TR - H - 193

Doubly constrained network for combinatorial optimization

Shin ISHII *Masa-aki SATO*

1996. 5. 21

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

Doubly constrained network for combinatorial optimization

Shin Ishii **Masa-aki Sato**

ATR Human Information Processing Research Laboratories

2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

(TEL) +81-774-95-1069 (FAX) +81-774-95-1008

(E-mail) ishii@hip.atr.co.jp

Abstract

In this paper, we propose a new approach for solving combinatorial optimization problems having two competing sets of constraints. In our approach, those constraints are treated as "hard," namely, they hold automatically. Experimental results for traveling salesman problems show that our approach can obtain better solutions even for large-scale problems than binary or Potts spin approaches.

1 Introduction

As parallel implementation of the continuation methods [21], analog neural approaches to combinatorial optimization problems have been widely studied.

The analog Hopfield network, which is equivalent to mean field theory (MFT) approximation [11, 12, 3] of the Boltzmann machine, globally converges to a local minimum of its Lyapunov function. When the slope of the sigmoidal output function becomes very large, the Lyapunov function is nearly equal to the quadratic energy function. Then, the Hopfield network can be used for solving combinatorial optimization problems, e.g., traveling salesman problems (TSPs), by designing a quadratic energy function whose local minima correspond to solutions of the problem [4]. The Hopfield network can be combined with a gradual slope increasing procedure of the sigmoidal output function [4], which corresponds to a deterministic (MFT) annealing procedure [18, 1] in MFT.

In the conventional neural approaches, each neural representation often involves redundancy. For example, when Ising (binary) spin approaches, i.e., the Hopfield network and the binary MFT, are applied to an N-city TSP, the number of possible spin configurations for neural representation is 2^{N^2}, while the number of possible TSP tours is $N!/2$. Thus, the domain space is likely to become much larger than that of the original combinatorial problem. In order to suppress the redundancy, in the Ising spin approaches, domain restrictions are designed as constraints, and the energy function includes penalty terms for constraint violations; these constraints are "soft" constraints. However, this soft constraint scheme produces a lot of infeasible solutions in general. This is one of the reasons why the Hopfield network is not a good algorithm when the problem scale becomes large [22, 7].

Some of the combinatorial optimization problems, such as TSPs, N-Queen problems, and so on, have two sets of constraints that restrict the redundant domain space. In TSPs, one set is for "a salesman visits one and only one city on one leg of a tour," and the other set is for "a city is visited once and only once in a complete tour." In such a problem, a Potts spin system [23] can be used, where one set of the constraints can be treated as "hard," namely, they hold automatically. A Potts spin is a generalization of an Ising spin so as to take more-than-two states. A combined algorithm of Potts spin MFT with a deterministic annealing procedure is Potts MFT annealing [13, 20], which is almost equivalent [17] to the elastic net [2]. By employing a Potts spin system, the domain space becomes smaller than that of the Ising spin approaches, and the obtained solutions by the Potts MFT annealing are consequently much improved so as to be comparable [14] to simulated annealing [8].

The above-mentioned two sets of constraints are consistent but competing with each other, and this makes the problem difficult to be solved by neural approaches. In the Potts spin approach, one set of the constraints is treated as "hard," while the other set maintains "soft" constraints. In this paper, we propose a new approach, where both of the two competing sets are treated as "hard." Therefore, the system searches for a local minimum of the energy function that does not include the constraint violation terms. This algorithm, which is called the doubly constrained network (DCN), makes the domain space smaller than the Potts spin approach. This paper also shows experimental results when our DCN is applied to relatively large-scale TSPs. Our approach is shown to be capable of obtaining better solutions than the Ising and Potts spin approaches.

This paper is organized as follows. In Section 2, the DCN equations are derived based

on the Lagrange method, and a discrete-time dynamical system that can obtain the solution is proposed. In Section 3, we experimentally compare our approach with the Ising and Potts spin MFT annealing for relatively large-scale TSPs. As a consequent, our DCN can obtain better solutions than conventional algorithms. In Section 4, we study the convergence property of the proposing algorithm. In Section 5, we study the bifurcation property in DCN annealing, which is important in analyzing the quality of the solution. Section 6 sums up the paper.

2 DCN algorithm

Let us consider the following combinatorial optimization problem that is defined by $N \times M$ binary variables (spins) $S_{a,n} \in \{0,1\}$.

$$\text{minimize} \quad E(\boldsymbol{S}) = \frac{1}{2} \sum_{a,n,b,m} W_{a,n;b,m} S_{a,n} S_{b,m} + \sum_{a,n} J_{a,n} S_{a,n}, \tag{1}$$

$$\text{subject to} \quad \sum_{a=1}^{N} S_{a,n} = r_n \ \ (n = 1, ..., M) \tag{2}$$

$$\sum_{n=1}^{M} S_{a,n} = s_a \ \ (a = 1, ..., N), \tag{3}$$

where parameters $\boldsymbol{W}$ and $\boldsymbol{J}$ are determined for each problem. r_n and s_a are constant integers such that $0 < r_n \leq N$, $0 < s_a \leq M$. We assume that these constraints are consistent with each other, namely, there is a solution for $\boldsymbol{S}$ under the constraints. In order to deal with this discrete problem, we introduce analog variables and the corresponding barrier function preventing the state from going beyond the domain [10]. The new continuous problem is

$$\text{minimize} \quad F(\boldsymbol{V}), \tag{4}$$

$$\text{subject to} \quad \sum_{a=1}^{N} V_{a,n} = r_n \ \ (n = 1, ..., M) \tag{5}$$

$$\sum_{n=1}^{M} V_{a,n} = s_a \ \ (a = 1, ..., N), \tag{6}$$

where $F(\boldsymbol{V})$ is the free energy function [11] that can be given by

$$\begin{aligned} F(\boldsymbol{V}) &= E(\boldsymbol{V}) - TH(\boldsymbol{V}) \\ &= \frac{1}{2} \sum_{a,b,n,m} W_{a,n;b,m} V_{a,n} V_{b,m} + \sum_{a,n} J_{a,n} V_{a,n} + T \sum_{a,n} V_{a,n} \log V_{a,n}. \end{aligned} \tag{7}$$

