Selected Topics in Algorithms 2009
NP-completeness of Minimum Fundamental Cycle Basis Problem

Kurt Mehlhorn

May 22, 2009

Let $G = (V,E)$ be an undirected graph. For a spanning tree T of G and vertices u and v, let $d_T(u,v)$ be the distance between u and v in T and let $P(T) = \sum_{\{u,v\}} d_T(u,v)$ be the sum over all distances. Here the sum is over all pairs of nodes. Let $W(T)$ be the weight of the fundamental cycle basis induced by T.

Exact 3-Cover Given 3-element subsets S_1 to S_s of $U = \{1,\ldots,3u\}$. Is there an index set I such that $U = \cup_{i \in I} S_i$ and $|I| = u$? We proved in Exercise Sheet 3 that Exact 3-Cover is NP-complete.

Shortest Total Path Length Spanning Tree (STPLST): Given an undirected graph G and a bound B. Is there a spanning tree T of G with $P(T) \le B$?

Minimum Fundamental Cycle Basis Problem (MFCB): Given an undirected graph G, a non-negative weight function $w : E \to \mathbb{N}$ and a bound B. Is there a spanning tree T of G with $W(T) \le B$?

Theorem 1 ([JLK78]) *Shortest Total Path Length Spanning Tree is NP-complete.*

Proof: We show Exact 3-Cover $\le$ STPLST. Consider an instance $S_1, \ldots, S_s$, $U = \{1,\ldots,3u\}$ of Exact 3-Cover. Construct the following graph, see Figure 1.

- $V = R \cup S \cup U$, where $R = \{v_0, v_1, \ldots, v_r\}$. We will fix r below.
- Edges
  - v_0 is connected to all v_i, $1 \le i \le r$ and to all S_i, $1 \le i \le s$.
  - each S_i is connected to the $j \in U$ with $j \in S_i$.

Assume Exact 3-Cover has a solution I and consider the following spanning tree T. It consists of edges v_0v_i for $1 \le i \le r$ and v_0S_i for $1 \le i \le s$. Each $j \in U$ is connected to the unique S_i with $j \in S_i$ and $i \in I$. The total path length of this solution is

$$B = B_{RR} + B_{RS} + B_{RU} + B_{SS} + B_{SU} + B_{UU}$$

Figure 1: . Reduction of Exact 3-Cover to STPLST.

where $B_{RR} = r + 2r(r-1)/2 = r^2$, $B_{RS} = s + 2rs$, $B_{RU} = 2 \cdot 3t + 3 \cdot r \cdot 3t = 6t + 9rt$, $B_{SS} = 2 \cdot s(s-1)/2 = s^2 - s$, $B_{SU} = 3u(s-1)3 + 3u \cdot 1 \cdot 1 = 9su - 6u$, and $B_{UU} = 4 \cdot 3u(3u-1)/2 - 2u3 = 18u^2 - 12u$.

So, if the instance of Exact 3-cover has a solution then there is a spanning tree with $P(T) \leq B$, where B is defined as above.

Assume now that there is a spanning tree with $P(T) \leq B$. We will show that Exact 3-Cover has a solution. For $X, Y \in \{R, S, U\}$ let P_{XY} be the cost of connecting X and Y in T. Clearly, all edges (v_0, v_i) are in T. Also, $P_{RR} \geq B_{RR}$, $P_{RU} \geq B_{RU}$ and $P_{RS} \geq B_{RS}$. Assume now that one of the edges (v_0, S_j) is NOT in T. Then $P_{RS} \geq B_{RS} + 2(r+1)$. For $r \geq (B_{SS} + B_{SU} + B_{UU})/2$ this implies $P(T) > B$. We fix r at $(B_{SS} + B_{SU} + B_{UU})/2$. So $(v_0, S_j) \in T$ for all j. Then $P_{RS} = B_{RS}$ and $P_{SS} = B_{SS}$.

For each $j \in U$, there is exactly one S_i with $(S_i, j) \in T$. For $0 \leq \ell \leq 3$, let k_ℓ be the number of S_i that are connected in T to exactly ℓ nodes in U. Then

$$P_{UU} = 4 \cdot 3u(3u-1)/2 - 2 \cdot 3k_3 - 2k_2$$

and hence $P_{UU} = B_{UU}$ only if $k_3 = u$. If $k_3 = u$, T encodes a solution to Exact 3-Cover. ∎

Theorem 2 ([DPK82]) *The Minimum Fundamental Cycle Basis Problem is NP-complete.*

Proof: We show STPLST $\leq$ MFCB. Let $G = (V, E)$ be an instance of STPLST. Consider the following instance of MFCB. We augment G to a complete graph $G' = (V, E')$ and set $w(e) = 1$ for $e \in E$ and $w(e) = L$ for $e \in E' \setminus E$. Here L is a large constant that we fix later. We call the edges in E light and the edges in $E' \setminus E$ heavy. We will show that a large value of L guarantees that the solution to MFCB will use only light edges for the spanning tree and that this spanning tree will be a solution to STPLST.

