

## A PARALLEL ALGORITHM FOR BOLTZMANN MACHINES

TAKASHI YOZAWA

(Received September 19, 1990)

### 1. Introduction

Recently, a neural network model called the *Boltzmann machine* is studied by many researchers [cf. 1, 2, 3]. One of its motivations comes from the fact that by this machinery we can approximately solve hard combinatorial optimization problems and that its processing units have the potential ability to do calculations locally and to permit massive parallelism.

The sequential Boltzmann machine calculates the *Gibbs distribution* of the "energy" function over the configuration space [cf. 3], whence we can solve the optimization problem of the energy function by a method so called *simulated annealing*. However this sequential algorithm does not fully exploit the potential parallelism of the Boltzmann machine.

When we apply unlimited and full parallelism in a simple way [cf. 3], the equilibrium distribution is not the Gibbs distribution and we can not use this to solve optimization problems. Although this parallel algorithm can be modified, using limited parallelism [cf. 3], to calculate the Gibbs distribution, the modification heavily depends on the connection pattern of the Boltzmann machine and needs a preprocessing which is difficult itself.

Does there exist a synchronously parallel algorithm for the Boltzmann machine, which calculates the Gibbs distribution and is independent of the connection pattern?

In this paper, we propose new synchronously parallel algorithms for the Boltzmann machines, which are independent of the connection pattern, and show that the equilibrium distributions are the Gibbs distributions, thereby guaranteeing that they solve the optimization problems.

The author wishes to thank Professor T. Tsujishita for valuable suggestions and encouragement.

### 2. Basic Concepts of the Boltzmann Machine

A Boltzmann machine  $\mathcal{B}$  is defined as a triplet  $(\mathcal{U}, \theta, W)$ , where  $\mathcal{U}$  is a finite set,  $\theta$  is a real valued function on  $\mathcal{U}$ , and  $W$  is a symmetric real valued

function on  $\{(u, v) \in \mathcal{U} \times \mathcal{U} \mid u \neq v\}$ . An element of  $\mathcal{U}$  is called a *unit* of  $\mathcal{B}$ , the value  $\theta_u := \theta(u)$  ( $u \in \mathcal{U}$ ) the *threshold* for the unit  $u$ , and  $W_{uv} := W(u, v)$  ( $u, v \in \mathcal{U}$ ) the *weight* between the units  $u$  and  $v$ . Note that  $W_{uv} = W_{vu}$ .

In this paper, the functions  $\theta$  and  $W$  are held constant, namely we do not consider the learning problem of the Boltzmann machines.

Let  $\mathcal{B} = (\mathcal{U}, \theta, W)$  be a Boltzmann machine. A function from  $\mathcal{U}$  to  $\{0, 1\}$  is called a *configuration* of  $\mathcal{B}$ . We denote by  $\mathcal{S}$  the set of all the configurations of  $\mathcal{B}$ , called the *configuration space*, i.e.

$$(1) \quad \mathcal{S} := \{0, 1\}^{\mathcal{U}}.$$

The *energy function*  $E: \mathcal{S} \rightarrow \mathbf{R}$  is defined by

$$(2) \quad E(k) := \frac{1}{2} \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{U} \setminus \{u\}} W_{uv} k(u) k(v) + \sum_{n \in \mathcal{U}} \theta_n k(n),$$

where  $k(u)$  is called the *state* of the unit  $u$  for the configuration  $k$ .

We can use the Boltzmann machine  $\mathcal{B}$  to solve the optimization problem of the energy  $E$ . In the next section, we recall one of such solutions, namely the sequential algorithm for the Boltzmann machine.

### 3. Sequential Boltzmann Machine

Let  $\mathcal{B} = (\mathcal{U}, \theta, W)$  be a Boltzmann machine. Fix a positive real  $c$ , called the *control parameter*.

The sequential algorithm for  $\mathcal{B}$  repeats the following two steps until certain criteria are met:

1. Select randomly a unit  $\alpha \in \mathcal{U}$ .
2. Change the current configuration  $k \in \mathcal{S}$  to  $k_\alpha$  with probability  $A_k(\alpha, c)$ .