The barrier function $H(\boldsymbol{V})$ is also called entropy function [11], and T is called temperature. The new problem (4), (5), (6) corresponds to the mean field theory (MFT) approximation [11] of the Gibbs distribution defined for the energy function (1) under the spin configuration constraints (2), (3) at the temperature T.

A local minimum solution of the problem (4), (5), (6) is a stationary point of the Lagrange function [10]

$$L(\boldsymbol{V}) = F(\boldsymbol{V}) + \sum_n P_n(\sum_a V_{a,n} - r_n) + \sum_a Q_a(\sum_n V_{a,n} - s_a), \tag{8}$$

where P_n $(n = 1, ..., N)$ and Q_a $(a = 1, ..., M)$ are Lagrange multipliers. A stationary condition of the Lagrange function (8) is given by

$$\frac{\partial L}{\partial V_{a,n}} = \sum_{b,m} W_{a,n;b,m} V_{b,m} + J_{a,n} + T(\log V_{a,n} + 1) + P_n + Q_a = 0 \tag{9}$$

$$\frac{\partial L}{\partial P_n} = \sum_a V_{a,n} - r_n = 0 \tag{10}$$

$$\frac{\partial L}{\partial Q_a} = \sum_n V_{a,n} - s_a = 0 \tag{11}$$

By defining auxiliary variables

$$U_{a,n} = -\frac{1}{T}\left(\sum_{b,m} W_{a,n;b,m} V_{b,m} + J_{a,n}\right), \tag{12}$$

(9) is rewritten as

$$V_{a,n} = \frac{\exp(U_{a,n})}{\mu_a \lambda_n}, \tag{13}$$

where $\mu_a = \exp(Q_a/T + 1)$ and $\lambda_n = \exp(P_n/T)$. From (11) and (13), we get

$$\mu_a = \frac{1}{s_a} \sum_n \frac{\exp(U_{a,n})}{\lambda_n}. \tag{14}$$

From (13) and (14), we get

$$V_{a,n} = \frac{s_a \exp(U_{a,n})/\lambda_n}{\sum_m (\exp(U_{a,m})/\lambda_m)}. \tag{15}$$

On the other hand, from (10) and (15), each λ_n must satisfy the following equation

$$\lambda_n = \frac{1}{r_n} \sum_a \frac{s_a \exp(U_{a,n})}{\sum_m (\exp(U_{a,m})/\lambda_m)}. \tag{16}$$

From (15) and (16), it can be shown that the constraints (5) and (6) are automatically satisfied. If λ_n takes an identical value for every n, (12) and (15) become a variant of Potts MFT equations [13]. In our derivation, however, λ_n varies with (16). It should be noted that (15) and (16) have scale invariance, namely, a scaling transformation of $\boldsymbol{\lambda} : \lambda_n \to \beta\lambda_n$ does not affect the solution $\boldsymbol{V}$. We call the nonlinear equations (12), (15), and (16), Doubly Constrained Network (DCN) equations [6].

Let us propose an algorithm to obtain a solution of the DCN equations. This algorithm is a synchronous discrete-time dynamical system, where t denotes the discrete-time.

[BASIC ALGORITHM]

1. For all a, n, calculate

$$U_{a,n}(t) = -\frac{1}{T}\left(\sum_{b,m} W_{a,n;b,m} V_{b,m}(t-1) + J_{a,n}\right). \tag{17}$$

2. Set $\boldsymbol{\lambda}^{old}$ to be $\boldsymbol{\lambda}(t-1)$.
Then, calculate

$$\lambda_n^{new} = \frac{1}{r_n}\sum_a \frac{s_a \exp(U_{a,n}(t))}{\sum_m (\exp(U_{a,m}(t))/\lambda_m^{old})}, \tag{18}$$

for all n, iteratively. If $\boldsymbol{\lambda}$ converges, set $\boldsymbol{\lambda}(t)$ so as to be the converged value, and rescale $\boldsymbol{\lambda}(t)$ to be $\sum_n \lambda_n(t) = 1$.

3. For all a, n, calculate

$$V_{a,n}(t) = \frac{s_a \exp(U_{a,n}(t))/\lambda_n(t)}{\sum_m (\exp(U_{a,m}(t))/\lambda_m(t))}. \tag{19}$$

4. Add 1 to t and go to Step 1 until $\boldsymbol{V}$ converges.

In Step 2, $\boldsymbol{\lambda}$ is rescaled in order to suppress any underflow of $\boldsymbol{\lambda}$. Because of the scale invariance of (18) and (19), the rescaling does not affect the solution of the algorithm.

The convergence of the basic algorithm will be discussed in Section 4. If it converges, the obtained solution satisfies the DCN equations, i.e., (12), (15), and (16), and it is a local minimum of the free energy function (7) subject to the constraints, (5) and (6).

