Proceedings of the 26th Conference
on Decision and Control
Los Angeles, CA • December 1987

The Embedding of the Traveling Salesman Problem in a Markov Decision Process*

Jerzy A. Filar
Department of Mathematics and Statistics
University of Maryland Baltimore County
Catonsville, Maryland 21228

Dmitry Krass
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, Maryland 21218

1. Introduction

In this paper we derive a new LP-relaxation of the Traveling Salesman Problem (TSP, for short). This formulation comes from first embedding the TSP in a Markov Decision Process (MDP, for short), and from perturbing this MDP appropriately.

A similar approach was employed earlier by Derman and Klein [2], and Derman [1], but apparently not for the purpose of analyzing the TSP. Indeed, Derman and Klein's [2] embedding was called the "Stochastic Traveling Salesman Problem" and it stimulated a number of more applied works. The relation of the interesting results of [2] and [1] to this paper can be summarized as follows:

- The perturbation used by Derman on page 136 of [1] is not the same one as that introduced in Section 3 below. Indeed, it can be shown that with Derman's perturbation our Theorem 3.1 would be invalid. Moreover, our 'additional' constraints ((C4)-(C5) in Section 3) are also different from those used in [2] and [1].
- At the time [2], and [1] were written the results of Hordijk and Kallenberg [3] were unavailable, thereby making this approach appear, perhaps, less promising for further theoretical investigations. Indeed, Derman and Klein [2] were apparently not interested in solving the TSP with the help of their model, which they appropriately regarded as interesting in its own right.

2. Definitions and Preliminaries

A discrete Markovian decision process Γ is observed at discrete time points $t = 1, 2, \ldots,$. The state space is denoted by $E = \{1, 2, \ldots, N\}$. With each state $i \in E$, we associate a finite set $A(i)$ of "actions". At any time point t the system is in one of the states and an action has to be chosen by the decision maker. If the system is in state i and action $a \in A(i)$ is chosen, then an immediate reward r_{ia} is earned and the process moves to a state $j \in E$ with transition probability p_{iaj}, where $p_{iaj} \geq 0$ and $\sum_{j=1}^{N} p_{iaj} = 1$.

Henceforth, the process Γ will be synonymous with the four-tuple $\langle E, A, r, p \rangle$, where $A = \{A(i) \mid i \in E\}$, $r = \{r_{ia} \mid (i,a) \in E \times A(i)\}$ and $p = \{p_{iaj} \mid (i,a,j) \in E \times A(i) \times E\}$. Sometimes p will be referred to as the *law of motion* of Γ.

A *decision rule* f^t at time t is a function which assigns a probability to the event that action a is taken at time t. In general f^t may depend on all realized states up to and including time t, and on all realized actions up to time t. A *policy* f is a sequence of decision rules: $f = (f^1, f^2, \ldots, f^t \ldots)$. A *Markov policy*, i.e. one in which f^t depends only on the "current" state at time t, is called *stationary* if all its decision rules are identical. A *deterministic policy* is a stationary policy with nonrandomized decision rules. In particular, we shall denote a stationary policy f by the collection of probability vectors $\mathbf{f(i)} = (f(i,1), f(i,2), \ldots, f(i,m_i))$, where $m_i = |A(i)|$ for $i = 1, \ldots, N$. Here $f(i,k)$ is the probability that action k is chosen in state i whenever that state is visited. If f is deterministic, each $f(i,a) \in \{0,1\}$ and hence we shall write $f = (f(1), \ldots, f(N))$, where $f(i)$ now denotes the action chosen whenever state i is visited.

