Journal of the Physical Society of Japan
Vol. 71, No. 8, August, 2002, pp. 1882–1904

On the Conditions for the Existence of Perfect Learning and Power Law Behaviour in Learning from Stochastic Examples by Ising Perceptrons

Tatsuya UEZU*

Graduate School of Human Culture, Nara Women's University, Nara 630-8506

(Received November 19, 2001)

In a previous work, we studied learning from stochastic examples by perceptrons with Ising weights in the framework of statistical mechanics. Employing the one-step replica symmetry breaking ansatz, types of behaviour of learning curves were classified according to a certain local property of the rules by which examples were drawn. Further, the conditions for the existence of the perfect learning, together with other behaviour of the learning curves, were given. In this paper, we give a detailed derivation of these results and a further argument regarding perfect learning. We also present the results of extensive numerical calculations.

KEYWORDS: Ising perceptron, perfect learning, replica method
DOI: 10.1143/JPSJ.71.1882

1. Introduction

In study of the problem of supervised learning from examples by feed forward networks, learning curves of the generalization error ϵ_g have been derived for various types of networks.[1] The generalization error is a false prediction of students, and is an indicator of the accomplishment of learning. From these studies, it came to be known that when the number of examples p is large relative to the number of synaptic weights N, that is, when $\alpha = p/N$ is large, the learning curves exhibit only a few types of behaviour[2–9] For example, learning curves of networks with continuous weights all exhibit power law behaviour as

$$(\epsilon_g - \epsilon_{min}) \propto \alpha^{-\gamma},$$

where γ depends on several properties of the system including architecture and type of weight vectors, and ϵ_{min} is the minimum generalization error. On the other hand, in the case of discrete weights, it has been shown that, in addition to power law behaviour, there exists perfect learning (PL) for deterministic and realizable cases.[10,11] Here, 'deterministic' means that there exists no noise in the process of learning, and 'realizable' means that the teacher's vector is also an element of the set of all students' vectors. Perfect learning is defined as the situation in which all of the students weight vectors coincide with the teacher's weight vector for a finite value of α.

If perfect learning does not take place, students can never realize the state of the teacher, but can only approach to it. In the case that perfect learning takes place, the learning procedure is completed at finite α. If we consider the situation that the dimension N of weight vectors is finite but sufficiently large, in order for students to complete learning, a very large number of examples $p(p \gg N)$ is necessary in the former case, while only a number of examples of order N is necessary in the latter case. Perfect learning can take place even in the presence of external noise. Therefore, it is important to determine the conditions for the existence of PL in the case of discrete weights and in the presence of external noise.

In a previous paper,[12] we derived these conditions. The results we obtained there are similar to those found by Seung,[13] who classified the learning behaviour of Ising networks by introducing two quantities y and z that characterize two important statistical properties of the system. We deduced a different meaning for them and determined the manner in which they are related. In addition, in that paper, we investigated the asymptotic behaviour of the learning curve.

The purpose of this paper is to present a detailed derivation of the results obtained in ref. 12. Using the replica method, we determine the necessary and sufficient conditions for the existence of perfect learning and conditions in the asymptotic region of $\alpha \gg 1$ on the appearance of power law learning curves in terms of δ, which represents a certain local property of the rules by which examples are drawn.

We now describe the basic features of our model and summarize our results.

We study supervised learning of stochastic relations by Ising perceptrons. We consider a stochastic target relation between an N-dimensional input vector $\boldsymbol{x}$ and a binary output $r \in \{1, -1\}$, which is represented by a conditional probability $p_r(r|\boldsymbol{x})$. Input vectors $\boldsymbol{x}$ are normalized as $|\boldsymbol{x}| = \sqrt{N}$, and $p_r(r|\boldsymbol{x})$ is defined as a function of the inner product of the input $\boldsymbol{x}$ and the optimal weight $\boldsymbol{w}^o$ as

$$p_r(+1|\boldsymbol{x}) = \mathscr{P}(u^o) = \frac{1 + P(u^o)}{2}, \quad u^o \equiv (\boldsymbol{x} \cdot \boldsymbol{w}^o)/\sqrt{N}. \tag{1}$$

The output produced with this stochastic relation is considered to be the output of a perceptron that is subject to external noise. We call this target perceptron the "teacher perceptron", with the weight vector $\boldsymbol{w}^o$. Students are pure perceptrons; that is, a student perceptron with a weight vector $\boldsymbol{w}$ answers $r = \pm 1$ to the input vector $\boldsymbol{x}$ according to the rule

$$r = \mathrm{sgn}(u),$$

where, $\boldsymbol{w}$ is an N-dimensional vector and $u \equiv (\boldsymbol{x} \cdot \boldsymbol{w})/\sqrt{N}$. In this paper, we consider Ising perceptrons, for which the components of the weight vectors $\boldsymbol{w}^o$ and $\boldsymbol{w}$ take discrete values ± 1. Further, we choose the function $P(u)$ to be non-

*E-mail: uezu@ki-rin.phys.nara-wu.ac.jp

decreasing w.r.t. u and to behave as

$$P(u) \simeq a\,\mathrm{sgn}(u)|u|^{\delta}(\delta \geq 0) \tag{2}$$

near $u = 0$. Further, $P(-u) = -P(u)$ is stipulated, for simplicity. The case $\delta = 0$ corresponds to the output noise model[4] in which the output of the target perceptron stochastically changes sign through the influence of noise according to some probability distribution. The case $\delta = 1$ corresponds to the input noise model,[10] in which the input of the target perceptron is corrupted by Gaussian noise with zero mean. The Gibbs algorithm with temperature T is used as the learning algorithm.

With the model described above, we obtained the following results using the statistical mechanical method, i.e. the replica method.

Conditions for PL

The necessary and sufficient conditions for the existence of perfect learning are the following.

(1) $0 \leq \delta < 1/2$ when $0 < \beta < \infty$, where β is the inverse temperature.
(2) Deterministic case. That is, the target relation is deterministic, obeying the perceptron rule, and the alogorithm is the minimum-error algorithm, i.e. the Gibbs algorithm with $\beta \to \infty$.

Behaviour of learning curves

Employing the replica symmetric (RS) ansatz and the one-step replica symmetry breaking (1RSB) ansatz, we found that the behaviour of the generalization error ϵ_g can be classified into the following three categories, according to the value of δ.

(1) For $0 \leq \delta < \frac{1}{2}$, at $\alpha = \alpha_{\max}$ a first-order phase transition from the RS solution with positive entropy or from the 1RSB solution to the PL solution takes place.
(2) For $\delta = \frac{1}{2}$ and large α, the 1RSB solution appears, and ϵ_g for the 1RSB solution decays exponentially according to

$$\Delta\epsilon_g \propto e^{-3F_0\alpha},$$

where $\Delta\epsilon_g \equiv \epsilon_g - \epsilon_{\min}$ and F_0 is a constant.
(3) For $\delta > \frac{1}{2}$ and large α, the 1RSB solution appears, and ϵ_g for this solution decays according to a power law with a logarithmic correction in accordance with

$$\Delta\epsilon_g \propto \left(\frac{\ln\alpha}{\alpha}\right)^{\frac{1+\delta}{2\delta-1}}.$$

In the following section, we formulate the problem. In §3, we analyse the RS solution. The conditions for the existence of PL are derived in §4. The 1RSB solution is studied in §5. The results of numerical calculations are given in §6. Section 7 is devoted to summary and discussion.

2. Formulation

A set of p examples $\xi_p = \{(\boldsymbol{x}^1, r_1^{\mathrm{o}}), (\boldsymbol{x}^2, r_2^{\mathrm{o}}), \ldots, (\boldsymbol{x}^p, r_p^{\mathrm{o}})\}$ is obtained as follows. The vectors $\boldsymbol{x}^\mu$ are chosen randomly and independently from a uniform distribution on a hypersphere of radius $\sqrt{N}$ centered at the origin in an N-dimensional Euclidean space, and $r_\mu^{\mathrm{o}} (= 1 \text{ or } = -1)$ is output with the conditional probability $p_r(r_\mu^{\mathrm{o}}|\boldsymbol{x}^\mu)$ for each $\boldsymbol{x}^\mu$. For a given realization of examples ξ_p, we define the energy "$E[w, \xi_p]$" of a student with a weight vector $\boldsymbol{w}$ as the number of false predictions, which is given as

$$E[w, \xi_p] = \sum_{\mu=1}^{p} \Theta(-r_\mu^{\mathrm{o}} u_\mu), \quad u_\mu \equiv (\boldsymbol{x}^\mu \cdot \boldsymbol{w})/\sqrt{N}, \tag{3}$$

where $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. The learning performance is represented by the generalization error ϵ_g, which is defined as

$$\begin{aligned}\epsilon_g &\equiv \langle \mathcal{P}(u^{\mathrm{o}})(1 - \Theta(u)) + (1 - \mathcal{P}(u^{\mathrm{o}}))\Theta(u)\rangle \qquad (4)\\ &= \epsilon_{\min} + 2\int_0^\infty DyP(y)H\left(\frac{Ry}{\sqrt{1-R^2}}\right),\\ \epsilon_{\min} &= \frac{1}{2} - \int_0^\infty DyP(y).\end{aligned}$$

Here, $\langle\cdots\rangle$ represents the average over examples and $\epsilon_{\min}$ is the minimum value of the generalization error, obtained with the optimal weight $\boldsymbol{w}^{\mathrm{o}}$. R is the overlap between the optimal weight vector and the weight vector of a student, $R = (\boldsymbol{w}^{\mathrm{o}} \cdot \boldsymbol{w})/N$. Also, as usual, we use $Dy = \exp(-y^2/2)dy/\sqrt{2\pi}$ and $H(x) = \int_x^\infty Dy$. From the eq. (4), we find that, in particular, when $\Delta R \equiv 1 - R$ is small, the relation

$$\Delta\epsilon_g = (\epsilon_g - \epsilon_{\min}) \simeq \frac{2s}{(1+\delta)\sqrt{2\pi}}(2\Delta R)^{\frac{1+\delta}{2}}, \tag{5}$$

where $s \equiv a\int_0^\infty Dyy^{1+\delta}$.

In this paper, we adopt the Gibbs algorithm with temperature T as the learning algorithm. The minimum-error algorithm, which minimizes the number of false predictions for the presented examples, is obtained by taking the $T \to 0$ limit.

From the "energy" defined by eq. (3), the partition function Z with inverse temperature β is given by

$$Z = \mathrm{Tr}_w\, e^{-\beta E[w,\xi_p]} = \mathrm{Tr}_w\, \Pi_{\mu=1}^{p}[e^{-\beta} + (1 - e^{-\beta})\Theta(r_\mu u_\mu)],$$

where Tr_w represents a summation over all possible $\boldsymbol{w}$. The average free energy f per synaptic weight is calculated using the standard recipe,

$$-\beta Nf = \langle \ln Z\rangle_{\xi_p, w^{\mathrm{o}}} = \lim_{n\to 0}\frac{1}{n}(\langle Z^n\rangle_{\xi_p, w^{\mathrm{o}}} - 1),$$

where $\langle\cdots\rangle_{\xi_p, w^{\mathrm{o}}}$ denotes the average over quenched variables.

The quantity $\langle Z^n\rangle_{\xi_p, w^{\mathrm{o}}}$ becomes a function of several replica order parameters, namely the overlap between weight vectors of students $q^{ab} = \frac{(\boldsymbol{w}^a \cdot \boldsymbol{w}^b)}{N}$, its conjugate $\hat{q}^{ab}$, the overlap between the weight vector of a student and the optimal weight vector $R^a = \frac{(\boldsymbol{w}^{\mathrm{o}} \cdot \boldsymbol{w}^a)}{N}$, and its conjugate $\hat{R}^a$. (See Appendix A for a derivation of the free energy.)

3. RS Solution

Let us consider the RS solution. For this solution, no quantity depends on the replica indices, and we write $q^{ab} = q, \hat{q}^{ab} = \hat{q}, R^a = R$ and $\hat{R}^a = \hat{R}$. Then, the RS free energy f_{RS} becomes

$$-\beta f_{\mathrm{RS}}(q, \hat{q}, R, \hat{R}, \beta) = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + \alpha K + I, \tag{6}$$

where

$$K \equiv \int Dy 2\mathcal{P}(y) \int Du \ln \tilde{H}\left(\frac{\sqrt{q-R^2}u - Ry}{\sqrt{1-q}}\right)$$
$$= \int Du \ln \tilde{H}(u/Q) E(u/Q), \tag{7}$$

$$I \equiv \int Dt \ln[2\cosh(\sqrt{\hat{q}}t + \hat{R})], \tag{8}$$

with

$$E(u) = \int Dy 2\mathcal{P}(-\Lambda) = 1 - e^{-v^2/2} \times \int_0^\infty Dy(e^{-vy} - e^{vy}) P(\xi y), \tag{9}$$
$$\tilde{H}(u) \equiv e^{-\beta} + (1 - e^{-\beta}) H(u),$$
$$\Lambda = \xi y + \sqrt{1-\xi^2} Q u = \xi(y - v),$$

and, finally,

$$\xi = \sqrt{1 - \frac{R^2}{q}}, \quad Q = \sqrt{\frac{1-q}{q}}, \quad v = -\frac{R}{\sqrt{q}\chi} u, \quad \chi = \frac{\xi}{Q}.$$

3.1 *Saddle point equations (S.P.E.)*

The saddle point equations are given by

$$q = \int Du \tanh^2(\sqrt{\hat{q}}u + \hat{R}), \tag{10}$$

$$R = \int Du \tanh(\sqrt{\hat{q}}u + \hat{R}), \tag{11}$$

$$\hat{q} = \frac{\alpha Q}{1-q} \int \tilde{D}u (\tilde{\varphi}(u))^2 E(u), \tag{12}$$

$$\hat{R} = -\frac{\alpha}{\sqrt{q-R^2}} \int \tilde{D}u \tilde{\varphi}(u) \int Dy y 2\mathcal{P}(\Lambda)$$
$$= -\frac{\alpha}{\sqrt{q-R^2}} \int \tilde{D}u \tilde{\varphi}(u) w(u), \tag{13}$$

where

$$w(u) \equiv \int Dy y P(\Lambda) = e^{-v^2/2} \times \int_0^\infty Dy P(\xi y)[(y+v)e^{-vy} + (y-v)e^{vy}], \tag{14}$$

$$\tilde{D}u = \frac{du}{\sqrt{2\pi}} e^{-Q^2u^2/2}, \quad \tilde{\varphi}(u) = \frac{\tilde{H}'(u)}{\tilde{H}(u)}.$$

For later use, we give the expression of the entropy $S_{\rm RS}$:

$$S_{\rm RS} = -\beta f_{\rm RS} - \alpha\beta e^{-\beta} J, \tag{15}$$

where

$$J = \int Dy 2\mathcal{P}(y) \int Du \frac{H\left(\frac{\sqrt{q-R^2}u - Ry}{\sqrt{1-q}}\right) - 1}{\tilde{H}\left(\frac{\sqrt{q-R^2}u - Ry}{\sqrt{1-q}}\right)}.$$

Defining L as $L = K - \beta e^{-\beta} J$, $S_{\rm RS}$ becomes

$$S_{\rm RS} = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I + \alpha L, \tag{16}$$

where L is expressed as

$$L = \int Du E(u/Q) \left\{ \ln[1 + (e^\beta - 1) H(u/Q)] - \beta \frac{H(u/Q)}{\tilde{H}(u/Q)} \right\}. \tag{17}$$

Also, the energy (training error) per synaptic weight is expressed as

$$e_t = -\alpha e^{-\beta} J. \tag{18}$$

3.2 *Numerical calculations of the S.P.E. for the RS solution*

Here, we give the results of numerical calculations for the RS solution.

Case (I): $\delta = 1$

Here, we considered the case $P(y) = 1 - 2H(y)$, for which $\epsilon_{\rm min} = \frac{1}{4}$. In Fig. 1, we display the α dependences for $T = 1$ of q, R and $S_{\rm RS}$, as well as Λ_1 and Λ_3, which are indicators of AT-stability. The RS solution is stable only when both Λ_1 and Λ_3 are negative. From the numerical results, it seems that as $\alpha \to \infty$, q and R tend to 1. In this case, the entropy $S_{\rm RS}$ becomes zero at some finite value of α, $\alpha_{\rm s}(T)$, and Λ_3 becomes zero at a different finite value of α, $\alpha_{\rm AT}(T)$, for any T.

Case (II): $\delta = 0$

In this case, for the numerical calculations, we considered $P(y) = \frac{1}{2}\,{\rm sgn}(y)$, for which $\epsilon_{\rm min} = \frac{1}{4}$. In Figs. 2–4, for several temperatures we display the α dependences of q, R, $S_{\rm RS}$, Λ_1 and Λ_3. The most interesting feature of these graphs is that there are no solutions for which q and R tend to 1 as $\alpha \to \infty$. As seen, there are two branches of solutions, which we call "branch I" and "branch II". Each solution is characterized

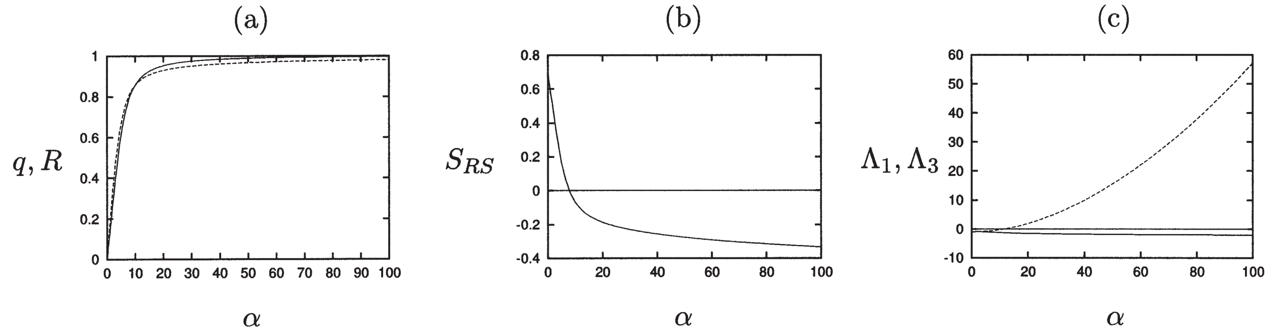

Fig. 1. The α dependences of several quantities for the RS solution with $\delta = 1$ and $T = 1$. (a) q and R (dotted curve). (b) $S_{\rm RS}$. (c) Λ_1 and Λ_3 (dotted curve).