Here  $k_\alpha \in \mathcal{S}$  is obtained from  $k$  by changing the state of the unit  $\alpha$ , i.e.

$$(3) \quad k_\alpha(u) = \begin{cases} k(u) & \text{if } u \neq \alpha \\ 1 - k(u) & \text{if } u = \alpha, \end{cases}$$

and  $A_k(\alpha, c)$ , called the *acceptance probability*, is given by

$$(4) \quad A_k(\alpha, c) = \frac{1}{1 + \exp\left(\frac{-\Delta_\alpha E(k)}{c}\right)},$$

where

$$(5) \quad \Delta_\alpha E(k) := E(k_\alpha) - E(k).$$

The Markov chain  $\{X(t) \in \mathcal{S} \mid t = 1, 2, \dots\}$  induced by this algorithm has the Gibbs distribution as the equilibrium distribution, i.e.

$$(6) \quad \lim_{t \rightarrow \infty} \Pr^{(c)}(X(t) = k) = \frac{1}{N_0(c)} \exp\left(\frac{E(k)}{c}\right), \quad \text{for } k \in \mathcal{S},$$

where  $\Pr^{(c)}(X(t) = k)$  is the probability of entering  $k$  at the  $t$ -th trial, and

$$(7) \quad N_0(c) = \sum_{l \in \mathcal{S}} \exp\left(\frac{E(l)}{c}\right).$$

Thus, by the annealing process, we obtain the uniform distribution concentrated on

$$(8) \quad \mathcal{S}_{opt} := \{k \in \mathcal{S} \mid E(k) = \max_{l \in \mathcal{S}} E(l)\}.$$

From the point of view of the efficiency this algorithm is unsatisfactory, since it does not take advantage of the structure of the Boltzmann machine, which admits parallelism in updating the configurations. In the next section, we shall recall two parallel algorithms: the one is simple but does not solve the optimization problem, the other solves it but heavily depends upon the connection pattern.

#### 4. Parallel Algorithms for the Boltzmann Machine

In a simple algorithm which synchronously updates configurations in completely parallel way, each unit  $u \in \mathcal{U}$  changes the state with the probability  $A_k(u, c)$  for each trial. The equilibrium distribution, calculated in [3], is given by

$$(9) \quad \lim_{t \rightarrow \infty} \Pr^{(c)}(X(t) = k) = \frac{1}{K_0(c)} \prod_{u \in \mathcal{U}} 2 \cosh\left(\frac{h_k(u)}{2c}\right) \exp\left(\frac{h_k(u) + 2\theta_u k(u)}{2c}\right),$$

where

$$(10) \quad K_0(c) = \sum_{l \in \mathcal{S}} \prod_{u \in \mathcal{U}} 2 \cosh\left(\frac{h_l(u)}{2c}\right) \exp\left(\frac{h_l(u) + 2\theta_u l(u)}{2c}\right),$$

and

$$(11) \quad h_k(u) = \sum_{v \in \mathcal{U} \setminus \{u\}} W_{uv} k(v) + \theta_u.$$

This distribution does not converge, as the control parameter  $c$  tends to zero, to the distribution concentrated and uniform over the set of optimal configurations  $\mathcal{S}_{opt}$ . From the point of view of solving the optimization problem, the above parallel algorithm is ineffective.

There exists a parallel algorithm, called limited parallelism, whose equilibrium distribution is the Gibbs distribution [cf. 3]. However, the algorithm applying the limited parallelism for the Boltzmann machine has the drawback as is seen from the following: In this algorithm we first do a preprocessing related

to the connection pattern, namely we partition the set of units  $\mathcal{U}$  into disjoint subsets so that the weights between the units belonging to the same subset equal zero. Then the limited parallelism is realized by selecting one of the subsets and then by allowing its units to change their states simultaneously for each trial. If the set of units  $\mathcal{U}$  is partitioned into a minimal number of disjoint subsets, this algorithm is most effective, namely it is the maximal speed-up procedure. But the problem of partitioning  $\mathcal{U}$  into a minimal number of disjoint subsets is equivalent to the graph coloring problem, which is a hard combinatorial optimization problem.

In the rest of this paper, we explain a new parallel algorithm IPBM—an abbreviation of Improved Parallel Boltzmann Machine—which is easily implemented and solves the optimization problem.

