Operations Research Letters 10 (1991) 291–295
North-Holland

July 1991

Analysis of Christofides' heuristic: Some paths are more difficult than cycles

J.A. Hoogeveen

Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, Netherlands

Received March 1990
Revised September 1990

For the traveling salesman problem in which the distances satisfy the triangle inequality, Christofides' heuristic produces a tour whose length is guaranteed to be less than $\frac{3}{2}$ times the optimum tour length. We investigate the performance of appropriate modifications of this heuristic for the problem of finding a shortest Hamiltonian path. There are three variants of this problem, depending on the number of prespecified endpoints: zero, one, or two. It is not hard to see that, for the first two problems, the worst-case performance ratio of a Christofides-like heuristic is still $\frac{3}{2}$. For the third case, we show that the ratio is $\frac{5}{3}$ and that this bound is tight.

traveling salesman problem; Hamiltonian cycle; Hamiltonian path; approximation algorithm; worst-case analysis.

1. Introduction

The traveling salesman problem (TSP) is defined as follows. Given a complete undirected graph G on n vertices and a distance c_{ij} for each edge $\{i, j\}$, find a Hamiltonian cycle (i.e., a cycle that traverses each vertex exactly once) of minimum total length. This problem is NP-hard, and much attention has been paid to the design and analysis of approximation algorithms for its solution. An indication for the quality of an approximation algorithm is its worst-case performance ratio. For an instance of the TSP, let C^* denote the optimal Hamiltonian cycle, and let C^{A} denote the Hamiltonian cycle produced by an algorithm A. For any edge set E, let $c(E)$ be the sum of the distances of all edges in E, so that $c(C)$ denotes the length of a cycle C. The worst-case performance ratio ρ of algorithm A is then defined as the supremum of $c(C^{\mathrm{A}})/c(C^*)$ over all instances, and A is said to be a *ρ-approximation algorithm*.

Sahni and Gonzalez [1976] have shown that, in the case of general distances, no polynomial-time algorithm for the TSP can have a constant worst-case performance ratio, unless P = NP. We will consider the case in which the distance satisfy the triangle inequality, i.e., $c_{ij} + c_{jk} \geqslant c_{ik}$ for all i, j, k. We will also assume that $c_{ij} > 0$ for all $i \neq j$. For this case, Christofides [1976] proposed a $\frac{3}{2}$-approximation algorithm that requires $O(n^3)$ time, and no polynomial-time algorithm with a better worst-case performance ratio is known.

We will be concerned with a problem closely related to the TSP, namely, the problem of finding a Hamiltonian *path* (i.e., a path that contains each vertex exactly once) of minimum total length. There are three variants of this problem, depending on the number of prespecified endpoints of the path. We introduce the following notation: P^* denotes an optimal Hamiltonian path without fixed endpoints, P_s^* denotes an optimal Hamiltonian path with a single fixed endpoint s, and P_{st}^* denotes an optimal Hamiltonian path with fixed endpoints s and t.

We formulate a Christofides-like algorithm for each of these problems. It is not hard to see that, for the first two problems, the heuristic is still a $\frac{3}{2}$-approximation algorithm. For the third case, we show that the ratio is $\frac{5}{3}$ and that this bound is tight. This answers a question posed by Johnson and Papadimitriou [1985].

We recall Christofides' heuristic for the TSP in Section 2. In Section 3, we present the three modified heuristics for finding Hamiltonian paths and analyze the two easy cases. Section 4 deals with the case of two fixed endpoints. In Section 5, we analyze a slightly different heuristic for the third case.

2. Hamiltonian cycles

Chritofides' heuristic for the determination of a short Hamiltonian cycle proceeds as follows:

(1) Construct a minimum spanning tree T of G.

(2) Construct a minimum perfect matching M on the set S of all odd-degree vertices in T.

(3) Find an Eulerian tour in the Eulerian graph that is the union of T and M. A graph is Eulerian if it contains a tour that traverses each edge exactly once, and such an Eulerian tour exists if and only if the graph is connected and each of its vertices is of even degree. Note that the union of T and M satisfies these requirements.

(4) Transform the Eulerian tour into a Hamiltonian cycle by applying shortcuts. A shortcut is a contraction of two edges $\{i, j\}$ and $\{j, k\}$ to a single edge $\{i, k\}$. This cycle will be denoted by C^C.