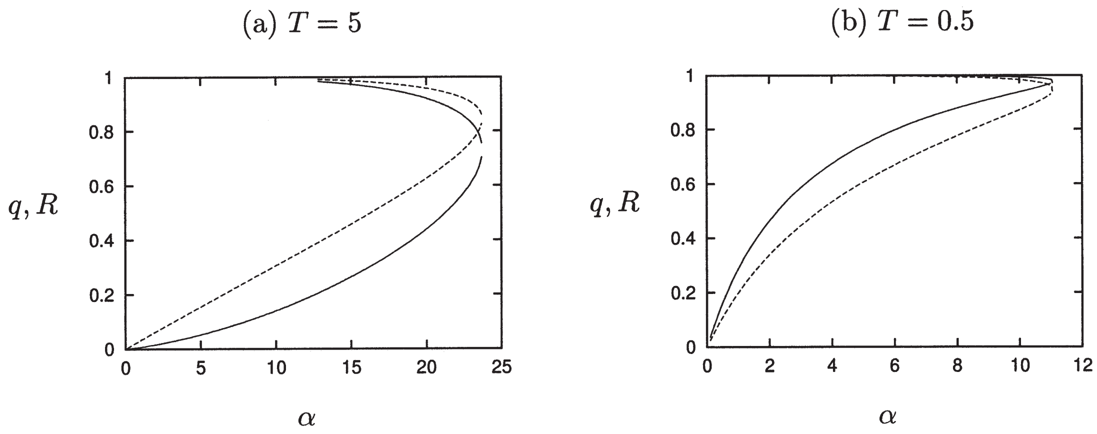

Fig. 2. The α dependences of q and R (dotted curve) for the RS solution with $\delta = 0$.

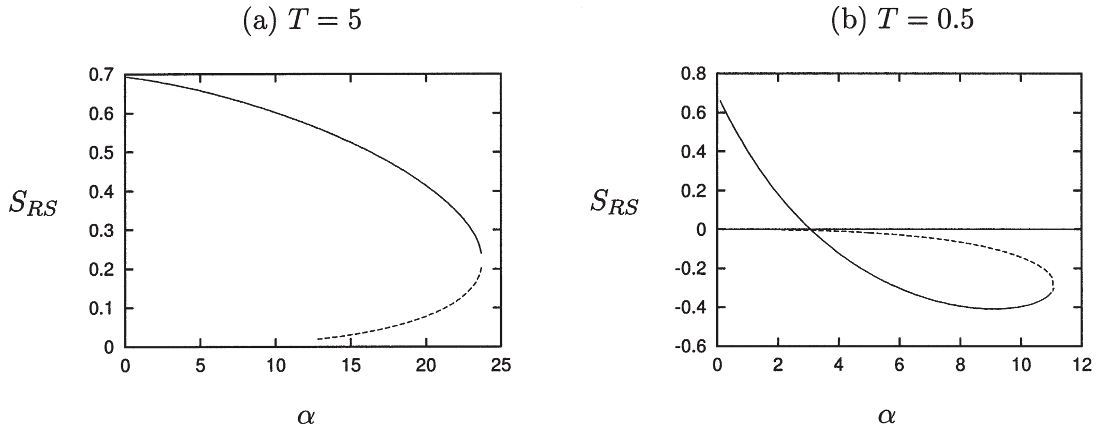

Fig. 3. The α dependence of the entropy for the RS solution with $\delta = 0$. The solid and dashed curves correspond to $S^{\mathrm{I}}_{\mathrm{RS}}$ and $S^{\mathrm{II}}_{\mathrm{RS}}$, respectively.

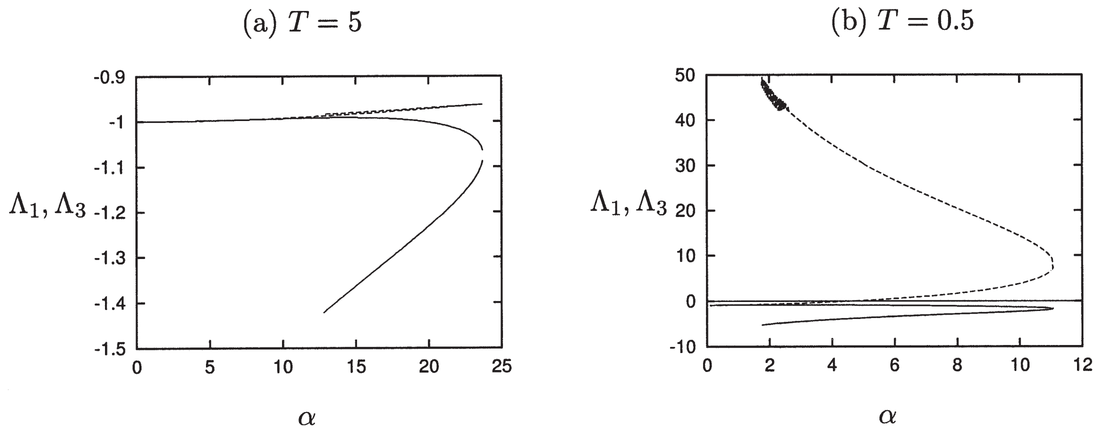

Fig. 4. The α dependences of Λ_1 and Λ_3 (dotted curve) for the RS solution with $\delta = 0$. Λ^{I}_1 and Λ^{I}_3 correspond to the curves starting from -1 for small α.

by its behaviour in the limit $\alpha \to 0$. In branch I, q and R tend to 0, while in branch II, q and R tend to 1. (We attach the superscript I and II to quantities evaluated in branches I and II.) From our numerical results, we found that when T is greater than some temperature, which we call T_{s}, solutions in both branches are AT-stable and their entropies are positive. However, when $T < T_{\mathrm{s}}$, the entropy of branch II, $S^{\mathrm{II}}_{\mathrm{RS}}$, is negative for any α. Since $S^{\mathrm{II}}_{\mathrm{RS}} = 0$ at $\alpha = 0$, T_{s} is determined by the condition that $S^{\mathrm{II}}_{\mathrm{RS}}$ changes sign for small α as T passes through T_{s}. Also, there exists a critical value of $\alpha = \alpha_{\mathrm{s}}(T)$, such that for $\alpha > \alpha_{\mathrm{s}}(T)$ we have $S^{\mathrm{I}}_{\mathrm{RS}} < 0$, and for $\alpha < \alpha_{\mathrm{s}}(T)$ we have $S^{\mathrm{I}}_{\mathrm{RS}} > 0$. In addition, we note that there is a second critical value of T, T_{AT}, and on branch I a corresponding critical value of α, $\alpha_{\mathrm{AT}}(T)$. When $T < T_{\mathrm{AT}}$, Λ^{II}_3 is positive for all α and Λ^{I}_3 is positive for $\alpha > \alpha_{\mathrm{AT}}(T)$. Contrastingly, Λ^{I}_1 and Λ^{II}_1 are negative for all T and α.

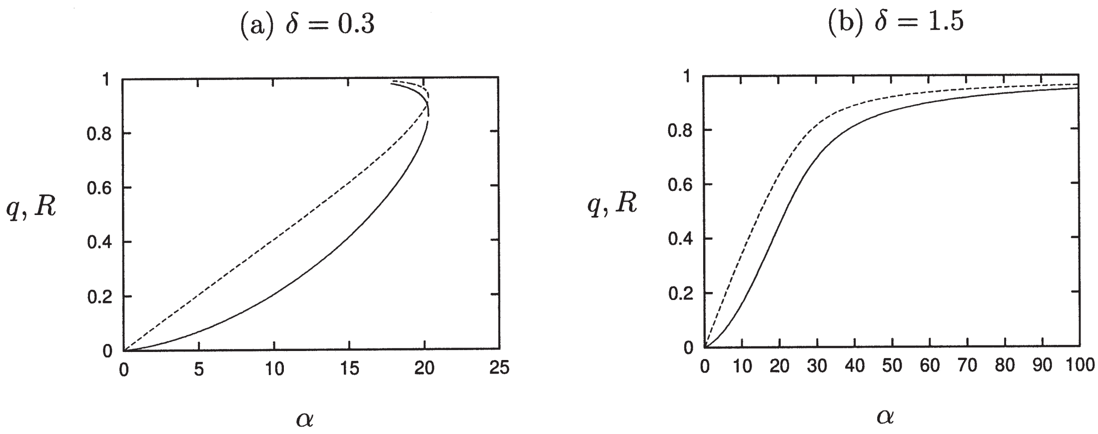

Fig. 5. The α dependences of q and R (dotted curve) for the RS solution with $T = 5$.

Case (III): general δ

In this case, numerical calculations were carried out for several values of δ and T. As a typical result, we found that when $T = 5$ and $\delta = 0.3$, q and R tend to 1 as $\alpha \to 0$. [See Fig. 5(a).] In this case, the RS solution is AT-stable, and its entropy is positive. There exists another case in which q and R tend to 1 as $\alpha \to \infty$. [See Fig. 5(b).] In this case, $S_{\rm RS}$ decreases and becomes 0 at the finite value $\alpha = \alpha_{\rm s}(T)$ for any T, and Λ_3 becomes 0 at the finite value $\alpha = \alpha_{\rm AT}(T)$ for any T.

For all the cases we considered in our numerical calculations, we found that for any value of δ and for $T < T_{\rm AT}$, the relation $\alpha_{\rm AT}(T) > \alpha_{\rm s}(T)$ holds, and $\alpha_{\rm AT}(T)$ and $\alpha_{\rm s}(T)$ are increasing functions of T, as long as these quantities are defined.

For any value of δ, the entropy becomes negative for small T. This implies that the RS solution is invalid. For this reason, we have to impose the replica symmetry breaking ansatz.

3.3 *The limiting forms of the expressions for $\hat{q}$ and $\hat{R}$ as $q \to 1$ and $R \to 1$.*

In this section, in order to derive the asymptotic learning curves and to allow for investigation of the conditions for the existence of PL, we determine the limiting forms of $\hat{q}$ and $\hat{R}$ for $q \to 1$ and $R \to 1$ as functions of α, β and χ by evaluating eqs. (12) and (13) in these limits. (See Appendix B for the derivations.) For $0 < \beta < \infty$, we find

$$\hat{q} \simeq \frac{\alpha}{\sqrt{\Delta q}} g_{1,\delta}(\chi, \beta) \quad \text{for} \quad \delta \geq 0, \tag{19}$$

$$\hat{R} \simeq \alpha \frac{\xi^{\delta}}{\sqrt{\Delta q}} g_{2,\delta}(\chi, \beta) \quad \text{for} \quad \delta \geq 0, \tag{20}$$

while for $\beta = \infty$, we have

$$\hat{q} \simeq \begin{cases} \dfrac{\alpha}{(\Delta q)^2} g_3 & (\text{for } \delta > 0, \text{ or for } \delta = 0 \text{ and } k < 1), \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_{3,D} & (\text{in the deterministic case}). \end{cases} \tag{21}$$

$$\hat{R} \simeq \begin{cases} \dfrac{\alpha}{\Delta q} g_4 & (\text{when } P(y) \text{ is not constant for } y > 0), \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_5(\chi) & (\text{when } P(y) \equiv k \text{ for } y > 0 \text{ and } k < 1), \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_{3,D} & (\text{in the deterministic case}). \end{cases} \tag{22}$$

Note that $\hat{R} = \hat{q}$ in the deterministic case. In these expressions we have used the following:

$$g_{1,\delta}(\chi, \beta) \equiv \frac{1}{\sqrt{2\pi}} \int du \tilde{\varphi}(u)^2 E_{\delta}(u, \chi),$$

$$E_{\delta}(u, \chi) = 1 \quad \text{for} \quad \delta > 0, \quad E_0(u, \chi) = 1 - k + 2kH(u/\chi),$$

$$g_{2,\delta}(\chi, \beta) \equiv \frac{a\delta}{\sqrt{2\pi}} (1 - e^{-\beta}) \frac{1}{\chi} (1 + \chi^{-2})^{(\delta-1)/2}$$
$$\times \int_0^{\infty} Dz z^{\delta-1} \int_{-\infty}^{\infty} Dt \left[\frac{1}{\tilde{H}\left(\dfrac{\chi t - z}{\sqrt{1+\chi^2}}\right)} + \frac{1}{\tilde{H}\left(-\dfrac{\chi t - z}{\sqrt{1+\chi^2}}\right)} \right], \quad \text{for} \quad \delta > 0,$$

$$g_{2,0}(\chi, \beta) = \frac{k(1 - e^{-\beta})}{\pi} \frac{1}{\sqrt{1+\chi^2}} \int_{-\infty}^{\infty} Dx \frac{1}{\tilde{H}\left(\dfrac{\chi x}{\sqrt{1+\chi^2}}\right)},$$

$$g_3 \equiv \int_0^\infty Dyy^2[1-P(y)], \quad g_{3,D} \equiv \frac{2}{\sqrt{2\pi}}\int Du\frac{h(u)}{H(u)},$$
$$g_4 \equiv \int_0^\infty DyP(y)(y^2-1), \quad g_5(\chi) \equiv \frac{k}{\pi}\frac{1}{\sqrt{1+\chi^2}}\int Dy\frac{1}{H\left(\frac{\chi y}{\sqrt{1+\chi^2}}\right)}.$$

Now, we examine the behaviour of $g_1 - g_5$ for later use. For $\delta \geq 0$, $g_{1,\delta}(\chi,\beta)$ is finite if $0<\beta<\infty$ for all χ, since $H(x)$ is bounded. For $g_{2,\delta}(\chi,\beta)$ with $\delta>0$, if $0<\beta<\infty$, we have

$$g_{2,\delta}(\chi,\beta) \sim \begin{cases} \nu_0\chi^{-\delta} & (\text{for } \chi \ll 1), \\ \nu_1\chi^{-1} & (\text{for } \chi \gg 1), \\ \mathcal{O}(1) & (\text{for finite } \chi), \end{cases} \tag{23}$$

where $\nu_0 = \frac{a\delta(1-e^{-\beta})}{\sqrt{2\pi}}\int_0^\infty Dzz^{\delta-1}\{\frac{1}{\tilde{H}(z)}+\frac{1}{\tilde{H}(-z)}\}$ and $\nu_1 = \frac{2s\beta}{\sqrt{2\pi}}$. In the case $\delta=0$, for $0<\beta<\infty$, $g_{2,0}(\chi,\beta)$ is finite, except when $\chi\to\infty$; that is,

$$g_{2,0} \sim \begin{cases} \frac{2k}{\pi}\tanh(\beta/2) & (\text{for } \chi \ll 1), \\ \frac{k\beta}{\pi}\chi^{-1} & (\text{for } \chi \gg 1), \\ \mathcal{O}(1) & (\text{for } \chi \text{ of } \mathcal{O}(1)). \end{cases} \tag{24}$$

The quantities $g_3, g_{3,D}$ and g_4 are all finite if $P(y)$ is not constant ($g_3=0$ for $P(y)\equiv 1$ for $y\geq 0$ and $g_4=0$ for $P(y)\equiv$ const. for $y\geq 0$). For g_5, we obtain

$$g_5(\chi) \sim \begin{cases} \frac{2k}{\pi} & (\text{for } \chi \ll 1), \\ \frac{k}{\pi}\chi & (\text{for } \chi \gg 1), \\ \mathcal{O}(1) & (\text{for } \chi \text{ of } \mathcal{O}(1)). \end{cases} \tag{25}$$

Here, for later use, we give the limiting form of the entropy S_{RS}. First, L is given by

$$L = \sqrt{\Delta q}\, r(\chi,\beta), \tag{26}$$

where

$$r(\chi,\beta) \equiv \frac{1}{\sqrt{2\pi}}\int duE_\delta(u,\chi)\left\{\ln[1+(e^\beta-1)H(u)] - \beta\frac{H(u)}{\tilde{H}(u)}\right\}. \tag{27}$$

The function $r(\chi,\beta)$ is finite for $0<\beta<\infty$ and for all δ and all χ. Then, as $q\to 1$ and $R\to 1$, for $0<\beta<\infty$ and for any δ and any χ, S_{RS} is given by

$$S_{\mathrm{RS}} = -\frac{\hat{q}\Delta q}{2} - (1-\Delta R)\hat{R} + \alpha\sqrt{\Delta q}\, r(\chi,\beta) + I. \tag{28}$$

3.4 Solutions of the S.P.E. when $q\to 1$ and $R\to 1$.

The equations for $\Delta q = 1-q$ and R are the following:

$$\Delta q = 2\frac{\partial I}{\partial \hat{q}}, \quad R = \frac{\partial I}{\partial \hat{R}}.$$

Therefore, we have to evaluate I when $q\to 1$ and $R\to 1$, which depends on $\mu \equiv \frac{\hat{R}}{2\hat{q}}$. We give expressions for I in Appendix C. Using these, after some algebra, we obtain the following types of behaviour.

