PERGAMON

Neural Networks 14 (2001) 217–230

Neural Networks

www.elsevier.com/locate/neunet

Contributed article

A globally convergent Lagrange and barrier function iterative algorithm for the traveling salesman problem

Chuangyin Dang[a,*], Lei Xu[b]

[a]*Department of Manufacturing Engineering & Engineering Management, City University of Hong Kong, Hong Kong*
[b]*Department of Computer Science & Engineering, Chinese University of Hong Kong, Hong Kong*

Received 27 January 1998; revised 19 October 2000; accepted 19 October 2000

Abstract

In this paper a globally convergent Lagrange and barrier function iterative algorithm is proposed for approximating a solution of the traveling salesman problem. The algorithm employs an entropy-type barrier function to deal with nonnegativity constraints and Lagrange multipliers to handle linear equality constraints, and attempts to produce a solution of high quality by generating a minimum point of a barrier problem for a sequence of descending values of the barrier parameter. For any given value of the barrier parameter, the algorithm searches for a minimum point of the barrier problem in a feasible descent direction, which has a desired property that the nonnegativity constraints are always satisfied automatically if the step length is a number between zero and one. At each iteration the feasible descent direction is found by updating Lagrange multipliers with a globally convergent iterative procedure. For any given value of the barrier parameter, the algorithm converges to a stationary point of the barrier problem without any condition on the objective function. Theoretical and numerical results show that the algorithm seems more effective and efficient than the softassign algorithm.

Keywords: Traveling salesman problem; Lagrange multiplier; Entropy-type barrier function; Descent direction; Iterative algorithm

1. Introduction

The traveling salesman problem (TSP) is an NP hard combinatorial optimization problem and has a variety of important applications. In order to solve it, several classic algorithms and heuristics have been proposed. An excellent survey of techniques for solving the TSP can be found in Lawler, Lenstra, Rinnoy Kan and Shmoys (1985).

In Hopfield and Tank (1985), the first combinatorial optimization neural network was proposed, which minimized an energy function in quadratic form and solves a system of ordinary differential equations. Since then, many combinatorial optimization neural networks have been developed. One of them we would like to mention here is an elastic network combinatorial optimization algorithm given by Durbin and Willshaw (1987). An extension of the neural network algorithm to solving the multiple TSP can be found in Wacholder, Han and Mann (1989). A systematic investigation of combinatorial optimization neural networks was carried out in van den Berg (1996). Some other optimization neural networks were studied in Cichocki and Unbehaunen (1993).

Instead of solving a system of ordinary differential equations, a Lagrange and barrier function iterative algorithm was proposed in Xu (1994) for combinatorial optimization problems of assignment type. It treats linear equality constraints with Lagrange multipliers and nonnegativity constraints with an entropy-type barrier function, respectively. Although the separate treatments of the linear equality constraints and the nonnegativity constraints with Lagrange multipliers and a barrier function can also be found in van den Berg (1996) and Fang and Tsao (1995), the algorithm (Xu, 1994) bears an interesting feature of the alternative minimization iterative procedure. Firstly, an iterative formula was proposed to generate an interior point within binary bounds. The interior point can be interpreted as the expectation of a binary distribution implicitly specified by the value of a Lagrange and barrier function, which is related to the statistical physics algorithms for optimization given by Yuille and Kosowsky (1994). Secondly, at the interior point, another iterative formula was proposed to obtain Lagrange multipliers that satisfy a system of special nonlinear equations induced from the linear equality constraints. It was shown experimentally in Lau, Chan and Xu (1995) that the algorithm (Xu, 1994) is frequently superior to the algorithm (Hopfield & Tank, 1985) with a doubled convergence speed and a higher rate of finding valid and better quality solutions. However,

* Corresponding author. Tel.: +852-2788-8429; fax: +852-2788-8423.
E-mail address: mecdang@cityu.edu.hk (C. Dang).

0893-6080/01/$ - see front matter
PII: S0893-6080(00)00092-7

whether the algorithm (Xu, 1994) converges still remains unknown. In Rangarajan, Gold and Mjolsness (1996), a softassign algorithm based on Sinkhorn's formula for updating Lagrange multipliers was proposed for the combinatorial optimization problems of assignment type. The softassign algorithm is the same as Xu's algorithm except that the objective function in Rangarajan et al. (1996) has an additional negative quadratic term. Under the assumption that the objective function is strictly concave on the null space of the constraint matrix, it was proved in Rangarajan, Yuille and Mjolsness (1999) that for any given value of the barrier parameter, the softassign algorithm converges to a stationary point of a barrier problem.

In this paper we propose a globally convergent Lagrange and barrier function iterative algorithm for approximating a solution of the TSP. The algorithm employs an entropy-type barrier function to deal with nonnegativity constraints and Lagrange multipliers to handle linear equality constraints, and attempts to produce a solution of high quality by generating a minimum point of a barrier problem for a sequence of descending values of the barrier parameter. For any given value of the barrier parameter, the algorithm searches for a minimum point of the barrier problem in a feasible descent direction, which has a desired property that the nonnegativity constraints are always satisfied automatically if the step length is a number between zero and one. At each iteration the feasible descent direction is found by updating Lagrange multipliers with a globally convergent iterative procedure. For any given value of the barrier parameter, the algorithm converges to a stationary point of the barrier problem without any condition on the objective function. Theoretical and numerical results show that the algorithm seems more effective and efficient than the softassign algorithm.

The rest of this paper is organized as follows. We introduce the entropy-type barrier function and derive some important properties in Section 2. We present the algorithm and show its convergence to a stationary point of the barrier problem for any given value of the barrier parameter in Section 3. We prove global convergence of the iterative procedure for updating Lagrange multipliers to find a feasible descent direction in Section 4. We report some numerical results in Section 5. We conclude the paper with some remarks in Section 6.

2. Entropy-type barrier function

Given n cities, we consider the problem of finding a tour such that each city is visited exactly once and that the total distance traveled is minimized. Let $v_{ik} = 1$ if city i is the kth city to be visited in a tour, 0 otherwise, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, n$, and

$$v = (v_{11}, v_{12}, \ldots, v_{1n}, \ldots, v_{n1}, \ldots, v_{n2}, \ldots, v_{nn})^{\mathrm{T}}.$$

In Hopfield and Tank (1985), the problem was formulated as

$$\begin{aligned} &\min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{ij} v_{ik} v_{j,k+1} \\ &\text{subject to } \sum_{j=1}^{n} v_{ij} = 1, i = 1, 2, \ldots, n, \qquad \sum_{i=1}^{n} v_{ij} = 1, \\ &j = 1, 2, \ldots, n, \qquad v_{ij} \in \{0, 1\}, \qquad i = 1, 2, \ldots, n, \\ &j = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where d_{ij} denotes the distance from city i to city j, and $v_{j,k+1} = v_{j1}$ for $k = n$. Clearly, for any given $\rho \geqslant 0$, (1) is equivalent to

$$\begin{aligned} &\min e_0(v) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n} d_{ij} v_{ik} v_{j,k+1} - \frac{1}{2}\rho v_{ij}^2\right) \qquad \text{subject to} \\ &\sum_{j=1}^{n} v_{ij} = 1, \qquad i = 1, 2, \ldots, n, \qquad \sum_{i=1}^{n} v_{ij} = 1, \\ &j = 1, 2, \ldots, n, \qquad v_{ij} \in \{0, 1\}, \qquad i = 1, 2, \ldots, n, \\ &j = 1, 2, \ldots, n, \end{aligned} \tag{2}$$

where the negative quadratic term was employed in the energy function given by Rangarajan et al. (1996). The continuous relaxation of (2) yields

$$\begin{aligned} &\min e_0(v) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n} d_{ij} v_{ik} v_{j,k+1} - \frac{1}{2}\rho_{ij}^2\right) \qquad \text{subject to} \\ &\sum_{j=1}^{n} v_{ij} = 1, \qquad i = 1, 2, \ldots, n, \qquad \sum_{i=1}^{n} v_{ij} = 1, \\ &j = 1, 2, \ldots, n, \qquad 0 \leq v_{ij}, \qquad i = 1, 2, \ldots, n, \\ &j = 1, 2, \ldots, n. \end{aligned} \tag{3}$$

When ρ is sufficiently large, one can see that an optimal solution of (3) is an integer solution. Thus, when ρ is sufficiently large, (3) is equivalent to (1). We remark that the size of ρ affects the quality of a solution produced by a deterministic annealing algorithm and it should be as small as possible.

Following Xu (1994), we introduce an entropy-type barrier term,

$$d(v_{ij}) = \int_0^{v_{ij}} \ln t \mathrm{d}t = v_{ij} \ln v_{ij} - v_{ij},$$

to incorporate $0 \leqslant x_{ij}$ into the objective function of (3), and obtain

$$\min e(v;\beta) = e_0(v) + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} d(v_{ij}) \qquad \text{subject to}$$

$$\sum_{j=1}^{n} v_{ij} = 1, \qquad i = 1,2,\ldots,n, \qquad \sum_{i=1}^{n} v_{ij} = 1, \tag{4}$$

$$j = 1,2,\ldots,n,$$

where β is a positive barrier parameter. Note that the barrier term can also be found in Eriksson (1980) and Erlander (1981). Instead of solving (3) directly, we consider a scheme, which obtains a solution of (3) from the solution of (4) at the limit of $\beta \downarrow 0$.