The triangle inequality implies that $c(C^C) \leqslant c(T) + c(M)$. It is obvious that $c(T) < c(C^*)$, and we will argue below that $c(M) \leqslant \frac{1}{2}c(C^*)$. It now follows that

$$c(C^*) \leqslant c(C^C) < \tfrac{3}{2}c(C^*).$$

Hence, Christofides' heuristic is a $\frac{3}{2}$-approximation algorithm. Cornuéjols and Nemhauser [1978] show that the precise worst-case ratio is equal to $(3\lfloor\frac{1}{2}n\rfloor - 1)/(2\lfloor\frac{1}{2}n\rfloor)$.

As to the length of the matching, let $1, \ldots, 2m$ be the odd-degree vertices in T, and suppose that they occur in C^* in this order. Consider the following two edge-disjoint subsets of C^*: E_1, containing all the edges between 1 and 2, 3 and 4, ..., and $2m-1$ and $2m$; and E_2, containing all the edges between 2 and $3, \ldots, 2m-2$ and $2m-1$, and $2m$ and 1. Taking shortcuts yields two perfect matchings

$$M_1 = \{\{1, 2\}, \{3, 4\}, \ldots, \{2m-1, 2m\}\}$$

and

$$M_2 = \{\{2, 3\}, \ldots, \{2m-2, 2m-1\}, \{2m, 1\}\}.$$

Due to the triangle inequality, we have

$$c(M_1) + c(M_2) \leqslant c(E_1) + c(E_2) = c(C^*).$$

Hence, $c(M) \leqslant \frac{1}{2}c(C^*)$.

3. Hamiltonian paths

For the determination of a Hamiltonian path, Christofides' heuristic has to be adapted to ensure that the union of the tree T and the matching M contains exactly two vertices of odd degree. In addition, any prespecified endpoint has to be among those odd-degree vertices. We present the following modification of Christofides' heuristic:

(1) Construct a minimum spanning tree T of the graph G.

(2) First, determine the set S of vertices that are of *wrong* degree in T, i.e., the collection of fixed endpoints of even degree and other vertices of odd degree. Next, construct a minimum matching M on S that leaves $2-k$ vertices exposed, where k is the number of fixed endpoints. We note that such a matching can be found by constructing a minimum perfect matching on S augmented with $2-k$ dummy vertices in an obvious fashion.

(3) Consider the graph that is the union of T and M. This graph is connected and has either two or zero odd-degree vertices. The latter case occurs only if there is a single fixed endpoint that belongs to S and is left exposed by M; in this case, delete an arbitrary edge incident to this vertex. Find an Eulerian path in the resulting graph. This path traverses each edge exactly once and has the two odd-degree vertices as its endpoints.

(4) Transform the Eulerian path into a Hamiltonian path by applying shortcuts. This path will be denoted by P^C, P_s^C, or P_{st}^C, depending on the number of prespecified endpoints.

We analyze the performance of this heuristic by establishing an upper bound on the length of the minimum matching in terms of the length of the optimal Hamiltonian path. For $k = 0$ or $k = 1$, the analysis is very similar to the one given in Section 2.

Theorem 1. $c(P^C)/c(P^*) < \frac{3}{2}$.

Proof. The theorem follows from the observations that $c(T) \leqslant c(P^*)$, which is obvious, and that $c(M) < \frac{1}{2}c(P^*)$, which we will prove. Let $1,\ldots,2m$ be the odd-degree vertices in T, and suppose that they occur in P^* in this order. Again, consider two edge-disjoint subsets E_1 and E_2 of P^*: E_1 contains all the edge between 1 and 2, 3 and $4,\ldots,2m-3$ and $2m-2$; and E_2 contains all the edges between 2 and 3, 4 and $5,\ldots,2m-2$ and $2m-1$. Taking shortcuts yields two matchings M_1 and M_2, containing $m-1$ edges and leaving two vertices exposed, and having total length $c(M_1)+c(M_2) < c(P^*)$. Hence, $c(M) < \frac{1}{2}c(P^*)$. □

Theorem 2. $c(P_s^C)/c(P_S^*) \leqslant \frac{3}{2}$.

Proof. We have to prove that $c(M) \leqslant \frac{1}{2}c(P_S^*)$. Suppose the endpoints of P_s^* are s and i. We distinguish two special vertices in P_s^*: j is the first vertex of odd degree after s, and k is the last vertex of odd degree. Consider two cases.