Let X_t be the state at time t, Y_t be the action at time t, and $P_f(X_t = j, Y_t = a \mid X_1 = i)$ be the conditional probability that at time t the state is j and the action taken is a, given that the initial state is i and the decision maker uses a policy f. Now if R_t denotes the reward at time t, then for any policy f and initial state i the expectation of R_t is given by

$$E_f(R_t, i) = \sum_{j \in E} \sum_{a \in A(j)} P_f(X_t = j, Y_t = a \mid X_1 = i) r_{ja}. \quad (2.1)$$

The manner in which we aggregate the resulting stream of expected rewards $\{E_f(R_t, i); t = 1, 2, \ldots\}$ defines the Markov Decision Process discussed in the sequel:

Average Reward Markovian Decision Process (AMD): Here the corresponding overall reward is defined by

$$\phi_i(f) = \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E_f(R_t, i).$$

A policy, f^*, is called *optimal* if for every $i \in E$

$$\phi_i(f^*) = \max_f \phi_i(f). \quad (2.2)$$

*This research was supported in part by the AFOSR and the NSF under the grant #ECS8704954

CH2505-6/87/0000-2255$1.00 © 1987 IEEE

We shall assume that the *initial distribution* on the states of Γ is the given vector $\gamma = (\gamma_1, \ldots, \gamma_N)^T$, with $\gamma_i = P(X_1 = i)$ and $\sum_{i=1}^N \gamma_i = 1$. The *overall payoff* resulting from the use of a policy f, if the initial distribution is γ, will be denoted by

$$\phi(f,\gamma) = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_i E_f(R_t, i).$$

Given a stationary policy f, let $p_{ij}(f) = \sum_{a\in A(i)} p_{iaj} f(i,a)$. It is now clear that f defines a Markov Chain with the probability transition matrix

$$P(f) = (p_{ij}(f))_{i,j=1}^{N}. \tag{2.3}$$

For any policy f, initial distribution γ, $j \in E$ and $a \in A(j)$, define

$$x_{ja}^T(f) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \gamma_i P_f(X_t = j, Y_t = a \mid X_1 = i). \tag{2.4}$$

Further, let $X(f)$ denote the set of all limit points of the vectors $\{x^T(f) \mid T = 1,2,\ldots\}$, where $x^T(f)$ is a $\sum_{i=1}^N |A(i)|$-dimensional vector with entries given by (2.4). If $X(f) = \{\mathbf{x}(f)\}$, a singleton, then the entries $x_{ja}(f)$ of $\mathbf{x}(f)$ can be interpreted as the *long-run expected state-action frequencies* induced by f. Similarly, the long-run expected frequencies of visits to any state $j \in E$ under f are given by

$$x_j(f) = \sum_{a\in A(j)} x_{ja}(f). \tag{2.5}$$

A Markov Decision Process is called *unichain* if for any deterministic policy f, the Markov chain induced by $P(f)$ has one ergodic set plus a (perhaps empty) set of transient states.

Consider the following linear program (LP1):

$$\max \sum_{i\in A} \sum_{a\in A(i)} r_{ia} x_{ia}$$

subject to:

$$\begin{aligned}
&(C1) \quad \sum_{i\in E} \sum_{a\in A(i)} (\delta_{ij} - p_{iaj}) x_{ia} = 0, \qquad j \in E \\
&(C2) \quad \sum_{i\in E} \sum_{a\in A(i)} x_{ia} = 1 \\
&(C3) \quad x_{ia} \geq 0, \qquad i \in E, a \in A(i),
\end{aligned}$$

where δ_{ij} is the Kronecker delta. Let X denote the feasible region of the above program, and $C(S)$ denote the class of stationary strategies of the unichain MDP. Now consider the map $T : X \to C(S)$, where $T(\mathbf{x}) = f_x$ is defined by

$$f_x(i,a) = \begin{cases} \dfrac{x_{ia}}{x_i}, & \text{if } x_i = \displaystyle\sum_{a\in A(i)} x_{ia} > 0 \\ 1, & \text{if } x_i = 0 \text{ and } a = 1 \\ 0, & \text{if } x_i = 0 \text{ and } a \neq 1, \end{cases} \tag{2.6}$$

for every $i \in E$ and $a \in A(i)$. Also consider the map $\hat{T} : C(S) \to X$ where $\hat{T}(f) = \mathbf{x}(f)$ is defined by (consistently with (2.5) and (2.6)):

$$x_{ia}(f) = p_i^*(f) f(i,a), \quad i \in E, \quad a \in A(i). \tag{2.7}$$

In the above, $p_i^*(f)$ is the i-th entry of the unique fixed probability vector of $P(f)$. The transformations T and $\hat{T}$ have been studied by a number of authors (e.g., see Derman [1], Kallenberg [4]). Those of their properties that we shall require in the sequel are summarized in the following result which can be reconstructed with the help of [1] and [4].