### 5. Algorithm IPBM

In this section we present a new updating procedure. We fix a positive real constant  $c$  as the control parameter. Let  $G(I)$  ( $I \in \mathcal{P}(\mathcal{U})$ ) be a probability distribution over the power set of  $\mathcal{U}$ , called the *generation probability distribution*. The algorithm repeats the following two steps until certain criteria are met:

**Algorithm IPBM.** Let  $k \in \mathcal{S}$  be the current configuration.

1. Select a subset  $I$  of  $\mathcal{U}$  according to the probability distribution  $G$ .
2. Change the configuration  $k$  to the configuration  $k_I \in \mathcal{S}$  with probability  $A_k(I, c)$ , where  $k_I$  is obtained from  $k$  by changing the states of all units of  $I$ , i.e.

$$(12) \quad k_I(u) = \begin{cases} k(u) & \text{if } u \notin I \\ 1 - k(u) & \text{if } u \in I. \end{cases}$$

The probability  $A_k(I, c)$ , called the *acceptance probability*, is defined by

$$(13) \quad A_k(I, c) = \prod_{i \in I} \bar{A}_k(I, i, c),$$

where

$$(14) \quad \bar{A}_k(I, i, c) = \left\{ 1 + \exp \left( \frac{-\Delta_i E(k) - S_k(I, i)}{c} \right) \right\}^{-1},$$

and

$$(15) \quad S_k(I, i) = \frac{1}{2} (1 - 2k(i)) \sum_{j \in I \setminus \{i\}} (1 - 2k(j)) W_{ij}.$$

Note that if  $I = \{i\}$ , then

$$S_k(I, i) = 0$$

and

$$A_k(I, c) = \left\{ 1 + \exp \left( \frac{-\Delta_i E(k)}{c} \right) \right\}^{-1}.$$

**6. The Markov Chain generated by Algorithm IPBM**

Algorithm IPBM in the previous section can be formulated as a time homogeneous Markov chain on the configuration space  $\mathcal{S}$  for the fixed control parameter  $c$ :

Let  $I_{kl}$  be the set of the units whose states for the configuration  $k$  differ from those of  $l$ , i.e.

$$(16) \quad I_{kl} = \{u \in \mathcal{U} \mid k(u) \neq l(u)\} .$$

Note that  $I_{kl} = I_{lk}$ . Then it is obvious that Algorithm IPBM generates a Markov chain on  $\mathcal{S}$  whose transition probability  $P_{kl}(c)$  from  $k \in \mathcal{S}$  to  $l \in \mathcal{S}$  is given by

$$(17) \quad P_{kl}(c) = \begin{cases} G(I_{kl}) \cdot A_k(I_{kl}, c) & \text{if } l \neq k \\ 1 - \sum_{h \in \mathcal{S} \setminus \{k\}} P_{kh}(c) & \text{if } l = k . \end{cases}$$

We assume the following condition on the generation probability distribution  $G$ :

ASSUMPTION. For any  $k, l \in \mathcal{S}$  there exists a number  $p \geq 1$  and  $k_0, k_1, \dots, k_p \in \mathcal{S}$  with  $k_0 = k, k_p = l$  such that

$$G(I_{k_i k_{i+1}}) > 0 \quad (i = 0, 1, \dots, p-1) .$$

It is easy to check that the following examples of the generation probability distributions over the power set of  $\mathcal{U}$  satisfy the above Assumption:

1. The number  $m$  of the selected set  $I$  is randomly chosen and then the  $m$  units belonging to  $I$  are randomly chosen from  $\mathcal{U}$ , i.e.

$$G(I) = \begin{cases} \frac{1}{|\mathcal{U}|} \cdot \binom{|\mathcal{U}|}{|I|}^{-1} & \text{if } I \neq \emptyset \\ 0 & \text{if } I = \emptyset . \end{cases}$$

2. Probability distributions concentrating on the set of the units of size  $\lceil |\mathcal{U}|/2 \rceil$ .
3. Each unit belongs to the selected units  $I$  with the probability  $p$ , i.e.

$$G(I) = p^{|I|} (1-p)^{|\mathcal{U}| - |I|} .$$

Note that if  $p=1/2$ , then  $G$  is the uniform distribution over the power set of  $\mathcal{U}$ .

**7. Equilibrium Distribution calculated by Algorithm IPBM**

In this section we determine the equilibrium distribution of the Markov

chain constructed in the previous section.

**Theorem 1.** *Let  $\{X(t) \in \mathcal{S} \mid t=1, 2, \dots\}$  be the Markov chain with the transition probability (17). Then its equilibrium distribution is given by the Gibbs distribution, i.e.*

$$\lim_{t \rightarrow \infty} \Pr^{(c)}(X(t) = k) = \frac{1}{N_0(c)} \exp\left(\frac{E(k)}{c}\right), \quad \text{for any } k \in \mathcal{S},$$

where

$$N_0(c) = \sum_{l \in \mathcal{S}} \exp\left(\frac{E(l)}{c}\right).$$

First we shall recall a theorem on the stationary distribution of a Markov chain.

