Mathematical Programming 59 (1993) 413-420
North-Holland

A note on the prize collecting traveling salesman problem

Daniel Bienstock
Department of Industrial Engineering and Research, Columbia University, New York, USA

Michel X. Goemans
Department of Mathematics, MIT, Cambridge, MA, USA

David Simchi-Levi
Department of Industrial Engineering and Operations Research, Columbia University, New York, USA

David Williamson
Laboratory for Computer Science, MIT, Cambridge, MA, USA

Received October 1990
Revised manuscript received 19 November 1991

We study the version of the prize collecting traveling salesman problem, where the objective is to find a tour that visits a subset of vertices such that the length of the tour plus the sum of penalties associated with vertices not in the tour is as small as possible. We present an approximation algorithm with constant bound. The algorithm is based on Christofides' algorithm for the traveling salesman problem as well as a method to round fractional solutions of a linear programming relaxation to integers, feasible for the original problem.

Key words: Linear programming, prize collecting, rounding fractional solutions, traveling salesman problem, worst-case analysis.

1. Introduction[1]

Let $G=(V, E)$ be a complete undirected graph with vertex set V and edge set E. Associated with each edge $e=\{i, j\}\in E$ is a cost c_e and with each vertex $i\in V$ a nonnegative penalty π_i. The edge costs are assumed to satisfy the triangle inequality, that is, $c_{\{i,j\}}\leq c_{\{i,k\}}+c_{\{k,j\}}$ for all $i, j, k\in V$. In this paper we consider a simplified version of "the prize collecting" traveling salesman problem, namely, to find a tour that visits a subset of the vertices such that the length of the tour plus the sum of penalties of all vertices not in the tour is as small as possible.

Correspondence to: Prof. D. Bienstock, Department of Industrial Engineering and Operations Research, Columbia University in the City of New York, New York, NY 10027-6699, USA.
Research supported in part by ONR contract N00014-90-J-1649 and NSF contract DDM-8922712.
[1] An earlier version of this paper, due to the first and the third authors, contained an analysis similar to the one described here with a slightly worse performance bound. This bound was improved by the second and the fourth authors, independently.

In the general version of the problem, introduced by Balas (1989), the edge costs are not assumed to satisfy the triangle inequality. Further, associated with each vertex there is a certain reward or prize, and in the optimization problem one must choose the subset of vertices to be visited so that the total reward is at least a given parameter W_0.

As is well-known, there is no polynomial time approximation algorithm for the traveling salesman problem, with bounded ratio, unless P = NP. Consequently, the same holds for the general prize collecting traveling salesman problem. However, one sub-class of the traveling salesman problem for which there is a fixed-bound polynomial time approximation is that where the edge-costs satisfy the triangle inequality (Christofides, 1976, or Johnson and Papadimitriou, 1985) and one wonders if a similar situation holds for the prize-collecting traveling salesman problem (with arbitrary vertex prizes and penalties).

As a a first step in the study of this problem we consider the variant of the problem where the minimum reward constraint is dropped and obtain a positive answer. Unlike the Christofides heuristic, our algorithm is not a combinatorial one: it uses the ellipsoid method. We remark that the version of the problem with the reward constraint appears much more difficult.

Let Z^* be the optimal solution to the prize collecting traveling salesman problem. For any S, $S \subseteq V$, let $L(S)$ be the length of the optimal traveling salesman tour through S and $L^C(S)$ be the length of the tour, that includes all vertices in S, produced by Christofides' algorithm. For any heuristic H for the prize collecting traveling salesman problem, let Z^H be the cost of the solution produced by that heuristic. We define the worst-case performance of a heuristic as an upper bound on the worst-case ratio of the cost of the heuristic solution to the cost of the optimal solution.