Theorem 2.1 *Let Γ be a unichain Markov Decision Process, and (LP1), $C(S), X, T$ and $\hat{T}$ be as defined above. Then*

(i) For all $f \in C(S)$ and initial state $i \in E$

$$\phi_i(f) = \sum_{i\in E} \sum_{a\in A(i)} r_{ia} x_{ia}(f).$$

(ii) If $\mathbf{x}^0$ is optimal in (LP1) then $f_{\mathbf{x}^0} = T(\mathbf{x}^0)$ is optimal in Γ. Conversely, if f^0 is an optimal stationary policy in Γ, then $\hat{T}(f^0) = \mathbf{x}(f^0)$ is optimal in (LP1).

(iii) For all $\mathbf{x} \in X$, $\hat{T}(T(\mathbf{x})) = \mathbf{x}$.

(iv) If $L(S) = \{\mathbf{x}(f) \mid f \in C(S)\}$, then $X = L(S)$, and the extreme points of X correspond to those $\mathbf{x}$ for which $f_{\mathbf{x}}$ is a deterministic policy.

We shall now describe the famous Traveling Salesman Problem (TSP, for short) which has been studied by many authors (see [5] for a recent survey).

A "traveling salesman" starts out at his home city and must visit each of $N-1$ other cities exactly once before returning home. His objective is to minimize the total distance traveled in making his tour. In graph theoretic terms, the problem is to find a minimum cost Hamiltonian cycle in a complete graph G with N nodes and with costs c_{ij} associated with the arcs (i,j).

The first MDP which we shall associate with the TSP will be the process $\Gamma =< E, A, r, p >$, $E = \{1,2,\ldots,N\}$ = set of nodes of G, $A(i) = E\backslash\{i\}$ for each $i \in E$ and $A = \bigcup_{i=1}^N A(i)$, $r = \{r_{ij} = -c_{ij} \mid i \in E, j \in A(i)\}$, and $p = \{p_{iaj} \mid (i,a,j) \in E \times A(i) \times E\}$ with $p_{iaj} = \delta_{aj}$, the Kronecker delta. Also, we assume that 1 is the initial state. We shall say that a deterministic policy f in Γ *is a tour* in the TSP if the state-sequence $i_1 = 1$, $i_2 = f(1)$, $i_3 = f(i_2), \ldots, i_{N+1} = f(i_N) = 1$ is a Hamiltonian cycle in G. If the above sequence contains sub-cycles, we shall say that f *has subtours in the TSP*. Note that if f is a tour, then $x_j(f) = \frac{1}{N}$ for every $j \in E$ (see (2.5)).

3. The ϵ-Perturbed Embedding of the TSP

The preceding embedding of the TSP in Γ suggests that analysis be carried out in the space of long-run state-action frequencies, the union of $\{x(f)\}$ over all policies f. A characterization of this space as a polyhedral set is now available (e.g., see Hordijk and Kallenberg [3]). However, it is known ([4]) that there are points in that space that cannot be obtained from any stationary policy via the transformation $\hat{T}$, and furthermore the long-run frequency $\mathbf{x}(f)$ is not continuous in f. These, and some other, technical difficulties would vanish if Γ were a unichain Markov Decision