**Theorem 2.** (cf. [4]). *If a finite Markov chain  $\{X(n) \in \Omega \mid n=1, 2, \dots\}$  with the transition matrix  $P = (P_{kl})_{k, l \in \Omega}$  is irreducible and aperiodic, then*

(a) *there exists a unique stationary distribution  $\mathbf{q} = (q_k)_{k \in \Omega}$ , and*

(b)

$$(18) \quad \lim_{n \rightarrow \infty} P_{lk}^{(n)} = q_k, \quad \text{for any } k, l \in \Omega,$$

where  $P_{lk}^{(n)}$  is the  $(l, k)$  component of the matrix  $P^n$ .

Recall that a probability distribution  $\mathbf{q}$  is called *stationary* if

$$(19) \quad q_k = \sum_{l \in \Omega} q_l P_{lk} \quad \text{for any } k \in \Omega.$$

**Corollary of Theorem 2.** *Under the same conditions as in Theorem 2, a distribution which is stationary is the equilibrium distribution.*

**Proof.** By (b) of Theorem 2 we have for any  $k \in \Omega$

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr(X(n) = k) &= \lim_{n \rightarrow \infty} \sum_l \Pr(X(n) = k \mid X(0) = l) \Pr(X(0) = l) \\ &= \lim_{n \rightarrow \infty} \sum_l P_{lk}^{(n)} \Pr(X(0) = l) \\ &= \sum_l q_l \Pr(X(0) = l) \\ &= q_k. \end{aligned}$$

To prove Theorem 1, we need the following two lemmas. The proofs will be given in Appendix.

**Lemma 1.** *For any  $k, l \in \mathcal{S}$*

$$(20) \quad E(l) - E(k) = \sum_{i \in I_{kl}} \{\Delta_i E(k) + S_k(I_{kl}, i)\}.$$

**Lemma 2.** *For any  $k, l \in \mathcal{S}$*

$$(21) \quad \Delta_i E(k) + S_k(I_{kl}, i) = -\Delta_i E(l) - S_l(I_{lk}, i), \quad \text{for any } i \in I_{kl}.$$

Proof of Theorem 1. The Markov chain with the transition matrix  $(P_{kl}(c))_{k,l \in S}$  given by (17) is irreducible by Assumption and aperiodic by the irreducibility and the following inequality:

$$(22) \quad A_k(I_{kl}, c) < 1, \quad \text{for any } k, l \in S$$

(cf. [3]). Thus, the conditions of Theorem 2 are satisfied and there exists a unique stationary distribution. Hence, by Corollary of Theorem 2, it remains only to show that the vector  $q(c) = (q_k(c))_{k \in S}$  defined by

$$q_k(c) = \frac{1}{N_0(c)} \exp\left(\frac{E(k)}{c}\right)$$

is the stationary distribution.

It is clear that  $q(c)$  is stochastic (i.e.  $\sum_{k \in S} q_k(c) = 1$ ).

Recall that if  $q(c)$  satisfies the detailed balance equation:

$$(23) \quad q_k(c)P_{kl}(c) = q_l(c)P_{lk}(c), \quad \text{for any } k, l \in S,$$

then the vector  $q(c)$  is stationary. Thus it remains to show that  $q(c)$  satisfies the detailed balance equation. For  $k \neq l \in S$ , using Lemmas 1 and 2, we obtain

$$\begin{aligned} q_k(c)P_{kl}(c) &= \frac{1}{N_0(c)} \exp\left(\frac{E(k)}{c}\right) G(I_{kl}) A_k(I_{kl}, c) \\ &= \frac{1}{N_0(c)} \exp\left(\frac{E(l)}{c}\right) \exp\left(\frac{-(E(l) - E(k))}{c}\right) G(I_{kl}) A_k(I_{kl}, c) \\ &= q_l(c) G(I_{lk}) \exp\left(\frac{-\sum_{i \in I_{kl}} (\Delta_i E(k) + S_k(I_{kl}, i))}{c}\right) \prod_{i \in I_{kl}} A_k(I_{kl}, i, c) \\ &= q_l(c) G(I_{lk}) \prod_{i \in I_{kl}} \exp\left(\frac{-(\Delta_i E(k) + S_k(I_{kl}, i))}{c}\right) A_k(I_{kl}, i, c) \\ &= q_l(c) G(I_{lk}) \prod_{i \in I_{kl}} \left\{ 1 + \exp\left(\frac{\Delta_i E(k) + S_k(I_{kl}, i)}{c}\right) \right\}^{-1} \\ &= q_l(c) G(I_{lk}) \prod_{i \in I_{lk}} \left\{ 1 + \exp\left(\frac{-\Delta_i E(l) - S_l(I_{lk}, i)}{c}\right) \right\}^{-1} \\ &= q_l(c) G(I_{lk}) \prod_{i \in I_{lk}} A_l(I_{lk}, i, c) \\ &= q_l(c) G(I_{lk}) A_l(I_{lk}, c) \\ &= q_l(c) P_{lk}(c). \end{aligned}$$