For our analysis, it is convenient to formulate the prize collecting traveling salesman problem in the following way: Let $Z^*(j)$ be the optimal solution to the prize collecting traveling salesman problem when vertex j must be in the tour. Clearly,

$$Z^* = \min\left\{ \sum_{i \in V} \pi_i, \min_{j \in V}\{Z^*(j)\} \right\}.$$

In what follows we use the integer program described below, whose solution is $Z^*(j)$. Let y_i be one if vertex $i \in V$ is in the tour and zero otherwise. Let x_e be one if edge e is in the tour and zero otherwise. For every subset of vertices S, let $\delta(S)$ be the set of edges with one end in S and the other in $V \backslash S$. Then, the prize collecting traveling salesman problem, when vertex j must be in the tour, can be formulated as follows:

Problem $P_1(j)$:

$$Z^*(j) = \text{minimize} \quad \sum_{e \in E} c_e x_e + \sum_{i \in V} \pi_i (1 - y_i)$$

$$\text{subject to} \quad \sum_{e \in \delta(\{i\})} x_e = 2 y_i \quad \forall i \in V, \tag{1.1}$$

$$\sum_{e \in \delta(S)} x_e \geq 2y_i \quad \forall i \in V,\ S \subset V$$

$$\text{such that } |S \cap \{i, j\}| = 1, \tag{1.2}$$

$$0 \leq x_e \leq 1 \text{ and integer}, \tag{1.3}$$

$$0 \leq y_i \leq 1 \text{ and integer} \quad \forall i \neq j, \tag{1.4}$$

$$y_j = 1. \tag{1.5}$$

Constraint (1.1) guarantees that if i is not visited y_i gets the value zero and therefore in the objective function we incurred a penalty π_i. Constraint (1.2) ensures that if $|S \cap \{i, j\}| = 1$, and i is also in the tour then at least two edges from the cut set $\delta(S)$ should have $x_e = 1$.

In this paper we present a three-step heuristic for the prize collecting traveling salesman problem: In the first step we solve the Linear Programming (LP) relaxation of Problem $P_1(j)$ (using the ellipsoid method—the separation problem is a min-cut problem) for each $j \in V$. In the second step we transform the solution of the LP relaxation of problem $P_1(j)$ into a feasible solution to the prize collecting traveling salesman problem, thus, constructing $|V|$ feasible solutions to it, each corresponding to a different $P_1(j)$. Finally, we choose the best of these $|V|$ possible solutions, or the solution in which no vertex is visited, whichever yields the better cost. The construction of the feasible solution in the second step is based on Christofides' algorithm for the traveling salesman problem as well as a method to round fractional solutions obtained in the first step to integers. The algorithm is called Modified LP relaxation (MLP) heuristic, and we prove that its worst-case performance is 2.5, that is, that

$$Z^{\text{MLP}}/Z^* \leq 2.5.$$

For another example of an algorithm that rounds fractional solutions of a linear program to integers that are "nearly" feasible see Lenstra, Shmoys and Tardos (1987).

2. Preliminaries

In what follows we make use of two recent results concerning a lower bound on the length of the optimal traveling salesman tour. Consider the traveling salesman problem defined on the set of vertices V. A well-known lower bound on the length of the optimal tour is given by Held and Karp (1971) and is the solution to the following LP:

Problem P_2:

$$Z_{\text{HK}} = \text{minimize} \quad \sum_{e \in E} c_e x_e$$

$$\text{subject to} \quad \sum_{e \in \delta(S)} x_e \geq 2 \quad \forall S \subset V,\ S \neq \emptyset, \tag{2.1}$$

$$\sum_{e \in \delta(\{i\})} x_e = 2 \quad \forall i \in V, \tag{2.2}$$

$$0 \leqslant x_e. \tag{2.3}$$

Note that Problem P_2 with the additional constraint that the x_e's are integers is the integer linear programming formulation of the traveling salesman problem. This is true since constraint (2.1) ensures that at least two edges from every cut set $\delta(S)$ must be in the solution while constraint (2.2) guarantees that exactly two edges are connected to every vertex.

Theorem 2.1 (Wolsey, 1980, Shmoys and Williamson, 1988).

$$Z_{\mathrm{HK}}/L(V) \geqslant Z_{\mathrm{HK}}/L^{\mathrm{C}}(V) \geqslant \tfrac{2}{3}. \qquad \square$$

For the next lemma we need to formulate the following LP. Associated with each vertex $i \in V$ is a given number r_i which is either zero or two. Let $V_2 = \{i \in V \mid r_i = 2\}$.

Problem P_3:

$$\text{minimize} \quad \sum_{e \in E} c_e x_e$$

$$\text{subject to} \quad \sum_{e \in \delta(S)} x_e \geqslant 2 \quad \forall S \subset V \text{ such that } V_2 \cap S \neq \emptyset,\ V_2 \cap (V \setminus S) \neq \emptyset, \tag{2.4}$$

$$\sum_{e \in \delta(\{i\})} x_e = r_i \quad \forall i \in V, \tag{2.5}$$

$$0 \leqslant x_e. \tag{2.6}$$

In the Appendix we provide a short proof of the following result:

Lemma 2.2 (Goemans and Bertsimas, 1990). *The optimal solution value to Problem P_3 is unchanged if we solve it without constraint* (2.5).