Process. In view of the above we now perturb the law of motion of Γ to $p(\epsilon) = \{p_{iaj}(\epsilon) \mid (i,a,j) \in E \times A(i) \times E\}$ where for any $\epsilon \in (0,1)$ we define

$$p_{iaj}(\epsilon) = \begin{cases} 1 & \text{if } i = 1 \text{ and } a = j \\ 0 & \text{if } i = 1 \text{ and } a \neq j \\ 1 & \text{if } i > 1 \text{ and } a = j = 1 \\ \epsilon & \text{if } i > 1,\ a \neq j, \text{ and } j = 1 \\ 1-\epsilon & \text{if } i > 1,\ a = j, \text{ and } j > 1 \\ 0 & \text{if } i > 1,\ a \neq j, \text{ and } j > 1. \end{cases}$$

The ϵ-perturbed process $\Gamma(\epsilon) = < E, A, r, p(\epsilon) >$ clearly tends to Γ as $\epsilon \to 0$. It has the following properties, that can be established by standard arguments.

Lemma 3.1 *(i) The Markov Decision Process $\Gamma(\epsilon)$ is unichain.*

(ii) Consider the Markov chain induced by a stationary policy f in $\Gamma(\epsilon)$ and let $C \subseteq E$ be an ergodic class in that chain. Then (a) $1 \in C$, and (b) if a state $j \notin C$, then j is transient.

Lemma 3.2 *Let f be a deterministic policy in $\Gamma(\epsilon)$ (and thereby also in Γ) which is a tour of the TSP, and assume that s is the k-th city of this tour (starting at 1). Now, if $\mathbf{x}(f) = \hat{T}(f)$, then*

$$x_{sa}(f) = \begin{cases} \dfrac{(1-\epsilon)^{k-2}}{d(\epsilon)}, & \text{if } k > 1 \text{ and } a = f(s) \\ \frac{1}{d(\epsilon)}, & \text{if } k = 1 \text{ and } a = f(s) \\ 0, & \text{otherwise,} \end{cases}$$

where $d(\epsilon) = 1 + \sum_{i=2}^{N}(1-\epsilon)^{i-2}$.

Proof: First we shall prove the result for f^1 which is the tour $\tau_0 = (1,2,3,\ldots,N,1)$. It should be clear from (2.7) that

$$x_{ka}(f^1) = \begin{cases} p_k^*(f^1), & \text{if } 1 < k \text{ and } a = f(k) \\ p_1^*(f^1), & \text{if } 1 = k \text{ and } a = f(k) \\ 0, & \text{otherwise,} \end{cases}$$

and hence that to determine $\mathbf{x}(f^1)$ we must solve the equations

$$\begin{aligned} \mathbf{y}^T P(f^1) &= \mathbf{y}^T \\ \mathbf{y}^T \mathbf{1} &= 1 \end{aligned} \tag{3.1}$$

where $\mathbf{1}$ is a vector with every entry equal to 1. Note that since $\Gamma(\epsilon)$ is unichain, the system of equations (3.1) possesses a unique solution. Simple computation using the definitions of $p(\epsilon)$ and f^1 yields:

$$y_2 = y_1, \qquad (1-\epsilon)y_{j-1} = y_j, \qquad j = 3,\ldots,N-1,$$

Consequently, the solution of (3.1) is of the form:

$$\mathbf{y}^T = \frac{1}{d(\epsilon)}(1, 1, (1-\epsilon), (1-\epsilon)^2, \ldots, (1-\epsilon)^{N-2}),$$

as required.

On the other hand, if $\tilde{f}$ is a tour $\tilde{\tau}$ different from τ_0, then $\tilde{\tau}$ can be obtained from τ_0 by a permutation of its entries. The corresponding permutation of the variables of (3.1) will yield the solution $\tilde{\mathbf{y}}$ whose entries are the appropriate permutation of the entries of $\mathbf{y}$ and satisfy the statement of the Lemma.

□

Remark 3.1

(i) Note that for any f which is a tour, $x_k(f) = \sum_{a \in A(k)} x_{ka}(f) \geq \frac{(1-\epsilon)^{N-2}}{d(\epsilon)}$, for every $k \in E$.