(1) For $\delta > \frac{1}{3}$ and $\alpha \gg 1$, we have

$$\mu \simeq \mu_0\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{2\delta}{3\delta-1}}, \quad \mu_0 = \left[\frac{\delta}{\hat{q}_0(3\delta-1)}\right]^{\frac{2\delta}{3\delta-1}}$$
$$\Delta q \simeq q_0\mu_0^{\frac{1+\delta}{\delta}}\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{2(1+\delta)}{3\delta-1}}, \quad \Delta R \simeq R_0\mu_0^{1/\delta}\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{2}{3\delta-1}},$$
$$\hat{q} \simeq \hat{q}_0\mu_0^{-\frac{1+\delta}{2\delta}}\alpha^{\frac{4\delta}{3\delta-1}}(\ln\alpha)^{-\frac{1+\delta}{3\delta-1}}, \quad \hat{R} \simeq \hat{R}_0\mu_0^{-\frac{1-\delta}{2\delta}}\alpha^{\frac{2\delta}{3\delta-1}}(\ln\alpha)^{-\frac{1-\delta}{3\delta-1}}.$$

(2) For $\delta = \frac{1}{3}$ and $\alpha \gg 1$, we have

$$\mu \simeq \mu_0\alpha^{-\frac{1}{6}}e^{-\frac{2}{3}\hat{q}_0\alpha}, \quad \mu_0 = \left[\frac{1}{2\sqrt{\hat{q}_0}}\right]^{\frac{1}{3}},$$
$$\Delta q \simeq q_0\mu_0^4\alpha^{-\frac{2}{3}}e^{-\frac{8}{3}\hat{q}_0\alpha}, \quad \Delta R \simeq R_0\mu_0^3\alpha^{-\frac{1}{2}}e^{-2\hat{q}_0\alpha},$$
$$\hat{q} \simeq \hat{q}_0\mu_0^{-2}\alpha^{\frac{1}{3}}e^{\frac{4}{3}\hat{q}_0\alpha}, \quad \hat{R} \simeq \hat{R}_0\mu_0^{-1}\alpha^{\frac{1}{6}}e^{\frac{2}{3}\hat{q}_0\alpha}.$$

(3) For $0<\delta<\frac{1}{3}$ and $\alpha \ll 1$, we have

$$\mu \simeq \mu_0\left(\frac{\ln\frac{1}{\alpha}}{\alpha}\right)^{-\frac{2\delta}{1-3\delta}}, \quad \mu_0 = \left[\frac{\delta}{\hat{q}_0(1-3\delta)}\right]^{-\frac{2\delta}{1-3\delta}}$$

$$\Delta q \simeq q_0 \mu_0^{\frac{1+\delta}{\delta}} \left(\frac{\ln \frac{1}{\alpha}}{\alpha} \right)^{-\frac{2(1+\delta)}{1-3\delta}}, \qquad \Delta R \simeq R_0 \mu_0^{1/\delta} \left(\frac{\ln \frac{1}{\alpha}}{\alpha} \right)^{-\frac{2}{1-3\delta}},$$

$$\hat{q} \simeq \hat{q}_0 \mu_0^{-\frac{1+\delta}{2\delta}} \alpha^{-\frac{4\delta}{1-3\delta}} \left(\ln \frac{1}{\alpha} \right)^{\frac{1+\delta}{1-3\delta}}, \qquad \hat{R} \simeq \hat{R}_0 \mu_0^{-\frac{1-\delta}{2\delta}} \alpha^{-\frac{2\delta}{1-3\delta}} \left(\ln \frac{1}{\alpha} \right)^{\frac{1-\delta}{1-3\delta}}.$$

(4a) For $\delta = 0$, $\alpha \ll 1$ and $0 \lesssim \mu < 1$, we have

$$\Delta R \simeq \frac{1}{\sqrt{2\pi}} \psi(\mu) \alpha^2 \left(\ln \frac{1}{\alpha} \right)^{-2}, \quad \Delta q = 2\mu \Delta R,$$

$$\hat{R} \simeq \frac{2}{\mu} \ln \left[\frac{1}{\alpha} \left(\ln \frac{1}{\alpha} \right)^{3/4} \right], \qquad \hat{q} = \frac{\hat{R}}{2\mu},$$

where μ is determined by the following equations through χ:

$$\frac{g_{2,0}(\chi, \beta)}{2g_{1,0}(\chi, \beta)} = \frac{1}{1+\chi^2}, \tag{29}$$

$$\mu = \frac{1}{1+\chi^2}. \tag{30}$$

(4b) For $\delta = 0$, $\alpha \ll 1$ and $1 \le \mu$, we have

$$\Delta R \simeq 2a_\mu \left(\frac{1}{\alpha} \ln \frac{1}{\alpha} \right)^{-2}, \qquad \Delta q = 2\Delta R,$$

$$\hat{R} \simeq \frac{2\mu}{2\mu - 1} \ln \left(\frac{1}{\alpha} \ln \frac{1}{\alpha} \right), \qquad \hat{q} = \frac{\hat{R}}{2\mu},$$

where μ is given by

$$\mu = \frac{g_{2,0}(0, \beta)}{2g_{1,0}(0, \beta)}. \tag{31}$$

Now, let us check the conditions under which the above forms of $\Delta q, \Delta R, \hat{q}$ and $\hat{R}$ are valid. First, we consider $\delta > 0$. Here, we only have to see the conditions for $\mu \ll 1, \Delta R \ll 1$ and $\hat{R} \gg 1$, because $\Delta q \ll 1$ and $\hat{q} \gg 1$ follow from the relations $\Delta q \simeq \frac{2\Delta R}{1+\chi^2}$ and $\hat{q} = \frac{\hat{R}}{2\mu}$. For $\delta > 1/3$, these conditions are

$$g_{1,\delta}^{(\delta-1)/\delta} \beta^{1/\delta} \gg \frac{\ln \alpha}{\alpha} \qquad \text{(for } \mu \ll 1\text{)},$$

$$g_{1,\delta}^{-2} \beta^3 \gg \frac{\ln \alpha}{\alpha} \qquad \text{(for } \Delta R \ll 1\text{)},$$

$$g_{1,\delta}^{2(\delta-1)} \beta^2 \gg \frac{(\ln \alpha)^{1-\delta}}{\alpha^{2\delta}} \qquad \text{(for } \hat{R} \gg 1\text{)}.$$

As long as $\alpha \gg 1$, these conditions are satisfied. For $\delta = 1/3$, the condition is $\hat{q}_0 > 0$, and this is automatically satisfied for $\beta > 0$. For $0 < \delta < 1/3$, the conditions are

$$g_{1,\delta}^{(\delta-1)/\delta} \beta^{1/\delta} \ll \frac{\ln \frac{1}{\alpha}}{\alpha} \qquad \text{(for } \mu \ll 1\text{)},$$

$$g_{1,\delta}^{-2} \beta^3 \ll \frac{\ln \frac{1}{\alpha}}{\alpha} \qquad \text{(for } \Delta R \ll 1\text{)},$$

$$g_{1,\delta}^{2(\delta-1)} \beta^2 \ll \frac{\left(\ln \frac{1}{\alpha} \right)^{1-\delta}}{\alpha^{2\delta}}$$

As long as $\alpha \ll 1$, these conditions are satisfied. We have thus found the situations in which $\mu \ll 1, \Delta R \ll 1$ and $\hat{R} \gg 1$ are satisfied. Next, let us consider $\delta = 0$. Here, for case (4a), the condition is that there is a positive solution χ of eq. (29), and for case (4b), the condition is $\mu \ge 1$, where μ is defined by the eq. (31). When $\beta \ll 1$, μ satisfies

$$\mu \sim \frac{k\sqrt{2}}{\beta\sqrt{1+\xi^2}} \qquad \text{[for the case (4a)]}$$

$$\sim \frac{k\sqrt{2}}{\beta} \qquad \text{[for the case (4b)]}$$

Thus, $\mu \gg 1$ holds in both cases (4a) and (4b). Therefore, Case (4a) does not exist for high temperatures. In the other extreme, when $\beta \gg 1$, in both cases (4a) and (4b), $g_{1,0}$ becomes very large, but $g_{2,0}$ does not. Thus, from eqs. (29), (30) and (31), we obtain $\mu \ll 1$ in both cases (4a) and (4b). Hence, the case (4b) is impossible for low temperatures.

The results obtained in this section suggest that for $\delta < 1/3$ PL exists, and the solution with $q < 1$ exists only for $\alpha \in [0, \alpha_{\max}]$, where $\alpha_{\max}$ is some positive finite number, and that for $\delta \ge 1/3$ PL does not exist, and the solution with $q < 1$ exists for any α. However, as is shown in the next two sections, this conclusion is incorrect. One reason that this conclusion is incorrect is that when the solution with $q < 1$ exists for any α, the entropy of the RS solution becomes negative for $T \to 0$. Another reason is that the actual condition for the existence of PL is not $\delta < 1/3$ but, rather, $\delta < 1/2$.

In the next section, we investigate the necessary and sufficient conditions for the existence of PL.

4. Perfect Learning

In PL, the weight vectors of the students coincides with the optimal weight vector (i.e. $\boldsymbol{w} = \boldsymbol{w}^{\mathrm{o}}$) for a finite value of α. In this case, $q = 1$ and $R = 1$. From eqs. (10) and (11), the necessary and sufficient conditions to realize $q = 1$ and $R = 1$ with a finite value of α are

$$\hat{R} \to \infty \quad \text{and} \quad \tau \equiv \frac{\hat{R}}{\sqrt{\hat{q}}} \to \infty. \tag{32}$$

In the case of PL, we impose the further condition $q = R$ when the limits $q \to 1$ and $R \to 1$ are taken, because in PL, the weight vectors of the teacher and students coincide. Therefore, in this case, we have $\chi = \sqrt{\frac{q-R^2}{1-q}} = \sqrt{q} = 1$, and thus $\xi = Q = \sqrt{\Delta q}$. Hence, for $0 < \beta < \infty$, we obtain from eqs. (19) and (20),

$$\hat{q} \simeq \frac{\alpha}{\sqrt{\Delta q}} g_{1,\delta}(1,\beta), \tag{33}$$

$$\hat{R} \simeq \alpha(\Delta q)^{(\delta-1)/2} g_{2,\delta}(1,\beta), \tag{34}$$

while for $\beta = \infty$, we have

$$\hat{q} \simeq \begin{cases} \dfrac{\alpha}{(\Delta q)^2} g_3 & \text{(for } \delta > 0\text{, or for } \delta = 0 \text{ and } k < 1\text{)}, \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_{3,D} & \text{(in the deterministic case).} \end{cases} \tag{35}$$

$$\hat{R} \simeq \begin{cases} \dfrac{\alpha}{\Delta q} g_4 & \text{(when } P(y) \text{ is not constant for } y > 0\text{)}, \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_5(1) & \text{(when } P(y) \equiv k \text{ for } y > 0 \text{ and } k < 1\text{)}, \\ \dfrac{\alpha}{\sqrt{\Delta q}} g_{3,D} & \text{(in the deterministic case).} \end{cases} \tag{36}$$

In the deterministic case $\hat{R} = \hat{q}$ holds.

Let us now determine what is derived when these conditions are imposed. First, we consider the case $0 < \beta < \infty$. In this case, both $g_{1,\delta}(1,\beta)$ and $g_{2,\delta}(1,\beta)$ are finite for $\delta \geq 0$. (See subsection 3.3.) Thus, from (33) and (34), the necessary and sufficient conditions to realize $\hat{q} \to \infty, \hat{R} \to \infty$ and $\tau \to \infty$ as $q \to 1$ and $R \to 1$ for any α are

$$\hat{q} \to \infty \text{ (for any } \delta \geq 0 \text{ and } 0 < \beta < \infty\text{)}, \tag{37}$$

$$\hat{R} \to \infty \text{ (for } 0 \leq \delta < 1 \text{ and } 0 < \beta < \infty\text{)}, \tag{38}$$

$$\tau \simeq \alpha^{1/2}(\Delta q)^{(2\delta-1)/4} \frac{g_{2,\delta}}{\sqrt{g_{1,\delta}}} \to \infty \text{ (for } 0 \leq \delta < 1/2 \text{ and } 0 < \beta < \infty\text{)}. \tag{39}$$

Hence, the necessary and sufficient condition for PL is $0 \leq \delta < 1/2$. Next, we consider the case $\beta = \infty$. Here, when $P(y)$ is not constant for $y > 0$, the quantities $g_1 - g_4$ are all finite. Then, from (35) and (36), we find that both $\hat{q}$ and $\hat{R}$ tend to infinity and $\tau \simeq \sqrt{\frac{\alpha}{g_3}} g_4$, which is finite. Thus, in this case PL does not exist. Then, when $P(y) \equiv k < 1$ for $y > 0$, $\tau \simeq \sqrt{\alpha \Delta q} \frac{g_5}{\sqrt{g_3}}$. Since g_5 is finite for finite χ, τ tends to 0. Thus, here again, PL does not exist. Finally, in the deterministic case, from (35) and (36), $\hat{q} = \hat{R}$ and $\tau = \sqrt{\hat{q}}$ tend to infinity, since $g_{3,D}$ is finite. Hence, here PL does exist.

Summarizing the above results, we conclude that PL exists in the case that $0 \leq \delta < 1/2$ with $0 < \beta < \infty$ and in the deterministic case.[14]

For the entropy S_{PL} and the free energy f_{PL} in the case of PL, we obtain the following reasonable results:

$$S_{\text{PL}} = 0, \quad f_{\text{PL}} = \alpha \epsilon_{\min}.$$

(See Appendix D.)

5. 1RSB Solution

Although we adopt the Gibbs algorithm as the learning strategy, we are also interested in the minimum-error algorithm. In the minimum-error algorithm we choose weights so as to minimize the number of errors, and for that reason we only have to take the limit $T \to 0$. However, as shown in §3 using the results of numerical calculations, for the RS solution, the entropy becomes negative for small T. Thus, we have to consider the breaking of the replica symmetry.[15] For the 1RSB solution, the matrix q^{ab} is divided into $(n/m)^2$ small matrices of dimension $m \times m$. All components of each of the off-diagonal such matrices are q_0, and all components of each of the diagonal matrices are q_1, except for the diagonal components, which are 0. The structure of the matrix $\hat{q}^{ab}$ is idential, with the values $\hat{q}_0$ and $\hat{q}_1$ replacing q_0 and q_1. Further, we stipulate $R^a = R$ and $\hat{R}^a = \hat{R}$. Then, the 1RSB free energy, f_{1RSB}, is given by

$$\begin{aligned} &-\beta f_{\text{1RSB}}(q_0, \hat{q}_0, q_1, \hat{q}_1, R, \hat{R}, m, \beta) \\ &= -\frac{\hat{q}_1}{2}(1 - q_1) + \frac{m}{2}(\hat{q}_0 q_0 - \hat{q}_1 q_1) - R\hat{R} \\ &\quad + \frac{\alpha}{m} \int Dy 2\mathcal{P}(y) \int Dz_0 \ln \int Dz_1 \left[\tilde{H}\left(\frac{\sqrt{q_0 - R^2} z_0 + \sqrt{q_1 - q_0} z_1 - Ry}{\sqrt{1 - q_1}} \right) \right]^m \\ &\quad + \frac{1}{m} \int Dz_0 \ln \int Dz_1 \left[2\cosh\left(\sqrt{\hat{q}_0} z_0 + \sqrt{\hat{q}_1 - \hat{q}_0} z_1 + \hat{R} \right) \right]^m. \end{aligned} \tag{40}$$

Next, following Krauth–Mézard,[16] we take the limits $q_1 \to 1$ and $\hat{q}_1 \to \infty$. Then we obtain

$$f_{\text{1RSB}}(q_0, \hat{q}_0, q_1 = 1, \hat{q}_1 = \infty, R, \hat{R}, m, \beta) = f_{\text{RS}}(q_0, m^2 \hat{q}_0, R, m\hat{R}, \beta m). \tag{41}$$

From this relation, the equations for $q_0, \hat{q}_0, R, \hat{R}$ and m become a coupled set of equations consisting of the saddle point equations for the RS solution, along with the equation $S_{\text{RS}} = 0$, where S_{RS} is the entropy for the RS solution. Let us denote the solutions of these coupled equations by $q = q_c$, $\hat{q} = \hat{q}_c$, $R = R_c$, $\hat{R} = \hat{R}_c$ and $\beta = \beta_c$. Then, the 1RSB solutions are expressed by $q_0 = q_c$, $\hat{q}_0 = (\frac{\beta}{\beta_c})^2 \hat{q}_c$, $R = R_c$, $\hat{R} = \frac{\beta}{\beta_c} \hat{R}_c$ and $m = \frac{\beta_c}{\beta}$. Thus, to obtain the $T \to 0$ limit, we only have to know the solution at $T = T_c \equiv \beta_c^{-1}$.

5.1 *Numerical calculation of the S.P.E. for the 1RSB solution*

Here, we give the results of numerical calculations for the 1RSB solution.

Case (I): $\delta > 0$

As a special case, we treated $P(y) = 1 - 2H(y)$, that is, the case $\delta = 1$. This is the same $P(y)$ as that used for the RS solution. In Fig. 6, we plot the α dependence of T_c, while in Fig. 11, we plot those of q_0, R and $\Delta\epsilon_g$. As is seen from

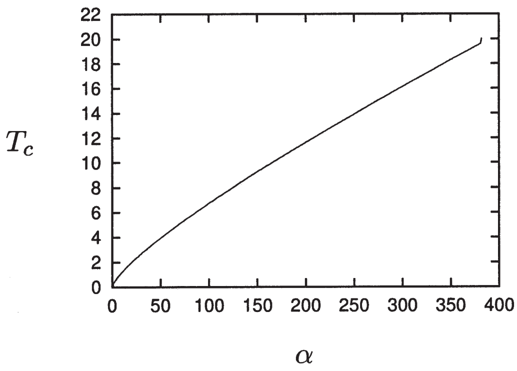

Fig. 6. Numerically obtained behaviour of T_c for $P(y) = 1 - 2H(y)(\delta = 1)$.