We can combine a deterministic annealing procedure, i.e., a gradual lowering of the temperature T, with the basic algorithm. First, the DCN equations are solved at high temperature and a solution is obtained. Then after slightly lowering the temperature, the DCN equations are solved again starting from the higher temperature solution. By continuing this process, one can get a low temperature solution of the DCN equations.

3 DCN for TSP

Let us consider an application to an N-city TSP, where $r_n = s_a = 1$ $(n, a = 1, ..., N)$. With $(N \times N)$-dimensional variable $V_{a,n} \in [0, 1]$, which represents the probability that the salesman visits city a at the n-th visit, a DCN energy function for a TSP is given by

$$E(\boldsymbol{V}) = \frac{1}{2}\sum_{a,b=1}^{N} D_{a,b} \sum_{n=1}^{N} V_{a,n}(V_{b,(n-1)} + V_{b,(n+1)}) + \frac{A}{2}\sum_{a,n=1}^{N} V_{a,n}(1 - V_{a,n}), \tag{20}$$

where $D_{a,b}$ is the distance between city a and city b, and A is a positive constant. Each suffix is modulo N. The first term on the right hand side of (20) corresponds to the total tour length to be minimized. The second term vanishes when the state variable $\boldsymbol{V}$ is close to a hypercube vertex at low temperature. However, this second term helps the basic algorithm converge, as will be seen in Section 4. In the experiments below, $A = 0.6$ is used.

Here, let us compare the DCN energy function (20) with the Ising spin TSP energy function [4]

$$E_I(\boldsymbol{V}) = \frac{1}{2}\sum_{a,b} D_{a,b} \sum_n V_{a,n}(V_{b,(n-1)} + V_{b,(n+1)}) + \frac{A}{2}\sum_{a,n} V_{a,n}(1 - V_{a,n}) + \frac{B}{2}\sum_a \left(\sum_n V_{a,n} - 1\right)^2 + \frac{C}{2}\sum_n \left(\sum_a V_{a,n} - 1\right)^2, \quad (21)$$

and the Potts spin TSP energy function [13]

$$E_P(\boldsymbol{V}) = \frac{1}{2}\sum_{a,b} D_{a,b} \sum_n V_{a,n}(V_{b,(n-1)} + V_{b,(n+1)}) + \frac{A}{2}\sum_{a,n} V_{a,n}(1 - V_{a,n}) + \frac{B}{2}\sum_a \left(\sum_n V_{a,n} - 1\right)^2. \quad (22)$$

In order to make the difference prominent, the energy functions, (21) and (22), are slightly different from those in the original papers [4, 13]. In the Ising spin energy function (21), constraints (5) and (6) are included as penalty terms. In the Potts spin energy function (22), a penalty term corresponding to the constraint (5) is removed since the constraint is designed to hold automatically by using a "soft-max" function (see (A4) in Appendix A) instead of the sigmoidal function used in the Ising spin approaches. However, in the DCN energy function (20), all of the soft constraint terms are removed.

With the energy function (20), the free energy function (7) is given by

$$F = \frac{1}{2}\sum_{a,b=1}^{N} D_{a,b} \sum_{n=1}^{N} V_{a,n}(V_{b,(n-1)} + V_{b,(n+1)}) + \frac{A}{2}\sum_{a,n=1}^{N} V_{a,n}(1 - V_{a,n}) + T\sum_{a,n} V_{a,n} \log V_{a,n}. \quad (23)$$

And the network parameters in (17) are given by

$$W_{a,n;b,m} = D_{a,b}(\delta_{n,m-1} + \delta_{n,m+1}) - A\delta_{a,b}\delta_{n,m} \quad (24)$$

$$J_{a,n} = A/2, \quad (25)$$

where $\delta_{i,j}$ is the Kronecker's delta.

In Table I, we compare our DCN annealing (DCA) with Ising spin MFT annealing (MFA) [1] and Potts spin MFT annealing (PMA) [13]. We prepared four testbeds for evaluation; they were 100 sets of 30-city problems, 100 sets of 50-city problems, 50 sets of 100-city problems, and 10 sets of 200-city problems. Each city allocation was randomly generated in a unit square. Brief algorithm descriptions and parameters for MFA and PMA are shown in Appendix A. In DCA, the annealing procedure was scheduled so that the temperature is lowered by $\delta T = 0.005$ and the convergence at each temperature was determined by $\delta\boldsymbol{\lambda} = 10^{-5}$ and $\delta\boldsymbol{V} = 10^{-5}$. In order to make the comparison fair, no additional heuristics, e.g., greedy heuristics, were employed in PMA, although they were used in the experiments done by Peterson [13, 14]. We can see that DCA was able to obtain a solution for every problem, while PMA and MFA sometimes failed in obtaining any solutions. For every testbed, DCA was able to obtain better result than the other two algorithms. The optimal tour length for an N-city problem can asymptotically be estimated as $0.765\sqrt{N}$ [15], and the results of DCN are about 11% larger than the estimation. Furthermore, the DCA results in

Table I are better than those [14] obtained by the PMA modified by the greedy heuristics and the simulated annealing.

