The Greedy Algorithm for the Symmetric TSP

Gregory Gutin* Anders Yeo†

Abstract

We corrected proofs of two results on the greedy algorithm for the Symmetric TSP and answered a question in Gutin and Yeo, Oper. Res. Lett. 30 (2002), 97–99.

Keywords: Symmetric TSP, greedy algorithm, heuristics

1 Introduction

Many combinatorial optimization problems can be formulated as follows [6]. We are given a pair $(E, \mathcal{F})$, where E is a finite set and $\mathcal{F}$ is a family of subsets of E, and a weight function c that assigns a real weight $c(e)$ to every element of E. The weight $c(S)$ of $S \in \mathcal{F}$ is defined as the sum of the weights of the elements of S. It is required to find a maximal (with respect to inclusion) set $B \in \mathcal{F}$ of minimum weight. The *greedy algorithm* starts from the element of E of minimum weight that belongs to a set in $\mathcal{F}$. In every iteration the greedy algorithm adds a minimum weight unconsidered element e to the current set X provided $X \cup \{e\}$ is a subset of a set in $\mathcal{F}$.

The sequence [4, 3, 1] of papers studied the greedy algorithm for the Asymmetric and Symmetric Traveling Salesman Problems (ATSP and STSP) and wide classes of combinatorial optimization problems that include both TSPs. It was proved in [4] that for each $n \geq 3$, there is an instance of ATSP on n vertices for which the greedy algorithm produces the unique worst possible solution. In [3] we introduced a wide class of optimization problems, which we called anti-matroids (they are defined later), for which the greedy algorithm similarly fails. The Assignment Problem and ATSP were proved to be anti-matroids by showing that they belong to a special family of anti-matroids (I-anti-matroids, also defined later). Erroneously, we also claimed that STSP is also an I-anti-matroid. However, this is not true as we show in Proposition 2.1 of this paper. We also prove, in Theorem 2.2, that STSP is an anti-matroid by giving a direct proof.

Both Proposition 2.1 and Theorem 2.2 answer an open question in [3] to provide a well-studied combinatorial optimization problem which is an anti-matroid, but not an I-anti-matroid. Another proof that STSP is anti-matroid is given in [1], but the proof there is indirect and relatively long. Our proof of Theorem 2.2 is also of interest since it can be used to correct the proof of Theorem 4.3 in [1] (see Theorem 2.3) for STSP, which is incorrect due to the fact that STSP is not an I-anti-matroid. (The proof of Theorem 4.3 in [1] is correct for ATSP.) Notice that the result of Theorem 2.3 is stronger (by a factor of n) than Theorem 4.3 in [1] for STSP.

*Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, gutin@cs.rhul.ac.uk

†Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, anders@cs.rhul.ac.uk

Notice that there are STSP heuristics which always find a tour that is better than a large number of other tours, see, e.g., [4, 5].

An *anti-matroid* is a pair $(E, \mathcal{F})$ such that there is an assignment of weights to the elements of E for which the greedy algorithm for finding a maximal set B in $\mathcal{F}$ of minimum weight constructs the unique maximal set of maximum weight.

An I-*independence family* is a pair consisting of a finite set E and a family $\mathcal{F}$ of subsets (called *independent sets*) of E such that (I1)-(I3) are satisfied.

(I1) the empty set is in $\mathcal{F}$;

(I2) If $X \in \mathcal{F}$ and Y is a subset of X, then $Y \in \mathcal{F}$;

(I3) All maximal sets of $\mathcal{F}$ (called *bases*) are of the same cardinality k.