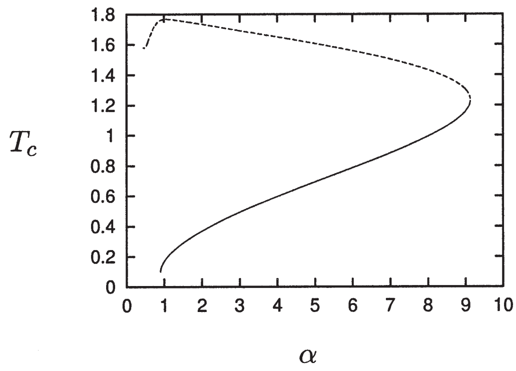

Fig. 8. The α dependence of T_c for $\delta = 0$. Solid curve: 1RSB in branch I. Dashed curve: 1RSB in branch II.

these figures, the 1RSB solution seems to extend to $\alpha = \infty$. To study asymptotic behaviour, representing each of the quantities Δq, ΔR, T_c and $\Delta\epsilon_g$ by A, we assume that each can be written as

$$\ln A = a_1 + b_1 \ln\alpha,$$
$$\ln A = a_2 + b_2 \ln\left(\frac{\alpha}{\ln\alpha}\right),$$

where a_1, a_2, b_1 and b_2 are to be determined for Δq, ΔR, T_c and $\Delta\epsilon_g$ separately. Similarly, representing $\hat{q}$ and $\hat{R}$ by A, we assume that each can be written as

$$A = a_1 + b_1 \ln\alpha,$$
$$A = a_2 + b_2 \ln\left(\frac{\alpha}{\ln\alpha}\right).$$

In Table I, we list the values of a_i and b_i for these quantities. In particular, we note that $T_c \to \infty$ as $\alpha \to \infty$. Further, we obtained $\frac{\Delta q}{\Delta R} \sim 1.7$ and $\frac{\hat{R}}{\hat{q}} \sim 2.5$. As an example, we display the asymptotic behaviour of $\Delta\epsilon_g$ in Fig. 7.

Table I. Coefficients a_i and expontents b_i evaluated for $200 \le \alpha \le 381$ and the theoretical value $b_{2,\mathrm{th}}$.

	Δq	ΔR	T_c	$\Delta\epsilon_g$	$\hat{R}$	$\hat{q}$
a_1	2.5	2.2	−1.8	1.1	−0.49	0.89
b_1	−1.5	−1.5	0.81	−1.5	1.1	0.28
a_2	1.1	0.88	−1.1	−0.27	0.50	1.1
b_2	−1.8	−1.9	0.99	−1.9	1.4	0.35
$b_{2,\mathrm{th}}$	−2	−2	1	−2	1	1

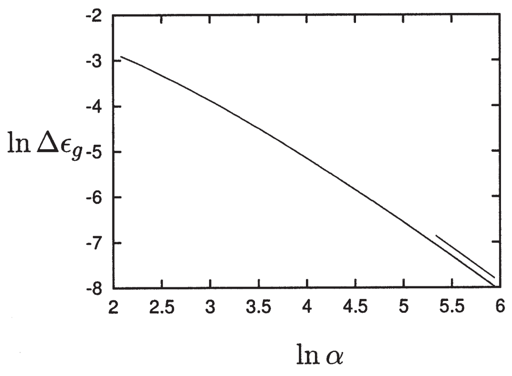

Fig. 7. Numerically obtained behaviour of $\Delta\epsilon_g$ for $P(y) = 1 - 2H(y)(\delta = 1)$. A line segment with slope b_1 determined by the least square methods is also plotted.

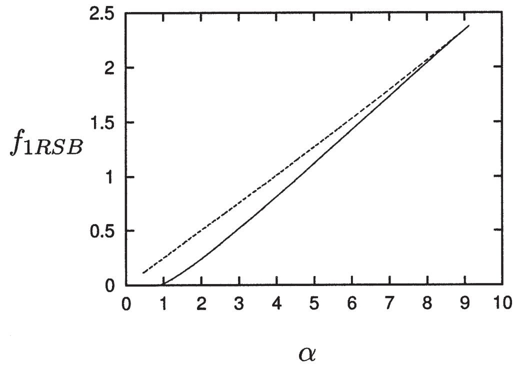

Fig. 9. The α dependence of the free energy. Solid curve: $f^{\mathrm{I}}_{1\mathrm{RSB}}$. Dashed curve: $f^{\mathrm{II}}_{1\mathrm{RSB}}$.

Case (II): $\delta = 0$

Here, we considered the same function $P(y)$ as in the $\delta = 0$ case for the RS solution, $P(y) = \frac{1}{2}\,\mathrm{sgn}\,(y)$. We display the α dependence of T_c in Fig. 8 and that of q_0, R and $\Delta\epsilon_g$ in Fig. 16. In the 1RSB solution, there exist two branches, I and II. In branch I all of these quantities become identical to those for the RS solution with $T = 0$ as $\alpha \to \alpha_s(0)$. In branch II, q and R tend to 1 as α tends to 0.

With the results of our numerical calculations, it is difficult to determine whether case of (a) $0 \lesssim \mu < 1$ or (b) $\mu \ge 1$ holds in branch II, because we could obtain solutions only for $\alpha \gtrsim 0.45$. As for T_c, it seems that T_c converges to a finite value as $\alpha \to 0$. In both branches, Λ_1 and Λ_3 are negative; that is, the RS solution at $T = T_c$ is AT-stable. For the free energy, the relation $f^{\mathrm{I}}_{1\mathrm{RSB}} < f^{\mathrm{II}}_{1\mathrm{RSB}}$ holds. (See Fig. 9.)

5.2 *Limiting behaviour as $q \to 1$ and $R \to 1$*

As is suggested by the above numerical results, and is considered below, the limiting behaviour of β_c as $q \to 1$ and $R \to 1$ differs in the cases $\delta = 0$ and $\delta > 0$. Therefore we discuss these cases separately.

Case (I): $\delta > 0$

As shown in Fig. 6, the numerical result suggests that $\beta_c(=\frac{1}{T_c}) \to 0$ as $\alpha \to \infty$. Therefore we consider the case $\beta_c \ll 1$. For $\beta \ll 1, f_{\mathrm{RS}}$ and S_{RS} take the following forms in the asymptotic region (see Appendix E):

$$-\beta f_{\mathrm{RS}} = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I - \alpha\beta\left[\epsilon_{\min} + \frac{2s}{(1+\delta)\sqrt{2\pi}}(2\Delta R)^{\frac{1+\delta}{2}} - \frac{\beta\sqrt{\Delta q}}{2\pi\sqrt{2}}\right], \tag{42}$$

$$S_{\mathrm{RS}} = -\frac{\hat{q}}{2}\Delta q - \hat{R}R + I - \alpha\beta^2\frac{\sqrt{\Delta q}}{2\pi\sqrt{2}}. \tag{43}$$

First, we show that for $0 \lesssim \mu < 1$, no consistent 1RSB solution exists in the presently considered case in which $\delta > 0$ and $\beta \ll 1$. The saddle point equations are

$$\Delta q \simeq \frac{2h(\tau)\psi(\mu)}{\sqrt{\hat{q}_c}}, \tag{44}$$

$$\Delta R \simeq \frac{\Delta q}{2\mu}, \tag{45}$$

$$\hat{q}_c = \hat{q}_0\alpha\beta_c^2/\sqrt{\Delta q}, \tag{46}$$

$$\hat{R}_c = \hat{R}_0\alpha\beta_c(\Delta R)^{(\delta-1)/2}, \tag{47}$$

$$\hat{q}_0 = \frac{1}{2\pi\sqrt{2}}, \quad \hat{R}_0 = \frac{s}{\sqrt{\pi}}2^{\delta/2},$$

where $\Delta q = 1 - q_c$ and $\Delta R = 1 - R_c$. Then, the condition that the entropy is zero becomes

$$\Delta q \simeq \frac{4\psi(\mu)h(\tau)}{\hat{R}\tau}. \tag{48}$$

From eqs. (44) and (48), we obtain $\tau^2 = 2$. However, since we consider the case $R \simeq 1$ and $q \simeq 1$, τ should be large. Thus, the case $0 \lesssim \mu < 1$ is not possible. Next, we consider the case $1 \le \mu$. In this case, $I \simeq \hat{R} + a_\mu e^{-2(\hat{R}-\hat{q})}$. Then, f_{RS} and S_{RS} become

$$-\beta_c f_{\mathrm{RS}} \simeq -\frac{\hat{q}_c}{2}\Delta q + \hat{R}_c\Delta R + a_\mu e^{-2(\hat{R}_c-\hat{q}_c)} - \alpha\beta_c\left(\epsilon_{\mathrm{g}} - \frac{\beta_c}{2\pi\sqrt{2}}\sqrt{\Delta q}\right),$$

$$S_{\mathrm{RS}} \simeq -\frac{\hat{q}_c}{2}\Delta q + \hat{R}_c\Delta R - \frac{\alpha\beta_c^2}{2\pi\sqrt{2}}\sqrt{\Delta q} + a_\mu e^{-2(\hat{R}_c-\hat{q}_c)}.$$

The saddle point equations for $\hat{q}$ and $\hat{R}$ here are the same as in the case $0 \lesssim \mu < 1$, while those for q and R and the condition for zero entropy are

$$\Delta R = 2a_\mu e^{-2(\hat{R}_c-\hat{q}_c)}, \tag{49}$$

$$\Delta q = 2\Delta R, \tag{50}$$

$$S_{\mathrm{RS}} \simeq -\frac{\hat{q}_c}{2}\Delta q + \hat{R}_c\Delta R - \alpha\beta_c^2\hat{q}_0\sqrt{\Delta q} + a_\mu e^{-2(\hat{R}_c-\hat{q}_c)} = 0. \tag{51}$$

Since $\hat{R}_c \gg 1$ and $\hat{q}_c \gg 1$, using eqs. (46), (49) and (50), we obtain from eq. (51) the relation

$$\hat{R}_c = 3\hat{q}_c. \tag{52}$$

That is, $\mu = 3/2$ and $a_\mu = 1$. Thus, from eqs. (46), (47) and (52), we obtain

$$\hat{q}_c = F_0\alpha e^{-2(2\delta-1)\hat{q}_c},$$

which implies

$$\ln\hat{q}_c = \ln\alpha - 2(2\delta-1)\hat{q}_c + \ln F_0,$$

where $F_0 = 4s^2 2^{2\delta}/9\sqrt{2}$. Thus,

$$\hat{q}_c \simeq \frac{\ln\alpha}{2(2\delta-1)}$$

for $2\delta - 1 \neq 0$. This implies that $\hat{q}_c$ tends to infinity when α tends to infinity for $\delta > 1/2$ or α tends to 0 for $\delta < 1/2$. For $\delta = 1/2$, $\hat{q}_c = F_0\alpha$. Therefore, we obtain the following results.

(1) In the case $\delta > \frac{1}{2}$, as $\alpha \to \infty$,

$$\Delta R \simeq 2\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{2}{2\delta-1}}, \quad \Delta q \simeq 2\Delta R, \tag{53}$$

$$\hat{R}_c \simeq \frac{3}{2(2\delta-1)}\ln\left(\frac{\alpha}{\ln\alpha}\right), \quad \hat{q}_c \simeq \hat{R}_c/3, \quad \beta_c \simeq \frac{4\sqrt{\pi}s}{3}2^\delta\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{\delta}{2\delta-1}}, \tag{54}$$

$$\Delta\epsilon_{\mathrm{g}} \simeq \epsilon_0(\Delta R)^{\frac{\delta+1}{2}} \simeq \epsilon_0 2^{\frac{\delta+1}{2}}\left(\frac{\ln\alpha}{\alpha}\right)^{\frac{\delta+1}{2\delta-1}}, \quad \epsilon_0 = \frac{2s}{(1+\delta)\sqrt{2\pi}}2^{\frac{1+\delta}{2}}. \tag{55}$$

(2) In the case $\delta = \frac{1}{2}$, as $\alpha \to \infty$,

$$\Delta R \simeq 2e^{-4F_0\alpha}, \quad \Delta q \simeq 2\Delta R, \tag{56}$$

$$\hat{R}_c \simeq 3F_0\alpha, \quad \hat{q}_c \simeq \hat{R}_c/3, \quad \beta_c \simeq \frac{4s}{3}\sqrt{2\pi}e^{-F_0\alpha}, \tag{57}$$

$$\Delta\epsilon_{\mathrm{g}} \simeq \epsilon_0 2^{\frac{3}{4}}e^{-3F_0\alpha}. \tag{58}$$

(3) In the case $0 < \delta < \frac{1}{2}$, as $\alpha \to 0$,

$$\Delta R \simeq 2\left(\frac{\alpha}{\ln\frac{1}{\alpha}}\right)^{\frac{2}{1-2\delta}}, \quad \Delta q \simeq 2\Delta R, \tag{59}$$

$$\hat{R}_c \simeq \frac{3}{2(1-2\delta)} \ln\left(\frac{1}{\alpha}\ln\frac{1}{\alpha}\right), \quad \hat{q}_c \simeq \hat{R}_c/3, \quad \beta_c \simeq \frac{4\sqrt{\pi}s}{3} 2^{\delta} \left(\frac{\alpha}{\ln\frac{1}{\alpha}}\right)^{\frac{\delta}{1-2\delta}}, \tag{60}$$

$$\Delta\epsilon_{\rm g} \simeq \epsilon_0 2^{\frac{\delta+1}{2}} \left(\frac{\alpha}{\ln\frac{1}{\alpha}}\right)^{\frac{1+\delta}{1-2\delta}}. \tag{61}$$

Thus, when $0 < \delta < 1/2$, for large α there is no solution such that $q_c \to 1$ and $R_c \to 1$. This implies that there is a value $\alpha = \alpha_{\rm max}$ such that for $\alpha > \alpha_{\rm max}$, when $0 < \delta < 1/2$, the only solution is the PL solution.

Next, we compare the theoretical and numerical results for the asymptotic behaviour in the case $\delta = 1$. As shown in Table I, the numerically obtained exponents b_2 agree fairly well with the theoretical exponents $b_{2,{\rm th}}$, except for $\hat{q}$ and $\hat{R}$. The reason for the somewhat poor agreement in these cases is that we have assumed that $\hat{q}$ and $\hat{R}$ depend on $\ln(\frac{\alpha}{\ln\alpha})$, and evaluating logarithmic dependence numerically is difficult. For this reason, it is more meaningful to compare the theoretical and numerical valus of $\frac{\hat{R}}{\hat{q}}$ and $\frac{\Delta R}{\Delta q}$. We find that for $\frac{\hat{R}}{\hat{q}}$, these values are 3 and 2.7, respectively, and for $\frac{\Delta R}{\Delta q}$, they are 2 and 1.7, respectively. We therefore conclude that the agreement between theoretical and numerical results is fairly good.

Now, we examine the case $\delta = 0$. Note that if we substitute $\delta = 0$ into the expressions of β_c in the eq. (60) for the case $0 < \delta < 1/2$, β_c is an $\mathcal{O}(1)$ constant. Thus, the condition that $\beta_c \ll 1$ is not satisfied. This is the reason we treat the cases $\delta > 0$ and $\delta = 0$ separately.

Case (II): $\delta = 0$

In §3, we examined the RS solution with β fixed. Now, we investigate the 1RSB solution imposing the condition $S_{\rm RS} = 0$.

(a) $0 \lesssim \mu < 1$

In this case, $S_{\rm RS}$ becomes

$$S_{\rm RS} \simeq \frac{\hat{q}\Delta q}{2} + \alpha\sqrt{\Delta q}r.$$

Then, from the condition $S_{\rm RS} = 0$, we obtain

$$g_{1,0}(\chi, \beta_c) = -2r(\chi, \beta_c). \tag{62}$$

If solutions $\chi > 0$ and $\beta_c > 0$ for eqs. (29) and (62) exist, then the 1RSB solution exists for $\beta > \beta_c$.

(b) $\mu \geq 1$

Here, $S_{\rm RS}$ is

$$S_{\rm RS} \simeq -\frac{\hat{q}\Delta q}{2} + \frac{\hat{R}\Delta q}{2} + \alpha\sqrt{\Delta q}r.$$

Then, the condition $S_{\rm RS} = 0$ becomes

$$g_{1,0}(0, \beta_c) = g_{2,0}(0, \beta_c) + 2r(0, \beta_c). \tag{63}$$

If a solution $\beta_c > 0$ of eq. (63) exists and satisfies the relation $g_{2,0}/g_{1,0} \geq 2$, the 1RSB solution exists for $\beta > \beta_c$.

The numerical calculations for $\delta = 0$ indicate that in the $\alpha \to 0$ limit χ tends to a finite constant, and hence in this limit, we have case (a) above. Thus, for $\alpha \ll 1$

$$\Delta R \simeq \frac{1}{\sqrt{2\pi}}\psi(\mu_c)\alpha^2\left(\ln\frac{1}{\alpha}\right)^{-2}, \quad \Delta q = 2\mu_c \Delta R,$$

$$\hat{R}_c \simeq \frac{2}{\mu_c}\ln\left[\frac{1}{\alpha}\left(\ln\frac{1}{\alpha}\right)^{3/4}\right], \quad \hat{q}_c = \frac{\hat{R}_c}{2\mu_c},$$

$$\Delta\epsilon_{\rm g} = \epsilon_0\sqrt{\frac{\psi(\mu_c)}{2\pi}}\alpha\left(\ln\frac{1}{\alpha}\right)^{-1}.$$

It should be noted that in the above limiting solutions for $\delta \geq 0$, no coefficient of any quantity contains β, and therefore there is no condition on the range of β for which these solutions are valid.