(1) The fixed endpoint s has odd degree in T. This means that $s \notin S$, the set of vertices of wrong degree. The set of edges in P_s^* between j and k can be partitioned into two disjoint subsets, either of which gives a matching of the desired form after shortcutting. Hence, we have again that $C(M) < \frac{1}{2}c(P_s^*)$, so that $c(P_s^C)/c(P_s^*) < \frac{3}{2}$.

(2) The fixed endpoint s has even degree in T. This means that $s \in S$. The set of edges in P_s^* between s and k can be partitioned into two disjoint subsets, which, after shortcutting, give matchings M_1 and M_2, respectively. Let $\{s, j\}$ be the edge contained in M_1.

First, suppose $C(M_1) \leqslant c(M_2)$. M_1 is a matching of the desired form, so that again $c(M) \leqslant c(M_1) \leqslant \frac{1}{2}c(P_s^*)$.

Next, suppose $c(M_2) \leqslant c(M_1)$. M_2 is a matching on all odd-degree vertices in T. Hence, all vertices in the union of T and M_2 are of even degree, so that there exists an Eulerian cycle. Removing an edge containing s yields an Eulerian path with endpoint s, which can be shortcut to obtain a Hamiltonian path with endpoint s. Hence, $c(P_s^C) < c(M_2) + c(T) \leqslant \frac{3}{2}c(P_s^*)$. □

Note that, in the above proof, the two subcases of (2) are not disjoint. If M_1 and M_2 are both optimal, then either can be chosen. This implies that the worst-case bound of $\frac{3}{2}$ can be attained. It is not hard, however, to take precautions so as to ensure that $c(P_s^C) < \frac{3}{2}c(P_s^*)$.

Consider the case that s has even degree in T. Determine the longest edge in T containing s, say $\{s, k\}$. Remove this edge from T, remove s from S and add a new vertex x to S, with $c_{xl} = c_{ls} + c_{sk}$ for all $l \in S$. Determine a minimum matching M on S that leaves one vertex exposed. If M leaves x exposed, then M remains the same; otherwise, if $\{x, m\}$ is the edge in M containing x, replace it by the edges $\{s, m\}$ and $\{s, k\}$. Now M_1 (the matching containing $\{s, i\}$) will be chosen if M_1 is at least c_{sk} shorter than M_2.

We leave it to the reader to show that the bounds of Theorems 1 and 2 are tight.

4. The case of two fixed endpoints

We have now come to the analysis of the Christofides-like algorithm for the case of two fixed endpoints. The main problem here is establishing an upper bound on the length of the minimum matching in terms of the length of the optimal Hamiltonian path with prespecified endpoints s and t. It is no longer true that the optimal path P_{st}^* can be partitioned into two edge-disjoint subsets such that either yields a matching on the wrong-degree vertices in T, as was the case for P^* and P_s^*. Consider the example given in Figure 1.

In this example, $0 < \varepsilon \ll 1$, and every edge that is not drawn in the figure has a length equal to the length of the shortest path between its endpoints. The optimal Hamiltonian path is $P_{st}^* = \{\{s, 1\},$

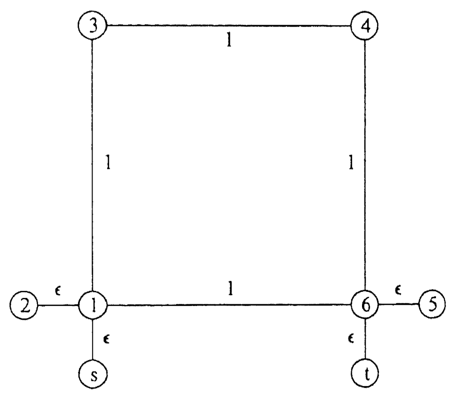

Fig. 1. Counterexample to $\rho(P_{st}^C) = \frac{3}{2}$.

$\{1, 2\}$, $\{2, 3\}$, $\{3, 4\}$, $\{4, 5\}$, $\{5, 6\}$, $\{6, t\}\}$, of length $3 + 6\varepsilon$. A minimum spanning tree is $T = \{\{s, 1\}, \{1, 2\}, \{1, 3\}, \{1, 6\}, \{4, 6\}, \{5, 6\}, \{6, t\}\}$, of length $3 + 4\varepsilon$. The set of vertices of wrong degree in T is $\{2, 3, 4, 5\}$. An optimal perfect matching on these vertices is $M = \{\{2, 3\}, \{4, 5\}\}$, of length $2 + 2\varepsilon$. An Eulerian path in the graph that is the union of T and M is $\{\{s, 1\}, \{1, 3\}, \{3, 2\}, \{2, 1\}, \{1, 6\}, \{6, 4\}, \{4, 5\}, \{5, 6\}, \{6, t\}\}$. Using shortcuts, the Hamiltonian path $P_{st}^C = \{\{s, 1\}, \{1, 3\}, \{3, 2\}, \{2, 5\}, \{5, 4\}, \{4, 6\}, \{6, t\}\}$ of length $5 + 6\varepsilon$ can be obtained. By choosing ε sufficiently small, we can get arbitrarily close to the bound $\frac{5}{3}$.

