

# *A priori* TSP in the Scenario Model

Martijn van Ee<sup>1</sup>(✉), Leo van Iersel<sup>2</sup>, Teun Janssen<sup>2</sup>, and René Sitters<sup>1,3</sup>

<sup>1</sup> Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

{m.van.ee,r.a.sitters}@vu.nl

<sup>2</sup> Delft University of Technology, Delft, The Netherlands

{l.j.j.vaniersel,t.m.l.janssen}@tudelft.nl

<sup>3</sup> Centrum voor Wiskunde en Informatica (CWI), Amsterdam, The Netherlands

r.a.sitters@cwi.nl

**Abstract.** In this paper, we consider the *a priori* traveling salesman problem (TSP) in the scenario model. In this problem, we are given a list of subsets of the vertices, called *scenarios*, along with a probability for each scenario. Given a tour on all vertices, the resulting tour for a given scenario is obtained by restricting the solution to the vertices of the scenario. The goal is to find a tour on all vertices that minimizes the expected length of the resulting restricted tour. We show that this problem is already NP-hard and APX-hard when all scenarios have size four. On the positive side, we show that there exists a constant-factor approximation algorithm in three restricted cases: if the number of scenarios is fixed, if the number of missing vertices per scenario is bounded by a constant, and if the scenarios are nested. Finally, we discuss an elegant relation with an *a priori* minimum spanning tree problem.

**Keywords:** Traveling salesman problem · A priori optimization · Master tour · Optimization under scenarios

## 1 Introduction

In universal and *a priori* routing, we extend our classical routing problems to the case that the set of clients is uncertain or changes regularly. Because reoptimizing over and over again might be inconvenient or impossible, we want to find a single tour. Given a tour and a set of clients, the active set, we shortcut the tour to the active set. In universal routing, the goal is to minimize the worst-case ratio of the value of the obtained solution and the deterministic optimal value. In *a priori* routing, we want to be good on average. The problem we consider in this paper is formally defined as follows.

In the *a priori* traveling salesman problem in the scenario model, we are given a complete weighted graph  $G = (V, E)$  and a set of scenarios  $\mathcal{S}$  with  $S_1, \dots, S_m \subseteq V$ . Scenario  $S_j$  has probability  $p_j$  of being the active set, where  $\sum_j p_j = 1$ . We begin by finding an ordering on  $V$ , called the first-stage tour. When an active set is released, the second-stage tour is obtained by shortcutting the first-stage tour on the vertices of the active set. The goal is to find a first-stage tour that minimizes the expected length of the second-stage tour.

This problem has, for example, a direct application to the photo-lithography processes used in semi-conductor manufacturing to transfer the geometric pattern of a chip onto a wafer [1]. This is done by putting UV light through a photomask on a photoresistant layer on top of the wafer. The entire wafer is not exposed at once, but one square at a time. If certain parts of the square do not need to be exposed, blades are moved in to block the UV light. Moving the blades is a time-consuming, and hence costly, process. Since it often influences the total processing time of a wafer in the lithography machine, minimizing the distance reduces the processing time. The blading positions are defined in a file. The blading positions are obtained from this file by reading it from top to bottom and the positions are used by the machine in order of appearance. A product will visit the photolithography machine multiple times during its fabrication. Every time it will use the same file that defines its blading positions, but it will not use all blading positions defined in the file in every visit. For each visit, there is a given subset of the blading positions that has to be used. Hence minimizing the movement of the blades comes down to finding an ordering of the blading positions such that the sum over all visits of the total distance between the blading-positions is minimized.

*A priori* TSP has already been considered in the independent decision and black-box model. In the independent decision model, vertex  $i$  is active with probability  $p_i$ , independent of the other vertices. Shmoys and Talwar [2] showed that a sample-and-augment approach gives a randomized 4-approximation, which can be derandomized to an 8-approximation algorithm. This factor was improved by van Zuylen [3] to 6.5. In the black-box model, we have no knowledge on the probability distribution over the vertices, but we are able to sample from it, i.e. to query the probability of any subset of the vertices. Schalekamp and Shmoys [4] showed that one can obtain a randomized  $O(\log n)$ -approximation even without sampling. A deterministic  $O(\log^2 n)$ -approximation can be obtained by using the result for universal TSP [5]. It was shown by [6] that there is an  $\Omega(\log n)$  lower bound for deterministic algorithms on general metrics. By using the result of [5] and Theorem 3 in [6], we also know that there is no deterministic algorithm with guarantee  $o\left(\sqrt[6]{\log n / \log \log n}\right)$  for planar metrics. For randomized algorithms, no lower bound is known for the black-box model.