Consider any spanning tree T of G' and assume that it uses q heavy edges. Then $n - 1 - q$ edges of T are light. Among the co-tree edges, $m - (n - 1 - q)$ are light and $r - q$ are heavy, where $r = n(n-1)/2 - m$. For a pair $\{u,v\} \in T$, we have $d_T(u,v) = w(u,v)$. We have

$$\begin{aligned}
W(T) &= \sum_{\{u,v\}\notin T} (d_T(u,v) + w(u,v)) \\
&= \sum_{\{u,v\}\notin T} d_T(u,v) + \sum_{\{u,v\}\notin T} w(u,v) \\
&= \sum_{\{u,v\}} d_T(u,v) - \sum_{\{u,v\}\in T} w(u,v) + \sum_{\{u,v\}\notin T} w(u,v) \\
&= \sum_{\{u,v\}} d_T(u,v) - 2\sum_{\{u,v\}\in T} w(u,v) + \sum_{\{u,v\}} w(u,v) \\
&= S(T) + rL + m - 2qL - 2(n-1-q) \\
&= S(T) + rL + m - 2(n-1) - 2qL + 2q\ .
\end{aligned}$$

Assume we knew that an optimal solution for MFCB is guaranteed to have $q = 0$. Then there is a tree with $S(T) \le B$ iff there is a tree with $W(T) \le B + rL + m - 2(n-1)$ and we have the desired reduction.

Consider now spanning trees T_1 and T_2, where T_1 uses only light edges and T_2 uses $q \ge 1$ heavy edges. We claim $W(T_1) < W(T_2)$. Indeed,

$$\begin{aligned}
W(T_1) < W(T_2) &\text{ iff } S(T_1) + rL + m - 2(n-1) < S(T_2) + rL + m - 2(n-1) - 2qL + 2q \\
&\text{ iff } S(T_1) + 2qL - 2q < S(T_2)\ .
\end{aligned}$$

Next observe that $S(T_1) \le n^2 n = n^3$. We need a lower bound of $S(T_2)$. Removal of the q heavy edges from T_2 decomposes T_2 into $q+1$ subtrees of sizes, say, k_1 to k_{q+1}. Then $S(T_2) \ge L\sum_{i<j} k_i k_j \ge L\max(k_1(k_2 + \ldots + k_{q+1}), (q+1)q/2) \ge L\max(n-1, q^2/2)$. For $L = n^4$, the inequality $n^3 + 2qL - 2q < L\max(n-1, q^2/2)$ holds for all values of q. ∎

There is no need for the use of weights in the proof of theorem 2. Simply replace the heavy edges by chains of length L and then the proof works for unweighted graphs.

The minimum fundamental cycle basis problem is not only NP-complete, it is also hard to approximate. It is APX-hard and hence has no PTAS (polynomial time approximation scheme) unless P = NP. The APX-hardness proof can be found in [KLM$^+$09]. A polynomial time approximation scheme for MFCB would have input (V,E,w) and parameter $\varepsilon > 0$. It would produce a fundamental cycle basis of cost no more than $(1+\varepsilon)$ times the optimal value and, for any fixed ε, run in polynomial time. The degree of the polynomial may depend on ε.

I close with two open problems:

- An approximation algorithm for the minimum fundamental cycle basis problem with approximation guarantee $O(\log n)$. See [KLM$^+$09] for what is known.

- The complexity status of the minimum integral cycle basis problem.

References

[DPK82] N. Deo, G.M. Prabhu, and M.S. Krishnamoorthy. Algorithms for generating fundamental cycles in a graph. *ACM Transactions on Mathematical Software*, 8:26–42, 1982.

[JLK78] D.S. Johnson, J.K. Lenstra, and A.H.G. Rinnoy Kan. The complexity of the network design problem. *Networks*, 8:279–285, 1978.

[KLM+09] T. Kavitha, Ch. Liebchen, K. Mehlhorn, D. Michail, R. Rizzi, T. Ueckerdt, and K. Zweig. Cycle Bases in Graphs: Characterization, Algorithms, Complexity, and Applications. 78 pages, submitted for publication, available at `www.mpi-inf.mpg.de/~mehlhorn/ftp/SurveyCycleBases.pdf`, March 2009.