Christofides's $\frac{3}{2}$-Approximation for Metric TSP

This is a polynomial-time $\frac{3}{2}$-approximation algorithm for the TSP in a metric space (X, d) due to N. Christofides [1]. Here X is a finite set and $d : X^2 \to \mathbb{R}$ is a distance function satisfying

- $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

If $S \subseteq X^2$, let

$$d(S) \stackrel{\text{def}}{=} \sum_{(x,y)\in S} d(x, y).$$

The object is to find a Hamiltonian cycle of minimum length.

First construct a minimum spanning tree M. Starting at an arbitrary point of X, go around M once counterclockwise, keeping your left hand on the wall. This tour traverses each edge once in each direction. It is not a Hamiltonian tour because it visits some vertices multiple times. However, we can fix this by short-circuiting past vertices that we have seen before. Starting at an arbitrary point, follow the tour, but mark vertices when you visit them as already visited. Later, when encountering a vertex previously visited, just skip over it and go directly to the next vertex on the tour. If that vertex has also already been seen, go on to the next, etc. Keep going until encountering the start vertex again. The resulting tour is a Hamiltonian tour, and its length is no greater than the original tour by the triangle inequality.

This construction gives a 2-approximation. Let T be the tour resulting from this construction, let T^* be an optimal TSP tour, and let e be an arbitrary edge on T^*. Then

$$\begin{aligned} d(T) &\leq 2d(M) && (1)\\ &\leq 2d(T^* - e) && (2)\\ &\leq 2d(T^*). && (3) \end{aligned}$$

The inequality (1) is from the fact that the length of the original tour is exactly twice the weight of M, since each edge of M is traverse exactly twice, and the final tour T obtained by skipping previously seen vertices is no worse by the triangle inequality. The inequality (2) is from the fact that $T^* - e$ is a spanning tree, therefore $d(M) \le d(T^* - e)$, since M is a minimum-weight spanning tree. The inequality (3) is from the fact that all distances are nonnegative.

Now we improve this to a $\frac{3}{2}$-approximation. Let O be the set of odd-degree vertices of M. The set O contains all leaves of M and perhaps some internal vertices as well. It is also the case that $|O|$ is even, because the sum of the degrees of all the nodes must be even, because it is exactly twice the number of edges.

Thus the complete graph on O has a perfect matching, and we can find a minimum weight perfect matching P in polynomial time. We claim that $d(P) \le d(T^*)/2$. Let N^* be an optimal TSP tour on just the odd-degree nodes O. Let N_1 and N_2 be the two perfect matchings on O obtained by taking the edges of N^* alternately. Then

$$
\begin{aligned}
d(P) &\le \min d(N_1),\ d(N_2) & (4)\\
&\le d(N^*)/2 & (5)\\
&\le d(T^*)/2. & (6)
\end{aligned}
$$

The inequality (4) is from the fact that N_1 and N_2 are perfect matchings on O, and P is a minimum-weight perfect matching on O. The inequality (5) is from the fact that the minimum of $d(N_1)$ and $d(N_2)$ is at most the average. For the inequality (6), get a TSP tour on O from T^* by skipping the even-degree vertices. By the triangle inequality, the length of the resulting tour is no worse than $d(T^*)$, and $d(N^*)$ is no worse than this because it is an optimal tour on O.

Now take the graph with edges E consisting of the edges P and M. This graph may contain multiple edges, since P and M may intersect. But this is ok—if an edge is in both P and M, we just count it twice. Two facts to notice are:

- All nodes have even degree, because we added just one edge of P incident to each node of O.
- $d(E) \le 3d(T^*)/2$, because $d(M) \le d(T^*)$ and $d(P) \le d(T^*)/2$.

Now because each node is of even degree, we can find an Eulerian tour, a tour traversing all the edges of E exactly once. The Eulerian tour is constructed by starting from an arbitrary point and following edges until encountering the start point again. It is always possible to continue, because each time a node is visited, two incident edges are used, so it is an invariant of the process that all nodes have even degree. The process may encounter the start vertex before traversing all the edges, but in that case, just start again from a new vertex and piece the resulting Eulerian cycles together at the end, using the fact that the graph is connected.

Now we have an Eulerian tour of length at most $3d(T^*)/2$. Compress the tour to a Hamiltonian tour by skipping vertices previously visited, as above.

References

[1] N. Christofides, Worst-case analysis of a new heuristic for the travelling salesman problem, Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.