(ii) Similarly, if we fix any $a \in A$ and consider any f which is a tour, then $x_{ka}(f) > 0$ for exactly one $k \in E$ (otherwise (2.7) implies that a city follows more than one city on the tour determined by f). Consequently, $\sum_{k \in E} x_{ka}(f) \geq \frac{(1-\epsilon)^{N-2}}{d(\epsilon)}$, for every $a \in A$.

We shall now consider the polytope $X(\epsilon)$ defined by the constraints corresponding to (C1)-(C3) in the perturbed process $\Gamma(\epsilon)$. Furthermore, we introduce additional constraints of the form

$$\begin{aligned} &(C4) \qquad \sum_{a \in A(i)} x_{ia} \geq c(\epsilon); \qquad i \in E \\ &(C5) \qquad \sum_{i \in E} x_{ia} \geq c(\epsilon); \qquad a \in A, \end{aligned}$$

where $c(\epsilon) = \frac{(1-\epsilon)^{N-2}}{d(\epsilon)}$. We now consider a subset of $X(\epsilon)$ defined by:

$$G(\epsilon) = \{\mathbf{x} \in X(\epsilon) \mid \mathbf{x} \quad satisfies \quad (C4)-(C5)\}. \tag{3.2}$$

The set $G(\epsilon)$ possesses a number of desirable properties with respect to the TSP which are captured in the following results.

Proposition 3.1 *Take any $\epsilon \in (0,1)$, and any deterministic policy f which is a tour, then $\mathbf{x}(f) \in G(\epsilon)$.*

Proof: By Theorem 2.1 we have that $\mathbf{x}(f) \in X(\epsilon)$. The satisfaction of constraints (C4)-(C5) follows from Remark 3.1.

□

Proposition 3.2 *Take any $\epsilon \in (0,1)$, and let $\mathbf{x} \in G(\epsilon)$ be such that $f_{\mathbf{x}} = T(\mathbf{x})$ is a deterministic policy in $\Gamma(\epsilon)$. Then $f_{\mathbf{x}}$ is a tour in the TSP.*

Proof: Let G be the underlying graph of the TSP (see Section 2), and $G_{f_{\mathbf{x}}}$ be the subgraph of G defined by:

$$arc(i,j) \in G_{f_{\mathbf{x}}} \longleftrightarrow f_{\mathbf{x}}(i) = j.$$

Note that by the definition of a deterministic policy as a function from E to A, it is sufficient to prove that $G_{f_{\mathbf{x}}}$ is a cycle. Now for each vertex $i \in E$ of $G_{f_{\mathbf{x}}}$ define $d^-(i)(d^+(i))$ to be the out (in)-degree of that vertex, namely the number of arcs emanating from (incident on) that vertex. Since $f_{\mathbf{x}}$ is a function on E we have

$$d^-(i) \equiv 1, \qquad i \in E, \tag{3.3}$$

and hence also

$$\sum_{i \in E} d^-(i) = \sum_{i \in E} d^+(i) = |E| = N. \tag{3.4}$$

Note, that $d^+(i)$ cannot be greater than 1 for any $i \in E$, because if k were such that $d^+(k) \geq 2$, then by (3.4) there

is some $j \in E$ such that $d^+(j) = 0$. That is, $f_{\mathbf{x}}(i) \neq j$ for all $i \in E$. By Theorem 2.1 part (iii), and (2.7) we now have that for all $i \in E$

$$x_{ij} = [\hat{T}(f_{\mathbf{x}})]_{ij} = p_i^*(f_{\mathbf{x}}) f_{\mathbf{x}}(i,j) = 0, \qquad (3.5)$$

where $[\mathbf{u}]_{ij}$ is the (ij)-th entry of the vector $\mathbf{u}$. However, (3.5) implies that constraints (C5) are violated by $\mathbf{x}$, contradicting the hypothesis: $\mathbf{x} \in G(\epsilon)$. Hence $d^+(i) \leq 1$ for all $i \in E$, which together with (3.4) yields

$$d^+(i) \equiv 1, \qquad i \in E. \qquad (3.6)$$

There are now two possibilities:

(a) $G_{f_{\mathbf{x}}}$ is a cycle, or

(b) $G_{f_{\mathbf{x}}}$ is a union of cycles $C_1, C_2, \ldots, C_m$.

