The Proceedings of the

IJCNN 2003

International Joint Conference on Neural Networks

The International Neural Network Society

IEEE

The IEEE Neural Networks Society

Doubletree Hotel – Jantzen Beach
Portland, Oregon
July 20 – 24, 2003

IEEE Catalog Number: 03CH37464C
ISBN: 0-7803-7899-7

Cover photo used with permission from Steve Terrill Photography, PO Box 55641, Portland, OR 97238, www.terrillphoto.com, 503-282-8587.

CD-ROM Support
by The Printing House, Inc.
Phone: +1-608-873-4500
www.printinghouseinc.com
e-mail: graphics@printinghouseinc.com

CD-ROM Help
Search
Close and Exit

Proceedings

of the

International Joint Conference on Neural Networks 2003

Co-Sponsored by:

The International Neural Network Society

The IEEE Neural Networks Society

Doubletree Hotel – Jantzen Beach

Portland, Oregon

July 20 – 24, 2003

Distribution Approximation, Combinatorial Optimization, and Lagrange-Barrier

Lei Xu
Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, Hong Kong
Email: lxu@cse.cuhk.edu.hk

***Abstract*— In this paper[1], typical analog combinatorial optimization approaches, such as Hopfield net, Hopfield-Lagrange net, Maximum entropy approach, Lagrange-Barrier approach, are systematically examined from the perspective of learning distribution. The minimization of a combinatorial cost is turned into a procedure of learning a simple distribution to approximate the Gibbs distribution induced from this cost such that both the distributions share a same global peak. From this new perspective, a new general guideline is obtained for developing analog combinatorial optimization approaches. Moreover, the Lagrange-Barrier iterative procedure proposed in Xu (1994, 1995a) is further elaborated with guaranteed convergence on a feasible solution that satisfies constraints.**

I. Introduction

Many combinatorial optimization problems can be usually formulated as follows:

$$\begin{aligned}&\min_{V} E_o(V),\ V=\{v_{ij}\}_{i=1,j=1}^{N,M},\\ &subject\ to\\ &C_e^{col}:\quad \sum_{i=1}^{N} v_{ij}=D_j^{col}, j=1,\cdots,M,\\ &C_e^{row}:\quad \sum_{j=1}^{M} v_{ij}=D_i^{row}, i=1,\cdots,N;\\ &C_b:\quad v_{ij}\ takes\ either\ 0\ or\ 1,\end{aligned} \tag{1}$$

where $D_j^{col}, j=1,\cdots,M$ and $D_i^{row}, i=1,\cdots,N$ are given constants. Moreover, C_b can also be a general binary constraint that v_{ij} takes either a constant θ_0 or θ_1. Without losing generality, it is always easy to use a one-to-one mapping to normalize it back to the case of C_b given in eq.(1).

A typical example is the traveling salesman problem (TSP) with $N=M, D_i^{row}=1, D_j^{col}=1$ and

$$E_o(V)=\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k=1}^{N} d_{ij}v_{i,k}(v_{j,k-1}+v_{j,k+1}) \tag{2}$$

where $d_{i,j}$ are known parameters. Other examples and a number of algorithms for the problem eq.(1) can be found in [9], [10], [13].

In the past decade, many efforts have been made in the literature of artificial neural networks on the problem eq.(1) of quadratic $E_o(V)$ by analog optimization since the work by [8]. Some detailed reviews of the efforts can be found in [2], [17], [1].

[1]The work described in this paper was fully supported by a grant from the Research Grant Council of the Hong Kong SAR (project No: CUHK4336/02E).

According to their features on dealing with the constraints C_e^{col}, C_e^{row} and C_b, these efforts can be roughly classified into three categories:

(1) The constraints C_e^{col} and C_e^{row} are transformed into a quadratic penalty term, e.g.,

$$G(V)=\sum_{j=1}^{M}(\sum_{i=1}^{N} v_{ij}-D_j^{col})^2+\sum_{i=1}^{N}(\sum_{j=1}^{M} v_{ij}-D_i^{row})^2, \tag{3}$$

that can be merged into $E_o(V)$ via a weighted sum and results in a new quadratic $E(V)$ from which the connection weights of a classic Hopfield network [8] are obtained. The network minimizes $E(V)$ by implementing a system of dynamic equations in a parallel way, with C_b realized via sigmoid functions that constrain v_{ij} between the interval $[0,1]$. This approach is referred as the classical Hopfield network approach. A disadvantage of this approach is that the penalty strength is difficult to control, which often results in unfeasible solutions that violate the constraints C_e^{col}, C_e^{row}.

(2) The connection weights of the Hopfield network are obtained based on the original $E_o(V)$, with the constraint C_b still realized by sigmoid functions. Instead, the constraints C_e^{col}, C_e^{row} are considered as the Lagrange terms which add on biasing constants to the Hopfield network that minimizes $E_o(V)$ and the Lagrange terms under the fixed Lagrange coefficients. Moreover, another set of dynamic equations implement the maximization of $E_o(V)$ to settle the appropriate values of the Lagrange coefficients [14], [1]. The approach is referred as the Hopfield-Lagrange network approach.