The scenario model may be relevant for applications where the vertices are not active independently, but we do have some knowledge on the distribution. The former results give us the first results for *a priori* TSP in the scenario model. First of all, we inherit the randomized  $O(\log n)$ -approximation. Secondly, we know that a deterministic algorithm that does not use the information given in the scenarios will not achieve an approximation guarantee better than  $O(\log n)$ . The main question is whether we can use the scenarios to improve upon the  $O(\log n)$  upper bound and which restrictions we can put on the scenarios in order to obtain constant-factor approximability.

The scenario model has not been studied extensively for other optimization problems. Immorlica et al. [7] investigated stochastic versions of Vertex Cover and Shortest Path. Ravi and Sinha [8] also looked at these problems and also

defined stochastic scenario versions of Bin Packing, Facility Location and Set Cover. The problems in [8] differ from our setting in the sense that the weights used in the instance differ between scenarios. On the other hand, the work of [9] investigates a two-stage stochastic scheduling problem, where the set of jobs to be processed is uncertain. Finally, in [10], the classical scheduling problem of minimizing the makespan on two machines is considered in the *a priori* model with scenarios. It would be interesting to consider other stochastic combinatorial optimization problems in this framework.

In this paper, we will first examine the most natural lower bound, called the master tour lower bound. We use this lower bound to show that there exists a constant factor approximation algorithm for the problem if the number of scenarios is fixed. However, we also show that this lower bound cannot be used to improve upon the  $O(\log n)$ -approximation. We then look at several natural restrictions on the scenarios, namely small, big and nested. For small scenarios, we give strong inapproximability results. After that, we analyze the performance of the optimal tour on  $V$  for big scenarios. For nested scenarios, we show that there exists a 9-approximation algorithm. Finally, we show that there exists an elegant connection to an *a priori* minimum spanning tree problem. We end with a discussion on some open problems.

## 2 Master Tour Lower Bound

In this section, we explore the master tour lower bound. Here, we use that the contribution of scenario  $S_j$  to the objective value of an optimal solution, denoted by  $\text{OPT}$ , is at least  $p_j T_j^*$ , where  $T_j^*$  is the length of the optimal tour on  $S_j$ , so  $\text{OPT} \geq \sum_j p_j T_j^*$ . Two natural algorithms for *a priori* TSP in the scenario model are as follows. For each scenario, find an  $\alpha$ -approximate tour, where  $\alpha$  is the best approximation ratio available for TSP, and sort the scenarios on their resulting tour lengths  $T_j$ . Rename the scenarios such that  $T_1 \leq T_2 \leq \dots \leq T_m$ . Now traverse the tours  $1, 2, \dots, m$ , skipping already visited vertices, resulting in tour  $\tau_1$ . Alternatively, rename the scenarios such that  $p_1 \geq p_2 \geq \dots \geq p_m$  and traverse the tours  $1, 2, \dots, m$ , skipping already visited vertices, resulting in tour  $\tau_2$ . We get the following result.

**Theorem 1.** *Tours  $\tau_1$  and  $\tau_2$  are  $(2m - 1)$ -approximations for a priori TSP in the scenario model, where  $m \geq 2$  is the number of scenarios.*

*Proof.* Let us analyze tour  $\tau_1$ . Consider an arbitrary scenario  $S_j$ . If  $D_j$  is the diameter of  $G$  restricted to  $S_j$ , we have  $T_j^* \geq 2D_j$ . Note that when analyzing the contribution of scenario  $S_j$ , it might happen that two tours, say  $T_x$  and  $T_y$ , with  $x, y < j$ ,  $S_x \cap S_j \neq \emptyset$  and  $S_y \cap S_j \neq \emptyset$ , belong to disjoint scenarios. In this case, we have to go from  $T_x$  to  $T_y$ . If  $d(A, B)$  denotes the maximum distance between a vertex in  $A$  and a vertex in  $B$ , then this move costs us at most an extra  $d(S_x \cap S_j, S_y \cap S_j)$ . In the worst case, all scenarios before  $S_j$  have a non-empty intersection with  $S_j$ . For  $j = 1$ , the contribution is just  $p_1 T_1 \leq \alpha p_1 T_1^*$ . For  $j \geq 2$ , the contribution of  $S_j$  to the objective value of our solution is at most

$$\begin{aligned}
 & p_j(T_1 + d(S_1 \cap S_j, S_2 \cap S_j) + T_2 + \dots + d(S_{j-2} \cap S_j, S_{j-1} \cap S_j) + T_{j-1} + T_j) \\
 & \leq p_j(jT_j + (j-2)D_j) \leq p_j(\alpha jT_j^* + (j-2)\frac{1}{2}T_j^*) = ((\alpha + \frac{1}{2})j - 1)p_jT_j^*.
 \end{aligned}$$