Lemma 2.2 implies that in Problem P_2 (Held and Karp lower bound on the length of the optimal traveling salesman tour) one can ignore constraint (2.2), without changing the value of the lower bound. This can be seen by choosing $r_i = 2$ for all $i \in V$ and applying the lemma.

3. Analysis of the MLP heuristic

The MLP heuristic generates $|V|$ different solutions to the prize collecting traveling salesman problem by solving the LP relaxation of Problem $P_1(j)$ for every $j \in V$. The jth solution associated with Problem $P_1(j)$ is generated in the following way.

Let $\bar{x}$ and $\bar{y}$ be the optimal solution to the LP relaxation of Problem $P_1(j)$. Define new vectors $\hat{x}$ and $\hat{y}$ as follows:

$$\hat{x}_e = \tfrac{5}{3}\bar{x}_e \quad \forall e \in E, \tag{3.1}$$

and for any $i \in V$

$$\hat{y}_i = \begin{cases} 1, & \text{if } \bar{y}_i \geq \frac{3}{5}, \\ 0, & \text{otherwise.} \end{cases}$$

Observe that by definition of $\hat{y}_i$ we have

$$1 - \hat{y}_i \leq \tfrac{5}{2}(1 - \bar{y}_i) \quad \forall i \in V. \tag{3.2}$$

Notice that we are not claiming that $\hat{x}$, $\hat{y}$ is a feasible solution to the LP relaxation of Problem $P_1(j)$.

Let $T = \{i \in V \mid \hat{y}_i = 1\}$. The MLP heuristic constructs a traveling salesman tour through all vertices in T using Christofides' algorithm and therefore charges penalty costs for all vertices not in T. Define

$$Z^{\mathrm{MLP}}(j) = L^{\mathrm{C}}(T) + \sum_{i \in V} \pi_i(1 - \hat{y}_i), \tag{3.3}$$

that is, $Z^{\mathrm{MLP}}(j)$ is the cost of the solution produced by the MLP heuristic, assuming j is in the tour.

The MLP heuristic chooses the best solution among all such solutions or the solution in which no vertex is visited, whichever yields the minimum cost. Hence,

$$Z^{\mathrm{MLP}} = \min\left\{ \sum_{i \in V} \pi_i, \min_{j \in V}\{Z^{\mathrm{MLP}}(j)\} \right\}.$$

Theorem 3.1. $Z^{\mathrm{MLP}}/Z^* \leq 2.5$.

Proof. It is sufficient to show that $Z^{\mathrm{MLP}}(j)/Z^*(j) \leq 2.5$, for every j. First, note that the following LP yields the Held and Karp lower bound on the length of the optimal traveling salesman tour through the subset of vertices T:

Problem P_4:

$$\text{minimize} \quad \sum_{e \in E} c_e x_e \tag{3.4}$$

$$\text{subject to} \quad \sum_{e \in \delta(S)} x_e \geq 2 \quad \forall S \subset V \text{ such that } T \cap S \neq \emptyset,\ T \cap (V \setminus S) \neq \emptyset, \tag{3.5}$$

$$\sum_{e \in \delta(\{i\})} x_e = 2 \quad \forall i \in T, \tag{3.6}$$

$$\sum_{e \in \delta(\{i\})} x_e = 0 \quad \forall i \notin T, \tag{3.7}$$

$$x_e \geq 0. \tag{3.8}$$

By Lemma 2.2, the solution value to Problem P_4 is unchanged when we take out constraints (3.6) and (3.7). Let Problem P_5 be (3.4), (3.5) and (3.8), and denote by $\acute{x}$ its optimal solution. Using Theorem 2.1, we have

$$L^{\mathrm{C}}(T) \leqslant \tfrac{3}{2} \sum_{e \in E} c_e \acute{x}_e. \tag{3.9}$$