(3) In a conference paper [17], a general cost $E_o(V)$, which can be either quadratic or nonquadratic, is considered and extended into a Lagrange-Barrier cost $E(V)$ by adding in the Lagrange terms for C_e^{col}, C_e^{row} and a barrier term

$$B(V)=\sum_{i,j} v_{ij}\ln v_{ij} \tag{4}$$

that constraints v_{ij} as an interior point between the interval $[0,1]$. Particularly, for a quadratic $E_o(V)$, this $E(V)$ becomes the cost used in [20] and the cost used under the name of maximum entropy approach [5], [7], [6]. Moreover, in [18] the barrier term

$$B(V)=\sum_{i,j}[v_{ij}\ln v_{ij}+(1-v_{ij})\ln(1-v_{ij})] \tag{5}$$

has been studied. Also, the link has been set up between the leaking energy term in the Hopfield network and a family of barriers that includes eq.(4) and eq.(5), and thus we know that

the above Hopfield-Lagrange network is one specific way for minimizing a Lagrange-Barrier cost. Furthermore, Xu in [17], [18] also proposed a new general approach for minimizing the Lagrange-Barrier cost. It consists of iterative updating equations obtained from $\frac{\partial E(V)}{\partial v_{ij}} = 0$, that updates v_{ij} with a particular emphasize on the satisfaction of the constraints C_e^{col}, C_e^{row}. This satisfaction is refined immediately after each updating on v_{ij}, through updating the Lagrange coefficients by an inner iterative loop. The approach is referred as the Lagrange-Transformation or the Lagrange-Barrier approach.

Though having the favorable parallel implementable feature, almost all the neural network motivated approaches share one unfavorable feature that these intuitive approaches have not been satisfactorily explained from a theoretical point of view. Basically, there is neither theory to support naturally the intuitive concept and to guarantee the convergence of a proposed algorithm, nor theory to relate this concept to computational complexity as traditionally developed in the discrete settings.

This paper aims at tackling partly these problems. The above neural network motivated approaches are systematically examined from a new perspective. The deterministic minimization of $E_o(V)$ is turned into a sequence of finding a simple distribution to approximate the Gibbs probabilistic distribution induced from $E_o(V)$ under the related constraints, such that the Gibbs distribution and the approximate distribution share a same global peak. In Sec. II, the basic idea of this distribution estimation theory is proposed. Particularly, it is shown that two special cases not only relate to the classical Metroplois sampling technique [12] but also interpret and justify the maximum entropy approach and the Lagrange-Barrier approach, as well as the Hopfield-Lagrange network via a link previously built [18]. In Sec.III, we further elaborate the Lagrange-Barrier iterative procedure proposed in [17], [18], with a guaranteed convergence on a feasible solution that satisfies constraints. Two simulation examples are demonstrated to illustrate the effectiveness of the iterative procedure. Finally, concluding remarks are given in Sec.IV.

II. Distribution Approximation and Optimization

A. The Basic Idea

The problem of finding a global minimization solution of $E_o(V)$ under a set of constraints is equivalent to the problem of finding a global peak of the following Gibbs distribution:

$$p(V) = \frac{e^{-\frac{1}{\beta}E_o(V)}}{Z_\beta}, \quad Z_\beta = \sum_V e^{-\frac{1}{\beta}E_o(V)}, \tag{6}$$

subject to the constraints, since $\max_V p(V)$ is equivalent to $\max_V \ln p(V)$ or $\min_V E_o(V)$.

Usually this $p(V)$ has many local maximums, it is difficult to the peak V_p. To avoid this difficult, we propose to use a simple distribution $q(V)$ to approximate $p(V)$ on a domain D_v such that the global peak of $q(V)$ is easy to find and that $p(V)$ and $q(V)$ share the same peak $V_p \in D_v$.

To do so, we adopt the following two types of measures for implementing this approximation:

$$\begin{aligned} Type\ (a): \quad & \min_q\ KL(p,q), \\ & KL(p,q) = \sum_{V\in D_v} p(V)\ln\frac{p(V)}{q(V)}, \\ Type\ (b): \quad & \min_p\ KL(q,p), \\ & KL(q,p) = \sum_{V\in D_v} q(V)\ln\frac{q(V)}{p(V)}. \end{aligned} \tag{7}$$

Given a D_v that contains V_p, the smaller the D_v is , the easier for a simple $q(V)$ to approximate $p(V)$, and thus the more likely $q(V)$ and $p(V)$ share a same global peak. This D_v is considered via the following support of $p(V)$:

$$D_\varepsilon(\beta) = \{V : p(V,\beta) > \varepsilon,\ a\ small\ constant\ \varepsilon > 0.\} \tag{8}$$

under the control of the parameter β. For a sequence $\beta_0 > \beta_1, \cdots > \beta_t$, we have $D_\varepsilon(\beta_t) \subset, \cdots, D_\varepsilon(\beta_1) \subset D_\varepsilon(\beta_0)$ that keep to contain the global minimization solution of $E_o(V)$, since the equivalence of $\max_V p(V)$ to $\min_V E_o(V)$ is irrelevant to β.