We have found that in the case $0 \leq \delta < \frac{1}{2}$, there exists no solution for $\alpha > \alpha_{\rm max}$. This is consistent with the result derived in §4 that PL exists for $0 \leq \delta < \frac{1}{2}$ when $0 < \beta < \infty$.

Combining the results obtained to this point, the learning behaviour exhibited by our model can be summarized as follows. When T is sufficiently small, there is a critical value of α, $\alpha_{\rm s}(T)$, above which the entropy of the RS solution becomes negative. Thus, for $\alpha > \alpha_{\rm s}(T)$, the 1RSB solution exists. With the 1RSB ansatz, we found that the behaviour of the generalization error $\epsilon_{\rm g}$ can be classified into the following three categories, according to the value of δ.

(1) If $0 \leq \delta < \frac{1}{2}$, solutions with $R < 1$ exist only for a finite range of α, $[0, \alpha_{\rm max}]$. In this case, there is a critical temperature $T_{\rm s}$. When $T > T_{\rm s}$, the entropy of the RS solution is positive, and this solution is AT-stable. When $T < T_{\rm s}$, for $\alpha > \alpha_{\rm s}(T)$, the 1RSB solution exists. In either case, at $\alpha = \alpha_{\rm max}$, a first-order phase transition from the RS solution with positive entropy or from the 1RSB solution to the PL solution takes place.

(2) If $\delta = \frac{1}{2}$, $\alpha_{\rm s}(T)$ is defined for any temperature T, and the 1RSB solution exists for $\alpha > \alpha_{\rm s}(T)$. Here, $\epsilon_{\rm g}$ for the 1RSB solution decays to $\epsilon_{\rm min}$ exponentially according to

$$\Delta\epsilon_{\rm g} \propto e^{-3F_0\alpha},$$

where F_0 is a constant.

(3) If $\delta > \frac{1}{2}$, for any temperature T, $\alpha_{\rm s}(T)$ is defined, and the 1RSB solution exists for $\alpha > \alpha_{\rm s}(T)$. Here, $\epsilon_{\rm g}$ for the 1RSB solution decays to $\epsilon_{\rm min}$ as a power law with a logarithmic correction according to

$$\Delta\epsilon_{\rm g} \propto \left(\frac{\ln\alpha}{\alpha}\right)^{\frac{1+\delta}{2\delta-1}}.$$

To check these theoretical results, we carried out numerical calculations. In the next section, we give the results of these calculations.

6. Numerical Calculations with the Exhaustive Method

We carried out numerical calculations using the exhaustive method for $\delta = 0$ and $\delta = 1$. We used the minimum-error algorithm and the Gibbs algorithm for several temperatures. We calculated the quantities q, q_0, R, ϵ_g, along with their standard deviations, and the distribution of q, $P(q)$. For example, q and its standard deviation δq were calculated using the formulas

$$q = \frac{1}{M_\xi} \sum_\xi q_\xi, \quad q_\xi = \sum_{a<b} q^{ab} P_a P_b / \sum_{a<b} P_a P_b, \tag{64}$$

$$\delta q = \sqrt{\left[\sum_\xi q_\xi^2 - \left(\sum_\xi q_\xi \right)^2 \Big/ M_\xi \right] \Big/ (M_\xi - 1)}, \tag{65}$$

$$P_a = e^{-\beta E_a} / \sum_a e^{-\beta E_a}, \tag{66}$$

where the index a runs over the 2^N configurations of the weight vectors, whose energies are E_a, q_ξ is the thermal average for a given example ξ, and M_ξ is the number of samples. The calculations were performed for several values of N up to 20 with $M_\xi = 200$.

6.1 $\delta = 1$

The i-th component of an example x_i is corrupted by the Gaussian noise η_i with mean 0 and standard deviation 1. This corresponds to the choice $P(y) = 1 - 2H(y)$. First, we give the results for the minimum-error algorithm. In Fig. 10, to elucidate the system size N dependence of quantities, we display the α dependences of $R(\alpha, N)$ and its standard deviation $\delta R(\alpha, N)$ for $N = 10$, 15 and 17. From these results, we found that the quantities $r(\alpha, N, N') \equiv |\frac{R(\alpha,N) - R(\alpha,N')}{R(\alpha,N)}|$ and $\delta r(\alpha, N, N') \equiv |\frac{\delta R(\alpha,N) - \delta R(\alpha,N')}{\delta R(\alpha,N)}|$ obtained by the exhaustive method for $N, N' = 15$ and 17 are at most several percent and it seems that the results obtained for $N = 15$ are sufficient to surmise the $N = \infty$ behaviour, at least for α up to 15. In Fig. 11, we display both the numerical results and the theoretial results for R, q, q_0 and $\Delta\epsilon_g$. The numerical results agree with the theoretical results within the numerical standard deviations. In Fig. 12 we display the α dependence of R for larger values of α in the case $N = 15$. We see that for $\alpha \gtrsim 85$, there exists only one state. Since on theoretical grounds it is known that R tends to 1 as α goes to infinity, we can determine whether the fact that there is just one solution is a finite size effect by considering the value of R and determining the value of α, α_{max}, at which R first exceeds $1 - \frac{1}{N}$. Then, if the value of α_{max} increases with N, we can conclude that the existence of one solution is indeed a finite size effect. We found that this is in fact the case.

Next, we give the results obtained using the Gibbs algorithm. With this algorithm, we considered the cases $N = 10$ and 12 for several temperatures, and we included all states in the calculation. We found that $r(\alpha, N, N')$ is less than 5% and $\delta r(\alpha, N, N')$ is less than 10% for $N, N' = 10$ and

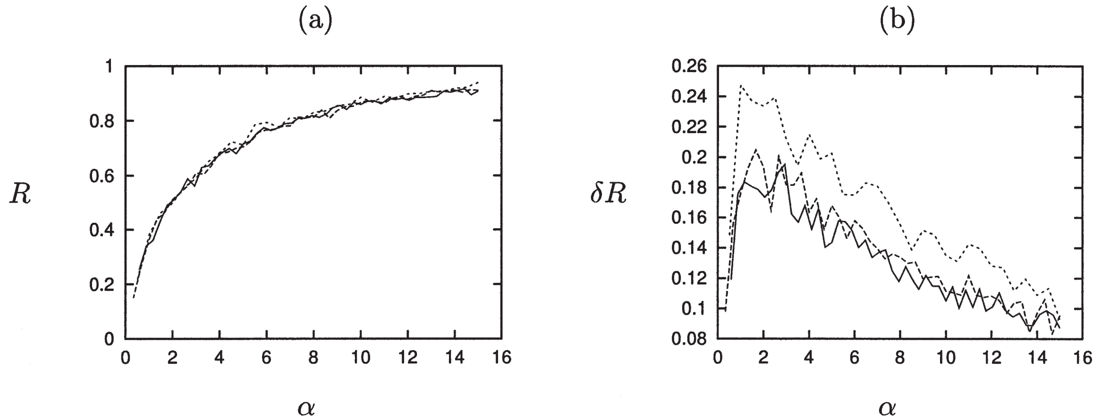

Fig. 10. The α dependences of R and δR obtained using the minimum-error algorithm for $\delta = 1$. Dotted curve: $N = 10$. Dashed curve: $N = 15$. Solid curve: $N = 17$.

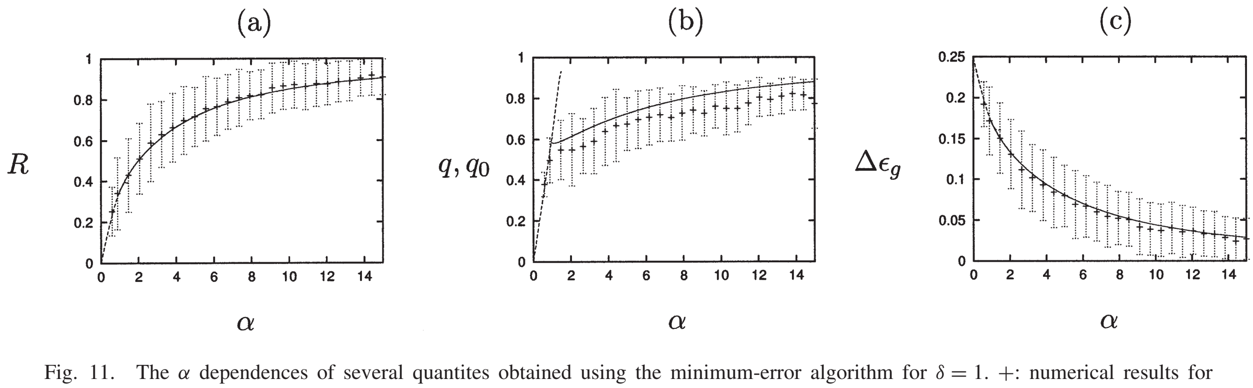

Fig. 11. The α dependences of several quantites obtained using the minimum-error algorithm for $\delta = 1$. $+$: numerical results for $N = 17$ (where the bars indicate the standard deviations). Dashed curve: RS solution ($T = 0$). Solid curve: 1RSB solution. In (b), q (for the RS solution) is represened by the dashed curve and q_0 (for the 1RSB solution) by the solid curve.

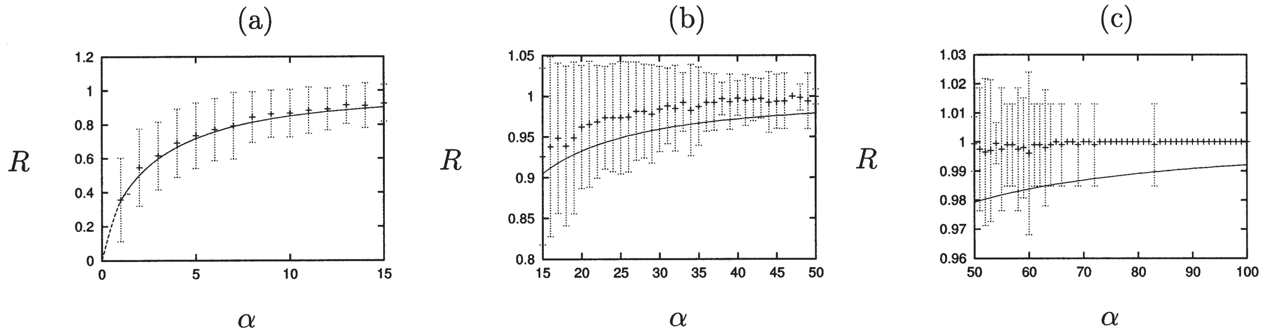

Fig. 12. The asymptotic behaviour of R obtained using the minimum-error algorithm for $\delta = 1$. +: numerical results for $N = 15$ (where the bars indicate the standard deviations). Dashed curve: RS solution. Solid curve: 1RSB solution. (a) $0 < \alpha < 15$. (b) $15 < \alpha < 50$. (c) $50 < \alpha < 100$.

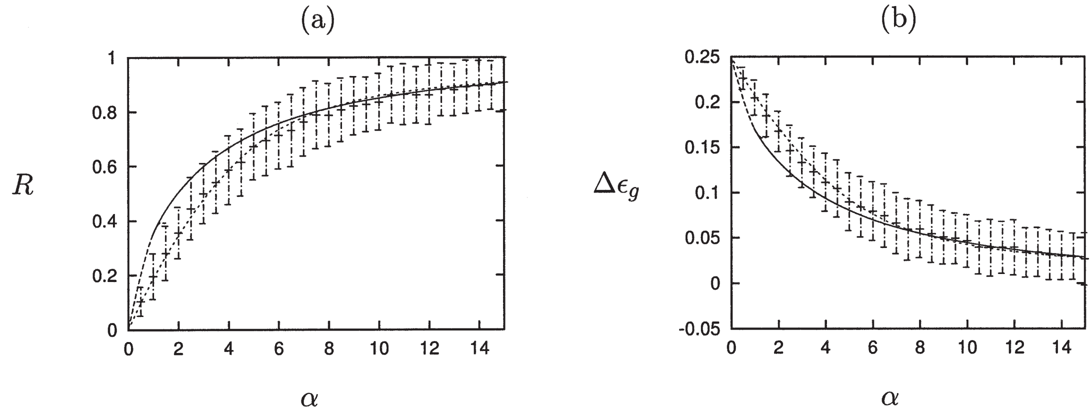

Fig. 13. The α dependences of R and $\Delta\epsilon_{\rm g}$ obtained using the Gibbs algorithm for $\delta = 1$. +: numerical results (with standard deviations) for $N = 12$ with $T = 1$. Dashed curve: RS solution with $T = 0$. Dotted curve: RS solution with $T = 1$. Solid curve: 1RSB solution.

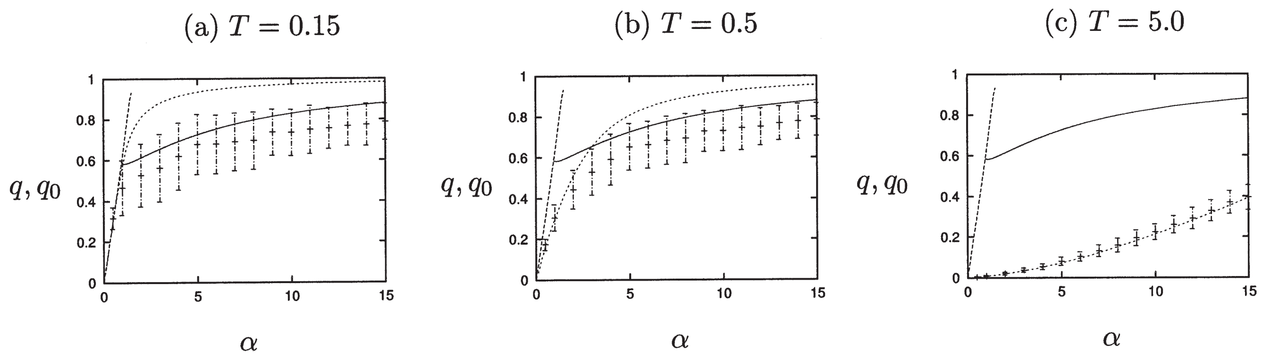

Fig. 14. The α dependences of q and q_0 obtained using the Gibbs algorithm for $\delta = 1$. +: numerical results (with standard deviations) for $N = 12$. Dashed curve: q for the RS solution with $T = 0$. Dotted curve: q for the RS solution with finite temperature. Solid curve: q_0 for the 1RSB solution.

12. We display both numerical and theoretical results in Fig. 13 for R and $\Delta\epsilon_{\rm g}$ and in Fig. 14 for q and q_0. First we explain the theoretical prediction for the α dependences of these quantities. (See Figs. 13 and 14.) $\alpha_{\rm s}$ is the value at which the dotted curve for the RS solution and the solid curve for the 1RSB solution coincides. For $\alpha < \alpha_{\rm s}$ the dotted curve is valid and for $\alpha \geq \alpha_{\rm s}$ the solid curve is valid. As for R and $\Delta\epsilon_{\rm g}$, the numerical results agree with the theoretical results within the numerical standard deviations. As for q and q_0, as shown in Fig. 14, the numerical results almost agree with the theoretical results within the numerical standard deviations, but the agreement is worse than that for other quantities. We calculate q_0 for each sample using formula (64). We find that in general, q_0 exhibits a large finite size effect, because it is calculated for pairs of states. Thus, when the number of states with the minimum energy becomes small, fluctuations of q_0 become very large. This is the reason why the agreement between the numerical and the theoretical results for q_0 is worse than that for other quantities. A relevant quantity is the distribution of q, $P(q)$. $P(q)$ is calculated as

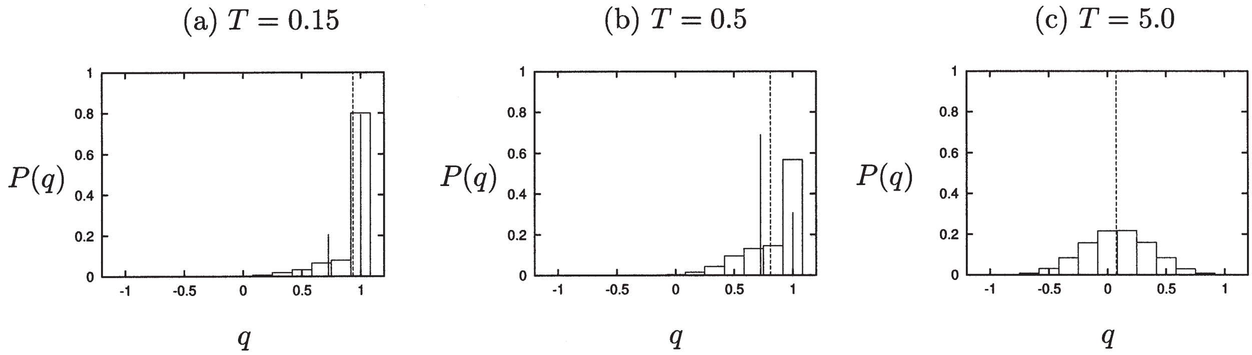

Fig. 15. The T dependence of $P(q)$ found using the Gibbs algorithm for $\delta = 1$. Histogram: numerical results for $N = 12$ and $p = 60$. Solid vertical segments: 1RSB solution. Dotted vertical segments: RS solution. For $T = 5$, no 1RSB solution exists.

