{2}} \frac{e^{\frac{1}{2}}}{(S_M)^{\frac{1}{M}}}. \qquad (262)
$$

Since $S_M^{\frac{1}{M}} = \sqrt{2\pi e}$, evaluating Eq. (262) at the saddle point, we obtain

$$
\begin{aligned}
\frac{\hat{G}_{2,l}}{n} &= \frac{1}{2}\{\ln\frac{2\pi}{E_l+F_l} + E_] + \frac{F_l-\bar{R}_l^2}{E_l+F_l}\} \qquad (263) \\
\frac{\tilde{G}_2}{n} &= M\sum_l \frac{\hat{G}_{2,l}}{n}. \qquad (264)
\end{aligned}
$$

Next, we calculate G_3:

$$\tilde{G_3} = iM \sum_{l\alpha} \bar{R}_l^{\alpha} R_l - M \sum_{l,\alpha<\beta} F_l^{\alpha\beta} q_l^{\alpha\beta}. \tag{265}$$

From the RS ansatz, we have

$$\tilde{G_3} = M\{i \sum_l \bar{R}_l R_l + \frac{1}{2} \sum_l F_l q_l\}. \tag{266}$$

We define $\tilde{G_3} = MG_3$. Therefore, we have

$$\langle \ln Z \rangle = \frac{1}{n}(\tilde{G_1} + \tilde{G_2} + \tilde{G_3}) = \frac{1}{n}\tilde{G}, \tag{267}$$

$$\frac{1}{M}\langle \ln Z \rangle = \frac{1}{nM}\tilde{G} = \frac{1}{n}(\frac{p}{M}G_1 + \sum_l \hat{G}_{2,l} + G_3). \tag{268}$$

Let us define

$$G = \tilde{G_1} + \tilde{G_2} + \tilde{G_3}. \tag{269}$$

Defining $N = MK$ and $\alpha = \frac{p}{N}$, we have

$$\begin{aligned} \frac{G}{M} &= \frac{p}{M}G_1 + \sum_l \hat{G}_{2,l} + G_3 = \frac{K}{N}pG_1 + \sum_l \hat{G}_{2,l} + G_3 \\ &= K\left[\alpha G_1 + \frac{1}{K}\sum_{l=1}^{K} \hat{G}_{2,l} + \frac{G_3}{K}\right] \end{aligned} \tag{270}$$

$$\frac{G}{nN} = \alpha\frac{G_1}{n} + \frac{1}{K}\sum_{l=1}^{K}\frac{\hat{G}_{2,l}}{n} + \frac{1}{n}\frac{G_3}{K}. \tag{271}$$

We define

$$G_2 = \frac{1}{K}\sum_{l=1}^{K} \hat{G}_{2,l} + \frac{G_3}{K}. \tag{272}$$

Thus, we obtain $\frac{G_1}{n}$ and $\frac{G_2}{n}$ as

$$\begin{aligned} \frac{G_1}{n} &= \Big[\int \prod_l Dt_l\Big] \sum_{\sigma=\pm 1} \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^{\sigma}(\{\sigma_l\}) \Big[\prod_l H\Big(\frac{R_l\sigma_l t_l}{\sqrt{q_l(1+\tau^2) - R_l^2}}\Big)\Big] \\ &\quad \times \ln \Phi_{\beta,s}^{\sigma}(\{\gamma_l t_l\}), \end{aligned} \tag{273}$$

$$\frac{G_2}{n} = \frac{1}{K}\sum_{l=1}^{K}\frac{1}{2}\left(\ln\frac{2\pi}{E_l+F_l} + E_l + \frac{F_l - \bar{R}_l^2}{E_l+F_l}\right) + \frac{1}{K}(i\sum_l \bar{R}_l R_l + \sum_l \frac{F_l q_l}{2}). \tag{274}$$

We calculate the saddle points for $\frac{G}{nN} = \frac{1}{n}\alpha G_1 + \frac{G_2}{n}$. Using the conditions for the saddle point of G_2 with respect to E_l, F_l and $\bar{R}$, we obtain

$$\frac{G_2}{n} = \frac{1}{2K}\sum_l^K \Big[\ln\Big(2\pi(1-q_l)\Big) + 1 + \frac{q_l - R_l^2}{1-q_l}\Big]. \tag{275}$$