Therefore, it follows from eq.(7) that we can find a sequence $q_0(V), q_1(V), \cdots, q_t(V)$ that approximates $p(V)$ on the shrinking domain $D_\varepsilon(\beta)$. For a large β_t, $p(V)$ has a large support and thus $q(V)$ adapts the overall configuration of $p(V)$ in a big domain $D_\varepsilon(\beta)$. As β_t reduces, $q_t(V)$ becomes more and more concentrating on adapting the detailed configuration of $p(V)$ around the global peak solution $V_p \in D_\varepsilon$. As long as β_0 is large enough and β reduces slowly enough towards to zero, we can finally find the global minimization solution of $E_o(V)$. Strictly speaking, whether the global solution can be found relates to the selection of the distribution form of $q(V)$, and the selection of β_0, the reducing rate of β as well as the value of β at which the searching procedure stops.

B. Type (a) in Two Typical Cases

For the problem eq.(1), we consider $q(V)$ in the following special cases

$$\begin{aligned} & q_1(V) = Z_1^{-1}\prod_{i,j} e^{v_{ij}\ln q_{ij}}, \quad 0 \le q_{ij} \le \infty, \\ & Z_1 = \sum_{i,j}\prod_{i,j} e^{v_{ij}\ln q_{ij}}, \\ & q_2(V) = \prod_{i,j} q_{ij}^{v_{ij}}(1-q_{ij})^{v_{ij}}, \quad 0 \le q_{ij} \le 1; \end{aligned} \tag{9}$$

and from the constraints in eq.(1) we have

$$\begin{aligned} & C_e^{col}: \quad \sum_{i=1}^{N} < v_{ij} >= D_j^{col}, j = 1, \cdots, M, \\ & C_e^{row}: \quad \sum_{j=1}^{M} < v_{ij} >= D_i^{row}, i = 1, \cdots, N; \\ & < v_{ij} >= \begin{cases} q_{ij}\frac{Z_{ij}}{Z_1}, & \text{for } q_1(V), \\ q_{ij}, & \text{for } q_2(V), \end{cases} \\ & Z_{ij} = \sum_{k\neq i, l\neq j}\prod_{k,l} e^{v_{kl}\ln q_{kl}}; \end{aligned} \tag{10}$$

where $< x >$ denotes the expectation of the random variable x. When N, M are large, we have $Z_{ij} \approx Z_1$, and we also have $< v_{ij} > \approx q_{ij}$ also for the case of $q_1(V)$.

If not considering the constraints C_e^{col}, C_e^{row}, the global peak solution of $q(V)$ by eq.(9) is simply estimated by

$$v_{ij} = \begin{cases} 1, & \text{if } q_{ij} > 0.5, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Moreover, we can either consider the constraints C_e^{col} and get the global peak solution by

$$v_{ij} = \begin{cases} 1, & \text{if } j = arg\max_k q_{ik}, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

or we can similarly consider the constraints C_e^{row}.

Next, we consider the problem of estimating q_{ij}, according to the two types of approximations given in eq.(7).

From eq.(7), Type (a) approximation is equivalent to the minimization of

$$\begin{aligned} & H_0(\{q_{ij}\}) = -\sum_{V \in D_v} p(V) \ln q(V) \\ & = \begin{cases} -\sum_{ij} p_{ij} \ln q_{ij}, & q_1(V), \\ -\sum_{ij} [p_{ij} \ln q_{ij} + (1 - p_{ij}) \ln (1 - q_{ij})], & q_2(V); \end{cases} \\ & p_{ij} = \sum_V v_{ij} p(V), \end{aligned}$$

under the constraints C_e^{col}, C_e^{row} given in eq.(10). Considering Lagrange approach, we have

$$\begin{aligned} & H(\{q_{ij}\}) = H_0(\{q_{ij}\}) + \sum_{j=1}^{M} \lambda_j^{col} [\sum_{i=1}^{N} q_{ij} - D_j^{col}] \\ & + \sum_{i=1}^{N} \lambda_i^{row} [\sum_{j=1}^{M} q_{ij} - D_i^{row}], \qquad (13) \\ & \frac{\partial H(V)}{\partial q_{ij}} = \begin{cases} -\frac{p_{ij}}{q_{ij}} + \lambda_j^{col} + \lambda_i^{row}, & \text{for } q_1(V), \\ -\frac{p_{ij}}{q_{ij}} + \frac{1-p_{ij}}{1-q_{ij}} + \lambda_j^{col} + \lambda_i^{row}, & \text{for } q_2(V). \end{cases} \end{aligned}$$

We have two methods for solving this problem:

(1) The Method I. If we have already get the mean field p_{ij} in eq.(13) under the Gibbs distribution $p(V)$, it follows from $\frac{\partial H(\{q_{ij}\})}{\partial q_{ij}} = 0$ that

$$\begin{aligned} & q_{ij} = \begin{cases} \frac{p_{ij}}{\lambda_j^{col} + \lambda_i^{row}}, & \text{for } q_1(V), \\ a\ root\ of\ the\ equation\ below, & \text{for } q_2(V). \end{cases} \\ & (\lambda_j^{col} + \lambda_i^{row}) q_{ij}^2 - (1 + \lambda_j^{col} + \lambda_i^{row}) q_{ij} + p_{ij} = 0, \\ & 0 \le q_{ij}^r \le 1 \end{aligned} \tag{14}$$

By putting it into the constraints C_e^{col} and C_e^{row} given by eq.(10), we update the Lagrange coefficients $\lambda_j^{col}, \lambda_i^{row}, j = 1, \cdots, M, i = 1, \cdots, N$ to satisfy :

$$C_e^{col} : \sum_{i=1}^{N} q_{ij} = D_j^{col}, \quad C_e^{row} : \sum_{j=1}^{M} q_{ij} = D_i^{row}. \tag{15}$$

Thus, we encounter a problem of solving $N + M$ nonlinear equations. Particularly, for the case of $q_1(V)$ with $D_j^{col} = D_i^{row} = 1$, we use the following iterative algorithm to solve it:

$$\begin{aligned} & (\lambda_i^{row})^{new} = (\lambda_i^{row})^{old} \sum_{j=1}^{M} q_{ij}^e (\{(\lambda_i^{row})^{old}, (\lambda_j^{col})^{old}\}), \\ & (\lambda_j^{col})^{new} = (\lambda_j^{col})^{old} \sum_{i=1}^{N} q_{ij}^e (\{(\lambda_i^{row})^{old}, (\lambda_j^{col})^{old}\}), \\ & i = 1, \cdots, N, j = 1, \cdots, M, \end{aligned} \tag{16}$$

where $q_{ij}^e(\{(\lambda_i^{row})^{old}, (\lambda_j^{col})^{old}\})$ denotes q_{ij} by eq.(14) for the case of $q_1(V)$.

The remaining problem is to get the mean field p_{ij} of the Gibbs distribution $p(V)$, which can be made by a well known classical stochastic approximation approach called Metroplois sampling technique [12]. That is, we let β to start from a value large enough and then reduces to a value low enough. Under each given β, the Metroplois sampling is used to get eq.(13) at each β. Finally, at the lowest β value we can obtain a solution by eq.(14) and eq.(16). In fact, this procedure can be regarded as a combination of the Lagrange approach and a variant of the widely studied simulated annealing technique [11].

(2) The Method II. Even using the Metroplois sampling technique, it still needs quite a long sampling period to get a stationary process that can be used for estimating p_{ij}, because the Gibbs distribution $p(V)$ in eq.(6) is usually complicated with many local minimums. Alternatively, we get a set of samples $V_t = \{v_{ij}^{(t)}\}, t = 1, \cdots, N_t$ from its approximation $q(V)$, and then estimate

$$p_{ij} = \sum_{t=1}^{N_t} v_{ij}^{(t)} p(V_t) / q(V_t), \tag{17}$$

which is much easier to compute. Next, we use the same approach given in the above eq.(14) and eq.(16) to get a new set of estimates on q_{ij} for a new estimate $q(V)$, based on which we can again get new estimates on p_{ij} by eq.(17). The process can be iterated until it converges.

C. Type (b) in Two Typical Cases

We further consider Type (b) approximation in eq.(7). From eq.(9), we have

$$\begin{aligned} & (a) \quad for\ q_1(V)\ \sum_V q(V) \ln q(V) = -\ln Z_1 + \\ & \sum_{ij} q_{ij} \frac{Z_{ij}}{Z_1} \ln q_{ij} \approx -\ln Z_1 + \sum_{ij} q_{ij} \ln q_{ij}; \\ & (b) \quad for\ q_2(V),\ \sum_V q(V) \ln q(V) = \\ & \sum_{ij} [q_{ij} \ln q_{ij} + (1 - q_{ij}) \ln (1 - q_{ij})]. \end{aligned} \tag{18}$$

Moreover, we consider the cases that satisfy the condition:

$$E_o(V)\ is\ quadratic\ with\ respect\ to\ V\ and\ \frac{\partial^2 E_o(V)}{\partial^2 v_{ij}} = 0. \tag{19}$$

which is satisfied by the TSP problem eq.(2) and also by a number of concave costs [9].