Let

$$b(v) = \sum_{i=1}^{n}\sum_{j=1}^{n} d(v_{ij}).$$

Then, $e(v;\beta) = e_0(v) + \beta b(v)$. Let

$$P = \left\{ v \in R^{n^2} \,\middle|\, \begin{array}{lll} \sum_{j=1}^{n} v_{ij} = 1, & i = 1,2,\ldots,n, & \\ \sum_{i=1}^{n} v_{ij} = 1, & j = 1,2,\ldots,n, & \\ 0 \le v_{ij}, & i = 1,2,\ldots,n, & j = 1,2,\ldots,n \end{array} \right\}.$$

Then, P is the feasible region of (3) and bounded. Let us define $d(0) = 0$. Since $\lim_{v_{ij}\to 0^+} d(v_{ij}) = 0$, hence, $b(v)$ is continuous on $R_+^{n^2} = \{v \in R^{n^2} | 0 \le v\}$. From $b(v)$, we obtain $\partial b(v)/\partial v_{ij} = \ln v_{ij}$ and $\lim_{v_{ij}\to 0^+}(\partial b(v)/\partial v_{ij}) = -\infty$. From $e_0(v)$ we get

$$\frac{\partial e_0(v)}{\partial v_{ij}} = \sum_{k=1}^{n} (d_{ki} v_{k,j-1} + d_{ik} v_{k,j+1}) - \rho v_{ij},$$

where $v_{k,j-1} = v_{kn}$ for $j = 1$, and $v_{k,j+1} = v_{k1}$ for $j = n$. Clearly, $\partial e_0(v)/\partial v_{ij}$ is bounded on P. Due to $\partial e(v;\beta)/\partial v_{ij} = \partial e_0(v)/\partial v_{ij} + \beta \partial b(v)/\partial v_{ij}$, we have $\lim_{v_{ij}\to 0^+} \partial e(v;\beta)/\partial v_{ij} = -\infty$.

Lemma 1. For any given $\beta > 0$, if v^* is a local minimum point of (4), v^* is an interior point of P, i.e. $0 < v_{ij}^*$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,n$.

Proof. Let v^0 be an interior point of P. Suppose that some component of v^*, say v_{ij}^*, equals 0. For any given number $\epsilon \in (0, 1]$, let $y^* = v^* + \epsilon(v^0 - v^*)$. Then y^* is an interior point of P. For any given $\delta \in (0, 1]$ satisfying $\epsilon + \delta \leqslant 1$, let $z^* = y^* + \delta(v^0 - v^*) = v^* + (\epsilon + \delta)(v^0 - v^*)$. Then z^* is an interior point of P, which can be made arbitrarily close to v^* through decreasing $\epsilon + \delta$. From the Taylor's expansion, we obtain

$$e(z^*;\beta) = e(y^*;\beta) + \delta(v^0 - v^*)^{\mathrm{T}} \nabla_v e(y^* + \eta\delta(v^0 - v^*);\beta), \tag{5}$$

where $\eta \in [0, 1]$ and $\nabla_v e(y^* + \eta\delta(v^0 - v^*);\beta)$ is the gradient of $e(v;\beta)$ at $v = y^* + \eta\delta(v^0 - v^*)$. Consider

$$(v^0 - v^*)^{\mathrm{T}} \nabla_v e(y^* + \eta\delta(v^0 - v^*);\beta)$$

$$= \sum_{k=1}^{n}\sum_{l=1}^{n} (v_{kl}^0 - v_{kl}^*) \frac{\partial e(y^* + \eta\delta(v^0 - v^*);\beta)}{\partial v_{kl}}.$$

Let $\theta = \epsilon + \eta\delta$. Then,

$$y^* + \eta\delta(v^0 - v^*) = v^* + (\epsilon + \eta\delta)(v^0 - v^*)$$

$$= v^* + \theta(v^0 - v^*).$$

1. If $v_{kl}^* = 0$, then $v_{kl}^0 - v_{kl}^* > 0$ and

$$\lim_{\theta\to 0} \frac{\partial e(y^* + \eta\delta(v^0 - v^*);\beta)}{\partial v_{kl}}$$

$$= \lim_{\theta\to 0} \frac{\partial e_0(y^* + \eta\delta(v^0 - v^*))}{\partial v_{kl}} + \beta \ln(\theta(v_{kl}^0 - v_{kl}^*)) = -\infty.$$

2. If $0 < v_{kl}^*$, then $\lim_{\theta\to 0} \dfrac{\partial e(y^* + \eta\delta(v^0 - v^*);\beta)}{\partial v_{kl}}$ is bounded.
   Since there is a component of v^*, v_{ij}^*, satisfying $v_{ij}^* = 0$, the above results imply

$$\lim_{\theta\to 0} (v^0 - v^*)^{\mathrm{T}} \nabla_v e(y^* + \eta\delta(v^0 - v^*);\beta) = -\infty.$$

Thus, when ϵ and δ are sufficiently small, from (5) we obtain

$$e(z^*;\beta) < e(y^*;\beta)$$

since $(v^0 - v^*)^{\mathrm{T}} \nabla_v e(y^* + \eta\delta(v^0 - v^*);\beta) < 0$. Therefore, using $\lim_{\epsilon\to 0} e(y^*;\beta) = e(v^*;\beta)$, we get $e(z^*;\beta) < e(v^*;\beta)$ when ϵ and δ are sufficiently small. It contradicts that v^* is a local minimum point of (4). Hence, no component of v^* equals 0. The lemma follows.□

Let

$$L(v, \lambda^r, \lambda^c) = e(v;\beta) + \sum_{i=1}^{n} \lambda_i^r \left(\sum_{j=1}^{n} v_{ij} - 1 \right) + \sum_{j=1}^{n} \lambda_j^c \left(\sum_{i=1}^{n} v_{ij} - 1 \right).$$

Lemma 1 indicates that if v^* is a local minimum point of (4) then there exist λ^{r*} and λ^{c*} satisfying

$$\nabla_v L(v^*, \lambda^{r^*}, \lambda^{c^*}) = 0, \qquad \sum_{j=1}^{n} v_{ij}^* = 1, \qquad i = 1,2,\ldots,n,$$

$$\sum_{i=1}^{n} v_{ij}^* = 1, \qquad j = 1,2\ldots,n,$$

where

$$\nabla_v L(v,\lambda^r,\lambda^c)$$

$$=\left(\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{11}},\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{12}},\dots,\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{1n}},\right.$$

$$\left.\dots,\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{n1}},\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{n2}},\dots,\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{nn}}\right)^{\mathrm{T}}$$

with

$$\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{ij}}=\frac{\partial e_0(v)}{\partial v_{ij}}+\lambda_i^r+\lambda_j^c+\beta\ln v_{ij},$$

$i=1,2,\dots,n,\qquad j=1,2,\dots,n.$

Let β_k, $k=1,2,\dots,$ be a sequence of positive numbers satisfying $\beta_1>\beta_2>\cdots$ and $\lim_{k\to\infty}\beta_k=0$. For $k=1,2,\dots,$ let $v(\beta_k)$ denote a global minimum point of (4) with $\beta=\beta_k$. Following a standard argument (Minoux, 1986), one can readily obtain the next therorem.

Theorem 1. $e_0(v(\beta_k))\geq e_0(v(\beta_{k+1}))$, $k=1,2,\dots,$ and every limit point of $v(\beta_k)$, $k=1,2,\dots,$ is a global minimum point of (3).

This theorem indicates that a global minimum point of (3) can be obtained if we are able to generate a global minimum point of (4) for a sequence of descending values of the barrier parameter with zero limit.

Theorem 2. For $k=1,2,\dots,$ let v^k be a local minimum point of (4) with $\beta=\beta_k$. For any limit point v^* of v^k, $k=1,2,\dots,$ if there are no $\lambda^r=(\lambda_1^r,\lambda_2^r,\dots,\lambda_n^r)^{\mathrm{T}}$ and $\lambda^c=(\lambda_1^c,\lambda_2^c,\dots,\lambda_n^c)^{\mathrm{T}}$satisfying

$$\frac{\partial e_0(v^*)}{\partial v_{ij}}+\lambda_i^r+\lambda_j^c=0,$$

$i=1,2,\dots,n$, $j=1,2,\dots,n$, then v^* is a local minimum point of (3).

Proof. Since v^k, $k=1,2,\dots,$ are contained in the bounded set P, we can extract a convergent subsequence. Let v^{k_q}, $q=1,2,\dots,$ be a convergent subsequence of v^k, $k=1,2,\dots.$ Assume $\lim_{q\to\infty}v^{k_q}=v^*$.