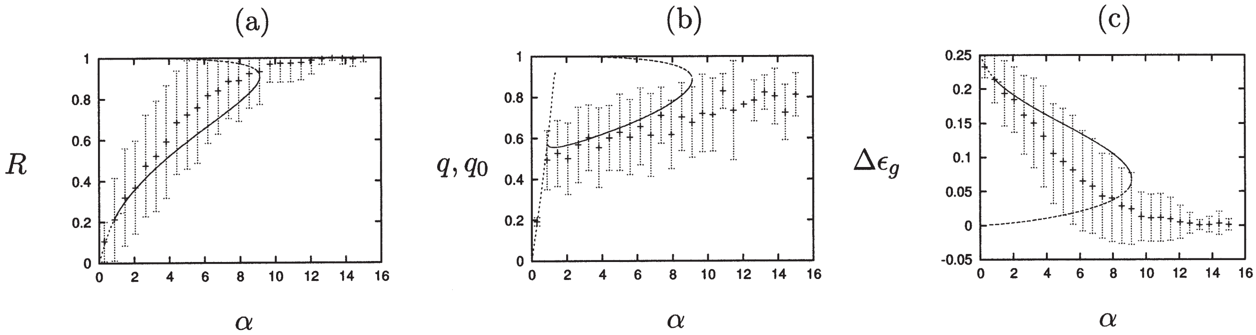

Fig. 16. The α dependences of several quantites obtained using the minimum-error algorithm for $\delta = 0$. $+$: numerical results for $N = 17$ (with bars indicating standard deviations). Dashed curve: RS ($T = 0$). Solid curve: 1RSB in branch I. Dashed curve: 1RSB in branch II. In (b), q is for the RS solution and q_0 is for the 1RSB solution.

$$P(q) = \left\langle \sum_{a,b} \delta(q, q^{ab}) P_a P_b \right\rangle,$$

where $\delta(q, q^{ab})$ is the Kronecker delta and $\langle\cdot\rangle$ represents the average over samples. $P(q)$ is also plotted for several temperatures, together with theoretical results, in Fig. 15. We see that for the theoretical and numerical results the peak values agree for $T = 0.15$ and $T = 5.0$, while they do not agree for $T = 0.5$. Since $T = 0.5$ is near to the transition temperature T_c from the RS to the 1RSB solutions, and then the standard deviations of numerical results become lagre, the disagreement of the peaks for $T = 0.5$ is considered to be the finite size effect.

6.2 $\delta = 0$

The output by the teacher is reversed with probability $(1 - k)/2$, where here we use $k = \frac{1}{2}$. This corresponds to the choice $P(y) = \frac{1}{2}\,\mathrm{sgn}(y)$. First, we give the results for the minimum-error algorithm. We investigated the N dependence of $R(\alpha, N)$ and $\delta R(\alpha, N)$ with $N = 10$, 15, 17 and 20 obtained by the exhaustive method and found that the quantities $r(\alpha, N, N')$ and $\delta r(\alpha, N, N')$ for $N, N' = 15, 17$ and 20 are at most 2% and 10%, respectively. In Fig. 16, the α dependences of several quantities are displayed for $N = 17$, together with theoretical results. The numerical results agree with the theoretical results within the numerical standard deviations up to $\alpha \lesssim \alpha_{\max}$. However, we note that numerical data take value for α above the theoretical upper bound, $\alpha_{\max}$. This tendency is remarkable for q_0. This is due to the finite size effect mentioned above. In Fig. 17 we display the behaviour of R for larger values of α in the case $N = 15$. It is seen that for $\alpha > 22$, there exists only one state. In the case $N = 10$, this is the case for $\alpha > 25$. To investigate whether or not PL exists, we numerically computed $\alpha_{\max}$ using the same method as in the case $\delta = 1$. We found that, in contrast to the case $\delta = 1$, here $\alpha_{\max}$ decreases as a function of N. With these results, we conclude that PL exists even for arbitrarily large N. Theoretically, $\alpha_{\max}$ is found to be approximately 9.13.

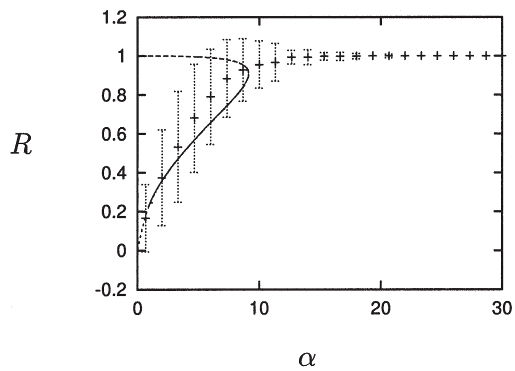

Fig. 17. The asymptotic behaviour of R obtained using the minimum-error algorithm for $\delta = 0$. $+$: numerical results for $N = 15$ (with standard deviations). Dashed curve: RS solution. Solid curve: 1RSB solution.

Next, we give the results for the Gibbs algorithm. In this

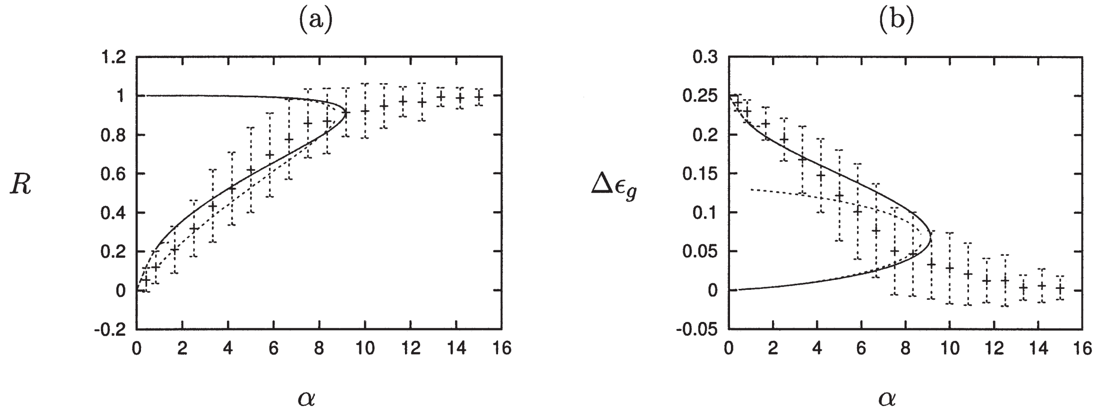

Fig. 18. The α dependences of R and $\Delta\epsilon_g$ found using the Gibbs algorithm for $\delta = 0$ and $T = 1.0$. +: numerical results for $N = 12$. Dashed curve: RS solution for $T = 0$. Dotted curve: RS solution for $T = 1.0$. Solid curve: 1RSB solution.

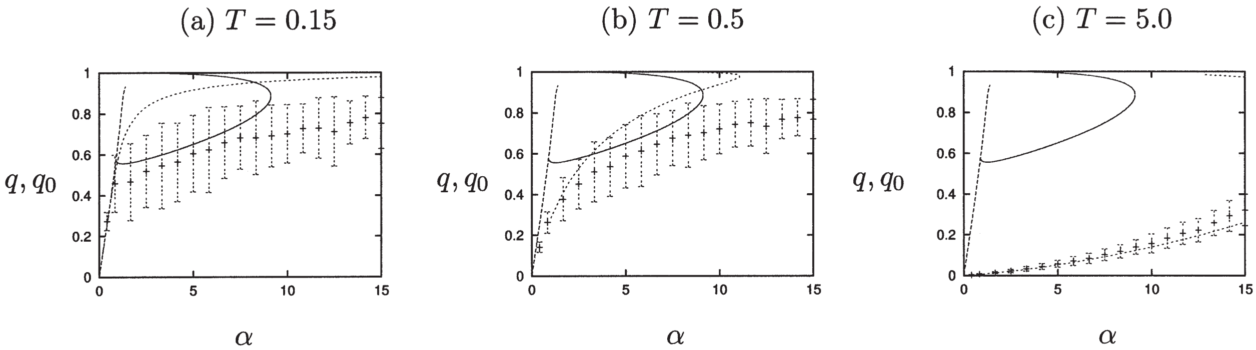

Fig. 19. The α dependences of q and q_0 obtained using the Gibbs algorithm for $\delta = 0$. +: numerical results (with standard deviations) for $N = 12$. Dashed curve: RS solution for $T = 0$. Dotted curve: RS solution for finite temperature.Solid curve: 1RSB solution.

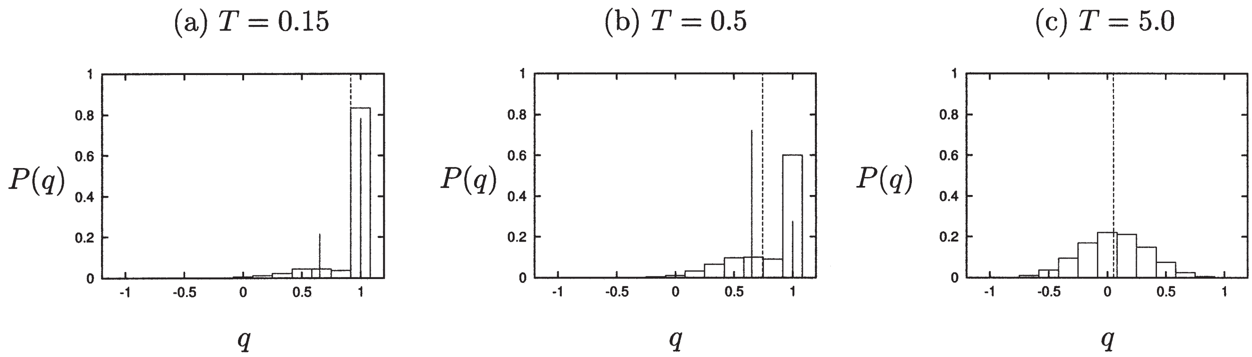

Fig. 20. The T dependence of $P(q)$ obtained using the Gibbs algorithm for $\delta = 0$. Here, $T_c \simeq 0.7$. Histogram: numerical results for $N = 12$ and $p = 60$. Solid line: 1RSB solution. Dotted line: RS solution. For $T = 5$, no 1RSB solution exists.

algorithm, we carried out calculations for $N = 10$ and 12 with several temperatures. We included all states in the calculation. We found that $r(\alpha, N, N')$ is less than 5% and $\delta r(\alpha, N, N')$ is less than 10% for $N, N' = 10$ and 12. In Fig. 18, the α dependences of R and $\Delta\epsilon_g$ are plotted for $T = 1.0$. The numerical results agree with the theoretical results within the numerical standard deviations up to $\alpha \lesssim \alpha_{max}$ for R and up to $\alpha \lesssim 5$ for $\Delta\epsilon_g$. Also, the α dependences of q and q_0 are displayed in Fig. 19 for $T = 0.15$, 0.5 and 5.0. As shown in Fig. 19 numerical results agree with the theoretical results within the numerical standard deviations as long as $\alpha \lesssim \alpha_{max}$ for $T = 0.15$ and 0.5 and any α in the figure for $T = 5.0$. The disagreement between numerical and theoretical results for α around and above α_{max} is due to the finite size effect as before. Concerning q_0, since a relevant quantity is $P(q)$, we calculated it at $\alpha = 5$ for several temperatures. In Fig. 20, we show the numerical results for $P(q)$ together with the theoretical results. We see that the positions of the peak values for $T = 0.15$ and 5.0 agree for the theoretical and numerical results. The disagreement for $T = 0.5$ is attributed to the finite size effect as before.

In conclusion, in both the cases $\delta = 1$ and 0, although there exist finite size effects, as is reflected in the behaviour of q_0, as a whole, the theoretical results and numerical results agree fairly well, in which theoretically we have employed the RS and the 1RSB ansatz.

7. Summary and Discussion

In this paper, we studied a model of supervised learning in which perceptrons with Ising weights learn from stochastic examples. Using the replica method, we obtained the necessary and sufficient conditions for the existence of PL and the conditions for learning curves that exhibit power law forms in the asymptotic region as $\alpha \to \infty$. These conditions are given in terms of δ, which characterizes a certain local property of the rules by which examples are drawn. First, let us summarize the results in more detail.

The basic ingredients of the model are as follows. When an input vector $\boldsymbol{x}$ is given, the probability $p_r(+1|\boldsymbol{x})$ that the teacher returns an output $+1$ is a function of the inner product between the input $\boldsymbol{x}$ and the teacher's weight $\boldsymbol{w}^{\mathrm{o}}$ and takes the form

$$p_r(+1|\boldsymbol{x}) = \mathcal{P}(u^{\mathrm{o}}) = \frac{1 + P(u^{\mathrm{o}})}{2},$$
$$u^{\mathrm{o}} = (\boldsymbol{x} \cdot \boldsymbol{w}^{\mathrm{o}})/\sqrt{N}, \quad |\boldsymbol{x}| = \sqrt{N}, \quad |\boldsymbol{w}^{\mathrm{o}}| = \sqrt{N}.$$

Further, we stipulate $P(y)$ to be a non-decreasing function that behaves near $y = 0$ as $P(y) \simeq a\,\mathrm{sgn}(y)|y|^{\delta}$ ($\delta \geq 0$). For simplicity, we choose $P(y)$ to be an odd function. As the learning algorithm, we used the Gibbs algorithm.

With these basic ingredients, we obtained the following results.

Conditions for PL

The necessary and sufficient conditions for the existence of perfect learning are the following.

(1) $0 \leq \delta < 1/2$ when $0 < \beta < \infty$, where β is the inverse temperature.

(2) Deterministic case. That is, the target relation is deterministic obeying the perceptron rule, and the alogorithm is the minimum-error algorithm, i.e. the Gibbs algorithm with $\beta \to \infty$.

Behaviour of learning curves

With the RS and the 1RSB ansatz, we found that the behaviour of the generalization error ϵ_{g} can be classified into the following three categories, according to the value of δ.

(1) For $0 \leq \delta < \frac{1}{2}$, at $\alpha = \alpha_{\max}$ there is a first-order phase transition from the RS solution with positive entropy or from the 1RSB solution to the PL solution.

(2) For $\delta = \frac{1}{2}$ and large α, the 1RSB solution appears, and ϵ_{g} for this solution decays exponentially according to

$$\Delta\epsilon_{\mathrm{g}} \propto e^{-3F_0\alpha},$$

where F_0 is a constant.

(3) For $\delta > \frac{1}{2}$ and large α, the 1RSB solution appears, and ϵ_{g} for this solution decays according to a power law with a logarithmic correction:

$$\Delta\epsilon_{\mathrm{g}} \propto \left(\frac{\ln\alpha}{\alpha}\right)^{\frac{1+\delta}{2\delta-1}}.$$

To check these results, we carried out several numerical calculations, in which we solved the saddle point equations for the RS and 1RSB solutions and directly computed the concerned quantities using enumeration methods. The latter results show fairly good agreement with the former results. As mentioned in the Introduction, Seung also investigated the existence of PL in the situation that the weights are Ising and the rule to be learnt is stochastic,[13] employing the annealed approximation. He classified the learning behaviour of Ising networks by introducing two exponents y and z used in the following manners. The exponent y is defined as follows. Let $\rho(\epsilon_{\mathrm{g}})$ be the logarithm of the number of weight vectors whose generalization errors have a value ϵ_{g}. Then, he assumed that when $\Delta\epsilon_{\mathrm{g}} = \epsilon_{\mathrm{g}} - \epsilon_{\min}$ is small, $\rho(\epsilon_{\mathrm{g}})$ increases as $\rho(\epsilon_{\mathrm{g}}) \sim \mathcal{O}((\Delta\epsilon_{\mathrm{g}})^y)$, where $\epsilon_{\min}$ is the minimum value of the generalization error, obtained with the unique optimal weight vector $\boldsymbol{w}^{\mathrm{o}}$. The exponent z characterizes $e_{\mathrm{d}}(\boldsymbol{w}, \boldsymbol{w}^{\mathrm{o}})$, which is the probability that the output for the weight vector $\boldsymbol{w}$ differs from that for the optimal weight vector $\boldsymbol{w}^{\mathrm{o}}$. He also assumed that $e_{\mathrm{d}}(\boldsymbol{w}, \boldsymbol{w}^{\mathrm{o}})$ scales as $e_{\mathrm{d}}(\boldsymbol{w}, \boldsymbol{w}^{\mathrm{o}}) \sim \mathcal{O}((\Delta\epsilon_{\mathrm{g}})^z)$. He derived upper bounds for the generalization errors and found that the behaviour of the learning curves depends on the values of y and z. His results are summarized as follows.

(1) If $y + z > 2$, there is a first-order transition.

(2) If $y + z < 2$, the generalization error decays according to a power law: $\Delta\epsilon_{\mathrm{g}} \sim \alpha^{-\frac{1}{2-y-z}}$.

(3) If $y + z = 2$, there is a second-order transition, or the generalization error decays exponentially.

The exponents y and z in Seung's model correspond to $\frac{2}{1+\delta}$ and $\frac{1}{1+\delta} = \frac{y}{2}$, respectively, in our model. Therefore, it is found that our results concerning typical learning behaviour are consisitent with those of Seung's results, which are the upper bounds of the learning curves.

Regarding the condition for the existence of PL, we note that for $\beta = \infty$ (i.e. $T = 0$), PL does not exist for learning from stochastic examples. This follows from the fact that for $T = 0$ and for large α there exists no student whose outputs are the same as the teacher's, because the teacher makes mistakes. Thus, the measure of weight vectors whose energies are 0 vanishes for large α. However, for $T \to 0$, we consider the weight vectors of the minimum energy, and there is at least one solution of $\boldsymbol{w} = \boldsymbol{w}^{\mathrm{o}}$ when α is sufficiently large. Hence, PL is possible in the limit $T \to 0$.