We now show that $\hat{x}$ is feasible for Problem P_5. Clearly $\hat{x}$ satisfies (3.8). To prove that it also satisfies (3.5) consider any $S \subset V$ such that $i \in T \cap S$ and $j \in T \backslash S$. By feasibility of $\bar{x}$ in Problem $P_1(j)$ and the definition of T we have, using constraint (1.2) and equation (3.1),

$$\sum_{e \in \delta(S)} \bar{x}_e \geqslant 2\bar{y}_i \geqslant 2(\tfrac{3}{5}) = \tfrac{6}{5} \quad \forall S \subset V \text{ such that } T \cap S \neq \emptyset,\ T \cap (V \backslash S) \neq \emptyset.$$

Hence, for any $S \subset V$ such that $T \cap S \neq \emptyset$ and $T \cap (V \backslash S) \neq \emptyset$ we have

$$\sum_{e \in \delta(S)} \hat{x}_e = \tfrac{5}{3} \sum_{e \in \delta(S)} \bar{x}_e \geqslant 2,$$

and therefore $\hat{x}$ satisfies (3.5). Consequently, since $\acute{x}_e$ is optimal

$$\sum_{e \in E} c_e \hat{x}_e \geqslant \sum_{e \in E} c_e \acute{x}_e. \tag{3.10}$$

Hence,

$$\begin{aligned}
Z^{\mathrm{MLP}}(j) &= L^{\mathrm{C}}(T) + \sum_{i \in V} \pi_i (1 - \hat{y}_i) \\
&\leqslant \tfrac{3}{2} \sum_{e \in E} c_e \acute{x}_e + \sum_{i \in V} \pi_i (1 - \hat{y}_i) \quad \text{(from (3.9))} \\
&\leqslant \tfrac{3}{2} \sum_{e \in E} c_e \hat{x}_e + \sum_{i \in V} \pi_i (1 - \hat{y}_i) \quad \text{(from (3.10))} \\
&\leqslant \tfrac{3}{2} \sum_{e \in E} c_e \tfrac{5}{3} \bar{x}_e + \tfrac{5}{2} \sum_{i \in V} \pi_i (1 - \bar{y}_i) \quad \text{(from (3.1), (3.2))} \\
&= \tfrac{5}{2} \left\{ \sum_{e \in E} c_e \bar{x}_e + \sum_{i \in V} \pi_i (1 - \bar{y}_i) \right\} \\
&\leqslant \tfrac{5}{2} Z^*(j). \qquad \square
\end{aligned}$$

4. Extensions

The method developed here can be used for other versions of the prize collecting problem. For instance, consider the Steiner tree version of the prize collecting problem. In it, the objective is to find a Steiner tree that spans a subset of the vertices such that its cost plus the penalty cost associated with all vertices not in the tree is minimized. Note that this problem, when vertex j must be in the tree, can be formulated similarly to Problem $P_1(j)$, without constraint (1.1) and by replacing the right hand side of (1.2) with y_i. Solving the LP relaxation of this problem and rounding fractional solutions in a similar way to the MLP heuristic provides a solution for which its worst-case performance is 3.

Appendix: A proof of Lemma 2.2.

Here we provide a short proof of Lemma 2.2. Goemans and Bertsimas (1989) have obtained a more general result, whose proof relies heavily on a powerful theorem of Lovasz (1976). Our proof is similar to theirs with the exception that we use a much simpler result of Lovasz (1979). In this book of problems (exercise No. 6.51) Lovasz presents the following result, together with a short proof.

Lemma A. *Let G be an Eulerian multigraph and $s \in V(G)$, such that G is k-connected between any two vertices different from s. Then, for any neighbor u of s, there exists another neighbor w of s, such that the multigraph obtained from G by removing $\{s, u\}$ and $\{s, w\}$, and adding a new edge $\{u, w\}$ (the* splitting-off *operation) is also k-connected between any two vertices different from s.* □

Lovasz's proof of Lemma A can be easily modified to yield the following:

Lemma B. *Let G be an Eulerian multigraph, $Y \subseteq V(G)$ and $s \in V(G)$, such that G is k-connected between any two vertices of Y different from s. Then, for any neighbor u of s, there exists another neighbor w of s, such that the multigraph obtained from G by removing $\{s, u\}$ and $\{s, w\}$, and adding a new edge $\{u, w\}$ is also k-connected between any two vertices of T different from s.* □

Proof of Lemma 2.2. Let $V_0 = V \setminus V_2$, that is, $V_0 = \{i \in V \mid r_i = 0\}$. Let Problem P_3' be Problem P_3 without (2.5). Finally, let $\tilde{x}$ be a rational vector feasible for Problem P_3', chosen such that

(1) $\tilde{x}$ is optimal for Problem P_3', and

(2) subject to (1), $\sum_{e \in E} \tilde{x}_e$ is minimized.