In these cases, it follows from eq.(9) that $< \sum_{k\neq i,l\neq j} v_{ij}v_{kl}w_{ijkl} >= \sum_{k\neq i,l\neq j} < v_{ij} >< v_{kl} > w_{ijkl}$ which is true due to the independence in $q_2(V)$ and can be regarded as being true approximately for $q_1(V)$. Thus, $-\sum_V q(V)\ln p(V)$ becomes

$$\ln Z + \frac{1}{\beta}\sum_V q(V)E_o(V) = \ln Z + \begin{cases} \frac{1}{\beta}E_o(\{q_{ij}\})\frac{Z_{ij}}{Z_1} \approx \frac{1}{\beta}E_o(\{q_{ij}\}), & \text{for } q_1(V), \\ \frac{1}{\beta}E_o(\{q_{ij}\}), & \text{for } q_2(V). \end{cases} \quad (20)$$

After ignoring the irrelevant terms $\ln Z_1, \ln Z$ and putting eq.(18) and eq.(20) together, we get that Type (b) approximation is equivalent to

$$\min_{q_{ij}} E(\{q_{ij}\}),\ subject\ to\ eq.(10),$$
$$E(\{q_{ij}\}) = \frac{1}{\beta}E_o(\{q_{ij}\}) + \begin{cases} \sum_{ij} q_{ij}\ln q_{ij}, & \text{for } q_1(V), \\ \sum_{ij}[q_{ij}\ln q_{ij} + (1-q_{ij})\ln(1-q_{ij})], & \text{for } q_2(V). \end{cases} \quad (21)$$

The case for $q_1(V)$ interprets the Lagrange-Barrier approach with the barrier eq.(4) discussed under the category (3) in Sec.I. In fact, it justifies the intuitive treatment of simply regarding the discrete v_{ij} as an analog variable between the interval $[0,1]$, previously used in [17] under the name of Lagrange-Transform (LT) approach, in [20] under the name of statistical physics, and in [7], [5], [6] under the name of maximum entropy approach.

From this new perspective, we do not regard the analog variables as the direct targets that we want to optimize. These analog variables are the parameters of the simple distribution that we use to approximate the Gibbs distribution induced from the cost $E_o(V)$ of the discrete variables. Instead, the discrete solution will be recovered from these analog parameters according to one of eq.(11) and eq.(12).

Similarly, the case for $q_2(V)$ interprets and justifies the Lagrange-Barrier approach with the barrier eq.(5), also discussed under the category (3) in Sec.I. In [18], this barrier is intuitively argued to be better than the barrier eq.(4) because it gives a U-shape curve. Here, this intuitive preference can also be justified from eq.(10), by noticing that there is an approximation $Z_{ij} \approx Z_1$ used for the case $q_1(V)$ during the transformation from the discrete random variable v_{ij} into the analog parameter q_{ij}, but no approximation for the case $q_2(V)$.

Moreover, the barriers eq.(4) and eq.(5) are respectively the special cases (a) $S(v_{ij}) = v_{ij}$ and (b) $S(v_{ij}) = v_{ij}/(1-v_{ij})$ of a family of barrier functions as follows [18]:

$$B(v_{ij}) = \int_0^{v_{ij}} \ln S(v_{ij})dv_{ij},\ S(0)=0, S(1)=+\infty,$$
$$S(v)\ monotonously\ increases\ within\ (0,1), \quad (22)$$

where $g^{-1}(v) = \ln S(v)$ has a shape similar to $tg(v)$, and thus its inverse $g(v)$ has a shape similar to $tg^{-1}(v)$. That is, $g(v)$ is a type of sigmoid function. In other words, such a barrier term has a form of

$$\sum_{i=1}^N\sum_{j=1}^M B(v_{ij}) = \beta\sum_{i=1}^N\sum_{j=1}^M \int_0^{v_{ij}} g^{-1}(v_{ij})dv_{ij}, \quad (23)$$

which is exactly the leaking energy term in the Hopfield network [8]. In other words, imposing a barrier term of the family eq.(22) is equivalent to minimizing the leaking energy in the classical Hopfield network. This link was first discovered in [18]. With this link, we can see that the Hopfield-Lagrange network studied by [14], [1] can also be interpreted as one specific implementation for minimizing $E(\{q_{ij}\})$ in eq.(21), though whether this implementation gives a converged solution remains an open question.

Being different from the specific implementation by the Hopfield-Lagrange network and also from the algorithms in [20], [7], [5], [6], a general iterative procedure is proposed firstly in [17] and then refined in [18] for minimizing $E(\{q_{ij}\})$ in eq.(21), with a guaranteed convergence on a feasible solution that satisfies constraints, as will be further recommended in Sec.III.

III. Lagrange-Barrier Iterative Procedure

A. Lagrange-Barrier Iterative Procedure

We consider the Lagrange-Barrier costs:

$$E(\{q_{ij}\}) = \frac{1}{\beta}E_o(\{q_{ij}\}) + \sum_{j=1}^M \lambda_j^{col}[\sum_{i=1}^N q_{ij} - D_j^{col}] + B(q_{ij}) + \sum_{i=1}^N \lambda_i^{row}[\sum_{j=1}^M q_{ij} - D_i^{row}],\quad B(q_{ij}) = \begin{cases} \sum_{ij} q_{ij}\ln q_{ij}, & q_1(V), \\ \sum_{ij}[q_{ij}\ln q_{ij} + (1-q_{ij})\ln(1-q_{ij})], & q_2(V). \end{cases} \quad (24)$$