As a student learns, its weight vector tends to $\boldsymbol{w}^{\mathrm{o}}$. Examples that import a crucial influence on learning are those for which $u^{\mathrm{o}} = (\boldsymbol{x} \cdot \boldsymbol{w}^{\mathrm{o}})/\sqrt{N} \sim 0$. The more slowly the probability $\mathcal{P}(u)$ varies around $u = 0$ for large δ, the more difficult it is for students to realize the optimal vector $\boldsymbol{w}^{\mathrm{o}}$. This is the reason that it becomes more difficult to realize PL as δ increases.

Acknowledgements

I am grateful to Y. Kabashima for valuable discussions and collaboration in the first part of this study. I am also grateful to A. C. C. Coolen and P. Sollich for valuable discussions and suggestions during his stay at King's College, University of London.

Appendix A: Derivation of the Free Energy

Here, we derive the free energy using the replica method. Introducing n replicas, the partition function Z^n becomes

$$Z^n = \mathrm{Tr}\prod_{a=1}^{n}\left[e^{-\beta}\int_{-\infty}^{0} d\lambda_\mu^a + \int_0^\infty d\lambda_\mu^a\right]\int_{-\infty}^{\infty}\frac{dy_\mu^a}{2\pi}\exp[-iy_\mu^a(r_\mu^{\mathrm{o}}u_\mu^a - \lambda_\mu^a)], \tag{A·1}$$

where $u_\mu^a = (\boldsymbol{x}^\mu \cdot \boldsymbol{w}^a)/\sqrt{N}$ and Tr represents the summation over all configurations of $\boldsymbol{w}^a, a = 1, \cdots n$. Defining the overlap between the weight vector of a student and the optimal weight vector as $R^a = \frac{1}{N}\sum_{j=1}^N w_j^a w_j^{\mathrm{o}}$ and the overlap between the weight vectors of students as $q^{ab} = \frac{1}{N}\sum_{j=1}^N w_j^a w_j^b$, and using the relations

$$\delta\left(\sum_{j=1}^N (x_j^\mu)^2 - N\right) = \int_{-i\infty}^{i\infty}\frac{d\tilde{K}_\mu}{2\pi i}\exp\left[-\tilde{K}_\mu\left(\sum_{j=1}^N (x_j^\mu)^2 - N\right)\right],$$

$$1 = \prod_a \int dR^a \int_{-i\infty}^{i\infty}\frac{Nd\hat{R}^a}{2\pi i}\exp\left[-N\hat{R}^a\left(R^a - \frac{1}{N}\sum_{j=1}^N w_j^a w_j^{\mathrm{o}}\right)\right],$$

$$1 = \prod_{a<b}\int dq^{ab}\int_{-i\infty}^{i\infty}\frac{Nd\hat{q}^{ab}}{2\pi i}\exp\left[-N\hat{q}^{ab}\left(q^{ab} - \frac{1}{N}\sum_{j=1}^N w_j^a w_j^b\right)\right],$$

we take the average over r_μ^{o} and $\boldsymbol{x}^\mu$ and obtain the following expression for $\langle Z^n\rangle_{\xi_p, w^{\mathrm{o}}}$:

$$\langle Z^n\rangle_{\xi_p, w^{\mathrm{o}}} = \int\left[\prod_{a<b} dq^{ab}\frac{Nd\hat{q}^{ab}}{2\pi i}\right]\left[\prod_a dR^a\frac{Nd\hat{R}^a}{2\pi i}\right]e^{\mathrm{NG}}, \tag{A·2}$$

where

$$G = \frac{p}{N}G_1(\{q^{ab}\},\{R^a\}) + G_2(\{\hat{q}^{ab}\},\{\hat{R}^a\}) - \sum_a \hat{R}^a R^a - \sum_{a<b}\hat{q}^{ab}q^{ab}, \tag{A·3}$$

with

$$e^{G_1} = \left[\prod_a\left(e^{-\beta}\int_{-\infty}^0 d\lambda^a + \int_0^\infty d\lambda^a\right)\int_{-\infty}^\infty\frac{dy^a}{2\pi}\right]$$
$$\times\exp\left[-\frac{1}{2}\sum_a (y^a)^2 - \sum_{a<b} q^{ab}y^a y^b + i\sum_a y^a\lambda^a\right]\Psi\left(\sum_a y^a R^a\right), \tag{A·4}$$

$$e^{G_2} = \mathrm{Tr}\exp\left[\sum_a \hat{R}^a w^a + \sum_{a<b}\hat{q}^{ab}w^a w^b\right], \tag{A·5}$$

$$\Psi(y) \equiv \frac{1}{\sqrt{2\pi}}\int d\xi e^{-\frac{1}{2}(\xi - iy)^2}\left\{1 - \frac{1}{2}[P(\xi) - P(-\xi)]\right\}.$$

Here Tr represents the summation over $w^a, a = 1, \ldots n$. In the above expressions, we set $\tilde{K}_\mu = 1/2$, which is the optimal value. When $P(-y) = -P(y)$, $\Psi(y)$ becomes

$$\Psi(y) = \frac{1}{\sqrt{2\pi}}\int d\xi e^{-\frac{1}{2}(\xi - iy)^2}2\mathcal{P}(-\xi). \tag{A·6}$$

The general form of the free energy per synaptic weight is given by

$$f = -\frac{\langle\ln Z\rangle_{\xi_p, w^{\mathrm{o}}}}{N\beta} = -\frac{G}{n\beta}. \tag{A·7}$$

Appendix B: Derivation of the Limiting Forms of the Expressions for $\hat{q}$ and $\hat{R}$ as $q \to 1$ and $R \to 1$

In this appendix, we briefly derive the asymptotic relations for $\hat{q}$ and $\hat{R}$. First, we consider the case $0 < \beta < \infty$. Equation (12) for $\hat{q}$ can be rewritten as

$$\hat{q} = \frac{\alpha Q}{1 - q}\frac{(1 - e^{-\beta})^2}{\sqrt{2\pi}}(A - B), \tag{B·1}$$

where

$$A = \int du e^{-Q^2u^2/2}\frac{h(u)^2}{\tilde{H}(u)^2}, \tag{B·2}$$

$$B = \frac{1}{\sqrt{2 + Q^2}}\frac{1}{\sqrt{2\pi}}\int_0^\infty DzP(\varepsilon z)\int_{-\infty}^\infty DtH_2\left[\kappa\left(t + \frac{1}{\chi}\sqrt{\frac{1 - \xi^2}{2 + Q^2}}z\right)\right], \tag{B·3}$$

with $H_2(u) = \frac{1}{\tilde{H}(u)^2} - \frac{1}{\tilde{H}(-u)^2}$, $\kappa = \frac{\chi}{\sqrt{1+2\chi^2}}$ and $\varepsilon = \sqrt{\frac{Q^2+2\xi^2}{2+Q^2}}$. It follows that $H_2(u)$ is a strictly increasing odd function and $0 < |H_2(u)| < e^{2\beta} - 1$ for $u \neq 0$. Thus, for $\delta > 0$, $P(\varepsilon z)$ can be replaced by $a(\varepsilon z)^\delta$ in eq. (B·3), and we obtain

$$B \simeq \frac{1}{2\sqrt{\pi}} \int_0^{1/\varepsilon} Dza(\varepsilon z)^\delta \int_{-\infty}^{\infty} DtH_2\left[\kappa\left(t + \frac{1}{\sqrt{2}\chi} z\right)\right] + \mathcal{O}(H(1/\varepsilon)).$$

As q and R tend to 1, ε tends to 0, and thus $B \to 0$. Also, for $T > 0$, A is finite in these limits. Thus, in these limits we have $A - B \simeq A$. Therefore, for $\delta > 0$ we obtain

$$\hat{q} \simeq \frac{\alpha}{\sqrt{\Delta q}} g_{1,\delta}(\chi, \beta), \tag{B·4}$$

where

$$g_{1,\delta}(\chi, \beta) = \frac{1}{\sqrt{2\pi}} \int du \tilde{\varphi}(u)^2, \tag{B·5}$$

with $\Delta q \equiv 1 - q$. For the case $\delta = 0$, from eq. (B·3), we obtain

$$B \simeq \frac{1}{2\sqrt{\pi}} \int_0^{\infty} Dzk \int_{-\infty}^{\infty} DtH_2\left[\frac{\chi}{\sqrt{1+2\chi^2}}\left(t + \frac{1}{\sqrt{2}\chi} z\right)\right] = k \int du \frac{h(u)^2}{\tilde{H}(u)^2}[1 - 2H(u/\chi)],$$

where $k = \lim_{y\to+0} P(y)$. Then, we obtain

$$A - B \simeq \int du \frac{h(u)^2}{\tilde{H}(u)^2}[1 - k + 2kH(u/\chi)],$$

thus,

$$\hat{q} \simeq \frac{\alpha}{\sqrt{\Delta q}} g_{1,0}(\chi, \beta), \tag{B·6}$$

where

$$g_{1,0}(\chi, \beta) \equiv \frac{1}{\sqrt{2\pi}} \int du \tilde{\varphi}(u)^2 [1 - k + 2kH(u/\chi)]. \tag{B·7}$$

Now, we derive approximate expressions for $\hat{R}$. The eq. (13) can be rewritten as

$$\hat{R} = \frac{\alpha}{q\xi} \frac{(1 - e^{-\beta})}{\sqrt{2\pi}} D, \tag{B·8}$$

$$D = \int due^{-Q^2u^2/2} \frac{h(u)}{\tilde{H}(u)} w(u) = \nu \int_{-\infty}^{\infty} DxH_1(\nu x) \int_0^{\infty} DyP(\xi y)e^{-\eta xy}(y + \eta x), \tag{B·9}$$

where

$$H_1(x) = \frac{1}{\tilde{H}(x)} + \frac{1}{\tilde{H}(-x)}, \quad \nu = \frac{\xi}{\sqrt{\xi^2 + Q^2}}, \quad \eta = \frac{Q}{\sqrt{\xi^2 + Q^2}} \frac{R}{\sqrt{q}}.$$

For $\delta > 0$, D is given by

$$D = \frac{\xi}{\sqrt{1 + Q^2}} \int_0^{\infty} DyP'(\zeta z)\psi\left(\frac{1}{\sqrt{1+\chi^2}} \frac{R}{\sqrt{q}} \frac{z}{\sqrt{1+Q^2}}\right),$$

where $\zeta = \sqrt{\frac{Q^2+\xi^2}{1+Q^2}}$ and $\psi(z) = \int_{-\infty}^{\infty} DtH_1(\nu t - z)$. Since $\psi(z)$ is bounded, D can be evaluated as follows.

$$D \simeq \xi\left[\int_0^{1/\zeta} Dza\delta(\zeta z)^{\delta-1}\psi\left(\frac{z}{\sqrt{1+\chi^2}}\right) + \mathcal{O}\{H(1/\zeta)\}\right] \simeq \xi a\delta\zeta^{\delta-1} \int_0^{\infty} Dzz^{\delta-1}\psi\left(\frac{z}{\sqrt{1+\chi^2}}\right).$$

Therefore, we obtain

$$\hat{R} \simeq \alpha \frac{\xi^\delta}{\sqrt{\Delta q}} g_{2,\delta}(\chi, \beta), \tag{B·10}$$

where

$$g_{2,\delta}(\chi, \beta) \equiv \frac{a\delta}{\sqrt{2\pi}}(1 - e^{-\beta})\frac{1}{\chi}(1 + \chi^{-2})^{(\delta-1)/2} \int_0^{\infty} Dzz^{\delta-1}\psi\left(\frac{z}{\sqrt{1+\chi^2}}\right). \tag{B·11}$$

For $\delta = 0$, from eq. (B·9), we obtain

$$D = \nu \int_{-\infty}^{\infty} DxH_1(\nu x)\left[\frac{k}{\sqrt{2\pi}} + \xi \int_0^{\infty} Dye^{-\eta xy}P'(\xi y)\right].$$

We assume that $|P'(y)|$ is bounded.[17] Then the second term in the parenthesis is $\mathcal{O}(\xi)$, and it can be ignored. This yields

$$D \simeq \frac{k\nu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} DxH_1(\nu x) = \frac{2k}{\sqrt{2\pi}} \frac{\chi}{\sqrt{1+\chi^2}} \int_{-\infty}^{\infty} Dx \frac{1}{\tilde{H}\left(\frac{\chi x}{\sqrt{1+\chi^2}}\right)}.$$

Thus, we obtain

$$\hat{R} \simeq \alpha \frac{1}{\sqrt{\Delta q}} g_{2,0}, \tag{B·12}$$

where

$$g_{2,0} = \frac{k(1-e^{-\beta})}{\pi} \frac{1}{\sqrt{1+\chi^2}} \int_{-\infty}^{\infty} Dx \frac{1}{\tilde{H}\left(\frac{\chi x}{\sqrt{1+\chi^2}}\right)}. \tag{B·13}$$

Now, let us consider the case $\beta = \infty$. In this case, $\tilde{\varphi}(u)$ becomes $\varphi(u) = \frac{H'(u)}{H(u)}$. Therefore, from eq. (12) we obtain

$$\hat{q} = \frac{\alpha}{\Delta q} \int Du\varphi(u/Q)^2 E(u/Q) \simeq \frac{\alpha}{\Delta q}(A' - B'),$$

where

$$A' = \int_0^{\infty} Du(u/Q)^2 = \frac{1}{2Q^2},$$

$$B' = \frac{\xi}{\sqrt{2\pi}Q^2} \int_0^{\infty} DyP(\xi y)$$
$$\times \left\{2\sqrt{1-\xi^2}\xi^2 y + \sqrt{2\pi}\left[\xi^2 + (1-\xi^2)\xi^2 y^2\right] e^{\frac{1-\xi^2}{2}y^2}\left[1 - 2H\left(\sqrt{1-\xi^2}y\right)\right]\right\} \tag{B·14}$$

$$\simeq \frac{a}{\pi} \frac{\xi^{3+\delta}}{Q^2} + \frac{1}{Q^2} \int_0^{\infty} DyP(y)y^2. \tag{B·15}$$

Thus, $A' - B' \simeq \frac{1}{Q^2}\int_0^{\infty} Dy(1-P(y))y^2$, which gives

$$\hat{q} \simeq \frac{\alpha}{(\Delta q)^2} g_3, \tag{B·16}$$

where

$$g_3 \equiv \int_0^{\infty} Dyy^2[1-P(y)]. \tag{B·17}$$

If $P(u) = 1$ (i.e. in the deterministic case), this is 0. We discuss this case later. Similarly, from eq. (13) for $\hat{R}$, we obtain

$$\hat{R} \simeq \frac{\alpha}{\xi Q^2}\left\{\frac{2\xi^4}{\sqrt{2\pi}} \int_0^{\infty} DyP(\xi y)y - \xi\sqrt{1-\xi^2} \int_0^{\infty} DyP(y)(1-y^2)\left[1 - 2H\left(\frac{\sqrt{1-\xi^2}}{\xi}y\right)\right]\right\}$$
$$\simeq \frac{\alpha}{\Delta q} g_4, \tag{B·18}$$

where

$$g_4 \equiv \int_0^{\infty} DyP(y)(y^2 - 1). \tag{B·19}$$

If $P(y)$ is not constant for $y > 0$, the integral here is positive. If $P(y)$ is constant for $y > 0$, which can be the case when $\delta = 0$, the integral is 0. In the latter case, i.e. in the case that $P(y) = k$ for $y > 0$, we have to consider higher order terms. From eq. (14), $w(u)$ is calculated as

$$w(u) = e^{-v^2/2} \int_0^{\infty} Dyk[(y+v)e^{-vy} + (y-v)e^{vy}] = e^{-v^2/2} \frac{2k}{\sqrt{2\pi}},$$

where $v = -\frac{R}{\sqrt{q}\chi}u$. Thererore, from eq. (13), $\hat{R}$ is calculated as

$$\hat{R} = -\frac{\alpha}{\sqrt{q-R^2}}\int \tilde{D}u\varphi(u)w(u) \simeq -\frac{\alpha}{\xi}\int \frac{du}{\sqrt{2\pi}}\varphi(u)e^{-\frac{u^2}{2\chi^2}}\frac{2k}{\sqrt{2\pi}}$$
$$\simeq \frac{\alpha}{\sqrt{\Delta q}}\frac{k}{\pi}\frac{1}{\sqrt{1+\chi^2}}\int Dy\frac{1}{H\left(\frac{\chi y}{\sqrt{1+\chi^2}}\right)}.$$

Thus, we obtain

$$\hat{R} \simeq \frac{\alpha}{\sqrt{\Delta q}}g_5(\chi), \tag{B·20}$$

where

$$g_5(\chi) \equiv \frac{k}{\pi}\frac{1}{\sqrt{1+\chi^2}}\int Dy\frac{1}{H\left(\frac{\chi y}{\sqrt{1+\chi^2}}\right)}. \tag{B·21}$$

Finally, we consider the deterministic ($P(u) = 1$) case. In this case, $\delta = 0$, $k = 1$, $q = R$ and $\hat{q} = \hat{R}$ hold. Then, we obtain $\frac{\sqrt{1-\xi^2}}{\xi} = \frac{1}{Q}$ and $E(u/Q) = 2H(u/Q)$. Thus, eq. (12) becomes

$$\hat{q} = \frac{2\alpha}{1-q}\int Du\frac{h(u/Q)^2}{H(u/Q)} \simeq \frac{\alpha}{\sqrt{\Delta q}}\frac{2}{\sqrt{2\pi}}\int Du\frac{h(u)}{H(u)} = \frac{\alpha}{\sqrt{\Delta q}}g_{3,D}, \tag{B·22}$$

where

$$g_{3,D} = \frac{2}{\sqrt{2\pi}}\int Du\frac{h(u)}{H(u)}. \tag{B·23}$$

Also, since $v = -u$, from eq. (12), we obtain

$$\hat{R} = \frac{2\alpha}{1-q}\int Du\frac{h(u/Q)^2}{H(u/Q)} = \hat{q}. \tag{B·24}$$