Table II shows the results obtained by the application of local heuristics, i.e., 2-opt heuristics [9], to the solutions obtained in the previous experiments. The 2-opt heuristics remove all crossing parts in a tour by exchanging the tour branches. The three numbers in each column of Table II are the average tour length improved by the 2-opt heuristics, the average rate of improvement in percent, and the average number of branch exchanges. In DCA, the improvement rate and the number of exchanges are smaller than those in the other two algorithms, which implies that the DCA solutions have better structures so that only a little local improvements can be done on them.

Figure 1 shows a solution obtained by DCN annealing for a 200-city problem. We can see that the obtained solution is fairly good.

Table I Algorithm comparison. The testbeds are 100 sets of 30-city problems, 100 sets of 50-city problems, 50 sets of 100-city problems, and 10 sets of 200-city problems. Each city allocation was randomly generated in a unit square. The number in each column denote the average tour length for the valid tours. The number in the parenthesis is the number of valid obtained tours.

	30	50	100	200
DCA	4.69 (100)	5.98 (100)	8.48 (50)	11.98 (10)
PMA	4.70 (100)	6.01 (100)	9.64 (48)	17.80 (4)
MFA	5.42 (99)	7.34 (98)	10.35 (48)	15.77 (9)

Table II Improvement by 2-opt heuristics. The three numbers in each column are the average tour length improved by the 2-opt heuristics, the average rate of improvement in percent (%), and the average number of branch exchanges.

	30			50			100			200		
DCA	4.65	0.8	1.6	5.88	1.6	3.5	8.21	3.2	11.3	11.23	6.2	33.8
PMA	4.64	1.3	1.9	5.88	2.1	3.8	8.16	15.4	32.3	11.38	36.1	163.3
MFA	4.73	12.9	9.8	6.05	17.6	20.4	8.26	20.2	44.0	11.36	18.0	121.4

4 Convergence

Let us discuss the convergence of the basic DCN algorithm described in Section 2. The inner iteration loop described in Step 2 is a discrete-time dynamical system. This sub-system has a Lyapunov function

$$G(\boldsymbol{\lambda}) = \sum_n r_n \log \lambda_n + \sum_a s_a \log \sum_m (\exp(U_{a,m})/\lambda_m), \tag{26}$$

whose stationary condition, $\partial G/\partial \lambda_n = 0$, is equivalent to (16). By defining auxiliary variables

$$Y_n = \frac{1}{r_n} \sum_a s_a \frac{\exp(U_{a,n})/\lambda_n^{old}}{\sum_m (\exp(U_{a,m})/\lambda_m^{old})}, \tag{27}$$

$$Z_a = \frac{1}{s_a} \sum_n r_n \frac{s_a \exp(U_{a,n})/\sum_m (\exp(U_{a,m})/\lambda_m^{old})}{\sum_b s_b \exp(U_{b,n})/\sum_m (\exp(U_{b,m})/\lambda_m^{old})}, \tag{28}$$

we can prove this as follows:

$$\begin{aligned} G(\boldsymbol{\lambda}^{new}) - G(\boldsymbol{\lambda}^{old}) &= \sum_n r_n \log Y_n + \sum_a s_a \log Z_a \\ &= \log \Pi_n (Y_n)^{r_n} + \log \Pi_a (Z_a)^{s_a} \\ &\leq \log \left(\frac{\sum_n r_n Y_n}{\sum_n r_n} \right) + \log \left(\frac{\sum_a s_a Z_a}{\sum_a s_a} \right) = 0. \end{aligned} \tag{29}$$

Therefore, the sub-system globally converges to a solution of (16).

However, the whole algorithm does not necessarily converge. Actually, when it is applied to TSPs with $A = 0$, the basic algorithm oscillates in a two-cycle periodicity, which implies that the largest absolute eigenvalue of its Jacobi matrix is negative. Therefore, if some positive values are added to diagonal elements of the Jacobi matrix, they play a role to suppress the oscillation. Hence, if the energy curvature matrix $\boldsymbol{W}$ has negative diagonal elements, i.e., $W_{a,n;a,n} = -A < 0$, the system is likely to converges. Negative diagonal elements of the energy curvature matrix correspond to positive diagonal elements of the Jacobi matrix of the dynamical system. In addition, when $A > 0$, the converged value is close to a hypercube vertex at a relatively high temperature so that it can be regarded as a binary solution of the combinatorial optimization problem. This feature can be seen as follows. The second right hand side term in the energy function (20), $(A/2) \sum V_{a,n}(1 - V_{a,n})$, is a convex quadratic function whose minimum is located at $V_{a,n} = 1/2$. Therefore, positive A enlarges the energy distance between the internal points of the domain hypercube and the hypercube vertices, and this makes the converged value close to a hypercube vertex.

If we employ a differential equation system

$$\tau \dot{U}_{a,n} = -U_{a,n} - \frac{1}{T} \left(\sum_{b,m} W_{a,n;b,m} V_{b,m} + J_{a,n} \right), \tag{30}$$

instead of the discrete-time equation (17), together with (15) and (16), we can show that the free energy function (7) is its Lyapunov function, which means that the new dynamical system globally converges to a local minimum of (7). Here, $\dot{U}_{a,n}$ denotes the time-derivative of $U_{a,n}$. The convergence does not depend on the network parameters, i.e., $\boldsymbol{W}$ and $\boldsymbol{J}$. This proof can be seen in Urahama [19] and is also given in Appendix B. In order to employ (30) in a computer implementation, let us consider that (17) in the basic algorithm is replaced by a Euler difference equation of (30):

$$U_{a,n}(t) = (1 - k) U_{a,n}(t-1) - \frac{k}{T} \left(\sum_{b,m} W_{a,n;b,m} V_{b,m}(t-1) + J_{a,n} \right), \tag{31}$$

where k is a time-interval parameter and $0 < k \ll 1$. When this new algorithm is applied to TSPs with $A = 0$, it does not work either, although the procedure theoretically converges for sufficiently small k. The state variable $\boldsymbol{V}$ does not become close to the hypercube vertices until the temperature gets very small. On the other hand, if the time-interval parameter k is not very small, the state variable at a very low temperature tends to oscillate in a two-cycle periodicity. Likewise in the basic algorithm, a positive A makes this procedure converge. Thus, in actual implementation, the convergence property with the differential equation (30) is not better than that of the basic algorithm.