By defining the quantities

$$\eta = 1+\tau^2,\ \tilde{\zeta}_l = \frac{R_l}{\sqrt{\eta q_l - R_l^2}},\ \tilde{X}_l = \tilde{\zeta}_l t_l, \tag{276}$$

$$\gamma_l = \sqrt{\frac{q_l}{1-q_l}},\ Y_l = \gamma_l t_l, \tag{277}$$

from Eq. (273) we obtain

$$\frac{G_1}{n} = \int [\prod_l Dt_l] \sum_{\sigma=\pm 1} [\mathrm{Tr}_{\{\sigma_l\}}\ \triangle_t^\sigma\ (\{\sigma_l\}) \prod_l H(\sigma_l \tilde{X}_l)] \ln \Phi^\sigma_{\beta,s}(\{Y_l\}). \tag{278}$$

$\Phi^\sigma_{\beta,s}(\{Y_l\})$ is expressed as

$$\Phi^\sigma_{\beta,s}(\{Y_l\}) = \mathrm{Tr}_{\{\sigma_l'\}} \tilde{\triangle}^{\ \sigma}_{s,\beta}(\{\sigma_l'\}) \prod_l H(\sigma_l' Y_l). \tag{279}$$

9.2 Output noise

In the output noise case, the sign of the output σ is reversed with a nonzero probability. We assume that only the teacher suffers from output noise. Let λ be the probability that the teacher's output is reversed. Then, the probability that teacher's outputs a value of σ is

$$\begin{aligned} p(\sigma|\{h_l^t\}) &= (1-\lambda)\ \triangle_t^\sigma\ \big(\{\mathrm{sgn}(h_l^t)\}\big) + \lambda\{1 - \triangle_t^\sigma\big(\{\mathrm{sgn}(h_l^t)\}\big)\} \\ &= \lambda + (1-2\lambda)\mathrm{Tr}_{\{\sigma_l\}}\ \triangle_t^\sigma\ (\{\sigma_l\}) \prod_l \Theta(\sigma_l h_l^t). \end{aligned} \tag{280}$$

Let us calculate e^{G_1} in Eq. (233). We obtain

$$\begin{aligned} \int dh_l^t e^{L_l} \Theta(\sigma_l h_l^t) &= \sqrt{\frac{2\pi(q_l - R_l^2)}{q_l}} e^{\frac{q_l - R_l^2}{2q_l}\left(\frac{(1-q_l)R_l}{q_l - R_l^2} b_l\right)^2} H(-\sigma_l \tilde{X}_l) \\ &= \sqrt{\frac{2\pi(q_l - R_l^2)}{q_l}} e^{A_l} H(-\sigma_l X_l), \end{aligned} \tag{281}$$

where A_l is given by Eq. (242). Therefore, we have

$$[\prod_l \int dh_l^t e^{L_l}] p(\sigma|\{h_l^t\}) = \left\{\prod_l \sqrt{\frac{2\pi(q_l - R_l^2)}{q_l}} e^{A_l}\right\} \Big[\lambda + (1-2\lambda)\mathrm{Tr}_{\{\sigma_l\}}\Big(\triangle_t^\sigma(\{\sigma_l\}) \prod_l H(-\sigma_l X_l)\Big)\Big]. \tag{282}$$

As in the case with the input noise model, making a variable transformation from b_l to $t_l = \dfrac{1-q_l}{\sqrt{q_l}} b_l$ yields

$$e^{G_1} = [\prod_l \int Dt_l] \sum_{\sigma=\pm 1} \left(\lambda + (1-2\lambda) \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(-\sigma_l X_l) \right) \left(\Phi^\sigma_{\beta,s}(\{-\sqrt{\frac{q_l}{1-q_l}} t_l\}) \right)^n \quad (283)$$

By defining the quantities

$$\zeta_l = \frac{R_l}{\sqrt{q_l - R_l^2}}, \quad (284)$$

$$Y_l = \gamma_l t_l \quad , \quad \gamma_l = \sqrt{\frac{q_l}{1-q_l}}, \quad (285)$$

$X_l = \zeta_l t_l$ follows and we have

$$e^{G_1} = [\prod_l \int Dt_l] \sum_{\sigma=\pm 1} \{ \lambda + (1-2\lambda) \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(-\sigma_l X_l) \} \{ \Phi^\sigma_{\beta,s}(\{-Y_l\}) \}^n \quad (286)$$

By taking the limit $n \to 0$, we obtain

$$\frac{G_1}{n} = [\prod_l \int Dt_l] \sum_{\sigma=\pm 1} \{ \lambda + (1-2\lambda) \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(-\sigma_l X_l) \} \ln \Phi^\sigma_{\beta,s}(\{-Y_l\}). \quad (287)$$

Making a variable change from t_l to $-t_l$ yields

$$\frac{G_1}{n} = [\prod_l \int Dt_l] \sum_{\sigma=\pm 1} \{ \lambda + (1-2\lambda) \mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(\sigma_l X_l) \} \ln \Phi^\sigma_{\beta,s}(\{Y_l\}) \quad (288)$$

The expression for G_2 is the same as that for the input noise model.