Since v^{k_q} is a local minimum point of (4) with $\beta=\beta_{k_q}$, using Lemma 1 and the first-order necessary optimality condition, we obtain that there are $\lambda^{r,k_q}=(\lambda_1^{r,k_q},\lambda_2^{r,k_q},\dots,\lambda_n^{r,k_q})^{\mathrm{T}}$ and $\lambda^{c,k_q}=(\lambda_1^{c,k_q},\lambda_2^{c,k_q},\dots,\lambda_n^{c,k_q})^{\mathrm{T}}$ satisfying

$$\frac{\partial e_0(v^{k_q})}{\partial v_{ij}}+\lambda_i^{r,k_q}+\lambda_j^{c,k_q}+\beta_{k_q}\ln v_{ij}^{k_q}=0,$$

$i=1,2,\dots,n, j=1,2,\dots,n$. Thus,

$$\frac{\partial e_0(v^*)}{\partial v_{ij}}=\lim_{q\to 0}\frac{\partial e_0(v^{k_q})}{\partial v_{ij}}=-\lim_{q\to\infty}(\lambda_i^{r,k_q}+\lambda_j^{c,k_q}+\beta_{k_q}\ln v_{ij}^{k_q}),\tag{6}$$

$i=1,2,\dots,n$, $j=1,2,\dots,n$. Let v be an interior point of P. Then

$$\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})\frac{\partial e_0(v^{k_q})}{\partial v_{ij}}$$

$$=-\left(\sum_{i=1}^n\lambda_i^{r,k_q}\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})+\sum_{j=1}^n\lambda_j^{c,k_q}\sum_{i=1}^n(v_{ij}-v_{ij}^{k_q})\right.$$

$$\left.+\beta_{k_q}\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})\ln v_{ij}^{k_q}\right)$$

$$=-\beta_{k_q}\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})\ln v_{ij}^{k_q}.$$

Let $K=\{(i,j)|v_{ij}^*=0\}$. Then, for any $(i,j)\notin K$,

$$\lim_{q\to\infty}\beta_{k_q}(v_{ij}-v_{ij}^{k_q})\mathrm{In}v_{ij}^{k_q}=0.$$

Consider $(i,j)\in K$. We have $v_{ij}-v_{ij}^*>0$ and $\lim_{q\to\infty}v_{ij}^{k_q}=0$. Then, when q is sufficiently large,

$$\beta_{k_q}(v_{ij}-v_{ij}^{k_q})\ln v_{ij}^{k_q}<0.$$

From (6) and the assumption, we obtain that $K\neq 0$ and at least one of

$$\lim_{q\to\infty}\beta_{k_q}\ln v_{ij}^{k_q},(i,j)\in K,$$

is not equal to zero. Thus, at least one of

$$(v_{ij}-v_{ij}^*)\lim_{q\to\infty}\beta_{k_q}\ln v_{ij}^{k_q},(i,j)\in K,$$

is negative, and all of them are not positive. Therefore,

$$\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^*)\frac{\partial e_0(v^*)}{\partial v_{ij}}$$

$$=\lim_{q\to\infty}\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})\frac{\partial e_0(v^{k_q})}{\partial v_{ij}}$$

$$=-\lim_{q\to\infty}\beta_{k_q}\sum_{i=1}^n\sum_{j=1}^n(v_{ij}-v_{ij}^{k_q})\ln v_{ij}^{k_q}\tag{7}$$

$$=-\lim_{q\to\infty}\beta_{k_q}\sum_{(i,j)\in K}(v_{ij}-v_{ij}^{k_q})\ln v_{ij}^{k_q}$$

$$=-\sum_{(i,j)\in K}(v_j-v_{ij}^*)\lim_{q\to\infty}\beta_{k_q}\ln v_{ij}^{k_q}\qquad>0.$$

Observe that $e_0(v)$ is a quadratic function and can be rewritten in a matrix form as $e_0(v)=(1/2)v^{\mathrm{T}}Qv$. From this matrix form, we get $\nabla e_0(v) = Qv$ and

$$e_0(v) - e_0(v^*) = \frac{1}{2} v^{\mathrm{T}} Q v - \frac{1}{2} (v^*)^{\mathrm{T}} Q v^*$$

$$= (v - v^*)^{\mathrm{T}} Q v^* + \frac{1}{2}(v - v^*)^{\mathrm{T}} Q (v - v^*).$$

Then, when v is an interior point of P sufficiently close to v^*, using (7), we obtain

$$e_0(v) - e_0(v^*) > 0$$

since $(v - v^*)^{\mathrm{T}} Q v^* = \sum_{i=1}^{n} \sum_{j=1}^{n} (v_{ij} - v_{ij}^*)(\partial e_0(v^*)/\partial v_{ij}) > 0$ and $(1/2)(v - v^*)^{\mathrm{T}} Q (v - v^*)$ goes to zero twice as fast as $(v - v^*)^{\mathrm{T}} Q v^*$ if v approaches v^*. It implies that v^* is a local minimum point of (3). The theorem follows.□

This theorem indicates that at least a local minimum point of (3) can be obtained if we are able to generate a local minimum point of (4) for a sequence of descending values of the barrier parameter with zero limit.

3. The algorithm

In this section we develop an algorithm for approximating a solution of (3). Given any $\beta > 0$, consider the first-order necessary optimality condition for (4),

$$\nabla_v L(v, \lambda^r, \lambda^c) = 0, \qquad \sum_{j=1}^{n} v_{ij} = 1, \qquad i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} v_{ij} = 1, \qquad j = 1, 2, \ldots, n.$$

From

$$\frac{\partial L(v, \lambda^r, \lambda^c)}{\partial v_{ij}} = \frac{\partial e_0(v)}{\partial v_{ij}} + \lambda_i^r + \lambda_j^c + \beta \ln v_{ij} = 0,$$

we obtain

$$v_{ij} = \frac{1}{\exp\left((\frac{\partial e_0(v)}{\partial v_{ij}} + \lambda_i^r + \lambda_j^c)/\beta \right)}.$$

Let $r_i = \exp(\lambda_i^r/\beta)$ and $c_j = \exp(\lambda_j^c/\beta)$. Then,

$$v_{ij} = \frac{1}{r_i c_j \exp\left(\frac{\partial e_0(v)}{\partial v_{ij}} / \beta \right)}.$$

For convenience of the following discussions, let

$$\alpha_{ij}(v) = \exp\left(\frac{\partial e_0(v)}{\partial v_{ij}} / \beta \right).$$

Then,

$$v_{ij} = \frac{1}{r_i c_j \alpha_{ij}(v)}. \tag{8}$$

Substituting (8) into $\sum_{j=1}^{n} v_{ij} = 1, i = 1, 2, \ldots, n,$ and $\sum_{i=1}^{n} v_{ij} = 1, j = 1, 2, \ldots, n$, we obtain

$$\begin{aligned} &\sum_{j=1}^{n} \frac{1}{r_i c_j \alpha_{ij}(v)} = 1, i = 1, 2, \ldots, n, \\ &\sum_{i=1}^{n} \frac{1}{r_i c_j \alpha_{ij}(v)} = 1, j = 1, 2, \ldots, n. \end{aligned} \tag{9}$$

Based on the above notations, an algorithm was proposed by Xu (1994) for approximating a solution of (3) without the negative quadratic term.

Let

$$h_{ij}(v, r, c) = \frac{1}{r_i c_j \alpha_{ij}(v)}, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n,$$

and

$$h(v, r, c) = (h_{11}(v, r, c), h_{12}(v, r, c),$$
$$\ldots, h_{1n}(v, r, c), \ldots, h_{n1}(v, r, c), h_{n2}(v, r, c), \ldots, h_{nn}(v, r, c))^{\mathrm{T}}.$$

When $v > 0$, the next lemma shows that $h(v, r, c,) - v$ is a descent direction of $L(v, \lambda^r, \lambda^v)$.

Lemma 2. Assume $0 < v$.

1. $\dfrac{\partial L(v, \lambda^r, \lambda^v)}{\partial v_{ij}} > 0 \quad$ if $\quad h_{ij}(v, r, c) - v_{ij} < 0.$
2. $\dfrac{\partial L(v, \lambda^r, \lambda^v)}{\partial v_{ij}} < 0 \quad$ if $\quad h_{ij}(v, r, c) - v_{ij} > 0.$
3. $\dfrac{\partial L(v, \lambda^r, \lambda^v)}{\partial v_{ij}} = 0 \quad$ if $\quad h_{ij}(v, r, c) - v_{ij} = 0.$
4. $(h(v, r, c) - v)^{\mathrm{T}} \nabla_v L(v, \lambda^r, \lambda^c) < 0$ if $h(v, r, c) - v \neq 0.$
5. $(h(v, r, c) - v)^{T} \nabla_v e(v; \beta) < 0$ if $h(v, r, c) - v \neq 0$ and $\sum_{k=1}^{n} (h_{ik}(v, r, c) - v_{ik}) = \sum_{k=1}^{n} (h_{kj}(v, r, c) - v_{kj}) = 0$, $i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n.$

Proof. We only need to show that $\partial L(v, \lambda^r, \lambda^v)/\partial v_{ij} > 0$ if $h_{ij}(v, r, c) - v_{ij} < 0$. The rest can be obtained similarly or straightforward. From

$$h_{ij}(v, r, c) - v_{ij} = \frac{1}{r_i c_j \alpha_{ij}(v)} - v_{ij} < 0,$$

we obtain

$$1 < r_i c_j \alpha_{ij}(v) v_{ij}. \tag{10}$$

Applying the natural logarithm, ln, to both sides of (10), we get

$$0 < \ln(r_i c_j \alpha_{ij}(v) v_{ij}) \qquad = \ln\alpha_{ij}(v) + \ln r_i + \ln c_j + \ln v_{ij}$$

$$= \frac{1}{\beta}\frac{\partial e_0(v)}{\partial v_{ij}} + \frac{1}{\beta}\lambda_i^r + \frac{1}{\beta}\lambda_j^c + \ln v_{ij} \qquad = \frac{1}{\beta}\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{ij}}.$$

Thus,

$$\frac{\partial L(v,\lambda^r,\lambda^c)}{\partial v_{ij}} > 0.$$

The lemma follows.□

Since $0 < h_{ij}(v,r,c)$, we remark that the descent direction $h(v,r,c) - v$ has a desired property that any point generated along $h(v,r,c) - v$ is always positive automatically if $v > 0$ and the step length is a number between zero and one.