In TSPs, if it is possible to set $A = 0$, the energy function (20) is composed only by the tour distance term. Although a positive A value makes the energy function (20) slightly different from the term representing the total tour length for $T > 0$, it enables the algorithm to work well. According to our experiments for TSPs, obtained solutions were found to be better with the basic algorithm than with the procedure employing (30). Furthermore, the experimental results described in Section 3, where an optimal or semi-optimal solution is obtained for every problem with the basic algorithm, indicate that the algorithm actually converges in every case with a positive A value. In addition, the basic algorithm is much faster than the procedure employing (30).

Figure 2 shows the average of DCN annealing solutions for various parameter A values, when applied to 10-, 20-, 30-, and 40-city problems. In every case, the "badness" of the solution for low A values is due to the fact that the basic algorithm is dominated by the two-cycle periodicity and it often fails to obtain a solution. Except for 10-city problems, $A = 0.6$ is found to be the best parameter value. We have experimentally confirmed that this parameter value $A = 0.6$ is the best value up to 200-city problems.

5 Bifurcations

In this section, let us study bifurcations of the DCN solutions.

The curvature of the entropy function H is given by $\partial^2 H/\partial V_{a,n}\partial V_{b,m} = -\delta_{ab}\delta_{nm}/V_{a,n}$. Since $1/V_{a,n} \geq 1$, the curvature matrix $\nabla\nabla H \equiv (\partial^2 H/\partial V \partial V)$ can be written as

$$\nabla\nabla H = -\boldsymbol{I} + (\text{negative semi-definite matrix}), \tag{32}$$

where $\boldsymbol{I}$ is the identity matrix. The curvature matrix of the free energy function (7) is given by

$$\nabla\nabla F = \boldsymbol{W} - T\nabla\nabla H. \tag{33}$$

Let ξ_{min} denote the minimum eigenvalue of the energy curvature matrix $\boldsymbol{W}$. When temperature T is larger than $-\xi_{min}$, the curvature matrix of the free energy function is positive definite, implying that the free energy function is convex and has a unique minimum. At the low temperature limit, however, the free energy function has in general a lot of local minima that correspond to solutions of the problem. Therefore, at some critical temperature $T_c(\leq -\xi_{min})$, a bifurcation of the minimum solution occurs.

These bifurcation processes are dependent on the structurally stable symmetry of the problem, like in the Ising and Potts spin MFT cases [16, 5]. Without structurally stable symmetries in the problem, one can generically expect only saddle-node bifurcations to

occur [16]. However, the free energy function for a TSP (23) is invariant under the N-th order cyclic transformation:

$$V_{a,n} \longrightarrow V_{a,n+m} \qquad (m = 1, ..., N-1), \tag{34}$$

and the N-th order reverse transformation:

$$V_{a,n} \longrightarrow V_{a,m-n} \qquad (m = 0, 1, ...N-1). \tag{35}$$

The cyclic and reverse transformation invariances correspond to the fact that the tour length does not depend on the starting city, and the fact that it does not change when the tour direction is reversed, respectively. When a solution is transformed to itself by the N-th order cyclic (or, reverse) transformation, the solution is said to have the N-th order cyclic (or, reverse) symmetry. In TSPs, the unique minimum at high temperature has the N-th order cyclic and reverse symmetries [16]. Furthermore, since the condition $\sum_n V_{a,n} = 1$ holds automatically for any a, the unique symmetric minimum is found to be $V_{a,n} = 1/N$ $(a, n = 1, ..., N)$. Since the low temperature minima are likely to have neither of the above-mentioned two symmetries, during the course of an annealing procedure, bifurcations that break the symmetries occur. They are called cyclic symmetry breaking bifurcations and reverse symmetry breaking bifurcations [16].

Figure 3 is a typical example showing a bifurcation diagram of a 5-city TSP, where $V_{1,i}$ $(i = 1, ..., 5)$ for every minimum are plotted against temperature. When $T > 0.35$, there is a unique symmetric minimum. At $T \approx 0.35$ and 0.32, a cyclic symmetry breaking bifurcation and a reverse symmetry breaking bifurcation can be observed, respectively. Several saddle-node bifurcations can also be observed at $T \approx 0.24, 0.22$ and 0.17.

These local bifurcation structures imply that the DCN annealing procedure may also have several adverse properties like in the Ising and Potts spin MFT annealing procedures [16, 5]. If an annealing solution disappear by a bifurcation at some temperature and there are more than two distinctive minima at this temperature, whatever minimum is obtained by the annealing procedure is not unique due to the instability at the disappearance point. This implies that the annealing solution may not unique, even if we employ a "deterministic" annealing procedure. When new minima are generated, their free energy levels are higher than that of the global minimum at that temperature. However, the free energy levels of some minimum solutions may cross one another as the temperature is lowered. Therefore, the deterministic annealing procedure does not always give the optimal solution. As a consequence, a solution obtained by a deterministic annealing procedure may nonunique and nonoptimal.