For the case of $q_1(V)$, by following the methods used in [17], [18], from

$$\frac{\partial E(\{q_{ij}\})}{\partial q_{ij}} = \frac{1}{\beta}\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}} + \lambda_j^{col} + \lambda_i^{row} + 1 + \ln q_{ij} = 0,$$

we get

$$q_{ij}^e = \frac{1}{a_i b_j exp(\frac{1}{\beta}\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}})},\quad a_i = exp(\lambda_i^{row} + 0.5),\quad b_j = exp(\lambda_j^{col} + 0.5). \quad (25)$$

Similarly, for the case of $q_2(V)$, from

$$\frac{\partial E(\{q_{ij}\})}{\partial q_{ij}} = \frac{1}{\beta}\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}} + \lambda_j^{col} + \lambda_i^{row} + \ln\frac{q_{ij}}{1-q_{ij}} = 0,$$
$$\frac{\partial E(\{q_{ij}\})}{\partial q_{ij}} = \frac{1}{\beta}\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}} + \lambda_j^{col} + \lambda_i^{row} - \ln(q_{ij}^{-1} - 1) = 0,$$

we get

$$q_{ij}^e = \frac{1}{1 + a_i b_j exp(\frac{1}{\beta}\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}})},$$

$$a_i = exp(\lambda_i^{row}), \quad b_j = exp(\lambda_j^{col}). \tag{26}$$

Under the condition eq.(19), $\frac{\partial E_o(\{q_{ij}\})}{\partial q_{ij}}$ is irrelevant to q_{ij} and $\frac{\partial^2 E(\{q_{ij}\})}{\partial^2 q_{ij}} > 0$. Conditioning on that other variables are fixed at their old values, $E(\{q_{ij}\})$ is minimized at $q_{ij} = q_{ij}^e$ given by eq.(25) or eq.(26). Therefore, we can update q_{ij} in either of the following two manners:

(a) *The sequential manner* That is, with other variables fixed, we update q_{ij} one by one with

$$q_{ij}^{new} = q_{ij}^e, \tag{27}$$

which will reduce $E(\{q_{ij}\})$ monotonically, as discussed as above.

(b) *The parallel manner* To the current values $\{q_{ij}^{old}\}$, each $q_{ij}^e - q_{ij}^{old}$ is obviously a descent direction of $E(\{q_{ij}\})$. Thus, with a stepsize $\eta > 0$ small enough, we can move all $\{q_{ij}\}$ in parallel along the direction $\{q_{ij}^e - q_{ij}^{old}\}$ that is

$$q_{ij}^{new} = q_{ij}^{old} + \eta(q_{ij}^e - q_{ij}^{old}), \tag{28}$$

which will also reduce $E(\{q_{ij}\})$ monotonically.

From eq.(27) and eq.(28), we can see that the constraints C_e^{col}, C_e^{row} given by eq.(10) are satisfied by q_{ij}^{new} as long as they are satisfied by both q_{ij}^{old} and q_{ij}^e. Therefore, what we need to do is to enforce that C_e^{col}, C_e^{row} are satisfied by q_{ij}^e through updating the Lagrange coefficients $a_i, b_j, j = 1, \cdots, M, i = 1, \cdots, N$, as follows:

$$C_e^{col}: \quad \sum_{i=1}^{N} q_{ij}^e = D_j^{col}, \quad C_e^{row}: \quad \sum_{j=1}^{M} q_{ij}^e = D_i^{row}. \tag{29}$$

That is, we encounter a problem of solving $N + M$ nonlinear equations of $a_i, b_j, j = 1, \cdots, M, i = 1, \cdots, N$, or equivalently a problem of finding the global minimum zero of the penalty

$$P(\{q_{ij}\}) = \sum_{j=1}^{M}(\sum_{i=1}^{N} q_{ij}^e - D_j^{col})^2 + \sum_{i=1}^{N}(\sum_{j=1}^{M} q_{ij}^e - D_i^{row})^2. \tag{30}$$

To solve it, we need another inner iteration loop, which is denoted as ENFORCING-LAGRANGE.

In a summary, a problem of eq.(1) that satisfies the condition eq.(19) can be solved by the following *Lagrange-Barrier Iterative Procedure*:

Step 0: Initialize $\{q_{ij}\}$ such that they satisfy all the constraints.

Step 1: Update q_{ij}^{old} into q_{ij}^{new}
(a) Either sequentially $q_{ij}^{new} = q_{ij}^e$ by using eq.(25) for the case of $q_1(V)$ or eq.(26) for the case $q_2(V)$;
(b) Or in parallel by eq.(28).

Step 2: Update $a_i, b_j, j = 1, \cdots, M, i = 1, \cdots, N$ by an ENFORCING-LAGRANGE loop to ensure the satisfaction of eq.(29).

Step 3: Check whether the procedure is converged, if yes, stop; otherwise, go to Step 1.

Since Step 1 reduces $E(\{q_{ij}\})$ monotonically, as long as the ENFORCING-LAGRANGE loop in Step 2 can ensure the satisfaction of eq.(29), the whole procedure will reduce $E(\{q_{ij}\})$ monotonically with the constraints C_e^{col}, C_e^{row} satisfied, until it converges to a local minimum of $E(\{q_{ij}\})$.