For any given point v, we use $(r(v), c(v))$ to denote a positive solution of (9). Let v be an interior point of P. In order for $h(v,r,c) - v$ to become a feasible descent direction of (4), we need to compute a positive solution $(r(v), c(v))$ of (9). Let

$$f(r,c) = \frac{1}{2}\left(\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{1}{r_i c_j \alpha_{ij}(v)} - 1\right)^2 + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{1}{r_i c_j \alpha_{ij}(v)} - 1\right)^2\right).$$

Then, $f(r,c)$ equals zero only at a solution of (9). For $i = 1, 2, \ldots, n$, let

$$x_i(r,c) = r_i\left(\sum_{j=1}^{n}\frac{1}{r_i c_j \alpha_{ij}(v)} - 1\right),$$

and for $j = 1, 2, \ldots, n$, let

$$y_j(r,c) = c_j\left(\sum_{i=1}^{n}\frac{1}{r_i c_j \alpha_{ij}(v)} - 1\right).$$

Let

$$x(r,c) = (x_1(r,c), x_2(r,c), \ldots, x_n(r,c))^{\mathrm{T}}$$

and

$$y(r,c) = (y_1(r,c), y_2(r,c), \ldots, y_n(r,c))^{\mathrm{T}}.$$

It is proved in the next section that

$$\begin{pmatrix} x(r,c) \\ y(r,c) \end{pmatrix}$$

is a descent direction of $f(r,c)$. For any given v, based on this descent direction, the following iterative procedure is proposed for computing a positive solution $(r(v), c(v))$ of (9).

Take (r^0, c^0) to be an arbitrary positive vector, and for $k = 0, 1, \ldots$, let

$$r^{k+1} = r^k + \mu_k x(r^k, c^k), \qquad c^{k+1} = c^k + \mu_k y(r^k, c^k), \quad (11)$$

where μ_k is a number in [0, 1] satisfying

$$f(r^{k+1}, c^{k+1}) = \min_{\mu\in[0,1]} f(r^k + \mu_k x(r^k,c^k), c^k + \mu_k y(r^k,c^k)).$$

Clearly, $(r^k, c^k) > 0$, $k = 0, 1, \ldots$. There are many ways to determine μ_k (Minoux, 1986). For example, one can simply choose μ_k to be any number in (0, 1] satisfying $\sum_{l=0}^{k}\mu \to \infty$ and $\mu_k \to 0$ as $k \to 0$. We have found in our numerical tests that when μ_k is any fixed number in (0, 1], the iterative procedure (11) converges to a positive solution of (9). Global convergence of the iterative procedure (11) will be given in the next section.

Based on the feasible descent direction, $h(v, r(v), c(v)) - v$, and the iterative procedure (11), we have developed an algorithm for approximating a solution of (3), which is as follows.

Step 0: Let $\epsilon > 0$ be a given tolerance. Let β_0 be a sufficiently large positive number satisfying that $e(v;\beta_0)$ is convex. Choose an arbitrary point $\bar{v}$ satisfying $0 < \bar{v}_{ij} < 1, i = 1, 2, \ldots, n, j = 1, 2\ldots, n$, and two arbitrary positive vectors, r^0 and c^0. Take an arbitrary positive number $\eta \in (0, 1)$ (in general, η should be close to one). Given $v = \bar{v}$, use (11) to obtain a positive solution $(r(\bar{v}), c(\bar{v}))$ of (9). Let $r^0 = r(\bar{v})$ and $c^0 = c(\bar{v})$. Let

$$v^0 = (v_{11}^0, v_{12}^0, \ldots, v_{1n}^0, \ldots, v_{n1}^0, v_{n2}^0, \ldots, v_{nn}^0)^{\mathrm{T}}$$

with

$$v_{ij}^0 = \frac{1}{r_i(\bar{v})c_j(\bar{v})\alpha_{ij}(\bar{v})},$$

$i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n$. Let $q = 0$ and $k = 0$. Go to Step 1.

Step 1: Given $v = v^k$, use (11) to obtain a positive solution $(r(v^k), c(v^k))$ of (9). Let $r^0 = r(v^k)$ and $c^0 = c(v^k)$. Go to Step 2.

Step 2: Let $h(v^k, r(v^k), c(v^k)) = (h_{11}(v^k, r(v^k), c(v^k)), h_{12}(v^k, r(v^k), c(v^k)), \ldots, h_{1n}(v^k, r(v^k), c(v^k)), \ldots, h_{n1}(v^k, r(v^k), c(v^k)), h_{n2}(v^k, r(v^k), c(v^k)), \ldots, h_{nn}(v^k, r(v^k), c(v^k)))^{\mathrm{T}}$ with

$$h_{ij}(v^k, r(v^k), c(v^k)) = \frac{1}{r_i(v^k)c_j(v^k)\alpha_{ij}(v^k)},$$

$i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n$. If $\|h(v^k, r(v^k),\ c(v^k)) - v^k\| < \epsilon$, do as follows:

- If β_q is sufficiently small, the algorithm terminates.
- Otherwise, let $v^{*,q} = v^k$, $v^0 = v^k$, $\beta_{q+1} = \eta\beta_q, q = q + 1$, and $k = 0$. Go to Step 1.

If $\|h(v^k, r(v^k), c(v^k)) - v^k\| \geq \epsilon$, do as follows: Compute

$$v^{k+1} = v^k + \theta_k(h(v^k, r(v^k), c(v^k)) - v^k), \qquad (12)$$

where θ_k is a number in [0, 1] satisfying

$$e(v^{k+1};\beta_q) = \min_{\theta\in[0,1]} e(v^k + \theta(h(v^k, r(v^k), c(v^k)) - v^k);\beta_q).$$

Let $k = k + 1$ and go to Step 1.

We remark that an exact positive solution $(r(v^k), c(v^k))$ of (9) for $v = v^k$ and an exact solution of $\min_{\theta\in[0,1]} e(v^k + \theta(h(v^k, r(v^k), c(v^k)) - v^k); \beta_q)$ are not required in the implementation of the algorithm, and their approximate solutions will do. There are many ways to determine θ_k (Minoux, 1986). For example, one can simply choose θ_k to be any number in (0, 1] satisfying $\sum_{i=0}^{k} \theta_i \to \infty$ and $\theta_k \to 0$ as $k \to \infty$. In our implementation of the algorithm, θ_k is determined with the Armijo-type search. Since $e(v; \beta_0)$ is convex, hence, the algorithm is insensitive to the starting point.

Theorem 3. For $\beta = \beta_q$, every limit point of v^k, $k = 0, 1, \ldots$, generated by (12) is a stationary point of (4).

Proof. Let $a_{\max} = \max_{1\le i\le n, 1\le j\le n} \max_{v\in P} \alpha_{ij}(v) r_i(v) c_j(v)$. Since $\{\partial e_0(v)/\partial v_{ij} | v \in P\}$ and $\{r_i(v)c_j(v) | v \in P\}$ are bounded, hence, $a_{\max}$ is finite. Let

$$v^{\min} = (v_{11}^{\min}, v_{12}^{\min}, \ldots, v_{1n}^{\min}, \ldots, v_{n1}^{\min}, v_{n2}^{\min}, \ldots, v_{nn}^{\min})^{\mathrm{T}}$$

with $v_{ij}^{\min} = 1/a_{\max}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. Then, for any $v \in P$, $0 < v_{ij}^{\min} \le h_{ij}(v, r(v), c(v))$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, and $v^{\min} \le v^k$, $k = 0, 1, \ldots$. Therefore, no limit of v_{ij}^k, $k = 0, 1, \ldots$, is equal to 0 for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. From Lemma 2, we obtain that $h(v^k, r(v^k), c(v^k)) - v^k$ is a feasible descent direction of (4).

Let $X = \{v \in P | v^{\min} \le v\}$ and

$$\Omega = \{v \in X | h(v, r(v), c(v)) - v = 0\}.$$

For any $v \in X$, let

$$A(v) = \left\{ v + \theta^*(h(v, r(v), c(v)) - v) \;\middle|\; \begin{array}{l} \theta^* \in [0, 1] \\ e(v + \theta^*(h(v, r(v), c(v)) - v); \beta) \\ = \min_{\theta\in[0,1]} e(v + \theta(h(v, r(v), c(v)) - v); \beta) \end{array} \right\}$$

In the following we prove that $A(v)$ is closed at every point $v \in X\backslash\Omega$.

Let $\bar{v}$ be an arbitrary point of $X\backslash\Omega$. Let $v^q \in X\backslash\Omega$, $q = 1, 2, \ldots$, be a sequence convergent to $\bar{v}$, and $y^q \in A(v^q)$, $q = 1, 2, \ldots$, a sequence convergent to $\bar{y}$. To prove that $A(\bar{v})$ is closed, we only need to show $\bar{y} \in A(\bar{v})$. From $v^q \in X\backslash\Omega$ and $\bar{v} \in X\backslash\Omega$, we have $h(v^q, r(v^q), c(v^q)) - v^q \ne 0$ and $h(\bar{v}, r(\bar{v}), c(\bar{v})) - \bar{v} \ne 0$. Due to continuity of $h(v, r(v), c(v))$, $h(v^q, r(v^q), c(v^q))$ converges to $h(\bar{v}, r(\bar{v}), c(\bar{v})$ as $q \to \infty$. Since $y^q \in A(v^q)$, hence, there is some number $\theta_q^* \in [0, 1]$ satisfying $y^q = v^q + \theta_q^*(h(v^q, r(v^q), c(v^q)) - v^q)$. From $h(v^q, r(v^q), c(v^q)) - v^q \ne 0$, we obtain that

$$\theta_q^* = \frac{\|y^q - v^q\|}{\|h(v^q, r(v^q), c(v^q)) - v^q\|},$$

and as $q \to \infty$,

$$\theta_q^* \to \bar{\theta}^* = \frac{\|\bar{y} - \bar{v}\|}{\|h(\bar{v}, r(\bar{v}), c(\bar{v})) - \bar{v}\|}$$

with $\bar{\theta}^* \in [0, 1]$. Therefore, $\bar{y} = \bar{v} + \bar{\theta}^*(h(\bar{v}, r(\bar{v}), c(\bar{v})) - \bar{v})$. Furthermore, since $y^q \in A(v^q)$, we have $e(y^q; \beta) \le e(v^q + \theta(h(v^q, r(v^q), c(v^q)) - v^q); \beta)$ for any $\theta \in [0, 1]$. It implies that $e(\bar{y}; \beta) \le e(\bar{v} + \theta(h(\bar{v}, r(\bar{v}), c(\bar{v})) - \bar{v}); \beta)$ for any $\theta \in [0, 1]$, which proves that

$$e(\bar{y}; \beta) = \min_{\theta\in[0,1]} e(\bar{v} + \theta(h(\bar{v}, r(\bar{v}), c(\bar{v})) - \bar{v}); \beta).$$

According to the definition of $A(v)$, it follows that $\bar{y} \in A(\bar{v})$.