However, these properties are much improved in our DCN approach. In DCN annealing, the disappearance of the annealing solution is theoretically possible but experimentally not so probable, while it is a typical phenomenon in the Ising spin MFT annealing approach. For comparison, Figure 4 shows the bifurcation diagram of the Ising spin MFT for the same 5-city problem as used in Figure 3, where the disappearance of the annealing solution can be observed at $T \approx 0.61$. If we remove soft constraint terms from the MFT energy function (21), the disappearance barely occurs, implying that the disappearance is due to the soft constraint terms. Since the DCN energy function (20) contains no soft constraint terms, the disappearance barely occurs, and the annealing solutions are likely to be unique. Thus, the

DCN annealing procedure is much improved in its nonuniqueness. This is a merit of the DCN approach.

Since the unique high temperature solution is given by $V_{a,n} = 1/N$, the first bifurcation temperature turns out to be lower than $-\xi_{min}/N$ (See (33)). Therefore, it is sufficient that we set the starting temperature of the annealing procedure at $-\xi_{min}/N$. Due to the above-mentioned symmetries of the TSPs, the eigenvalues of the energy curvature can be rigorously obtained by solving a reduced eigen equation whose order is N [16], instead of solving the original eigen equation whose order is N^2. Peterson and Söderberg [13] also proposed an approximation method for the first bifurcation temperature for their Potts MFT approach. In the experiment described in Section 3, however, we fixed the starting temperature for implementation convenience.

6 Conclusion

Our DCN approach treats two competing sets of constraints as hard, and the minimizing energy function does not include the constraint terms. Most important merit of DCN is that the domain space becomes smaller than that of the binary and Potts spin approaches. Another merit is that the energy function does not contain constraint terms. Further merit is that its bifurcation structure is improved so that an adverse property of the deterministic annealing is improved. Consequently, our DCN approach can obtain better solutions than those possible with the Ising and Potts spin approaches. Our basic algorithm is based on a synchronous discrete-time dynamical system, which is much faster than those based on differential equation systems. Although its convergence is not assured, the algorithm has been experimentally shown to work well with excitatory self-loops.

References

[1] Bilbro, G., Mann, R., Miller, T. K., Snyder, W. E., Van den Bout, D. E., & White, M. (1989). Optimization by mean field annealing. In *Advances in Neural Information Processing Systems I* (Denver 1988), D. S. Touretzky (ed.), 91-98. San Mateo: Morgan Kaufmann.

[2] Durbin, R., Szeleski, R., & Yuille, A. (1989). An analysis of the elastic net approach to the traveling salesman problem, *Neural Computation*, **1**, 348-358.

[3] Hinton, G. E. (1989). Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation* **1**, 143-150.

[4] Hopfield, J. J., & Tank, D. W. (1985). "Neural" computations of decisions in optimization problems. *Biological Cybernetics*, **52**, 141-152.

[5] Ishii, S., & Sato, M. (1995). Chaotic Potts spin. *ATR Technical Report*, **TR-H-175**, Kyoto: ATR.

[6] Ishii, S., & Sato, M. (1995). Doubly constrained network model for combinatorial optimization problems. In *Proceedings of the 1995 International Symposium on Nonlinear Theory and its Applications* (Las Vegas 1995), 371-374. Japan: IEICE.

[7] Kamgar-Parsi, B., & Kamgar-Parsi, B. (1990). On problem solving with Hopfield neural networks. *Biological Cybernetics*, **62**, 415-423.

[8] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, **220**, 671-680.

[9] Lin, S., & Kernighan, B. W. (1973). An efficient heuristic algorithm for the traveling salesman problem. *Operations Research*, **21**, 498-516.

[10] Luenberger, D. G. (1989). *Linear and Nonlinear Programming* (2nd edition). Reading: Addison-Wesley.

[11] Parisi, G. (1988). *Statistical Field Theory*. Reading: Addison-Wesley.

[12] Peterson, C., & Anderson, J. R. (1988). Neural networks and NP-complete optimization problems: A performance study on the graph bisection problem. *Complex Systems*, **2**, 59-89.

[13] Peterson, C., & Söderberg, B. (1989). A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, **1**, 3-22.

[14] Peterson, C. (1990). Parallel distributed approaches to combinatorial optimization: benchmark studies on traveling salesman problem. *Neural Computation*, **2**, 261-269.

[15] Platzman, L. K., & Bartholdi, J. J. (1989). Spacefilling curves and the planar travelling salesman problem. *Journal of the ACM*, **36**, 719-737.

[16] Sato, M., & Ishii, S. (1996). Bifurcations in mean-field-theory annealing. *Physical Review E*, in press.

[17] Simic, P. D. (1990). Statistical mechanics as the underlying theory of "elastic" and "neural" optimization. *Networks*, **1**, 89-103.

[18] Soukoulis, C. M., Levin, K., & Grest, G. S. (1983). Irreversibility and metastability in spin-glasses. I. Ising model. *Physical Review B*, **28**, 1495-1509.

[19] Urahama, K. (1994). Analog method for solving combinatorial optimization problems. *IEICE Transactions Fundamentals*, **E77-A**, 302-308.

[20] Van den Bout, D. E., & Miller III, T. K. (1989). Improving the performance of the Hopfield-Tank neural network through normalization and annealing. *Biological Cybernetics*, **62**, 129-139.