However, there remain two important problems. First, the satisfaction of the condition eq.(19) is required. Second, an effective algorithm for the ENFORCING-LAGRANGE loop needs to be designed. Though a direct use of the existing techniques in literature for solving nonlinear equations eq.(29) or for minimizing eq.(30) can be considered, we highly expect a simple iterative procedure that can be implemented in a parallel way.

Both the two problems have been solved in [3], [4]. First, the direction by $\{q_{ij}^e - q_{ij}^{old}\}$ is also proved to be a descent direction of $E(\{q_{ij}\})$ for a general $E_0(\{q_{ij}\})$ that the condition eq.(19) may not be satisfied, and thus the above Step 1(b) applies to any general cases. Second, a surprisingly simple and parallel implementable algorithm for the ENFORCING-LAGRANGE loop has been proposed as follows:

$$\begin{aligned} a_i^{new} &= a_i^{old} + \mu a_i^{old}(\sum_{j=1}^{M} q_{ij}^e(\{a_i^{old}, b_j^{old}\}) - D_i^{row}), \\ b_j^{new} &= b_j^{old} + \mu b_j^{old}(\sum_{i=1}^{N} q_{ij}^e(\{a_i^{old}, b_j^{old}\}) - D_j^{col}), \\ i &= 1, \cdots, N, \ j = 1, \cdots, M, \end{aligned} \tag{31}$$

where $q_{ij}^e(\{a_i^{old}, b_j^{old}\})$ denotes q_{ij}^e by eq.(25) for the case of $q_1(V)$ or eq.(26) for the case $q_2(V)$ at $\{a_i^{old}, b_j^{old}\}$. Moreover, it has been also mathematically proved in [3], [4] that the iterative algorithm eq.(31) guarantees to converge a solution of nonlinear equations eq.(29).

B. Simulation Examples

A large number of simulation examples on TSP instances have been provided in [3] on the cost $E(\{q_{ij}\})$ in eq.(21) for the case $q_1(V)$ and in [4] on the cost $E(\{q_{ij}\})$ in eq.(21) for the case $q_2(V)$, through the Lagrange-Barrier Iterative Procedure that consists of Step 1(b) and Step 2 given by eq.(31). Here, we only demonstrate the results on two TSP instances obtained from the well-known TSPLIB on WWW, in order to illustrate the effectiveness of the iterative procedure and to further confirm our preference on the cost $E(\{q_{ij}\})$ for the case p_2 over that for the case p_1 with the reason discussed in Sec.II.B.

In the simulations, the parameter β starts at 100 and is reduced by a factor of $\frac{8}{10}$ gradually, μ in eq.(31) is taken to be one, and η in eq.(28) is obtained with a line search. The iteration terminates as soon as a feasible solution is generated.

The first example is made by an instance (bays29.tsp) of 29 cities. The iterative procedures are made both on the cost $E(\{q_{ij}\})$ in eq.(21) for the case $q_1(V)$ and the case $q_2(V)$, resulting a same near optimal tour. The ratio of the distance of the near optimal tour to that of an optimal tour is equal to 1.02, as shown in Fig. 1. However, the convergence on the cost $E(\{q_{ij}\})$ for the case $q_1(V)$ is much slower than on

that for the case $q_2(V)$, by 1778 iterations in comparison with 344 iterations. Thus, it confirmed our preference discussed in Sec.II.B. Here, one iteration consists of one circle from Step 1 to Step 3.

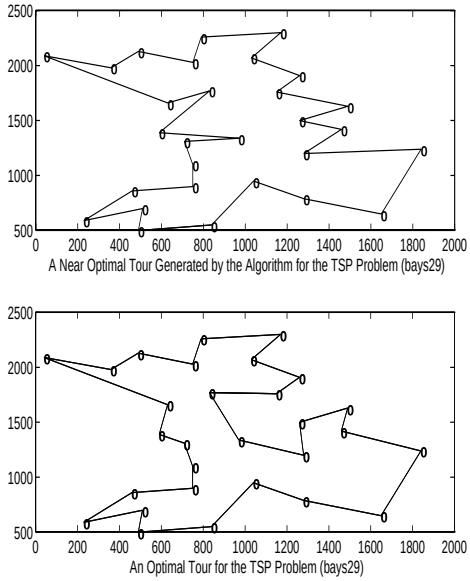

Fig. 1 Near Optimal Tour versus Optimal Tour (bays29.tsp)

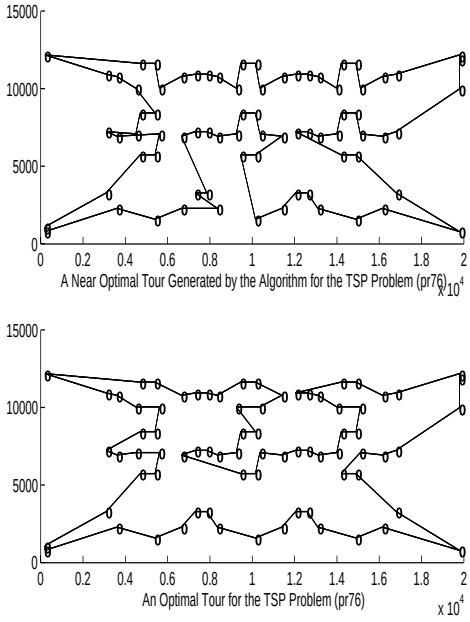

Fig. 2 Near Optimal Tour versus Optimal Tour (pr76.tsp)