The following theorem states that this bound is tight.

Theorem 3. $c(P_{st}^C)/c(P_{st}^*) \leqslant \frac{5}{3}$.

Proof. As P_{st}^* is a tree, $c(T) \leqslant c(P_{st}^*)$. The lemma below asserts that the multi-set containing all edges belonging to T and P_{st}^* can be split in three edge-disjoint subsets, each yielding a perfect matching after shortcutting. Hence, $c(M) \leqslant \frac{2}{3}c(P_{st}^*)$, so that

$$c(P_{st}^C) \leqslant c(T) + c(M) \leqslant \tfrac{5}{3}c(P_{st}^*). \quad \square$$

Lemma 1. *The multi-set Q, containing the edges belonging to the minimum spanning tree T plus the optimal Hamiltonian path with two fixed endpoints s and t, can be partitioned into three disjoint subsets E_1, E_2 and E_3, each yielding a perfect matching on the set of odd-degree vertices in T after shortcutting.*

Proof. Every vertex of wrong degree in T has odd degree in Q, so S contains all vertices of odd degree in Q. Let the number of vertices in S be equal to $2m$. Renumber the vertices according to their order of occurrence in P_{st}^*. After renumbering, P_{st}^* consists of the edges $\{s, 1\}, \{1, 2\}, \ldots, \{n-2, t\}$, where n is the number of vertices.

The first subset, E_1, contains the edges between the $(2k-1)$-st and the $(2k)$-th vertex in S for $k = 1, \ldots, m$ (so if s and 2 are the first two vertices in S, then E_1 contains the edges $\{s, 1\}$ and $\{1, 2\}$). The first perfect matching on the vertices in S is obtained by applying shortcuts to E_1.

Consider the graph defined by the vertex set $\{s, 1, \ldots, n-2, t\}$ and the edge set $Q \setminus E_1$. This graph is connected (it still contains T) and all of its vertices have even degree (due to the removal of E_1), so it contains an Eulerian cycle. This cycle can be partitioned into the subsets E_2 and E_3. Taking shortcuts yields two perfect matchings on S. $\square$

5. A further modification of the Christofides-like heuristic for the third case

The problem of determining a Hamiltonian path with prespecified endpoints s and t can also be regarded as the problem of determining a Hamiltonian *cycle* that contains a dummy edge d of length 0 connecting the vertices s and t. There are two possibilities to ensure that d is an edge of the cycle. It can be added to the Eulerian path, or it can be incorporated into the tree, whereafter a matching on the set of odd-degree vertices is determined.

The first possibility boils down to adding d to the Eulerian path that was determined in the previous section. In this section, we analyze the second possibility. Begin with a minimum spanning tree $\overline{T}$ that contains the dummy edge. $\overline{T}$ is obtained from T as follows: add d to T and remove the longest edge e from the unique cycle in $T \cup \{d\}$;

$$\overline{T} = (\{d\} \cup T) \setminus \{e\}$$

is the tree found in Step 1 of the algorithm. Application of Christofides' heuristic for the determination of a Hamiltonian cycle, starting with $\overline{T}$, leads to a Hamiltonian cycle containing d. Removal of this dummy edge yields a Hamiltonian path with endpoints s and t.

A straightforward calculation shows that the length of $\overline{M}$, the perfect matching of minimum length on the vertices of odd degree in $\overline{T}$, is no more than the length of the matching M in the previous section plus c_e. This implies that the length of the Eulerian cycle, obtained by applying the modified heuristic, is no more than the length of the Eulerian path, obtained by the Chritofides-like heuristic for the third case.