This completes the proof of Theorem 1.

## 8. Discussions and Conclusions

1. As we stated at the end of section 5, our algorithm IPBM is a generalization of the sequential algorithm for the Boltzmann machine which select only one unit to be updated for each trial. In fact if the probability distribution  $G$  satisfies the following condition:

$$G(I) = \begin{cases} 1/|\mathcal{U}| & \text{if } |I| = 1 \\ 0 & \text{otherwise,} \end{cases}$$

then IPBM is the same as the sequential Boltzmann machine.

2. We may schedule the number of selected units for each trial: For example, the number of selected units is large for early trials. As trials go on, it decreases, and finally it is fixed to one. Then the generation probability distribution  $G$  depends on the time, namely  $G$  is different for each trial. Let  $G := G_t (t = 1, 2, \dots)$  denote the generation probability distribution over the power set of  $\mathcal{U}$  for each trial  $t$ . The similar algorithm corresponding to this  $G$  can be formulated as a time inhomogeneous Markov chain whose transition probability depends on the time. The equilibrium distribution is again the Gibbs distribution if  $G_t$  satisfies the following condition:

*For any number  $m \geq 1$  and any  $k, l \in S$  there exists a number  $p \geq 1$  and  $k_0, \dots, k_p \in S$  with  $k_0 = k, k_p = l$  such that*

$$G_{m+i}(I_{k_i, k_{i+1}}) > 0 \quad (i = 0, 1, \dots, p-1).$$

The above statement can be proved by using the results in the theory of time inhomogeneous Markov chains (cf. [5,6]).

3. In regard to the acceptance probability  $A_k(I, c)$ , if the number of the selected units  $I$  is large,  $A_k(I, c)$  is small since  $\bar{A}_k(I, i, c) < 1$  for any  $i \in I$ , whence it is rare for the states of the units  $I$  to be changed. When we replace  $\bar{A}_k(I, i, c)$  with

$$(24) \quad \bar{A}_k(I, i, c) = \min \left\{ 1, \exp \left( \frac{\Delta_i E(k) + S_k(I, i)}{c} \right) \right\},$$

$A_k(I, c)$  is larger than the previous one. And by the computation like the proof of Theorem 2 we can prove that the Gibbs distribution is again the stationary distribution of the Markov chain which is formulated by this new  $A_k(I, c)$ .

4. Remark on the implementation.  $\bar{A}_k(I, i, c)$  is locally computed by the states of the units with which the unit  $i$  is connected and by the weights among their units, and global informations are necessary only in the computation of the product  $A_k(I, c) = \prod_{i \in I} \bar{A}_k(I, i, c)$ .

5. In conclusion I guess our algorithm IPBM is useful to solve the problem whose units and local maximums are many, since the set of the configurations, to which the current configuration may be changed, is large for each trial, whence



as compared with the sequential algorithm, the possibility of trapping to the local maximum seems small.

We are currently testing the effectiveness of our new algorithm IPBM by computer simulation, which will be reported in near future.