9.3 Noiseless case

From Eqs. (276), (277), (284) and (285), γ_l is the same as the input and output noise models. ζ_l is obtained by setting $\tau = 0$ in Eq. (276) for the input noise case. This is also the same as in the output noise case. On the other hand, the expression for G_1 is obtained by substituting $\lambda = 0$ in the expression for the output noise case,

$$\frac{G_1}{n} = [\prod_l \int Dt_l] \sum_{\sigma=\pm 1} \left(\mathrm{Tr}_{\{\sigma_l\}} \triangle_t^\sigma (\{\sigma_l\}) \prod_l H(\sigma_l X_l) \right) \ln \Phi^\sigma_{\beta,s}(\{Y_l\}). \quad (289)$$

This is also the same as in the input noise case with $\eta = 0$. G_2 is the same as that in the input and output noise cases.

10. Appendix B. Derivation of the SPEs for Parity Machine

10.1 Input noise

Since $\sigma = 1$ for $\sigma_1 = \sigma_2$ and $\sigma = -1$ for $\sigma_1 = -\sigma_2$, $\frac{G_1}{n}$ in Eq. (278) becomes

$$\frac{G_1}{n} = \int Dt_1 Dt_2 \{ H(\tilde{X}_1) H(\tilde{X}_2) + H(-\tilde{X}_1) H(-\tilde{X}_2) \} \ln \Phi_+(Y_1, Y_2)$$

$$+\{H(-\tilde{X}_1)H(\tilde{X}_2)+H(\tilde{X}_1)H(-\tilde{X}_2)\}\ln\Phi_-(Y_1,Y_2) \quad (290)$$

$$= \int Dt_1Dt_2[H_2(\tilde{X}_1,\tilde{X}_2)\ln\Phi_+(Y_1,Y_2)+H_2(\tilde{X}_1,-\tilde{X}_2)]\ln\Phi_-(Y_1,Y_2). \quad (291)$$

Here, we define

$$\Phi_+(Y_1,Y_2) = \Phi^1_{\beta,s}(Y_1,Y_2) \quad (292)$$

$$\Phi_-(Y_1,Y_2) = \Phi^{-1}_{\beta,s}(Y_1,Y_2), \quad (293)$$

$$H_2(x,y) = H(x)H(y)+H(-x)H(-y). \quad (294)$$

The following relations hold:

$$H_2(x,y)=H_2(y,x)=H_2(-x,-y),\ H_2(x,-y)=1-H_2(x,y)=H_2(-x,y). \quad (295)$$

Using these relations, from Eq. (279), we obtain

$$\begin{aligned}\Phi_+(Y_1,Y_2) &= H(Y_1)H(Y_2)+H(-Y_1)H(-Y_2)+e^{-\beta}\{H(Y_1)H(-Y_2)+H(-Y_1)H(Y_2)\}\\ &= H_2(Y_1,Y_2)+e^{-\beta}H_2(Y_1,-Y_2) \quad (296)\\ \Phi_-(Y_1,Y_2) &= H_2(-Y_1,Y_2)+e^{-\beta}H_2(Y_1,Y_2)=\Phi_+(-Y_1,Y_2)=\Phi_+(Y_1,-Y_2)\\ &= 1+e^{-\beta}-\Phi_+(Y_1,Y_2). \quad (297)\end{aligned}$$

Then, $\frac{G_1}{n}$ becomes

$$\frac{G_1}{n} = \int Dt_1Dt_2[H_2(\tilde{X}_1,\tilde{X}_2)\ln\Phi_+(Y_1,Y_2)+H_2(\tilde{X}_1,-\tilde{X}_2)\ln\Phi_+(Y_1,-Y_2)]. \quad (298)$$