Let M be a positive integer, large enough so that $\tilde{v} = 2M\tilde{x}$ is a vector of even integers. We may regard $\tilde{v}$ (with a slight abuse of notation) as the incidence vector of the edge-set $\tilde{E}$ of a multigraph $\tilde{G}$ with vertex set V. Clearly, $\tilde{G}$ is Eulerian, and by (2.4), it is $4M$-connected between any two elements of V_2.

Now suppose that for some vertex s, $\sum_{e \in \delta(\{s\})} \tilde{x}_e > r_s$ (i.e., s has a degree larger than $2Mr_s$ in $\tilde{G}$). Let us apply Lemma B to s and any neighbor u of s (where $Y = V_2$), and let $\tilde{H}$ be the resulting multigraph, with incidence vector $\tilde{z}$.

Clearly,

$$\sum_{e \in E} c_e \tilde{z}_e \leqslant \sum_{e \in E} c_e \tilde{v}_e,$$

and so

$$\sum_{e \in E} c_e \frac{\tilde{z}_e}{2M} \leqslant \sum_{e \in E} c_e \tilde{x}_e.$$

Moreover,

$$\sum_{e \in E} \frac{\tilde{z}_e}{2M} = \sum_{e \in E} \tilde{x}_e - \frac{1}{2M}.$$

Hence, by the choice of $\tilde{x}$, $z = \tilde{z}/(2M)$ cannot be feasible for Problem P_3'.

If $s \in V_0$, then by Lemma B, z is feasible for Problem P_3'. Thus, we must have $s \in V_2$, and, in fact $\sum_{e \in \delta(\{t\})} \tilde{x}_e = 0$ for all $t \in V_0$. In other words $\tilde{E}$ spans precisely V_2, $\tilde{G}$ is $4M$-connected, and $\sum_{e \in \delta(\{s\})} \tilde{x}_e \geqslant 4M+2$. But we claim now that the multigraph $\tilde{H}$ is $4M$-connected. For by Lemma B, it could only fail to be $4M$-connected between s and some other vertex, but the only possible cut of size less than $4M$ is the one separating s from $V \backslash \{s\}$. Since this cut has at least $4M$ edges, the claim is proved as desired. Consequently, again we obtain that z is feasible for Problem P_3', a contradiction.

In other words, $\sum_{e \in E} \tilde{v}_e = 2Mr_i$ for all i, that is, (2.5) holds as required. □

References

[1] E. Balas, "The prize collecting traveling salesman problem," *Networks* 19 (1989) 621–636.

[2] N. Christofides, "Worst-case analysis of a new heuristic for the traveling salesman problem," Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University (Pittsburgh, PA, 1976).

[3] M.X. Goemans and D.J. Bertsimas, "On the parsimonious property of connectivity problems," in: *Proceeding of the 1st ACM–SIAM Symposium on Discrete Algorithms* (San Francisco, CA, 1990).

[4] M. Held and R.M. Karp, "The traveling salesman problem and minimum spanning trees," *Operations Research* 18 (1970) 1138–1162.

[5] D.S. Johnson and C.H. Papadimitriou, "Performance guarantees for heuristics," in: E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys, eds., *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization* (Wiley, New York, 1985) pp. 145–180.

[6] J.K. Lenstra, D.B. Shmoys and E. Tardos, "Approximation algorithms for scheduling unrelated parallel machines," in: *Proceedings of the 28th Annual IEEE Symposium on the Foundation of Computer Science* (Computer Society Press of the IEEE, New York, 1987) pp. 217–224.

[7] L. Lovasz, "On some connectivity properties of Eulerian multigraphs," *Acta Mathematica Academiae Scientiarum Hungaricae* 28 (1976) 129–138.

[8] L. Lovasz, *Combinatorial Problems and Exercises* (North-Holland, Amsterdam, 1979).

[9] D. Shmoys and D. Williamson, "Analyzing the Held–Karp TSP bound: A monotonicity property with application," to appear in: *Information Processing Letters* (1988).

[10] L. Wolsey, "Heuristic analysis, linear programming and branch and bound," *Mathematical Programming Study* 13 (1980) 121–134.