We shall show that (b) is impossible, because otherwise there exists some $i \in E$ which belongs to a different cycle than the initial state 1. Without loss of generality assume that $1 \in C_1$. Now, in the Markov Chain induced by $f_{\mathbf{x}}$ in $\Gamma(\epsilon)$ (and starting at 1), the state i is not accessible from 1. Then, by similar argument as above $\sum_{j \in E} x_{ij} = 0$, which by violating C(4) implies that $\mathbf{x} \notin G(\epsilon)$, contradicting the hypotheses.

□

The above Propositions now lead to the following characterization of tours in the underlying Traveling Salesman Problem.

Theorem 3.1 *(i) Let $\epsilon \in (0,1)$ and f be a deterministic policy of $\Gamma(\epsilon)$. Then f is a tour in the TSP if and only if $\mathbf{x}(f) \in G(\epsilon)$.*

(ii) Let $\mathbf{x}$ be an extreme point of $X(\epsilon)$ which is also in $G(\epsilon)$. Then $f_{\mathbf{x}} = T(\mathbf{x})$ is a tour in the TSP.

Proof:

(i) Necessity follows immediately from Proposition 3.1. To establish sufficiency, Proposition 3.2 shows that we need only prove that

$$f = T(\mathbf{x}(f)) = T(\hat{T}(f)). \qquad (3.7)$$

Note that for a general deterministic policy in $\Gamma(\epsilon)$, equation (3.7) could be false, since $\hat{T}$ is not a 1:1 map of $C(S)$ onto $X(\epsilon)$. However, here f is such that $\mathbf{x}(f) \in G(\epsilon)$, and hence (with the help of (2.7)) we have that for all $i \in E$

$$0 < x_i(f) = \sum_{a \in A(i)} x_{ia}(f) = p_i^*(f).$$

Thus for every $(i,a) \in E \times A(i)$ we have

$$f(i,a) = \frac{x_{ia}(f)}{p_i^*(f)} = \frac{x_{ia}(f)}{x_i(f)}$$

which yields (3.7) as required.

(ii) C(4) implies that $x_i = \sum_{a \in A(i)} x_{ia} > 0$ for all $i \in E$. Since the matrix of the constraints (C1)-(C2) in $\Gamma(\epsilon)$ is not of full row rank, the extreme point $\mathbf{x}$ can have at most N positive entries. Thus for each $i \in E$ there is exactly one $a \in A(i)$ such that $x_{ia} > 0$. Hence $f_{\mathbf{x}}$ is a deterministic policy and the result follows by Proposition 3.2.

Remark 3.2 As a consequence of the above Theorem we know that all the tours of the TSP have representations as vertices of $G(\epsilon)$. The latter statement is valid since for every deterministic f, $\mathbf{x}(f)$ is a vertex of $X(\epsilon)$ (e.g., see Kallenberg [4], p. 115), and thereby of $G(\epsilon)$.

4. A New LP-Relaxation of the Traveling Salesman Problem

In this section we demonstrate that the linear program (LP2):

$$\max \sum_{k \in E} \sum_{a \in A(k)} r_{ka} x_{ka}$$

$$\text{subject to :} \qquad \mathbf{x} \in G(\epsilon),$$

for ϵ sufficiently small can be viewed as an LP-relaxation of the TSP. Towards this end we shall need the following notation: let f be a deterministic policy which is a tour, and let

$$v(f) = \frac{1}{N} \sum_{k \in E} r_{kf(k)} = \sum_{k \in E} \sum_{a \in A(k)} r_{ka} \left(\frac{\delta_{af(k)}}{N} \right),$$

that is, $(v(f))$ is the cost of this tour scaled by $\frac{1}{N}$. Also, let

$$v_\epsilon(f) = \sum_{k \in E} \sum_{a \in A(k)} r_{ka} x_{ka}(f),$$

that is, the value of the objective function of (LP2) corresponding to $\mathbf{x}(f)$ induced by f.