By the variable change t_2 to $-t_2$, the second term on the right-hand side of this equation is equal to the first term. Thus, we have

$$\frac{G_1}{n} = 2\int Dt_1Dt_2H_2(\tilde{X}_1,\tilde{X}_2)\ln\Phi_+(Y_1,Y_2). \quad (299)$$

G_2 is given by Eq. (275). Thus, we have the following SPEs:

$$q_l^2-R_l^2 = 4\alpha(1-e^{-\beta})\sqrt{q_l(1-q_l)}I_{1,l},\ l=1,2, \quad (300)$$

$$R_l(\eta q_l-R_l^2)^{\frac{3}{2}} = -4\alpha\eta q_l(1-q_l)I_{2,l},\ l=1,2. \quad (301)$$

$$I_{1,1} = \int Dt_1Dt_2t_1H_2(\tilde{X}_1,\tilde{X}_2)h(Y_1)H_a(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)}, \quad (302)$$

$$I_{2,1} = \int Dt_1Dt_2t_1h(\tilde{X}_1)H_a(\tilde{X}_2)\ln\Phi_+(Y_1,Y_2), \quad (303)$$

$$I_{1,2} = \int Dt_1Dt_2t_2H_2(\tilde{X}_1,\tilde{X}_2)h(Y_2)H_a(Y_1)\frac{1}{\Phi_+(Y_1,Y_2)}, \quad (304)$$

$$I_{2,2} = \int Dt_1Dt_2t_2h(\tilde{X}_2)H_a(\tilde{X}_1)\ln\Phi_+(Y_1,Y_2). \quad (305)$$

We assume that $q_l = q$ and $R_l = R$ $(l = 1, 2)$. Then, we have $\tilde{X}_l = \tilde{\zeta} t_l$, $\tilde{\zeta} = \frac{R}{\sqrt{\eta q - R^2}}$, $Y_l = \gamma t_l$ and $\gamma = \sqrt{\frac{q}{q-1}}$. By exchanging integration variables $t_1 \leftrightarrow t_2$ in $I_{1,2}$, we have $I_{1,1} = I_{1,2}$. Similarly, in $I_{2,2}$, by using $\Phi_+(Y_1, Y_2) = \Phi_+(Y_2, Y_1)$, we have

$$I_{2,2} = \int Dt_1 Dt_2 t_2 h(\tilde{X}_2) H_a(\tilde{X}_1) \ln \Phi_+(Y_2, Y_1). \tag{306}$$

By exchanging integration variables $t_1 \leftrightarrow t_2$ in $I_{2,2}$, we have $I_{2,1} = I_{2,2}$. We define

$$I_1^{(\mathrm{i})} \equiv 2I_{1,1} = 2 \int Dt_1 Dt_2 t_1 h(Y_1) H_a(Y_2) \frac{H_2(\tilde{X}_1, \tilde{X}_2)}{\Phi_+(Y_1, Y_2)}, \tag{307}$$

$$I_2^{(\mathrm{i})} \equiv 2I_{2,1} = 2 \int Dt_1 Dt_2 t_1 h(\tilde{X}_1) H_a(\tilde{X}_2) \ln \Phi_+(Y_1, Y_2). \tag{308}$$

Thus, from Eqs. (300) and (301), the SPEs become

$$q^2 - R^2 = 2\alpha(1 - e^{-\beta}) \sqrt{q(1-q)}\ I_1^{(\mathrm{i})}, \tag{309}$$

$$R(\eta q - R^2)^{3/2} = -2\alpha\eta q(1-q)\ I_2^{(\mathrm{i})}. \tag{310}$$

10.2 Output noise

From Eq. (278), $\frac{G_1}{n}$ is

$$\begin{aligned} \frac{G_1}{n} &= \int Dt_1 Dt_2 \Big\{ \big\{ \lambda + (1-2\lambda) H_2(X_1, X_2) \big\} \ln \Phi_+(Y_1, Y_2) \\ &\quad + \big\{ \lambda + (1-2\lambda) H_2(X_1, -X_2) \big\} \ln \Phi_-(Y_1, Y_2) \Big\} \\ &= \int Dt_1 Dt_2 \Big\{ \big\{ \lambda + (1-2\lambda) H_2(X_1, X_2) \big\} \ln \Phi_+(Y_1, Y_2) \\ &\quad + \big\{ \lambda + (1-2\lambda) H_2(X_1, -X_2) \big\} \ln \Phi_+(Y_1, -Y_2) \Big\}. \end{aligned} \tag{311}$$