If $S \in \mathcal{F}$, then let $I(S) = \{x : \ S \cup \{x\} \in \mathcal{F}\} - S$. This means that $I(S)$ contains all elements (different from S), which can be added to S, in order to have an independent set. An I-independence family $(E, \mathcal{F})$ is an I-*anti-matroid* if

(I4) There exists a base $B' \in \mathcal{F}$, $B' = \{x_1, x_2, \ldots, x_k\}$, such that the following holds for every base $B \in \mathcal{F}$, $B \neq B'$,

$$\sum_{j=0}^{k-1} |I(x_1, x_2, \ldots, x_j) \cap B| < k(k+1)/2.$$

Consider STSP on n vertices, $n \geq 3$. We recall that STSP is the problem of finding a minimum weight Hamilton cycle in a weighted complete graph K_n. We view STSP as an I-independence family whose independent sets are collections of disjoint paths of K_n and Hamilton cycles in K_n. We will represent independent sets of STSP as sets of their edges. We denote the vertices of K_n $1, 2, \ldots, n$ and call Hamilton cycles *tours*.

2 Results

Proposition 2.1 *For $n \geq 4$, STSP is not an I-anti-matroid.*

Proof: Let $T' = \{e_1, e_2, \ldots, e_n\}$, where $e_i = \{i, i+1\}$ for $i < n$ and $e_n = \{n, 1\}$. Let $T = T' \cup \{f_1, f_3\} - \{e_1, e_3\}$, where $f_1 = \{1, 3\}$ and $f_3 = \{2, 4\}$. Since we can consider T' as an arbitrary tour (i.e., a base) of STSP, if STSP was an I-anti-matroid, we would have

$$\sum_{j=0}^{n-1} |I(e_1, e_2, \ldots, e_j) \cap T| < n(n+1)/2.$$

However, $|T| = |I(e_1) \cap T| = n$, $|I(e_1, e_2) \cap T| = n - 3$, and $|I(e_1, e_2, \ldots, e_j) \cap T| = n - j$ for each $j \geq 3$. Hence, $\sum_{j=0}^{n-1} |I(e_1, e_2, \ldots, e_j) \cap T| = n(n+1)/2$. □

Theorem 2.2 *STSP is an anti-matroid.*

Proof: Let $T' = \{e_1, \ldots, e_n\}$ be the tour, where $e_i = \{i, i+1\}$ for $i < n$ and $e_n = \{n, 1\}$. We first show the following:

Claim A: For every tour $T \neq T'$, we have the following:

$$\sum_{j=0}^{n-1} |I(e_1,e_2,\ldots,e_j)\cap T| \le \begin{cases} \frac{n(n-1)}{2} & \text{if } e_1 \notin T \\ \frac{n(n-1)}{2} - 1 & \text{if } e_1 \in T \end{cases}$$

Proof of Claim A: We will first prove that $|I(e_1,e_2,...,e_j)\cap T| \le n-j$ for all $j = 0,1,2,\ldots,n-1$, except for $e_1 \notin T$ and $j=1$, in which case $|I(e_1)\cap T| = n$. This statement is clearly true when $j=0$, $j=1$ and $j=n-1$, so now assume that $2 \le j < n-1$, and note that all edges of T belong to $I(e_1,e_2,...,e_j)$, except for those with one end-vertex in $J=\{2,3,...,j\}$ and the edge $\{1,j+1\}$. Let Q_j denote the set of edges in T with at least one end-vertex in J. Let m_j denote the number of edges in T in which both end-vertices belong to the set J and observe that $|Q_j| = 2(j-1)-m_j$ (since each vertex in J has degree 2 in T and every edge of T between vertices in J 'cancels' one degree unit). Observe that $m_j \le j-2$, which implies that $|I(e_1,e_2,...,e_j)\cap T| \le n-|Q_j| = n-2(j-1)+m_j \le n-j$.