Since X is bounded and $v^k \in X$, $k = 1, 2, \ldots$, we can extract a convergent subsequence from the sequence, v^k, $k = 1, 2, \ldots$. Let v^{k_j}, $j = 1, 2, \ldots$, be a convergent subsequence of the sequence, v^k, $k = 1, 2, \ldots$. Let v^* be the limit point of the subsequence. We show $v^* \in \Omega$ in the following. Clearly, as $k \to \infty$, $e(v^k; \beta)$ converges to $e(v^*; \beta)$ since $e(v; \beta)$ is continuous and $e(v^{k+1}; \beta) < e(v^k; \beta)$, $k = 1, 2, \ldots$. Consider the sequence, v^{k_j}, $j = 1, 2, \ldots$. Note that $v^{k_j+1} = v^{k_j} + \theta_{k_j}(h(v^{k_j}, r(v^{k_j}), c(v^{k_j})) - v^{k_j})$ and

$$e(v^{k_j+1}; \beta) = \min_{\theta\in[0,1]} e(v^{k_j} + \theta(h(v^{k_j}, r(v^{k_j}), c(v^{k_j})) - v^{k_j}); \beta).$$

According to the definition of $A(v)$, we have $v^{k_j} \in A(v^{k_j})$. Since v^{k_j+1}, $j = 1, 2, \ldots$, are bounded, we can extract a convergent subsequence from the sequence v^{k_j+1}, $j = 1, 2, \ldots$. Let v^{k_j+1}, $j \in K$, be a convergent subsequence extracted from the sequence, v^{k_j+1}, $j = 1, 2, \ldots$. Let $v^{\#}$ be the limit point of the subsequence, v^{k_j+1}, $j \in K$. Suppose that $v^* \notin \Omega$. Since $A(v^*)$ is closed, we have $v^{\#} \in A(v^*)$. Thus, $e(v^{\#}; \beta) < e(v^*; \beta)$, which contradicts that $e(v^k; \beta)$ converges as $k \to \infty$. Therefore, $v^* \in \Omega$. The theorem follows.□

Although it is difficult to prove that for any given $\beta > 0$, a limit point of v^k, $k = 0, 1, \ldots$, generated by (12) is at least a local minimum point of (4), in general, it is indeed at least a local minimum point of (4). Theorem 2 implies that every limit point of $v^{*,q}$, $q = 0, 1, \ldots$, is at least a local minimum point of (3) if $v^{*,q}$ is a minimum point of (4) with $\beta = \beta_q$.

We remark that for $\beta = \beta_q$, our algorithm converges to a stationary point of (4) for any given ρ, however, the softassign algorithm proposed in Rangarajan et al. (1996) converges to a stationary point of (4) only if ρ is sufficiently large so that $e_0(v)$ is strictly concave on the null space of the constraint matrix. Thus, for the softassign algorithm to converge, one has to determine the size of ρ through estimating the maximum eigenvalue of the matrix of the objective function of (1), which requires some extra computational work. As we pointed out before, the size of ρ affects quality of a solution generated by a deterministic annealing algorithm and it should be as small as possible. Since our algorithm converges for any ρ, hence, one can start with ρ being a smaller positive number and then increase ρ if the solution generated by the algorithm is not a near integer solution. In this aspect, our algorithm is better than the softassign algorithm. Numerical results will further support this argument.

4. Global convergence of the iterative procedure

In this section we prove that for any given v, the iterative procedure (11) converges to a positive solution (r^*, c^*) of (9).

We verify first that $(x(r,c), y(r,c))$ is a descent direction of $f(r, c)$. Let

$$u_i(r,c) = \sum_{j=1}^{n} \frac{1}{r_i c_j \alpha_{ij}(v)} - 1,$$

$i = 1, 2, \ldots, n$, and $u(r,c) = (u_1(r,c), u_2(r,c), \ldots, u_n(r,c))^{\mathrm{T}}$. Let

$$w_j(r,c) = \sum_{i=1}^{n} \frac{1}{r_i c_j \alpha_{ij}(v)} - 1,$$

$j = 1, 2, \ldots, n$, and $w(r,c) = (w_1(r,c), w_2(r,c), \ldots, w_n(r,c))^{\mathrm{T}}$. Computing the partial derivative of $f(r, c)$ with respect to r_l, we obtain

$$\begin{aligned}\frac{\partial f(r,c)}{\partial r_l} &= -\sum_{h=1}^{n} \frac{c_h \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}\left(\sum_{p=1}^{n} \frac{1}{r_l c_p \alpha_{lp}(v)} - 1\right.\\ &\quad \left. + \sum_{p=1}^{n} \frac{1}{r_p c_h \alpha_{ph}(v)} - 1\right)\\ &= -\sum_{h=1}^{n} \frac{c_h \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}(u_l(r,c) + w_h(r,c)).\end{aligned}$$

Computing the partial derivative of $f(r, c)$ with respect to c_h, we obtain

$$\begin{aligned}\frac{\partial f(r,c)}{\partial c_h} &= -\sum_{l=1}^{n} \frac{r_l \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}\left(\sum_{p=1}^{n} \frac{1}{r_l c_p \alpha_{lp}(v)} - 1\right.\\ &\quad \left. + \sum_{p=1}^{n} \frac{1}{r_p c_h \alpha_{ph}(v)} - 1\right)\\ &= -\sum_{l=1}^{n} \frac{r_l \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}(u_l(r,c) + w_h(r,c)).\end{aligned}$$

Let $\nabla f(r,c) = (\partial f(r,c)/\partial r_1, \partial f(r,c)/\partial r_2, \ldots, \partial f(r,c)/\partial r_n, \partial f(r,c)/\partial c_1, \partial f(r,c)/\partial c_2, \ldots, \partial f(r,c)/\partial c_n)^{\mathrm{T}}$. The next lemma shows that $(x(r,c), y(r,c))$ is a descent direction of $f(r, c)$.

Lemma 3. If $(r, c) > 0$ and $(x(r,c), y(r,c)) \neq 0$ then

$$\nabla f(r,c)^{\mathrm{T}} \begin{pmatrix} x(r,c) \\ y(r,c) \end{pmatrix} < 0.$$

Proof. Note that

$$x_l(r,c) = r_l\left(\sum_{p=1}^{n} \frac{1}{r_l c_p \alpha_{lp}(v)} - 1\right) = r_l u_l(r,c),$$

$l = 1, 2, \ldots, n$, and

$$y_h(r,c) = c_h\left(\sum_{p=1}^{n} \frac{1}{r_p c_h \alpha_{ph}(v)} - 1\right) = c_h w_h(r,c),$$

$h = 1, 2, \ldots, n$. Then,

$$\begin{aligned}\nabla f(r,c)^{\mathrm{T}} \begin{pmatrix} x(r,c) \\ y(r,c) \end{pmatrix} &= -\sum_{l=1}^{n}\sum_{h=1}^{n} \frac{r_l c_h \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}((u_l(r,c))^2\\ &\quad + 2u_l(r,c)w_h(r,c) + (w_h(r,c))^2)\\ &= -\sum_{l=1}^{n}\sum_{h=1}^{n} \frac{r_l c_h \alpha_{lh}(v)}{(r_l c_h \alpha_{lh}(v))^2}(u_l(r,c) + w_h(r,c))^2.\end{aligned} \tag{13}$$

We show in the following that if $u_l(r,c) + w_h(r,c) = 0$, $l = 1, 2, \ldots, n$, $h = 1, 2, \ldots, n$, then $u_l(r,c) = 0$, $l = 1, 2, \ldots, n$, and $w_h(r,c) = 0$, $h = 1, 2, \ldots, n$. From $u_l(r,c) + w_h(r,c) = 0$, $h = 1, 2, \ldots, n$, we obtain that $w_h(r,c)$, $h = 1, 2, \ldots, n$, are equal. From $u_l(r,c) + w_h(r,c) = 0$, $l = 1, 2, \ldots, n$, we get that $u_l(r,c)$, $l = 1, 2, \ldots, n$, are equal. Let $u_l(r,c) = \phi$, $l = 1, 2, \ldots, n$, and $w_h(r,c) = \varphi$, $h = 1, 2, \ldots, n$. Then, $\phi + \varphi = 0$. Note that

$$\begin{aligned}\sum_{l=1}^{n} u_l(r,c) &= \sum_{l=1}^{n}\left(\sum_{p=1}^{n} \frac{1}{r_l c_p \alpha_{lp}(v)} - 1\right)\\ &= \sum_{p=1}^{n}\left(\sum_{l=1}^{n} \frac{1}{r_l c_p \alpha_{lp}(v)} - 1\right)\\ &= \sum_{h=1}^{n}\left(\sum_{p=1}^{n} \frac{1}{r_p c_h \alpha_{ph}(v)} - 1\right) = \sum_{h=1}^{n} w_h(r,c).\end{aligned}$$