Lemma 4.1 *Consider $\epsilon \in (0,1)$, then*

$$\lim_{\epsilon \to 0^+} v_\epsilon(f) = v(f)$$

for every deterministic policy which is a tour.

Proof: Without loss of generality assume that $f(k) \equiv k+1$ (with $N+1$ defined to be 1). Now, from Lemma 3.2 we have

$$\begin{aligned} v_\epsilon(f) - v(f) &= r_{12} \left(\frac{1}{d(\epsilon)} - \frac{1}{N} \right) \\ &+ \sum_{k=2}^{N} r_{k,k+1} \left(\frac{(1-\epsilon)^{k-2}}{d(\epsilon)} - \frac{1}{N} \right) \qquad (4.1) \end{aligned}$$

from which the Lemma follows trivially.

□

We can now establish the main result of this paper.

Theorem 4.1 *There exists $\epsilon^* \in (0,1)$ such that if $\mathbf{x}^*$ is an optimal solution to (LP2) with $\epsilon < \epsilon^*$ for which $f^* = T(x^*)$ is a deterministic policy of $\Gamma(\epsilon)$, then f^* is an optimal tour in the TSP.*

Proof: Let v^* be the cost of an optimal tour scaled by $\frac{1}{N}$, and choose $\delta > 0$ and such that if f is any suboptimal tour, then

$$v^* - v(f) \geq \delta > 0. \tag{4.2}$$

Further, let $r^* = \{\max |r_{ka}| : (k,a) \in E \times A(k)\}$ and define the sequence $\beta^k(\epsilon)$; $k = 1, \ldots, N$ by

$$\beta^1(\epsilon) = \beta^2(\epsilon) \quad and \quad \beta^k(\epsilon) = \left| \frac{(1-\epsilon)^{k-2}}{d(\epsilon)} - \frac{1}{N} \right| for$$

k = 2,...,N, and let

$$\beta(\epsilon) = r^* \sum_{k=1}^{N} \beta^k(\epsilon).$$

It should be clear that $\beta(\epsilon) \to 0$ as $\epsilon \to 0^+$. Now suppose that f^* is a suboptimal tour, and let g be a deterministic policy that is an optimal tour of the TSP, then by Theorem 2.1 and optimality of $\mathbf{x}^*$

$$v_\epsilon(g) = \phi_1(g) = \textstyle\sum_{i \in E} \sum_{a \in A(i)} r_{ia} x_{ia}(g)$$

$$\leq \textstyle\sum_{i \in E} \sum_{a \in A(i)} r_{ia} x^*_{ia} = \phi_1(f^*) = v_\epsilon(f^*).$$

On the other hand, if ϵ^* is chosen so that $\beta(\epsilon*) < \frac{\delta}{2}$, then from (4.1)-(4.2) we have for $\epsilon < \epsilon*$ that

$$\begin{aligned} v_\epsilon(f^*) &\leq v(f^*) + \beta(\epsilon) \\ &\leq v^* - \delta + \beta(\epsilon) \\ &< v^* - \beta(\epsilon) = v(g) - \beta(\epsilon) \\ &\leq v_\epsilon(g), \end{aligned}$$

which contradicts (4.1). Thus f^* is an optimal tour. □

Remark 4.3

From the above proof it should be clear that if all the data are integer, it is not hard to compute an ϵ^* for which Theorem 4.2 will hold. In particular, any $\epsilon \in (0,1)$ such that

$$\beta(\epsilon) = r^* \sum_{k-1}^{N} \beta^k(\epsilon) < \frac{1}{2} \tag{4.3}$$

will do the job.