The second example is made by an instance (pr76.tsp) of 76 cities. The iterative procedure is made on the cost $E(\{q_{ij}\})$ in eq.(21) for the case $q_2(V)$ only, resulting in a near optimal tour within 509 iterations. The ratio of the distance of the near optimal tour to that of an optimal tour is equal to 1.05, as shown in Fig.2.

IV. Concluding Remarks

The minimization of a combinatorial cost is turned into a process of learning a simple distribution to approximate the Gibbs distribution induced from this cost such that the Gibbs distribution and the resulted approximation distribution give a same global solution. From this new perspective, the treatment of simply regarding a binary v_{ij} as an analog variable, intuitively used in all the existing analog optimization approaches, can be interpreted and justified. Particularly, the Lagrange-Barrier iterative procedure is recommended because it always gives a guaranteed convergence on a feasible solution that satisfies constraints. Moreover, it has been shown via experiments in [3], [4] that this Lagrange-Barrier iterative procedure is significantly superior to the modified soft-assign algorithm [15], [16] in computational time, with the quality of solutions being similarly or even slightly better.

References

[1] J. van den Berg (1996), "Neural Relaxation Dynamics", PhD Thesis, Erasmus University of Rotterdam, The Netherlands.

[2] A. Cichocki and R. Unbehaunen (1993) *Neural Networks for Optimization and Signal processing*, John Wiley & Sons, New York.

[3] C. Dang and L. Xu, (2001a), "A globally convergent Lagrange and barrier function iterative algorithm for the traveling salesman problem", *Neural Networks*, Vol.14, No.2, pp217-230, 2001.

[4] C. Dang and L. Xu (2001b), "A Lagrange Multiplier and Hopfield-Type Barrier Function Method for the Traveling Salesman Problem", *Neural Computation*, Vol. 14 , No. 2 , pp303 - 324 .

[5] J. Eriksson (1980), "A note on solution of large sparse maximum entropy problems with linear equality constraints", *Mathematical Programming 18*, 146-154.

[6] S. Erlander (1981), "Entropy in linear programs", *Mathematical Programming 21*, 137-151.

[7] S. C. Fang and H.-S. J. Tsao (1995), "Linearly-constrained entropy maximization problem with quadratic cost and its applications to transportation planning problems", *Transportation Science 29*, 353-365.

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

[9] R. Horst and P. M. Pardalos (1995), *Handbook of Global Optimization, Nonconvex Optimization and its Applications 2*, Kluwer Academic Publishers.

[10] B. Kalantari and J. B. Rosen (1987), " An algorithm for global minimization of linearly constrained concave quadratic functions", *Mathematics of Operations Research 12*, 544-561.

[11] S.Kirkpatrick, C.G.Gelatt Jr., and M.P. Vecchi (1983), "Optimization by Simulated Annealing", *Science 220*, pp671-680.

[12] N. Metroplois, et al, (1953), "Equation of State Calculations for fast Computing machines", *Journal of Chemical Physics 21*, 1087-1092.

[13] P. M. Pardalos and J. B. Rosen (1987). "Constrained Global Optimization: Algorithms and Applications", *Lecture Notes in Computer Science 268*, Springer-Varlag.

[14] E. Wacholder, J. Han and R. C. Mann (1989). "A neural network algorithm for the traveling salesman problem", *Biological Cybernetics 61*, 11-19.

[15] Rangarajan, A., Gold, S., & Mjolsness, E. (1996). "A novel optimizing network architecture with applications". *Neural Computation, 8*, 1041-1060.

[16] Rangarajan, A., Yuille, A., & Mjolsness, E. (1999). "Convergence properties of the softassign quadratic assignment algorithm", *Neural Computation, 11*, 1455- 1474.

[17] L. Xu (1994), "Combinatorial optimization neural nets based on a hybrid of Lagrange and transformation approaches", *Proc. of World Congress on Neural Networks*, San Diego, 399-404.

[18] L. Xu (1995a). "On the hybrid LT combinatorial optimization: new U-shape barrier, sigmoid activation, least leaking energy and maximum entropy", *Proc. of Intl. Conf. on Neural Information Processing (ICONIP'95)*, Beijing, 309-312.

[19] L. Xu. (1995b), "YING-YANG Machine: a Bayesian-Kullback scheme for unified learnings and new results on vector quantization", *Proc. of Intl. Conf. on Neural Information Processing (ICONIP'95)*, 977-988.

[20] A. L. Yuille & J. J. Kosowsky (1994), "Statistical physics algorithms that converge", *Neural Computation 6*, 341-356.