Thus, $\phi = \varphi$. From $\phi + \varphi = 0$ and $\phi = \varphi$, we obtain $\phi = \varphi = 0$. Therefore, when $(x(r,c), y(r,c)) \neq 0$, at least one of $u_l(r,c) + w_h(r,c)$, $l = 1, 2, \ldots, n$, $h = 1, 2, \ldots, n$, is not equal to zero. Thus, from (13), we get

$$\nabla f(r,c)^{\mathrm{T}} \begin{pmatrix} x(r,c) \\ y(r,c) \end{pmatrix} < 0.$$

The lemma follows.□

We show next that for any i, no subsequence of r_i^k, $k = 0, 1, \ldots$, approaches zero or infinity and that for any j, no subsequence of c_j^k, $k=0,1,\ldots,$ approaches zero or infinity. Let $\bar{v}$ be an interior point of P. Then,

$$\sum_{j=1}^{n} \bar{v}_{ij} = 1,\ i=1,2,\ldots,n, \qquad \sum_{i=1}^{n} \bar{v}_{ij} = 1, j=1,2,\ldots,n,$$

$$0 < \bar{v}_{ij}, i=1,2,\ldots,n, j=1,2,\ldots,n.$$

Thus, for any $r > 0$ and $c > 0$,

$$\begin{aligned}\sum_{i=1}^{n} \ln r_i + \sum_{j=1}^{n} \ln c_j &= \sum_{i=1}^{n} \ln r_i \sum_{j=1}^{n} \bar{v}_{ij} + \sum_{j=1}^{n} \ln c_j \sum_{i=1}^{n} \bar{v}_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \bar{v}_{ij}(\ln r_i + \ln c_j).\end{aligned} \tag{14}$$

Consider

$$s(r,c) = -\sum_{i=1}^{n}\sum_{j=1}^{n} \int_0^{\ln r_i + \ln c_j} \frac{1}{e^t \alpha_{ij}(v)}\,dt + \sum_{i=1}^{n} \ln r_i + \sum_{j=1}^{n} \ln c_j. \tag{15}$$

We have

$$\nabla s(r,c) = -(Q(r,c))^{-1}\begin{pmatrix} u(r,c) \\ w(r,c)\end{pmatrix},$$

where

$$Q(r,c) = \begin{pmatrix} r_1 & & & & & \\ & \ddots & & & & \\ & & r_n & & & \\ & & & c_1 & & \\ & & & & \ddots & \\ & & & & & c_n \end{pmatrix}.$$

Using

$$\begin{pmatrix} x(r,c) \\ y(r,c)\end{pmatrix} = Q(r,c)\begin{pmatrix} u(r,c) \\ w(r,c)\end{pmatrix}$$

we obtain that

$$\nabla s(r,c)^{\mathrm{T}}\begin{pmatrix} x(r,c) \\ y(r,c)\end{pmatrix} = -(u(r,c)^{\mathrm{T}}u(r,c) + w(r,c)^{\mathrm{T}}w(r,c)) < 0$$

when $(u(r,c), w(r,c)) \neq 0$. Thus, $(x(r,c), y(r,c))$ is a descent direction of $s(r,c)$. Therefore, $s(r^k,c^k)$ is not increasing as k is increasing since $r^{k+1} = r^k + \mu_k x(r^k,c^k)$ and $c^{k+1} = c^k + \mu_k y(r^k,c^k)$.

From the mean-value integration theorem, we get

$$\int_0^{\ln r_i + \ln c_j} \frac{1}{e^t \alpha_{ij}(v)}\,dt = \frac{1}{e^{t_{ij}(r,c)}\alpha_{ij}(v)}(\ln r_i + \ln c_j), \tag{16}$$

where $t_{ij}(r,c)$ is a number between zero and $\ln r_i + \ln c_j$ satisfying that $t_{ij}(r,c) \to \infty$ as $\ln r_i + \ln c_j \to \infty$ and that $t_{ij}(r,c) \to -\infty$ as $\ln r_i + \ln c_j \to -\infty$. Substituting (14) and (16) into (15), we obtain

$$s(r,c) = -\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{e^{t_{ij}(r,c)}\alpha_{ij}(v)} - \bar{v}_{ij}\right)(\ln r_i + \ln c_j). \tag{17}$$

Suppose that there is a pair of i and j, say (i_0, j_0), for which a subsequence of $r_{i_0}^k c_{j_0}^k$, $k=0,1,\ldots,$ approaches zero or infinity. Let $r_{i_0}^{k_q} c_{i_0}^{k_q}$, $q=0,1,\ldots$ be the subsequence. From (17), we get that $s(r^{k_q}, c^{k_q}) \to \infty$ as $q \to \infty$. It contradicts that $s(r^k,c^k)$ is not increasing as k is increasing. Thus, for any i and j, no subsequence of $r_i^k c_j^k$, $k=0,1,\ldots,$ approaches zero or infinity.

Suppose that there is some index l for which a subsequence of r_l^k, $k=0,1,\ldots,$ approaches zero. Let $r_l^{k_q}$, $q=0,1,\ldots,$ be the subsequence. Using the above results, we derive that as $q \to \infty$, $r_i^{k_q} \to 0$ for $i=1,2,\ldots,n$, and $c_j^{j_q} \to \infty$ for $j=1,2,\ldots,n$. Consider

$$g(r,c) = \sum_{i=1}^{n} \ln r_i - \sum_{j=1}^{n} \ln c_j.$$

Clearly, $g(r^{k_q}, c^{k_q}) \to \infty$ as $q \to \infty$. From $g(r,c)$, we obtain

$$\nabla g(r,c) = \left(\frac{1}{r_1}, \frac{1}{r_2}, \ldots, \frac{1}{r_n}, -\frac{1}{c_1}, -\frac{1}{c_2}, \ldots, -\frac{1}{c_n}\right)^{\mathrm{T}}.$$

Note that

$$\begin{aligned}\nabla g(r,c)^{\mathrm{T}}\begin{pmatrix} x(r,c) \\ y(r,c)\end{pmatrix} &= \sum_{i=1}^{n}\left(\sum_{p=1}^{n}\frac{1}{r_i c_p \alpha_{ip}(v)} - 1\right) \\ &\quad - \sum_{j=1}^{n}\left(\sum_{p=1}^{n}\frac{1}{r_p c_j \alpha_{pj}(v)} - 1\right) = 0.\end{aligned}$$

Then, $(x(r,c), y(r,c))^{\mathrm{T}}$ is perpendicular to the gradient of $g(r,c)$. Thus, $g(r^{k_q}, c^{k_q})$ cannot approach minus infinity as $q \to \infty$ since $r^{k+1} = r^k + \mu_k x(r^k,c^k)$ and $c^{k+1} = c^k + \mu_k y(r^k c^k)$, which yields a contradiction. Therefore, for any i, no subsequence of r_i^k, $k=0,1,\ldots,$ approaches zero. In the same way, one can prove that for any j, no subsequence of c_j^k, $k=0,1,\ldots,$ approaches zero, for any i, no subsequence of $k=1,2,\ldots,$ approaches infinity.

Note that $f(r,c) \geq 0$ and $f(r^k,c^k)$ decreases strictly and monotonically. From Lemma 3 and the boundedness of $\ln r_i^k$ and $\ln c_j^k$, we derive that

$$\nabla f(r^k,c^k)^{\mathrm{T}}\begin{pmatrix} x(r^k,c^k) \\ y(r^k,c^k)\end{pmatrix} \to 0$$

as $k \to \infty$. Then, from (13), we obtain that for any i and j, $u_i(r^k,c^k) + w_j(r^k,c^k) \to 0$ as $k \to \infty$. Thus, $u(r^k,c^k) \to 0$ and $w(r^k,c^k) \to 0$ as $k \to \infty$. Therefore, every limit point of (r^k,c^k), $k=0,1,\ldots,$ is a positive solution of (9). Hence, the next theorem follows.

Theorem 4. For any given v, every limit point of (r^k,c^k),

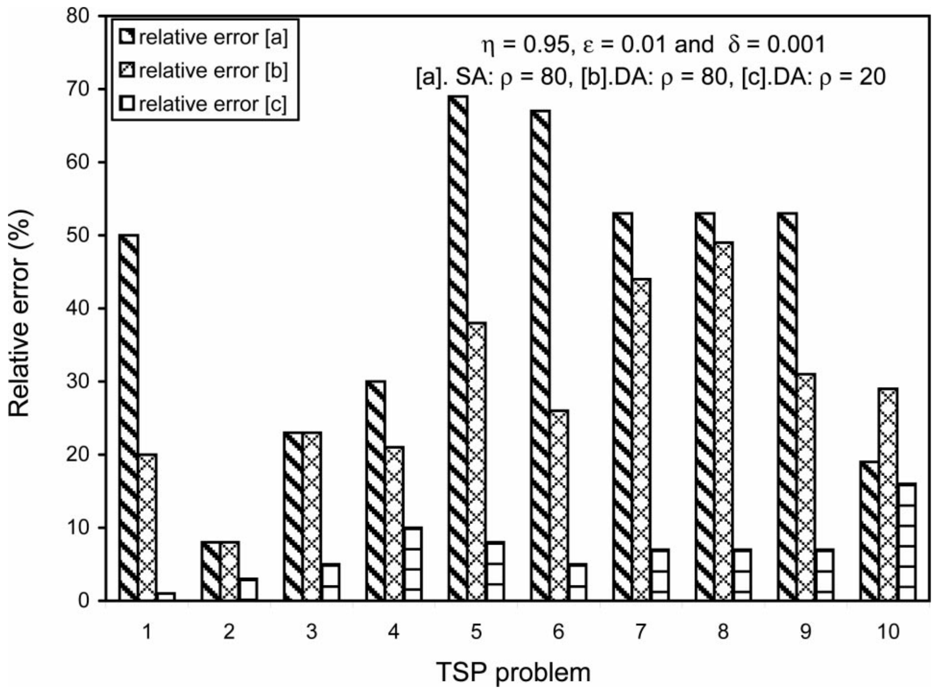

Fig. 1. Relative error to optimal tour: 1, bays29; 2, att48; 3, eil51; 4, berlin52; 5, st70; 6, eil76; 7, pr76; 8, rd100; 9, eil101; and 10, lin105.