The preceding results demonstrate that for ϵ sufficiently small the following mathematical program (MP1) solves the Traveling Salesman Problem:

$$\max \sum_{k \in E} \sum_{a \in A(i)} r_{ka} x_{ka}$$

s.t.

1. $\mathbf{x} \in G(\epsilon)$
2. $x_{ka} / \sum_{a \in A(k)} x_{ka} \in \{0,1\}, \quad k \in E, \quad a \in A(k).$

Of course, an optimal solution of (MP1) can be obtained from an optimal solution to the mixed linear-integer program (MP2) below:

$$\max \sum_{k \in E} \sum_{a \in A(k)} r_{ka} x_{ka}$$

s.t.

1. $\mathbf{x} \in G(\epsilon)$
2. $x_{ka} \leq i_{ka}; \quad k \in E, \quad a \in A(k)$
3. $\sum_{a \in A(k)} i_{ka} = 1; \quad k \in E$
4. $i_{ka} \in \{0,1\}; \quad k \in E, \quad a \in A(k).$

Remark 4.4

It now follows that for ϵ sufficiently small, the linear program (LP2) can be regarded as a well-solved relaxation of either (MP1) or (MP2). If its optimal solution $\mathbf{x}^*$ yields a deterministic policy $f^* = T(\mathbf{x}^*)$, then the TSP is solved.

5. Practical Considerations

The previous section presents some new formulations of the TSP. It is not known what computational advantages can be gained from these. It is hoped that some methods used for standard formulation of the TSP can be adapted advantageously for our formulations. One immediate application of the above formulation is suggested below for accelerating some existing TSP algorithms.

A lot of successful TSP algorithms produce intermediate solutions which might consist of several disjoint subtours rather than a single tour (for example modern polyhedral algorithms). When such a solution is obtained it must then be determined whether it represents a single tour or a collection of subtours. The latter can be computationally intensive. Below we propose an algorithm to do this which promises to be efficient.

Let f be a deterministic policy in the MDP, $\Gamma(\epsilon)$.

It follows from Section 3 that f represents a tour if and only if $p_i^*(f) > 0$ for all i. Moreover, if f is *not* a tour then

(i) if state 1 is contained in a cycle (in the subgraph G_f) then $p_i^*(f) > 0$ for all i in this cycle, and is identically 0 elsewhere

(ii) If state 1 is not in a given cycle, then $p_i^*(f) = 0$ for all i in this cycle.

The following algorithm identifies the subtour containing state 1;

Step 1: Form $P(f)$ in $\Gamma(\epsilon)$.

Step 2: Solve $\pi^T P(f) = \pi^T$, $\sum_i \pi_i = 1$; denote the unique solution by $p^*(f)$.

Step 3: Let C_1 denote the subtour containing state 1. $C_1 = \{i : p_i^*(f) > 0\}$. If $C_1 = V$ (set of vertices of the TSP), then f is a tour.

Remarks: Note that

1. To identify other subtours, pick j s.t. $p_j^*(f) = 0$. Rename the states so that j is now state 1 and apply the algorithm. It will yield a subtour containing j.
2. Nearly all the computational work is in **Step 2** which involves solving a very sparse $n \times n$ linear system.

6. References

1. Derman, C. (1970). Finite Markovian Decision Processes. Academic Press, New York.

2. Derman, C. and Klein, M. (1966). Surveillance of Multicomponent Systems: A Stochastic Traveling Salesman Problem. NRLQ, 13, pp. 103-111.

3. Hordijk, A. and Kallenberg, L. C. M. (1984). Constrained Undiscounted Stochastic Dynamic Programming. Math. Oper. Res., 9 pp. 276-289.

4. Kallenberg, L. C. M. (1983). Linear Programming and Finite Markovian Control Problems. Mathematical Centre Tracts #148, Amsterdam.

5. Lawler, E. O., Lenstra, J. K., Rinnoy Kan, A. H. G. and Schmoys, D. B. (1985). The Traveling Salesman Problem. Wiley, New York.