Appendix C: Asymptotic Form of I for $\tau \gg 1$ and $\hat{R} \gg 1$

I is given by

$$I = \int Dt\ln[2\cosh(\sqrt{\hat{q}}t + \hat{R})] \simeq \hat{R} + \frac{h(\tau)}{\tau\mu} + I_1, \tag{C·1}$$

where

$$I_1 = I_1^- + I_1^+,$$
$$I_1^{\pm} = \int_{\pm\tau}^{\infty} Dt\ln[1+e^{-2\sqrt{\hat{q}}(t\mp\tau)}] = \sqrt{2\pi}h(\tau)\int_0^{\infty} Dxe^{\mp x\tau}\ln(1+e^{-2\sqrt{\hat{q}}x}),$$

with $\tau = \hat{R}/\sqrt{\hat{q}}$ and $\mu = \frac{\hat{R}}{2\hat{q}}$. The following relations can be demonstrated rigorously:

$$I_1^{\pm} = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}e^{2\hat{q}n^2\pm 2\hat{R}n}H\left[\tau\left(\frac{n}{\mu}\pm 1\right)\right]. \tag{C·2}$$

For $0 < \mu < 1$, $H(\tau(\frac{n}{\mu}-1))$ and $H(\tau(\frac{n}{\mu}+1))$ in (C·2) can be approximated by $\frac{h(\tau(\frac{n}{\mu}-1))}{\tau(\frac{n}{\mu}-1)}$ and $\frac{h(\tau(\frac{n}{\mu}+1))}{\tau(\frac{n}{\mu}+1)}$, respectively. Therefore, I_1 is given approximately as

$$I_1 \simeq \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}h(\tau)\left[\frac{1}{\tau\left(\frac{n}{\mu}-1\right)} + \frac{1}{\tau\left(\frac{n}{\mu}+1\right)}\right] = \frac{2\mu}{\tau}h(\tau)c(\mu),$$

where $c(\mu) \equiv \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^2}\frac{1}{1-(\mu/n)^2}$. When μ is not an integer, we have

$$c(\mu) = \frac{\pi}{2\mu\sin(\mu\pi)} - \frac{1}{2\mu^2}.$$

Thus, we obtain

$$I \simeq \hat{R} + \frac{h(\tau)}{\tau\mu}[1+2\mu^2c(\mu)] = \hat{R} + \frac{\psi(\mu)h(\tau)}{\tau\mu} \quad \text{for} \quad 0 < \mu < 1,$$

where

$$\psi(\mu) \equiv 1 + 2\mu^2 c(\mu) = \frac{\pi\mu}{\sin(\pi\mu)}.$$

Noting that $c(0) = \pi^2/12$ and $\psi(0) = 1$ we find that I behaves near $\mu = 0$ as

$$I \simeq \hat{R} + \frac{h(\tau)}{\tau\mu} \text{ (for } \mu \approx 0).$$

For $\mu \geq 1$, I_1^- can be expressed as

$$I_1^- \simeq \sum_{n=1}^{n_0} \frac{(-1)^{n-1}}{n} e^{2\hat{q}n^2 - 2\hat{R}n} \left\{1 - H\left[-\tau\left(\frac{n}{\mu} - 1\right)\right]\right\} + \frac{h(\tau)}{\tau} \sum_{n=n_0+1}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{\dfrac{n}{\mu} - 1}, \tag{C·3}$$

where $n_0 = [\mu]$ (i.e., n_0 is the largest integer that does not exceed the value of μ). Let us compare the terms in the eq. (C·3). Let us assume $1 \leq n_1 < n_2 \leq n_0$. Then we have

$$e^{2\hat{q}n_1^2 - 2\hat{R}n_1} / e^{2\hat{q}n_2^2 - 2\hat{R}n_2} = e^{4\hat{q}(n_2 - n_1)(\mu - (n_2 - n_1)/2)} > e^{4\hat{q}(n_2 - n_1)(n_0 - n_1)}.$$

Thus, we obtain

$$e^{2\hat{q}n_1^2 - 2\hat{R}n_1} \gg e^{2\hat{q}n_2^2 - 2\hat{R}n_2} \quad \text{for} \quad \hat{q} \gg 1.$$

Note that $\hat{q} \gg 1$ is satisfied when $\hat{R} \gg 1$ as long as $\mu = \frac{\hat{R}}{2\hat{q}}$ is bounded from above. Further, since $e^{-\tau^2/2}/e^{2\hat{q}n^2 - 2\hat{R}n} = e^{-2\hat{q}(n-\mu)^2}$, each term in the first summation in the eq. (C·3) is of lower order than $h(\tau)/\tau$ for $\tau \gg 1$. Thus, I_1^+ and the second summation in I_1^- are of higher order than the terms in the first summation in I_1^-. Therefore, for $1 < \mu < 2$, we find

$$I_1^- \simeq e^{-2(\hat{R}-\hat{q})} + \frac{2h(\tau)\mu}{\tau} c^-(\mu), \quad I_1^+ \simeq \frac{2h(\tau)\mu}{\tau} c^+(\mu),$$

where

$$c^{\pm}(\mu) \equiv \frac{1}{2\mu} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{\dfrac{n}{\mu} \pm 1}.$$

Thus,

$$I_1 \simeq e^{-2(\hat{R}-\hat{q})} + \frac{2h(\tau)\mu}{\tau} c(\mu),$$

where $c(\mu) = c^-(\mu) + c^+(\mu)$. Therefore,

$$I \simeq \hat{R} + e^{-2(\hat{R}-\hat{q})} + \frac{h(\tau)\psi(\mu)}{\tau\mu} \quad \text{for} \quad 1 < \mu < 2.$$

For $\mu = 1$, we obtain

$$I_1 \simeq e^{-2(\hat{R}-\hat{q})}/2 + \frac{2h(\tau)\mu}{\tau} c_2(\mu),$$

where $c_2(\mu)$ is defined as

$$c_2(\mu) \equiv \frac{1}{2\mu} \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{\dfrac{n}{\mu} - 1} + c^+(\mu) = c(\mu) - \frac{1}{2(1-\mu)}.$$

Thus,

$$I \simeq \hat{R} + e^{-2(\hat{R}-\hat{q})}/2 \quad \text{for} \quad \mu = 1.$$

Here, we have used the facts that $c_2(\mu)$ is analytic at $\mu = 1$ and that $c_2(1) = 0$.

For $\mu \geq 2$, we obtain

$$I_1^- \simeq e^{-2(\hat{R}-\hat{q})} - \frac{1}{2} e^{-4(\hat{R}-2\hat{q})} H\left[\tau\left(\frac{2}{\mu} - 1\right)\right].$$

Therefore we have

$$I_1 \simeq I_1^- \simeq e^{-2(\hat{R}-\hat{q})} - \frac{1}{2} b_\mu e^{-4(\hat{R}-2\hat{q})},$$

where

$$b_\mu = 1 \quad \text{for} \quad \mu > 2, \quad b_2 = 1/2.$$

Thus,

$$I \simeq \hat{R} + e^{-2(\hat{R}-\hat{q})} - \frac{1}{2} b_\mu e^{-4(\hat{R}-2\hat{q})} \quad \text{for} \quad \mu \geq 2.$$

In summary, up to second order in I, we obtain

$$I \simeq \begin{cases} \hat{R} + \dfrac{h(\tau)\psi(\mu)}{\tau\mu} & (\text{for } 0 \lesssim \mu < 1), \\ \hat{R} + a_\mu e^{-2(\hat{R}-\hat{q})} & (\text{for } 1 \leq \mu), \end{cases} \tag{C·4}$$

where

$$\psi(\mu) \equiv 1 + 2\mu^2 c(\mu) = \frac{\pi\mu}{\sin(\pi\mu)} \tag{C·5}$$

and $a_1 = 1/2$ and $a_\mu = 1$ for $\mu > 1$.

Appendix D: Demonstration that $S_{\mathrm{PL}} = 0$ and $f_{\mathrm{PL}} = \alpha\epsilon_{\min}$.

As seen from eq. (28), when $q \to 1$ and $R \to 1$ for $0 < \beta < \infty$ and for any δ and any χ, S_{RS} can be expressed as

$$S_{\mathrm{RS}} = -\frac{\hat{q}\Delta q}{2} - (1 - \Delta R)\hat{R} + \alpha\sqrt{\Delta q} r(\chi, \beta) + I. \tag{D·1}$$

We consider the case in which $0 < \beta < \infty$ and $0 \leq \delta < 1/2$. For PL, $R = q = 1, \chi = 1$ and $\xi = Q = \sqrt{\Delta q}$. Therefore, here we have

$$\hat{q} = \frac{\alpha}{\sqrt{\Delta q}} g_{1,\delta}(1, \beta), \tag{D·2}$$

$$\hat{R} = \alpha(\Delta q)^{(\delta-1)/2} g_{2,\delta}(1, \beta). \tag{D·3}$$

For $0 < \beta < \infty$ and $\delta \geq 0$, $g_{1,\delta}(1, \beta)$ and $g_{2,\delta}(1, \beta)$ are finite.

Hence, with $\mu = \frac{\hat{R}}{2\hat{q}} = \frac{g_{2,\delta}}{2g_{1,\delta}}(\Delta q)^{\delta/2}$, we find for $\delta > 0$, $\mu \simeq 0$ and for $\delta = 0$, $\mu = \frac{g_{2,0}}{2g_{1,0}}$, which is finite. Next, we determine an approximate expression for I. In the case $\delta > 0$, since $\mu = 0$, we obtain from (C·4)

$$I \simeq \hat{R} + \frac{h(\tau)}{\tau\mu} = \hat{R} + \frac{2\hat{R}}{\tau^3}h(\tau).$$

Thus, we find

$$S = -\frac{1}{2}g_{1,\delta}\alpha\sqrt{\Delta q} + g_{2,\delta}\alpha(\Delta q)^{\frac{1+\delta}{2}} + \frac{2\hat{R}}{\tau^3}h(\tau) + \alpha r\sqrt{\Delta q}.$$

From eqs. (D·2) and (D·3), we obtain

$$\hat{q}^{\delta-1}\hat{R} = (\alpha g_{1,\delta})^{\delta-1}\alpha g_{2,\delta} \equiv C.$$

From this we find

$$\frac{\hat{R}}{\tau^3} = C^{-\frac{1}{1-2\delta}}\tau^{\frac{4\delta-1}{1-2\delta}}.$$

Then, since $\Delta q = 0, \tau = \infty$ and $g_{1,\delta}, g_{2,\delta}$, r and C are finite, we obtain $S = 0$. For the case $\delta = 0$, we have to determine the value of μ. As discussed in §3, when χ is finite, $\mu = \frac{1}{1+\chi^2}$ [as seen from eq. (30)]. In the case considered presently, $\chi = 1$, and therefore $\mu = 1/2$. Thus, we obtain

$$I \simeq \hat{R} + \frac{\pi h(\tau)}{\tau}.$$

For $\delta = 0$, using an argument similar to that used above for $\delta > 0$, we can again obtain $S = 0$. Now, let us determine $\langle e_t \rangle = -\alpha e^{-\beta}J$. J is given by

$$\begin{aligned} J &= \int Du\frac{H(u/Q)-1}{\tilde{H}(u/Q)}\int Dy[1-P(\xi y+\sqrt{1-\xi^2}u)] \\ &= \int Du\frac{H(u/Q)-1}{\tilde{H}(u/Q)}[1-P(u)] \\ &= -e^{\beta}\int_0^{\infty}Du[1-P(u)] = -e^{\beta}\epsilon_{\min}. \end{aligned}$$

Thus, $\langle e_t \rangle = -\alpha e^{-\beta}J = \alpha\epsilon_{\min}$. Finally, we obtain $f_{\rm PL} = \langle e_t \rangle - TS_{\rm PL} = \alpha\epsilon_{\min}$.

Appendix E: Asymptotic Forms of $f_{\rm RS}$ and $S_{\rm RS}$ for $\beta \ll 1$ ($T \gg 1$).

In this appendix, we derive the asymptotic forms of the free energy and the entropy for the RS solution in the case $\beta \ll 1$. The free energy $f_{\rm RS}$ is expressed as

$$-\beta f_{\rm RS}(q,\hat{q},R,\hat{R},\beta) = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + \alpha K + I, \tag{E·1}$$

$$K \equiv \int Dy2\mathcal{P}(y)\int Du\ln\tilde{H}(Y), \quad I \equiv \int Dt\ln[2\cosh(\sqrt{\hat{q}}t+\hat{R})],$$

where $Y = \frac{\sqrt{q-R^2}u-Ry}{\sqrt{1-q}}$. By defining K_a and K_b as

$$K_a \equiv \int Dy2\mathcal{P}(y)\int DuH(-Y) = \epsilon_{\min} + 2\int_0^{\infty}DyP(y)H\left(\frac{Ry}{\sqrt{1-R^2}}\right) = \epsilon_{\rm g},$$

$$K_b \equiv \int Dy2\mathcal{P}(y)\int DuH(Y)H(-Y) = Q\int\tilde{D}uH(u)H(-u),$$

K and $f_{\rm RS}$ can be expressed as

$$\begin{aligned} K &= -\beta\left(K_a - \frac{\beta}{2}K_b\right) + \mathcal{O}(\beta^3), \\ -\beta f_{\rm RS} &= -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I - \alpha\beta\left(K_a - \frac{\beta}{2}K_b\right) + \mathcal{O}(\beta^3) \\ &\simeq -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I - \alpha\beta\epsilon_{\rm g} + \frac{\alpha\beta^2}{2}K_b. \end{aligned}$$

The entropy $S_{\rm RS}$ is expressed as

$$S_{\rm RS} = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I + \alpha K - \alpha\beta e^{-\beta}J, \tag{E·2}$$

$$J = \int Dy2\mathcal{P}(y)\int Du\frac{H(Y)-1}{\tilde{H}(Y)}.$$

Then, defining L as $L = K - \beta e^{-\beta}J$, we have

$$L = -\frac{\beta^2}{2}K_b + \mathcal{O}(\beta^3).$$

Thus, we obtain

$$S_{\rm RS} = -\frac{\hat{q}}{2}\Delta q - R\hat{R} + I - \frac{\alpha\beta^2}{2}K_b + \mathcal{O}(\beta^3).$$

For $\Delta q \ll 1$ and $\Delta R \ll 1$, the following relations hold:

$$K_a = \epsilon_{\rm g} \simeq \epsilon_{\min} + \frac{2s}{(1+\delta)\sqrt{2\pi}}(2\Delta R)^{\frac{1+\delta}{2}}, \quad K_b \simeq \frac{\sqrt{\Delta q}}{\pi\sqrt{2}}.$$

Therefore, K, $f_{\rm RS}$ and $S_{\rm RS}$ can be expressed as

$$K = -\beta\left[\epsilon_{\min} + \frac{2s}{(1+\delta)\sqrt{2\pi}}(2\Delta R)^{\frac{1+\delta}{2}} - \frac{\beta\sqrt{\Delta q}}{2\pi\sqrt{2}}\right], \tag{E·3}$$

$$-\beta f_{\rm RS} = -\frac{\hat{q}}{2}(1-q) - R\hat{R} + I - \alpha\beta\left[\epsilon_{\min} + \frac{2s}{(1+\delta)\sqrt{2\pi}}(2\Delta R)^{\frac{1+\delta}{2}} - \frac{\beta\sqrt{\Delta q}}{2\pi\sqrt{2}}\right], \tag{E·4}$$

$$S_{\rm RS} = -\frac{\hat{q}}{2}\Delta q - \hat{R}R + I - \alpha\beta^2\frac{\sqrt{\Delta q}}{2\pi\sqrt{2}}. \tag{E·5}$$

1) See, for example, T. L. H. Watkin, R. Rau and M. Biehl: Rev. Mod. Phys. **65** (1993) 499.
2) H. S. Seung, H. Sompolinsky and N. Tishby: Phys. Rev. A **45** (1992) 6056.
3) G. Györgyi and N. Tishby: *Neural Networks and Spin Glasses*, ed. W. K. Theumann and R. Köberle (World Scientific, Singapore,1990) p. 3.
4) M. Opper and D. Haussler: Phys. Rev. Lett. **66** (1991) 2677.
5) Y. Kabashima and S. Shinomoto: Neural Comput. **4** (1992) 712.
6) J. Kim and D. Pollard: Ann: Stat. **18** (1990) 191.
7) T. Uezu, Y. Kabashima, K. Nokura and N. Nakamura: J. Phys. Soc. Jpn. **65** (1996) 3797.
8) T. Uezu and Y. Kabashima: J. Phys. A **29** (1996) L55.
9) S. Amari, N. Fujita and S. Shinomoto: Neural Comput. **4** (1992) 605.
10) G. Györgyi: Phys. Rev. A **41** (1990) 7097.
11) H. Sompolinsky, N. Tishby and H. S. Seung: Phys. Rev. Lett. **65** (1990) 1683.
12) T. Uezu and Y. Kabashima: J. Phys. A **29** (1996) L439.
13) H. S. Seung: *Proc. CTP-PBSRI Joint Workshop on Theoretical Physics*, ed. J.-H. Oh *et al.* (World Scientific, Singapore, 1995) p. 32.
14) We can obtain the same results for finite χ.
15) G. Parisi: J. Phys. A **13** (1980) 1101; *ibid.* **13** (1980) L115; *ibid.* **13** (1980) 1887.
16) W. Krauth and M. Mézard: J. Phys. (France) **50** (1989) 3057.
17) Here, we have assumed the boundedness of $|P'(y)|$. However, we can obtain the same results without using $P'(y)$.