Now let ab be an edge in T such that $a < b-1 < n-1$ and a is as small as possible. Observe that ab exists since $T' \ne T$. If $a=1$ then we have $|I(e_1,e_2,...,e_{b-1})\cap T| \le n-(b-1)-1$, as $ab \notin Q_{b-1}$, but $ab \notin I(e_1,e_2,...,e_{b-1})\cap T$. If $a>1$ then $123\ldots a$ is a path in T and $m_{a+1} = (a+1)-3$ as $e_a \notin T$. Hence, $|I(e_1,e_2,...,e_{a+1})\cap T| \le n-(a+1)-1$. Therefore the following holds.

$$\sum_{j=0}^{n-1} |I(e_1,e_2,\ldots,e_j)\cap T| \le \begin{cases} (n+n+\sum_{j=2}^{n-1}(n-j))-1 & \text{if } e_1 \notin T \\ (n+(n-1)+\sum_{j=2}^{n-1}(n-j))-1 & \text{if } e_1 \in T \end{cases}$$

This proves Claim A.

Now let $M>n$ and assign weights to the edges of K_n as follows:

$$\begin{array}{ll} c(e_1) = 2M & \\ c(e_i) = iM & \text{for all } i \ge 2 \\ c(e) = 1+jM & \text{if } e \notin T',\ e \in I(e_1,e_2,\ldots,e_{j-1}) \text{ but } e \notin I(e_1,e_2,\ldots,e_j) \end{array} \quad (1)$$

Observe that $c(e) \ge 2M$ for each e. By this remark and the definition of costs, the greedy algorithm constructs T' and $c(T') = Mn(n+1)/2 + M$.

Let $T = \{f_1,f_2,\ldots,f_n\}$ be an arbitrary tour distinct from T'. By the choice of c made above, we have that $c(f_i) \in \{a_iM, a_iM+1\}$ for some positive integer a_i. First assume that $f_i \ne e_1$, in which case $f_i \in I(e_1,e_2,\ldots,e_{a_i-1})$ but $f_i \notin I(e_1,e_2,\ldots,e_{a_i})$. Therefore $f_i \in I(e_1,e_2,\ldots,e_j)\cap T$ exactly when $j \le a_i - 1$, which implies that f_i is counted a_i times in the sum in Claim A. So if $e_1 \notin T$ then, by Claim A, the following holds:

$$\frac{n(n+1)}{2} \ge \sum_{j=0}^{n-1} |I(e_1,e_2,\ldots,e_j)\cap T| = \sum_{i=1}^{n} a_i.$$

Since $c(T) \le \sum_{i=1}^{n}(a_iM+1) \le Mn(n+1)/2 + n < c(T')$, we are done in the case when $e_1 \notin T$. If $f_i = e_1$ then $a_i = 2$ and $f_i \in I(e_1,e_2,\ldots,e_j)\cap T$ only when $j=0$, which, by Claim A, implies the following:

$$\frac{n(n+1)}{2} - 1 \ge \sum_{j=0}^{n-1} |I(e_1,e_2,\ldots,e_j)\cap T| = \left(\sum_{i=1}^{n} a_i\right) - 1.$$

As above we note that $c(T) \le \sum_{i=1}^{n}(a_iM+1) \le Mn(n+1)/2 + n < c(T')$, which completes the proof. □

Theorem 2.3 *For each even $n \geq 4$ there exists an instance of STSP that has $\Omega(\frac{(n-1)!}{2^n n^{1/2}})$ optimal tours, each of which is at least $f(n)$ times shorter than the unique worst tour, where $f(n) \geq 1$ is an arbitrary function in n, and yet the greedy algorithms produces the unique worst tour.*

Proof: Let K_n be a complete graph on vertices $\{1, 2, \ldots, n\}$ and let edge $\{i, i+1\}$ be denoted by e_i for $i = 1, 2, \ldots, n$, where $n+1 = 1$. Then $T' = \{e_1, e_2, \ldots, e_n\}$ is a base. Assign weights $c(e)$ for each edge e of K_n as in the proof of Theorem 2.2, see (1), with $M > n$. It is proved in Theorem 2.2 that T' is the unique heaviest tour in K_n; let $P(n)$ be the weight of T'. Let $L = \{2, 3, \ldots, \frac{n}{2} + 1\}$ and $R = \{\frac{n}{2} + 2, \frac{n}{2} + 3, \ldots, n\} \cup \{1\}$. We define the new weights of edges e of K_n as follows: $w(e) = c(e)$ unless both end-vertices of e are in R, in which case $w(e) = c(e) + f(n)P(n)$.