$k = 0, 1, \ldots,$ generated by the iterative procedure (11) is a positive solution of (9).

5. Numerical results

The algorithm has been used to approximate solutions of a number of TSP instances. The algorithm succeeds in finding a tour of high quality for each of the TSP instances. In our implementation of the algorithm,

1. $\beta_0 = 200$ (β_0 can be any positive number satisfying that $e(v; \beta_0)$ is convex);
2. $r^0 = (r_1^0, r_2^0, \ldots, r_n^0)^{\mathrm{T}}$ and $c^0 = (c_1^0, c_2^0, \ldots, c_n^0)^{\mathrm{T}}$ are two random vectors satisfying $0 < r_i^0 < 1$ and $0 < c_i^0 < 1$, $i = 1, 2, \ldots, n$;
3. $\mu_k = 0.95$ (μ_k can be any number in (0, 1)), and for any given v, the iterative procedure (11) terminates as soon as $\sqrt{f(r^k, c^k)} < \delta$;
4. we replace $e(x; \beta)$ with $L(v, \lambda^r, \lambda^c)$ in the algorithm since $(r(v^k), c(v^k))$ is an approximate solution of (9);
5. θ_k is determined with the following Armijo-type line search:

   $$\theta_k = \xi^{m_k}$$

   with m_k being the smallest nonnegative integer satisfying

   $$L(v^k + \xi^{m_k}(h(v^k, r(v^k), c(v^k)) - v^k), \lambda^{r,k}, \lambda^{c,k})$$
   $$\leq L(v^k, \lambda^{r,k}, \lambda^{c,k}) + \xi^{m_k}\gamma(h(v^k, r(v^k), c(v^k))$$
   $$- v^k)^T \nabla_v L(v^k, \lambda^{r,k}, \lambda^{c,k}),$$

   where

   $$\lambda^{r,k} = \beta_q(\ln r_i(v^k), \ln r_2(v^k), \ldots, \ln r_n(v^k))^T,$$

   $$\lambda^{c,k} = \beta_q(\ln c_1(v^k), \ln c_2(v^k), \ldots, \ln c_n(v^k))^T,$$

   and ξ and γ can be any numbers in (0, 1) (we set $\xi = 0.6$ and $\gamma = 0.8$, but there is no rule for selecting ξ and γ).

The algorithm terminates as soon as $\beta_q < 1$. To produce a solution of higher quality, the size of ρ should be as small as possible. However, a small ρ may lead to a fractional solution $v^{*,q}$. To make sure that an integer solution is generated, we continue the following procedure,

Step 0: Let $\beta = 1$, $v^0 = v^{*,q}$, and $k = 0$. Go to Step 1.

Step 1: Let $v^* = (v_{11}^*, v_{12}^*, \ldots, v_{1n}^*, \ldots, v_{n1}^*, v_{n2}^*, \ldots, v_{nn}^*)^{\mathrm{T}}$ with

$$v_{ij}^* = \begin{cases} 1 \text{ if } v_{ij}^k \geq 0.9, \\ 0 \text{ if } v_{ij}^k < 0.9, \end{cases}$$

$i = 1, 2, \ldots, n, j = 1, 2, \ldots, n$. If $v^* \in P$, the procedure terminates. Otherwise, let $\rho = \rho + 2$ and go to Step 2.

Step 2: Given $v = v^k$, use (11) to obtain a positive solution $(r(v^k), c(v^k))$ of (9). Let $r^0 = r(v^k), c^0 = c(v^k)$,

$$\lambda^{r,k} = (\ln r_1(v^k), \ln r_2(v^k), \ldots, \ln r_n(v^k))^{\mathrm{T}},$$

and

$$\lambda^{c,k} = (\ln c_1(v^k), \ln c_2(v^k), \ldots, \ln c_n(v^k))^{\mathrm{T}}.$$

Go to Step 3.

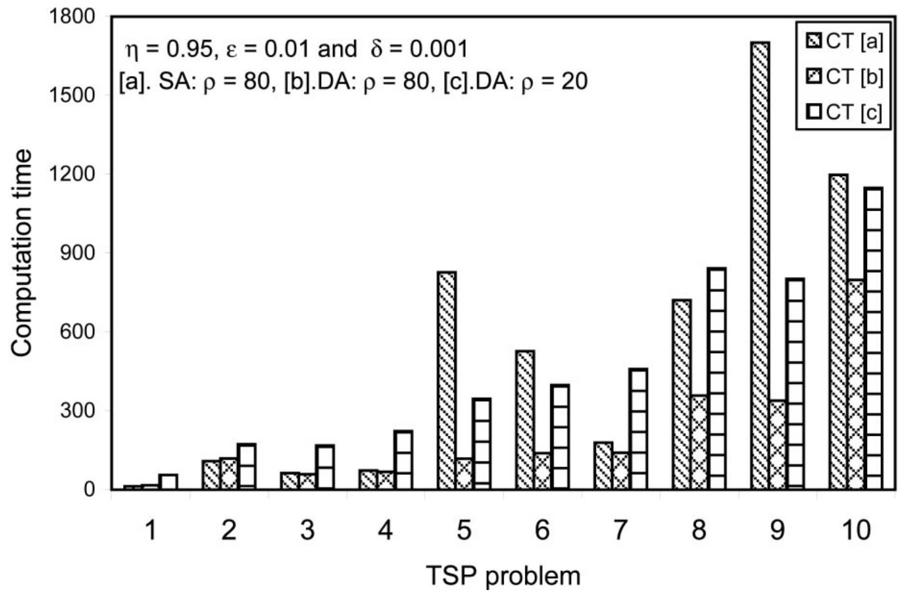

Fig. 2. Computation time for different algorithms: 1, bays29; 2, att48; 3, eil51; 4, berlin52; 5, st70; 6, eil76; 7, pr76; 8, rd100; 9, eil101; and 10, lin105.

Step 3: Let

$$h(v^k, r(v^k), c(v^k)) = (h_{11}(v^k, r(v^k), c(v^k)), h_{12}(v^k, r(v^k), c(v^k)), \ldots, h_{1n}(v^k, r(v^k), c(v^k)), \ldots, h_{n1}(v^k, r(v^k), c(v^k)), h_{n2}(v^k, r(v^k), c(v^k)), \ldots, h_{nn}(v^k, r(v^k), c(v^k)))^{\mathrm{T}}$$

with

$$h_{ij}(v^k, r(v^k), c(v^k)) = \frac{1}{r_i(v^k)c_j(v^k)\alpha_{ij}(v^k)},$$

$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. If $\|h(v^k, r(v^k), c(v^k)) - v^k\| < \epsilon$, let $v^0 = v^k$ and $k = 0$, and go to Step 1. Otherwise, do as follows: Compute

$$v^{k+1} = v^k + \theta_k(h(v^k, r(v^k), c(v^k)) - v^k),$$

where θ_k is determined with the Armijo-type line search. Let $k = k + 1$ and go to Step 2.

The algorithm is programmed in MATLAB. To compare the algorithm with the softassign algorithm proposed in Rangarajan et al. (1996, 1999), the softassign algorithm is also programmed in MATLAB. All our numerical tests are done on a PC. In the presentations of numerical results, DA stands for our algorithm, SA the softassign algorithm, CT the computation time in seconds, OPT the length of an optimal tour, OBJ the length of a tour generated by an algorithm, OBJD the length of the tour generated by our algorithm, OBJSA the length of the tour generated by the softassign, and

$$\mathrm{RE} = \frac{\mathrm{OBJ} - \mathrm{OPT}}{\mathrm{OPT}}.$$

To show the robustness of our algorithm, we have taken two different values for each of η, ϵ and δ. Numerical results are as follows.

Example 1. These ten TSP instances are from a well-known website, TSPLIB. We have used our algorithm and the softassign algorithm to approximate solutions of these TSP instances. Note that the softassign algorithm fails to converge when $\rho = 20$. For $\eta = 0.95$, $\epsilon = 0.01$ and $\delta = 0.001$, the computation time and the relative errors to the optimal tours of the tours generated by our algorithm and the softassign algorithm are compared in Figs. 1 and 2.

1. It is shown in Fig. 1 that the tour generated by our algorithm is closer to the optimal tour than that generated by the softassign algorithm. The maximum relative error in the optimal tour for our algorithm ($\rho = 20$) is 16%, whereas that for the softassign algorithm is 69%. Especially for bays29, the relative error to the optimal tour for our algorithm ($\rho = 20$) is only 1%, whereas the relative error to the optimal tour for the softassign algorithm is 50 times larger than that for our algorithm.
2. It is clearly indicated in Fig. 2 that our algorithm is much more efficient than the softassign algorithm. Considering the computation time of our algorithm and the softassign algorithm for eil101, one can see that the computation time of our algorithm ($\rho = 20$) is 802 s, whereas the computation time of the softassign algorithm is twice that of our algorithm.