[21] Wasserstrom, E. (1973). Numerical solutions by the continuation method. *SIAM Reviews*, **15**, 89-119.

[22] Wilson, G. W., & Pawley, G. S. (1988). On the stability of the travelling salesman problem algorithm of Hopfield and Tank. *Biological Cybernetics*, **58**, 63-70.

[23] Wu, F. Y. (1982). The Potts model. *Reviews of Modern Physics*, **54**, 235-268.

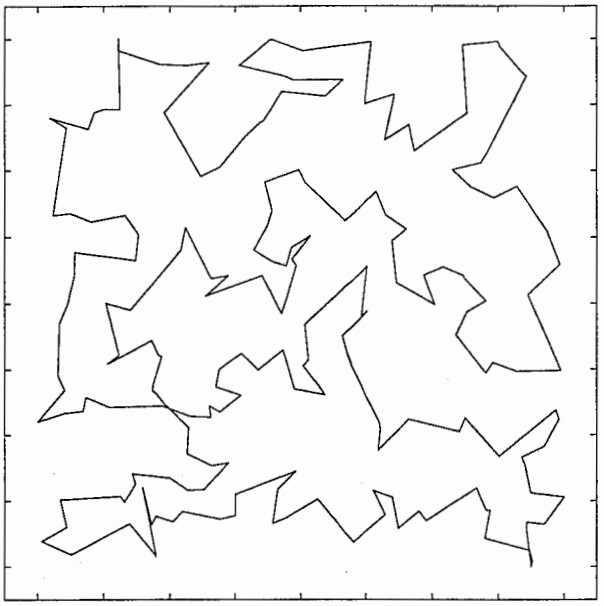

Figure 1

A solution obtained by DCN annealing for a 200-city problem.

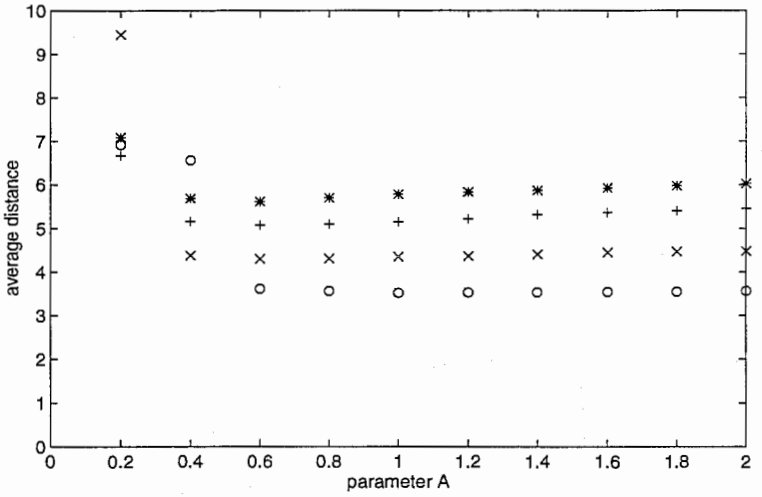

Figure 2

Average distance of the DCN annealing solutions for various values of a parameter A, where it is applied to 10-, 20-, 30-, and 40-city testbeds. Each testbed consists of 100 randomly generated city allocations. The solutions are averaged for successful results, whose number becomes small as A becomes small. For example, for 20-city TSPs with $A = 0.2$, 48 tours among 100 trials are obtained. "o," "×," "+," and "*" denote the results for 10-, 20-, 30-, and 40-city problems, respectively.

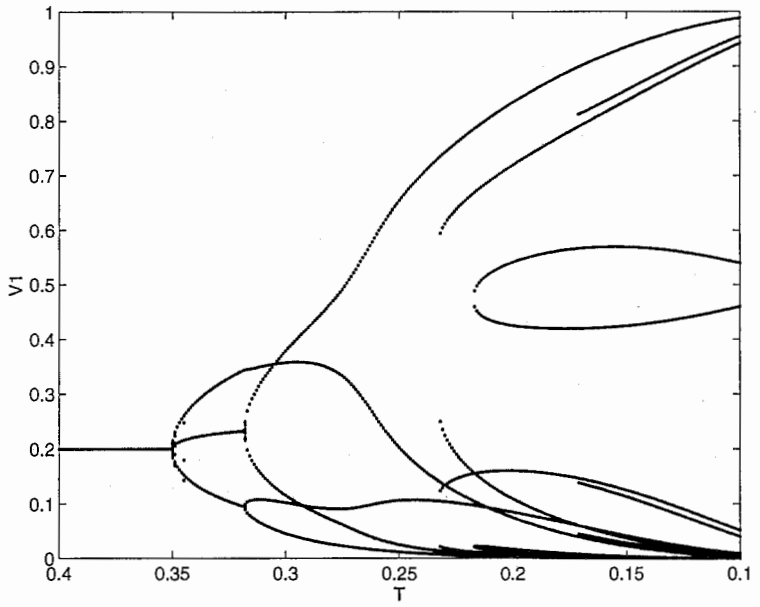

Figure 3

A DCN bifurcation diagram over temperature T. At each temperature 200 sets of random initial states are prepared, and each $V_{1,i}$ $(i = 1, ..., 5)$ after convergence is plotted. The criteria for convergence were; $\delta\lambda = 10^{-5}$ and $\delta \boldsymbol{V} = 10^{-5}$.