Here, we use $\Phi_-(Y_1, Y_2) = \Phi_+(Y_1, -Y_2)$. As in the input noise case, this is rewritten as

$$\frac{G_1}{n} = 2 \int Dt_1 Dt_2 \big\{ \lambda + (1-2\lambda) H_2(X_1, X_2) \big\} \ln \Phi_+(Y_1, Y_2). \tag{312}$$

From Eq. (312), we have

$$\frac{G_1}{n} = 2 \int Dt_1 Dt_2 \big\{ \lambda + (1-2\lambda) H_2(X_1, X_2) \big\} \ln \Phi_+(Y_1, Y_2). \tag{313}$$

We assume $q_l = q$ and $R_l = R$ for $l = 1, 2$. Then, we have $X_l = \zeta t_l$, $\zeta = \frac{R}{\sqrt{q - R^2}}$, $Y_l = \gamma t_l$ and $\gamma = \sqrt{\frac{q}{1-q}}$. We define $I_1^{(\mathrm{o})}$ and $I_2^{(\mathrm{o})}$ as

$$I_1^{(\mathrm{o})} = 2 \int Dt_1 \int Dt_2 \{ \lambda + (1-2\lambda) H_2(X_1, X_2) \} t_1 \frac{h(Y_1) H_a(Y_2)}{\Phi_+(Y_1, Y_2)}, \tag{314}$$

$$I_2^{(\mathrm{o})} = 2 \int Dt_1 \int Dt_2 (1-2\lambda) h(X_1) H_a(X_2) t_1 \ln \Phi_+(Y_1, Y_2). \tag{315}$$

The SPEs are given by

$$q^2 - R^2 = 2\alpha(1-e^{-\beta})\sqrt{q(1-q)}\ I_1^{(\mathrm{o})}, \tag{316}$$

$$R(q-R^2)^{3/2} = -2\alpha q(1-q)\ I_2^{(\mathrm{o})}. \tag{317}$$

10.3 Noiseless case

In this case, quantities are obtained by setting $\tau = 0$ in the expressions for the input noise, or setting $\lambda = 0$ in the output noise. Thus, assuming $q_l = q$ and $R_l = R$ for $l = 1, 2$, we obtain

$$I_1^{(\mathrm{n})} = 2\int Dt_1 \int Dt_2 H_2(X_1, X_2) t_1 \frac{h(Y_1)H_a(Y_2)}{\Phi_+(Y_1, Y_2)}, \tag{318}$$

$$I_2^{(\mathrm{n})} = 2\int Dt_1 \int Dt_2 h(X_1) H_a(X_2) t_1 \ln \Phi_+(Y_1, Y_2), \tag{319}$$

$$q^2 - R^2 = 2\alpha(1-e^{-\beta})\sqrt{q(1-q)}\ I_1^{(\mathrm{n})}, \tag{320}$$

$$R(q-R^2)^{3/2} = -2\alpha q(1-q)\ I_2^{(\mathrm{n})}. \tag{321}$$

11. Appendix C. Derivation of the SPEs for the And Machine

In this Appendix, we derive the SPEs for the And machine. We assume $q_l = q$ and $R_l = R$ for $l = 1, 2$. $\tilde{X}_l$, X_l and Y_l are the same as those for the Parity machine.

11.1 Input noise

$\frac{G_1}{n}$ in Eq. (278) is

$$\begin{aligned}\frac{G_1}{n} &= \int Dt_1 Dt_2 \{H(\tilde{X}_1)H(\tilde{X}_2)\ln\Phi_+(Y_1,Y_2) \\ &\quad +(H(\tilde{X}_1)H(-\tilde{X}_2)+H(-\tilde{X}_1)H(\tilde{X}_2)+H(-\tilde{X}_1)H(-\tilde{X}_2))\ln\Phi_-(Y_1,Y_2)\} \\ &= \int Dt_1 Dt_2 \{H(\tilde{X}_1)H(\tilde{X}_2)\ln\Phi_+(Y_1,Y_2) + \hat{H}(\tilde{X}_1,\tilde{X}_2)\ln\Phi_-(Y_1,Y_2)\},\end{aligned} \tag{322}$$

where $\hat{H}(\tilde{X}_1, \tilde{X}_2) = 1 - H(\tilde{X}_1)H(\tilde{X}_2)$. From Eq. (279), we obtain

$$\begin{aligned}\Phi_+(Y_1,Y_2) &= H(Y_1)H(Y_2) + e^{-\beta}\hat{H}(Y_1,Y_2) \\ &= e^{-\beta} + (1-e^{-\beta})H(Y_1)H(Y_2) = \Phi_+(Y_2,Y_1),\end{aligned} \tag{323}$$

$$\Phi_-(Y_1,Y_2) = \hat{H}(Y_1,Y_2) + e^{-\beta}H(Y_1)H(Y_2) = 1 + e^{-\beta} - \Phi_+(Y_1,Y_2). \tag{324}$$