Table 1
Numerical results

TSP		SA	DA	DA	SA	DA	DA	SA	DA	DA	SA	DA	DA	SA	DA	DA
		$\rho = 80$	$\rho = 80$	$\rho = 20$	$\rho = 80$	$\rho = 80$	$\rho = 20$	$\rho = 80$	$\rho = 80$	$\rho = 20$	$\rho = 80$	$\rho = 80$	$\rho = 20$	$\rho = 80$	$\rho = 80$	$\rho = 20$
$\eta = 0.9$		bays29; $n = 29$			att48; $n = 48$			eil51; $n = 51$			berlin52; $n = 52$			st70; $n = 70$		
$\epsilon = 0.01$	CT	16	14	42	109	101	138	60	50	96	63	50	176	762	97	276
$\delta = 0.001$	OBJ	2227	2289	2130	36,943	35,025	34,273	636	515	460	10,882	10,087	8183	1094	1105	727
	RE (%)	10	13	5	10	4	2	48	20	7	44	34	8	61	63	7
		eil76; $n = 76$			pr76; $n = 76$			rd100; $n = 100$			eil101; $n = 101$			lin105; $n = 105$		
	CT	386	153	271	467	118	306	1627	242	780	1183	263	699	978	967	783
	OBJ	895	734	571	181,910	133,721	114,768	13,728	11,534	9195	1009	856	673	17,852	18,838	17,529
	RE (%)	64	35	5	68	24	6	74	46	16	57	33	5	24	31	22
$\eta = 0.95$		bays29; $n = 29$			att48; $n = 48$			eil51; $n = 51$			berlin52; $n = 52$			st70; $n = 70$		
$\epsilon = 0.01$	CT	13	17	56	109	119	173	63	59	168	73	68	223	827	118	346
$\delta = 0.001$	OBJ	3030	2427	2038	36,061	36,293	34,661	527	530	451	10,739	9573	8318	1145	935	735
	RE (%)	50	20	1	8	8	3	23	23	5	30	21	10	69	38	8
		eil76; $n = 76$			pr76; $n = 76$			rd100; $n = 100$			eil101; $n = 101$			lin105; $n = 105$		
	CT	527	139	398	179	141	459	721	358	842	1701	339	802	1198	798	1148
	OBJ	912	687	571	165,664	155,388	115,926	12,122	11,803	8427	979	842	687	17,155	18,515	16,734
	RE (%)	67	26	5	53	44	7	53	49	7	53	31	7	19	29	16
$\eta = 0.9$		bays29; $n = 29$			att48; $n = 48$			eil51; $n = 51$			berlin52; $n = 52$			st70; $n = 70$		
$\epsilon = 0.001$	CT	141	33	150	660	254	430	1154	256	329	1098	219	523	1429	255	1072
$\delta = 0.0001$	OBJ	2537	2335	2038	45,706	36,818	34,218	635	537	459	10,856	9143	8052	980	876	731
	RE (%)	26	16	1	36	10	2	48	25	7	44	31	7	44	29	8
		eil76; $n = 76$			pr76; $n = 76$			rd100; $n = 100$			eil101; $n = 101$			lin105; $n = 105$		
	CT	10,615	457	1147	2597	725	1277	2876	1233	2609	15,872	1127	2710	13,403	2362	3561
	OBJ	837	710	574	166,218	150,899	115,002	9765	12,544	8935	1046	870	689	28,809	16,974	16,718
	RE (%)	54	30	5	54	40	6	23	58	13	66	36	7	100	17	16
$\eta = 0.95$		bays29; $n = 29$			att48; $n = 48$			eil51; $n = 51$			berlin52; $n = 52$			st70; $n = 70$		
$\epsilon = 0.001$	CT	80	49	184	508	199	586	636	205	496	977	228	768	959	342	1035
$\delta = 0.0001$	OBJ	2473	2239	2295	36,541	36,462	34,218	644	522	451	10,776	9903	8084	987	895	724
	RE (&)	22	11	14	9	9	2	50	21	5	43	31	7	45	32	7
		eil76; $n = 76$			pr76; $n = 76$			rd100; $n = 100$			eil101; $n = 101$			lin105; $n = 105$		
	CT	3507	639	1740	6647	655	1531	2410	1136	2749	6979	1123	3704	14,827	2351	4383
	OBJ	818	736	581	197,776	152,284	112,935	9428	10,641	8736	956	802	672	26,827	18,384	16,001
	RE (%)	50	35	7	83	41	4	19	32	10	49	23	5	87	28	11

Table 2
Numerical results ($\epsilon = 0.01$ and $\delta = 0.001$)

Algorithm	ρ	TSP	$\eta = 0.9$ CT	OBJ	OBJD/OBJSA	TSP	$\eta = 0.95$ CT	OBJ	OBJD/OBJSA
SA	80	1	543	889		11	584	905	
DA	80		370	956	1.07		565	967	1.06
DA	20		706	845	0.95		906	809	0.89
SA	80	2	613	907		12	592	847	
DA	80		377	988	1.08		432	1004	1.18
DA	20		783	809	0.89		836	799	0.94
SA	80	3	578	934		13	688	994	
DA	80		483	904	0.97		417	1143	1.15
DA	20		618	856	0.92		913	869	0.87
SA	80	4	451	906		14	685	854	
DA	80		369	1020	1.12		600	932	1.09
DA	20		638	810	0.89		866	802	0.94
SA	80	5	577	955		15	557	937	
DA	80		251	1063	1.11		578	1052	1.12
DA	20		615	895	0.93		859	791	0.84
SA	80	6	416	952		16	577	998	
DA	80		486	1011	1.06		382	1029	1.03
DA	20		796	858	0.90		894	893	0.89
SA	80	7	776	895		17	590	945	
DA	80		241	1130	1.26		703	978	1.03
DA	20		811	859	0.90		929	847	0.90
SA	80	8	439	862		18	688	861	
DA	80		258	959	1.11		474	953	1.10
DA	20		688	816	0.94		852	833	0.96
SA	80	9	343	907		19	527	929	
DA	80		422	846	0.93		349	1065	1.14
DA	20		620	792	0.87		973	866	0.93
SA	80	10	448	897		20	780	870	
DA	80		269	1085	1.20		303	1113	1.27
DA	20		643	774	0.86		830	810	0.93

3. For two different values of η, ϵ and δ, the numerical results are similar to those mentioned above, which are presented in Table 1. This clearly shows that our algorithm is robust and outperforms the softassign algorithm.

Example 2. These (TSP) instances have 100 cities and are generated randomly. Every city is a point in a square with integer coordinates (x, y) satisfying $0 \leq x \leq 100$ and $0 \leq y \leq 100$. We have used our algorithm and the softassign algorithm to approximate solutions of 20 (TSP) instances (10 for $\eta = 0.9$ and 10 for $\eta = 0.95$). For two different values of η, the computation time and the quality of a tour generated by our algorithm and the softassign algorithm are compared in Table 2. Note that the softassign algorithm fails to converge when $\rho = 20$. Numerical results further confirm that our algorithm outperforms the softassign algorithm.

6. Conclusions

We have developed a globally convergent Lagrange multiplier and barrier function iterative algorithm for approximating a solution of the TSP. Some theoretical results have been derived. For any given value of the barrier parameter, we have proved that the algorithm converges to a stationary point of (4) without any condition on the objective function, which is stronger than the convergence result for the softassign algorithm. We have reported some numerical results, which show that our algorithm seems more effective and efficient than the softassign algorithm. The algorithm would be improved if one could propose a faster iterative procedure for updating Lagrange multipliers to obtain the feasible descent direction.

Acknowledgements

The authors would like to thank anonymous referees for their constructive comments and remarks, which have significantly improved quality of this paper. The preliminary version of this paper was completed when the authors were on leave at the Chinese University of Hong Kong from 1997 to 1998. The work was supported by CUHK Direct Grant 220500680, Ho Sin-Hang Education

Endowment Fund HSH 95/02, and Research Fellow and Research Associate Scheme of CUHK Research Committee.

References

Cichocki, A., & Unbehaunen, R. (1993). *Neural networks for optimization and signal processing*, New York: John Wiley & Sons.

Durbin, R., & Willshaw, D. (1987). An analogue approach to the traveling salesman problem using an elastic network method. *Nature*, *326*, 689–691.

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

Erlander, S. (1981). Entropy in linear programs. *Mathematical Programming*, *21*, 137–151.

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

Hopfield, J., & Tank, D. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, *52*, 141–152.

Lau, K. M., Chan, S. M., & Xu, L. (1995). Comparison of the Hopfield scheme to the hybrid of Lagrange and transformation approaches for solving the traveling salesman problem. *Proceedings of International IEEE Symposium on Intelligence in Neural and Biological Systems, Hemdon*, 209–217.

Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G., & Shmoys, D. B. (1985). *The traveling salesman problem*, New York: John Wiley & Sons.

Minoux, M. (1986). *Mathematical programming: theory and algorithms*, New York: John Wiley & Sons.

Rangarajan, A., Gold, S., & Mjolsness, E. (1996). A novel optimizing network architecture with applications. *Neural Computations*, *8*, 1041–1060.

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

van den Berg, J. (1996). *Neural relaxation dynamics*. PhD Thesis, Erasmus University of Rotterdam, Rotterdam, The Netherlands.

Wacholder, E., Han, J., & Mann, R. (1989). A neural network algorithm for the multiple traveling salesman problem. *Biological Cybernetics*, *61*, 11–19.

Xu, L. (1994). Combinatorial optimization neural nets based on a hybrid of Lagrange and transformation approaches. *Proceedings of World Congress on Neural Networks, San Diego*, 399–404.

Yuille, A., & Kosowsky, J. (1994). Statistical physics algorithms that converge. *Neural Computation*, *6*, 241–356.