Bounds on the Traveling Salesman Problem

Sean Zachary Roberson

Texas A&M University

MATH 613, Graph Theory

A common routing problem is as follows: given a collection of stops (for example, towns, stations, facilities) and distances between them, what is the shortest distance to travel to all locations so that each location is visited once, returning to the origin? While simple in formulation, this problem is notoriously difficult to solve. In this paper, we examine some of the history behind this problem, known as the Traveling Salesman Problem (TSP), and explore some upper bounds in finding an optimal tour. The primary focus will be on the Christofides algorithm, which reduces the problem in finding both a minimum spanning tree and a minimum-weight perfect matching, thereby reducing this NP-hard problem into two easier polynomial-time problems.

The traveling salesman problem has its roots with Leonhard Euler and William Rowan Hamilton. An easy connection to Hamilton can be made as one can visualize Hamiltonian circuits - closed paths that visit each vertex of a graph. A natural question to ask is what conditions are necessary or sufficient for the existence of a Hamiltonian cycle. Two sufficient conditions can help determine if a feasible tour exists for TSP, which we give an explicit formulation later. The theorem below is due to Gabriel Dirac [7].

Theorem 1. *Let G be a simple graph with at least three vertices. If $\delta(G) \geq n(G)/2$, then G is Hamiltonian.*

A necessary and sufficient condition given by J. A. Bondy and Václav Chvátal generalizes the result above [6]. This relies on the notion of the closure of a graph. The closure of G, $\mathrm{cl}(G)$ is the (super)graph obtained by adding edges between non-adjacent vertices whose degrees sum to $n = |V|$, until no such edges can be added.

Theorem 2. *G has a Hamiltonian circuit if and only if its closure $\mathrm{cl}(G)$ has one.*

The above necessary and sufficient condition gives way to finding feasible tours for the Traveling Salesman Problem. Finding a tour while ignoring cost is easy, but finding a minimum-cost tour when the number of cities is large is difficult. Though large-scale tours have been solved, such as the record 85,900-city tour in 2006 as described in Cook's book on the history and devlopment of the Traveling Salesman Problem [4].

We now give an explicit formulation of the Traveling Salesman Problem. The formulation below of the symmetric problem (i.e., the cost from A to B is the same as that from B to A) is taken from Wolsey [8].

$$z = \min \sum_{e \in E} c_e x_e$$

$$\text{subject to} \sum_{e \in N(v)} x_e = 2, v \in V$$

$$\sum_{S \subset V,\ e \in E(S)} x_e \leq |S| - 1, \quad 2 \leq |S| \leq |V| - 1$$

$$x_e \in B^{|E|}$$

Here, the last sum is taken over all non-singleton, non-spanning vertex subsets, and each decision variable x_e is a Boolean variable - i.e., it takes the values either 0 or 1, where 1 means the edge is included in the tour.

Perhaps the most important constraint in the integer linear programming formulation is the final set. Without these constraints, this allows subtours. The objective of the traveling salesman problem is to find one Hamiltonian circuit of minimum cost. These constraints do not permit subtours, that is, other Hamiltonian circuits such that the union of all tours spans all the vertices. This is certainly not allowed, and the subtour elimination constraints prevent this from happening.

It is certainly of interest to determine how costly a tour can be. There is a naïve upper bound that can be given to the cost of any feasible tour. To establish these next bounds, we assume that, other than the obvious fact that the costs are nonnegative, the graph has a metric imposed upon it. That is, we assume the vertices of the graph live in a metric space (X, d). In particular, we assume that for distinct nodes $x,\ y$, and z, the triangle inequality

$$d(x, y) + d(y, z) \leq d(x, z)$$

is satisfied under the metric d. For simplicity, the Euclidean metric is typically used. The naïve bound can be stated as follows, as given in Kreher's text. The original result is given in Held and Karp's 1970 paper, *"The Traveling-Salesman Problem and Minimum Spanning Trees."*

Theorem 3. *Suppose T^* is the optimal tour for the TSP, and let M be the cost of the minimum spanning tree on G. Then*

$$C(T^*) \leq 2M$$

where C is the cost function on the graph G. [6]

Proof. We first establish an elementary lower bound, that for the minimum spanning tree T_M,

$$C(T_M) \leq C(T^*),$$

as any edge can be removed from T^* to yield a spanning tree. In particular, the most expensive edge can be removed from the tour to yield a minimum spanning tree. Let x be the root of this tree and a node incident to the most expensive edge removed from the tour (i.e., the tour started and ended at x). Then traveling around this tree and back yields a cost twice that of the minimum spanning tree. □

This bound, while not tight, gives an easy estimate for the upper cost of a feasible tour. Minimum spanning trees are typically easy to find by means of the Prim or Kruskal algorithms. The technique described by Held and Karp relies on a variant of spanning trees, called 1-trees. Such a tree is rooted at a vertex labeled 1 and contains only one simple cycle. This approach then allows one to solve the problem by considering a linear program and its dual. [5] The observations of the primal and dual LPs will not be considered in this work for brevity.