Clearly, the greedy algorithm constructs T' and T' remains the unique heaviest tour of K_n. Let $\mathcal{A}$ be the set of all tours alternating between L and R and containing the edge $e' = e_{n/2+1}$ and not containing the edge $e'' = e_1$. Observe that for each tour H in $\mathcal{A}$, we have $w(T')/w(H) \geq f(n)$. It remains to prove that every $H \in \mathcal{A}$ is an optimal tour and $|\mathcal{A}| = \Omega(\frac{(n-1)!}{2^n n^{1/2}})$.

Let $\mathcal{A}'$ be the set of tours alternating between L and R and containing the edge e'. Let $\mathcal{A}''$ be the set of tours in $\mathcal{A}'$ containing e''. Clearly, $\mathcal{A} = \mathcal{A}' \setminus \mathcal{A}''$. Let G be the induced subgraph of K_n obtained from K_n by deleting the vertices $\frac{n}{2} + 1$ and $\frac{n}{2} + 2$. Observe that there are $[(n/2 - 1)!]^2$ Hamilton paths in G. Each such path Q can be transformed into a tour in K_n by adding the edge e' and two more edges linking the end-vertices of Q with the end-vertices of e'. Thus, $|\mathcal{A}'| = [(n/2-1)!]^2$. Observe that there are $[(n/2-2)!]^2$ tours alternating between L and R and containing the edge e'' in the graph G. To form a tour containing e', e'' in K_n from a tour C containing e'' in G, it suffices to insert the edge e' into C such that e'' remains in the tour. This can be done in $n-3$ ways. Hence, $|\mathcal{A}''| = [(n/2-2)!]^2(n-3)$. So, we obtain that $|\mathcal{A}| = |\mathcal{A}'| - |\mathcal{A}''| = [(n/2 - 1)!]^2(1 - o(1)) = \Omega(\frac{(n-1)!}{2^n n^{1/2}})$.

Let $H \neq H' \in \mathcal{A}$. Observe that every tour C not alternating between L and R must contain an edge with both end-vertices in R. This, by the definition of w, implies that $w(H) < w(C)$. Let C be a tour alternating between L and R, but not in $\mathcal{A}$. To prove that $w(H) = w(H') < w(C)$, we add M to the weight of each edge incident to the vertex 1. Now observe that the sum of the weights of two edges of C incident to a vertex $i \in L$ equals $2iM+2$ provided none of the two edges coincides with e' or e''. Including e' into C, we decrease the weight of C by one and including e'' we increase it by M. Thus, $w(H) = w(H') < w(C)$. □

Acknowledgements We thank Ewa Drgas-Burchardt and Kamil Paraluch for bringing to our attention that the proof of Theorem 4.3 in [1] is incorrect.

References

[1] J. Bang-Jensen, G. Gutin and A. Yeo, When the greedy algorithm fails. *Discrete Optimization* **1** (2004), 121–127.

[2] G. Gutin and A. Yeo, Polynomial approximation algorithms for the TSP and the QAP with a factorial domination number. *Discrete Applied Mathematics* **119** (2002), 107–116.

[3] G. Gutin and A. Yeo, Anti-matroids. Oper. Res. Lett. **30** (2002), 97–99.

[4] G. Gutin, A. Yeo and A. Zverovich, Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discrete Applied Mathematics* **117** (2002), 81–86.

[5] A.P. Punnen, F. Margot and S.N. Kabadi, TSP heuristics: domination analysis and complexity. *Algorithmica* **35** (2003), 111-127.

[6] B. Korte and J. Vygen, Combinatorial Optimization. Springer-Verlag, Berlin, 3rd Ed., 2005.