Thus, $\frac{G_1}{n}$ is

$$\frac{G_1}{n} = \int Dt_1 Dt_2 \{H(\tilde{X}_1)H(\tilde{X}_2)\ln\Phi_+(Y_1,Y_2) + \hat{H}(\tilde{X}_1,\tilde{X}_2)\ln\Phi_-(Y_1,Y_2)\}. \tag{325}$$

G_2 is

$$\frac{G_2}{n} = \frac{1}{2}\Big(\ln(2\pi(1-q)) + \frac{1-R^2}{1-q} \Big). \tag{326}$$

The SPEs become

$$\begin{aligned}
q^2 - R^2 &= \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{i})}, && (327)\\
R(\eta q - R^2)^{\frac{3}{2}} &= -\alpha\eta q(1-q)J_2^{(\mathrm{i})}, && (328)\\
J_1^{(\mathrm{i})} &= -2\int Dt_1 Dt_2 t_1 h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\
&\quad \times\{\Phi_+(Y_1,Y_2) - (1+e^{-\beta})H(\tilde{X}_1)H(\tilde{X}_2)\}, && (329)\\
J_2^{(\mathrm{i})} &= -2\int Dt_1 Dt_2 h(\tilde{X}_1)H(\tilde{X}_2)t_1 \ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). && (330)
\end{aligned}$$

11.2 Output noise

Similar to the input noise case, from Eq. (278), we have

$$\begin{aligned}
\frac{G_1}{n} &= \int Dt_1 Dt_2 \Big[\big\{\lambda + (1-2\lambda)H(X_1)H(X_2)\big\} \ln \Phi_+(Y_1,Y_2) \\
&\quad + \big\{\lambda + (1-2\lambda)\hat{H}(X_1,X_2)\big\} \ln \Phi_-(Y_1,Y_2) \Big]. && (331)\\
& && (332)
\end{aligned}$$

The SPEs are

$$\begin{aligned}
q^2 - R^2 &= \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{o})}, && (333)\\
R(q - R^2)^{\frac{3}{2}} &= -\alpha q(1-q)J_2^{(\mathrm{o})}, && (334)\\
J_1^{(\mathrm{o})} &= 2\int Dt_1 Dt_2 t_1 h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\
&\quad \times[\Phi_+(Y_1,Y_2) - (1+e^{-\beta})\{\lambda + (1-2\lambda)H(X_1)H(X_2)], && (335)\\
J_2^{(\mathrm{o})} &= -2(1-2\lambda)\int Dt_1 Dt_2 h(X_1)H(X_2)t_1 \ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). && (336)
\end{aligned}$$

11.3 Noiseless case

Since the quantities for the noiseless case are obtained by substituting $\tau = 0$ in the expressions for the input noise or substituting $\lambda = 0$ in the expressions for the output noise, we have

$$\begin{aligned}
J_1^{(\mathrm{n})} &= -2\int Dt_1 Dt_2 t_1 h(Y_1)H(Y_2)\frac{1}{\Phi_+(Y_1,Y_2)\Phi_-(Y_1,Y_2)} \\
&\quad \times[\Phi_+(Y_1,Y_2) - (1+e^{-\beta})H(X_1)H(X_2)], && (337)
\end{aligned}$$

$$J_2^{(\mathrm{n})} = -2\int Dt_1 Dt_2 h(X_1)H(X_2)t_1 \ln\left(\frac{\Phi_-(Y_1,Y_2)}{\Phi_+(Y_1,Y_2)}\right). \tag{338}$$

The SPEs are

$$q^2 - R^2 = \alpha(1-e^{-\beta})\sqrt{q(1-q)}J_1^{(\mathrm{n})}, \tag{339}$$

$$R(q-R^2)^{\frac{3}{2}} = -\alpha q(1-q)J_2^{(\mathrm{n})}. \tag{340}$$