We now prepare to describe Christofides' bound via his algorithm. First, consider a TSP environment with n cities, represented by the graph K_n. Suppose that the weights of K_n obey the triangle inequality, as above. Build a minimum spanning tree T_M. Let S_1 be the set of vertices of odd degree in T_M. Note that $|S_1|$ is even. With the graph induced by S_1, we can find a perfect matching, by Hall's theorem. To that end, let $m(S_1)$ be the perfect matching of smallest cost on S_1. Build the graph $m(S_1) \cup T_M$ and find an Eulerian circuit. Finalize the algorithm by short-cutting - that is, replacing edges of smaller cost to build a Hamiltonian tour. [3]

We first state a lemma.

Lemma 1. *For an n-city TSP, where n is even,*

$$C(m(G)) \leq \frac{1}{2}C(T^*)$$

where T^* is the optimal tour and $m(G)$ is the perfect matching of minimum cost on G.

We now state the main theorem that provides a bound that has been unmatched.

Theorem 4. *There exists a Hamiltonian circuit T_H on G whose cost obeys the compound inequality*

$$C(T_H) \leq C(T_M) + C(m(S_1)) < C(T^*)$$

where T_H, T_M, S_1 are from above.

Proof. From another paper of Christofides (see [1]) it is known that for any G,

$$\text{minimum spanning tree} \leq \text{shortest Hamiltonian circuit} < \text{optimal tour}$$

where the last inequality is strict, for no trivial edges are allowed in G (that is, zero-cost edges). Now, the subgraph formed by joining the minimum-weight perfect matching and the minimum spanning tree of G is Eulerian, for no vertices are of odd degree (this is necessary and sufficient, see [7] or [2]). Any Eulerian circuit must visit each node at least once. Let $\mathcal{E}$ be the circuit of minimum cost, and C its cost function. Combining the lemma with the result stated at the start of this paragraph, we obtain

$$C(\mathcal{E}) \leq \frac{1}{2}C(T^*).$$

Adding this to the inequality

$$C(T_H) < C(T^*)$$

gives

$$C(\mathcal{E}) < \frac{3}{2}C(T^*).$$

As shortcutting via the triangle inequality improves the cost $C(\mathcal{E})$, the provided upper bound is best, as desired. □

Note that this doesn't give an explicit bound on any feasible tour. This does, however, describe how to place an upper bound on the cost of the optimal tour. When combined with other methods (for example, searching via a branch-and-bound tree or cutting planes), tight bounds can be found to solve the problem (or a relaxation of the problem). Another curious fact of note is that the preceding algorithm runs in cubic time, on average. This is the combination of a quadratic time algorithm for the minimum spanning tree, plus a cubic time algorithm for the perfect matching problem.

In closing, the Traveling Salesman Problem is a difficult problem to solve in combinatorial optimization. Different variations of the problem (such as the close-enough and asymmetric versions) yield different results, but the symmetric version is one of the most researched problems. With the many applications to vehicle routing and finding the "best road trip," the Traveling Salesman Problem will always have a practical application in many fields of mathematics, engineering, business, and manufacturing.

Bibliography

[1] Nicos Christofides. The shortest hamiltonian chain of a graph. *SIAM Journal on Applied Mathematics*, 19(4):689–696, December 1970.

[2] Nicos Christofides. *Graph Theory - An Algorithmic Approach*. Academic Press, 1975.

[3] Nicos Christofides. A worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, Carnegie-Mellon University, February 1976.

[4] William Cook. *In Pursuit of the Traveling Salesman*. Princeton University Press, 2012.

[5] Michael Held & Richard M. Karp. The traveling-salesman problem and minimum spanning trees. *Operations Research*, 18(6):1138–1162, December 1970.

[6] William Kocay & Donald L. Kreher. *Graphs, Algorithms, and Optimization*. Chapman & Hall, 2005.

[7] Douglas B. West. *Introduction to Graph Theory*. Prentice Hall, 2001.

[8] Laurence A. Wolsey. *Integer Programming*. Wiley, 1998.