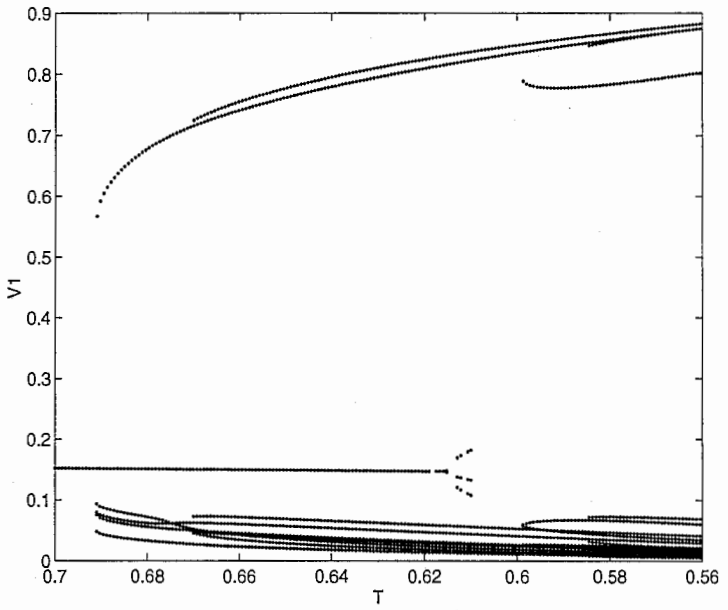

Figure 4

An MFT bifurcation diagram over temperature T. At each temperature 200 sets of random initial states are prepared, and each $V_{1,i}$ $(i = 1, ..., 5)$ after convergence is plotted. The criterion for convergence was $\delta \boldsymbol{V} = 10^{-5}$. Parameters in (21) were set at $A = 2B = 2C = 1.5$.

Appendix A

Ising spin MFT [1] is a discrete-time dynamical system, where a spin variable $V_{a,n} \in [0,1]$ is chosen in a random order and updated by the following equations asynchronously.

$$U_{a,n} = -\sum_{b,m} W_{a,n;b,m} V_{b,m} - J_{a,n}, \tag{A1}$$

$$V_{a,n} = \frac{1}{1+e^{-U_{a,n}/T}}, \tag{A2}$$

where $\boldsymbol{W}$ and $\boldsymbol{J}$ are determined by expanding the energy function (21). In the experiments described in Section 3, parameters in (21) were set at $A = 2B = 2C = 1.5$, the annealing procedure was scheduled as $\delta T = 0.005$, and the convergence of the equations (A1), (A2) was determined by $\delta \boldsymbol{V} = 10^{-5}$. Here, $\delta \boldsymbol{V}$ is the difference of the state variable $\boldsymbol{V}$ after all the units are updated once.

Potts spin MFT [13] is a discrete-time dynamical system, where a spin variable V_a, which is an N-dimensional vector in $[0,1]^N$, is chosen in a random order and updated by the following equations asynchronously.

$$U_{a,n}(t) = \sum_{b,m} W_{a,n;b,m} V_{b,m}(t-1) + J_{a,n} \tag{A3}$$

$$V_{a,n}(t) = \frac{\exp(-U_{a,n}(t)/T)}{\sum_m \exp(-U_{a,m}(t)/T)}, \tag{A4}$$

where $\boldsymbol{W}$ and $\boldsymbol{J}$ are determined by expanding the energy function (22). In the experiments described in Section 3, parameters in (22) were set at $2A = B = 1.0$, the annealing procedure was scheduled as $\delta T = 0.005$, and the convergence of the equations (A3), (A4) was determined by $\delta \boldsymbol{V} = 10^{-5}$.

Appendix B

Let us briefly show that the free energy function (7) is the Lyapunov function of the dynamical system, (29), (15), and (16). From (15),

$$\log V_{a,n} = U_{a,n} - \xi_n - \eta_a \tag{A5}$$

$$\dot{V}_{a,n}/V_{a,n} = \dot{U}_{a,n} - \dot{\xi}_n - \dot{\eta}_a \tag{A6}$$

holds, where $\eta_a = \log \sum_m (\exp(U_{a,m})/\lambda_m) - \log s_a$ and $\xi_n = \log \lambda_n$. From (7), (29), and (A6),

$$\partial F/\partial V_{a,n} = -\tau T \dot{U}_{a,n} - T(\xi_n + \eta_a - 1) \tag{A7}$$

holds. Then,

$$\begin{aligned} \dot{F} &= \sum_{a,n} (\partial F/\partial V_{a,n}) \dot{V}_{a,n} \qquad \text{(A8)} \\ &= \sum_{a,n} (-\tau T \dot{U}_{a,n} - T(\xi_n + \eta_a - 1)) \dot{V}_{a,n} \\ &= -\tau T \sum_{a,n} (\dot{V}_{a,n})^2 / V_{a,n} - T \sum_n (\tau \dot{\xi}_n + \xi_n) \sum_a \dot{V}_{a,n} - T \sum_a (\tau \dot{\eta}_a + \eta_a) \sum_n \dot{V}_{a,n} + T \sum_{a,n} \dot{V}_{a,n} \end{aligned}$$

holds. (A6) is used in this derivation. From (15) and (16), (10) and (11) are satisfied. Therefore, $\sum_a \dot{V}_{a,n} = \sum_n \dot{V}_{a,n} = 0$. Thus,

$$\dot{F} \leq 0 \tag{A9}$$

holds.