**Appendix**

Proof of Lemma 1. By direct computation we obtain

$$(25) \quad \Delta_\alpha E(k) = (1-2k(\alpha)) \left\{ \sum_{j \in \mathcal{U} \setminus \{\alpha\}} W_{\alpha j} k(j) + \theta_\alpha \right\}.$$

Using this, we have

$$\begin{aligned} E(l) - E(k) &= \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} W_{ij} \{(1-k(i))(1-k(j)) - k(i)k(j)\} \\ &\quad + \frac{1}{2} \sum_{i \in I_M} \sum_{j \notin I_M} (W_{ij} + W_{ji}) \{(1-k(i))k(j) - k(i)k(j)\} \\ &\quad + \sum_{i \in I_M} \theta_i \{(1-k(i)) - k(i)\} \\ &= \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} W_{ij} (1-k(i) - k(j)) \\ &\quad + \sum_{i \in I_M} (1-2k(i)) \sum_{j \notin I_M} W_{ij} k(j) + \sum_{i \in I_M} \theta_i (1-2k(i)) \\ &= \sum_{i \in I_M} (1-2k(i)) \left\{ \sum_{j \in \mathcal{U} \setminus \{i\}} W_{ij} k(j) + \theta_i \right\} \\ &\quad - \sum_{i \in I_M} (1-2k(i)) \sum_{j \in I_M \setminus \{i\}} W_{ij} k(j) \\ &\quad + \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} W_{ij} (1-k(i) - k(j)) \\ &= \sum_{i \in I_M} \Delta_i E(k) \\ &\quad - \frac{1}{2} \left\{ \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} (1-2k(i)) k(j) W_{ij} \right. \\ &\quad \left. + \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} (1-2k(i)) k(j) W_{ji} \right\} \\ &\quad + \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} W_{ij} (1-k(i) - k(j)) \\ &= \sum_{i \in I_M} \Delta_i E(k) \\ &\quad + \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} W_{ij} \{(2k(i) - 1)k(j) \\ &\quad + (2k(j) - 1)k(i) + (1-k(i) - k(j))\} \\ &= \sum_{i \in I_M} \Delta_i E(k) + \frac{1}{2} \sum_{i \in I_M} \sum_{j \in I_M \setminus \{i\}} (1-2k(i))(1-2k(j)) W_{ij} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i \in I_k} \left\{ \Delta_i E(k) + \frac{1}{2} (1-2k(i)) \sum_{j \in I_k \setminus \{i\}} (1-2k(j)) W_{ij} \right\} \\
&= \sum_{i \in I_k} (\Delta_i E(k) + S_k(I_k, i)).
\end{aligned}$$

Proof of Lemma 2. Using (15) and (25), we obtain for  $i \in I_k$

$$\begin{aligned}
&\Delta_i E(k) + S_k(I_k, i) \\
&= (1-2k(i)) \left\{ \sum_{j \in \mathcal{U} \setminus \{i\}} W_{ij} k(j) + \theta_i + \frac{1}{2} \sum_{j \in I_k \setminus \{i\}} (1-2k(j)) W_{ij} \right\} \\
&= (1-2k(i)) \left\{ \sum_{j \in \mathcal{U} \setminus I_k} W_{ij} k(j) + \theta_i + \frac{1}{2} \sum_{j \in I_k \setminus \{i\}} W_{ij} \right\} \\
&= -(1-2l(i)) \left\{ \sum_{j \in \mathcal{U} \setminus I_k} W_{ij} l(j) + \theta_i + \frac{1}{2} \sum_{j \in I_k \setminus \{i\}} W_{ij} \right\} \\
&= -(1-2l(i)) \left\{ \sum_{j \in \mathcal{U} \setminus \{i\}} W_{ij} l(j) + \theta_i + \frac{1}{2} \sum_{j \in I_k \setminus \{i\}} (1-2l(j)) W_{ij} \right\} \\
&= -\Delta_i E(l) - S_l(I_k, i).
\end{aligned}$$

#### References

- [1] D.H. Ackley, G.E. Hinton and T.J. Sejnowski: *A learning algorithm for Boltzmann machines*, Cognitive Science 9 (1985), 147-169.
- [2] D.E. Rumelhart, J.L. McClelland and the PDP Research group: *Parallel Distributed Processing Volume 1: Foundations*, The MIT Press, Cambridge(MA), 1986, 282-317.
- [3] E. Aarts and J. Korst: *Simulated Annealing and Boltzmann Machines*, Wiley, Chichester, 1989.
- [4] W. Feller: *An Introduction to Probability Theory and Its Applications 1*, Wiley, New York, 1950.
- [5] D. Isaacson and R. Madsen: *Markov Chains*, Wiley, New York, 1976.
- [6] E. Seneta: *Non-negative Matrices and Markov Chains*, 2nd ed., Springer-Verlag, New York, 1981.

Department of Mathematics  
Faculty of Science  
Osaka University  
Toyonaka, Osaka 560  
Japan

and

Matsushita Electric Industrial  
Co., Ltd.  
Kadoma, Osaka 571  
Japan