Theorem 4. *The modified approximation algorithm has a worst-case ratio $(5n-7)/(3n-3)$, and this bound is tight.*

Proof. Let $\overline{M}$ be a perfect matching of minimum length on the set of vertices of odd degree in $\overline{T}$, and let $\overline{P}_{st}$ denote the Hamiltonian path obtained by the modified heuristic. Let $\overline{Q}$ be the multi-set containing the edges belonging to $\overline{T}$ and P_{st}^*, augmented with two copies of the edge e. In the same fashion as in Lemma 1, the multi-set $\overline{Q}$ can be partitioned into three disjoint subsets E_1, E_2 and E_3, each containing a perfect matching on the set of vertices of odd degree in $\overline{T}$, after taking shortcuts. As $c(\overline{T}) = c(T) - c_e$, it follows immediately that $c(\overline{M}) \leqslant \frac{2}{3}c(P_{st}^*) + \frac{1}{3}c_e$.

Let $\{k, l\}$ be the longest edge in P_{st}^*. The removal of e splits T in two parts. Connecting these two parts with an edge from P_{st}^* yields a tree of length no more than $c(T) + c_{kl} - c_e \geqslant c(T)$, as T is a minimum spanning tree. This implies that $c_e \leqslant c_{kl}$. Furthermore, as $P_{st}^* \cup \{d\} \setminus \{k, l\}$ is a tree, $c(\overline{T}) \leqslant c(P_{st}^*) - c_{kl}$. Hence,

$$c(\overline{P}_{st}^C) \leqslant c(\overline{T}) + c(\overline{M}) \leqslant \tfrac{5}{3}c(P_{st}^*) - c_{kl} + \tfrac{1}{3}c_e \leqslant \tfrac{5}{3}c(P_{st}^*) - \tfrac{2}{3}c_{kl}.$$

The worst-case bound follows from the observation that $P_{st}^* \leqslant (n-1)c_{kl}$.

The example in Figure 2 shows that this bound is tight. All edges drawn in Figure 2 have length 1, the length of the other edges is equal to the length of the shortest path between the two endpoints of the edge. The optimal Hamiltonian path is $P_{st}^* = \{\{s, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 6\}, \{6, 7\}, \{7, 8\}, \{8, 9\}, \{9, t\}\}$, of length 10.

A possible tree $\overline{T} = \{d, \{s, 1\}, \{1, 2\}, \{s, 3\}, \{3, 4\}, \{4, 5\}, \{3, 6\}, \{6, 7\}, \{7, 8\}, \{6, 9\}\}$, of length 9. The set of vertices of odd degree in $\overline{T}$ is

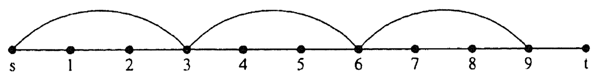

Fig. 2. Worst-case example for $\overline{P}_{st}^C$.

$\{s, 2, 3, 5, 6, 8, 9, t\}$. It is easy to check that $\overline{M} = \{\{s, 2\}, \{3, 5\}, \{6, 8\}, \{9, t\}\}$ is a perfect matching of minimum length on these vertices. The length of $\overline{M}$ is equal to 7. A possible Eulerian cycle in the union of $\overline{T}$ and $\overline{M}$ is $\{\{s, 2\}, \{2, 1\}, \{1, s\}, \{s, 3\}, \{3, 5\}, \{5, 4\}, \{4, 3\}, \{3, 6\}, \{6, 8\}, \{8, 7\}, \{7, 6\}, \{6, 9\}, \{9, t\}, d\}$. The Hamiltonian path $\overline{P}_{st}^C = \{\{s, 2\}, \{2, 1\}, \{1, 3\}, \{3, 5\}, \{5, 4\}, \{4, 6\}, \{6, 8\}, \{8, 7\}, \{7, 9\}, \{9, t\}\}$, of length 16, can be obtained by taking shortcuts and deleting the dummy edge d. □

Note that the worst-case bound can be attained for every number of vertices $n = 3p + 2$, by replacing the subgraph on the points $\{3, 4, 5\}$ by $p - 2$ subgraphs of the same form. For this augmented graph, a Hamiltonian path $\overline{P}_{st}^C$ with the same worst-case ratio can be found in a similar fashion.

Acknowledgement

The author would like to thank Jan Karel Lenstra for his helpful comments.

References

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

G. Cornuéjols and G.L. Nemhauser (1978), "Tight bounds for Christofides' traveling salesman heuristic. *Math. Programming* **14**, 116–121.

D.S. Johnson and C.H. Papadimitriou (1985), "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, Chichester, 145–180.

S. Sahni and T. Gonzalez (1976), "P-complete approximation problems", *J. Assoc. Comput. Mach.* **23**, 555–565.