Remarks on the traveling salesman problem

Rosalie Iemhoff

October 29, 2008

The Traveling Salesman Problem TSP can be stated as follows: we are given n cities with nonnegative distances d_{ij} between any two cities i and j, and are asked to find the shortest tour of the cities, that is, a tour which starts and ends in the same city, which passes every other city only once, and which has the shortest total length. Equivalently: we are asked to find a permutation π of $\{1, \ldots, n\}$ such that the sum of the distances, $\Sigma_{i=1}^{n-1} d_{\pi(i)\pi(i+1)} + d_{\pi(n)\pi(1)}$ is as small as possible.

It is clear that one can find such a shortest tour in $O(n!)$ time, far more than polynomial time or even NP. By using dynamic programming this bound can be improved to $O(2^n)$, but a proof of this fact falls outside the scope of this exposition. Considering space one can clearly do better: there is an $O(n)$-space algorithm to find an optimal solution. This algorithm simply computes the total distance of the tours one-by-one, reusing space, and keeping track of which is the smallest tour seen so far.

Although it is an open problem whether TSP is in NP, it can be shown to be NP-hard, that is, every problem in NP polynomial time reduces to it.

Recall from Sipser that for $k \geq 1$ an approximation algorithm is k-optimal when it produces a solution that is not more than k times the optimal solution in case it is a minimization problem, and not less than $1/k$ times the optimal solution if it is a maximization problem. One can show, and we will do so in the theorem below, that for no k there is a polynomial time k-optimal optimization algorithm for TSP, unless P=NP.

First we have to introduce a definition and a fact. A Hamilton cycle in a graph is a Hamilton path that begins and ends in the same node and passes all other nodes at most once. Note the similarity with TSP: a tour for TSP is in particular a Hamilton cycle. We define

$$HAMCYCLE = \{\langle G \rangle \mid \text{there is a Hamilton cycle in } G\}.$$

We leave the proof of the following fact to the reader:

Fact 1 HAMCYCLE is NP-complete.

Thus it follows that:

Corollary 1 If $\mathsf{P} \neq \mathsf{NP}$, then HAMCYCLE does not belong to P.

Theorem 1 Unless $\mathsf{P} \neq \mathsf{NP}$, for no $k \geq 1$ there exists a polynomial time k-optimal approximation for TSP.

Since it is generally believed that $\mathsf{P} \neq \mathsf{NP}$ this implies that TSP has no k-optimal solutions, for any k. Imagine how important this knowledge is for applications!

Proof of the theorem We prove the theorem as follows: we assume that for some k there is a polynomial time k-optimal approximation for TSP and then show that this implies that HAMCYCLE is in P.

Suppose M is a polynomial time TM that computes a k-optimal solution for TSP. Given M we produce the following polynomial decider N for HAMCYCLE as follows.

N: on input $\langle G \rangle$, where G is a graph with n nodes:

1. Construct a corresponding instance of TSP in the following way. It consists of n cities with the following distances: if there is an edge between i and j, then put $d_{ij} = 1$, and $d_{ij} = n \cdot k$ otherwise.
2. Run M on this instance of TSP, and consider the total distance of the tour that it produces as output.
3. If it is n, accept, otherwise, reject.

It is not difficult to see that N is a polynomial time TM. Observe that the optimal tour of the corresponding TSP problem has length n whenever the input of N is a graph with n nodes. We have to show that N decides HAMCYCLE. Thus

$$N \text{ accepts } \langle G \rangle \text{ if and only if } \langle G \rangle \in HAMCYCLE.$$

$\Rightarrow$: Assume N accepts $\langle G \rangle$ and graph G has n nodes. Thus the tour M has found has distance n. Since it passes n roads it follows that the distance between every two cities i and j for which the tour contains the direct road between them, is 1 ($d_{ij} = 1$). This again implies that for such i and j there is an edge between them. From this it follows that there is a Hamilton cycle in G.

$\Leftarrow$: Assume $\langle G \rangle \in HAMCYCLE$. The k-optimality of M implies that the tour that M outputs in the algorithm of N is at most k times the optimal, thus $\leq k \cdot n$. But if that is so, the tour cannot contain a rode of distance $n \cdot k$, since then the total distance would be $> k \cdot n$. Thus all rodes in the tour have distance 1, and thus the total distance is n, and thus N accepts $\langle G \rangle$. ♡

Although by the above theorem the Traveling Salesman Problem behaves "badly" in terms of complexity, there are certain instances of the problem that can be solved efficiently. For example, if the problem is restricted to those instances in which the distances between cities is 1 or 2, a 2-optimal approximation does exist, since the tour given by an approximation can be at most twice as large as an optimal tour.

The decision problem associated with TSP (given an instance of TSP and a number d, is there a tour with distance $\leq d$?) clearly is in NP, it is in fact NP-complete.