Note that you do not have to incur an extra distance from  $S_{j-1}$  to  $S_j$ , since they have a non-empty intersection. In general, this holds for the last scenario that intersects with  $S_j$ . The objective value is at most

$$\alpha p_1 T_1^* + \sum_{j=2}^m ((\alpha + \frac{1}{2})j - 1) p_j T_j^* \leq \left( \left( \alpha + \frac{1}{2} \right) m - 1 \right) \text{OPT}.$$

Since  $\alpha = 1.5$  [11], we get a  $2m - 1$ -approximation algorithm. The analysis for  $\tau_2$  is similar. □

It turns out that the master tour lower bound will not give a constant approximation for general metrics. This can be deduced from Theorem 2 in [6], which roughly states the following. Suppose you are given a  $d$ -regular Ramanujan graph  $G$  on  $n$  vertices with girth  $g \geq \frac{2}{3} \log_{d-1} n$ . Take a random walk of length  $70g$  in  $G$  and let  $S$  be the vertices visited in this walk. Now, fix a first-stage tour. Theorem 2 in [6] states that for each of the first  $g/2$  steps of the tour restricted to  $S$ , the probability that the edge has length  $\Omega(\log n)$  is bounded from below by a constant.

**Theorem 2.** *There is an instance such that  $\text{OPT} = \Omega(\log n) \sum_j p_j T_j^*$  and  $\text{OPT} = \Omega(\log m) \sum_j p_j T_j^*$ .*

*Proof.* As before, suppose you are given a  $d$ -regular Ramanujan graph  $G$  on  $n$  vertices with girth  $g \geq \frac{2}{3} \log_{d-1} n$ . The scenarios correspond to vertex sets induced by random walks of length  $70g$  in  $G$ . For a fixed first-stage tour, Theorem 2 in [6] states that in each of the first  $g/2$  steps of the second-stage tour, there is a constant fraction of the scenarios that use an edge of length  $\Omega(\log n)$ . This implies that the expected length of the first  $g/2$  steps of the tour have expected length  $\Omega(\log n)$ . Since this is true for a constant number of steps, the lower bound also holds for the entire tour. Hence, we have an instance such that  $\text{OPT} = \Omega(\log n) \sum_j p_j T_j^*$ . The number of scenarios is equal to the number of possible walks of length  $70g$ . This is equal to  $n \cdot d^{70g} = O(nd^{\log n}) = O(n^{\log d})$ . Since  $d$  is a constant, this number is polynomially bounded. Hence, we have  $\Theta(\log m) = \Theta(\log n)$ , which gives us the second lower bound. □

A similar question one can ask is whether a given instance has an optimal value that is equal to the master tour lower bound. Stated differently, is there a tour such that if we shortcut on the vertices of a scenario, we get the optimal solution for that scenario? This problem is known as the Master Tour problem. In the original problem, every subset of vertices was a scenario. Deineko et al. [12] showed that this problem is polynomially solvable. We can reformulate the problem to the case where we are given a set of scenarios and we only have to be optimal for these scenarios. It turns out that this problem is  $\Delta_2^P$ -complete [13].

### 3 Small Scenarios

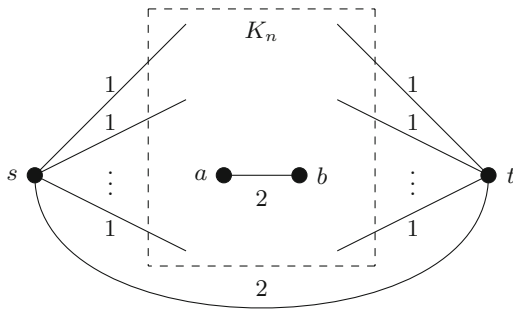
We start with showing that *a priori* TSP is still NP-complete when all scenarios are very small. We reduce from the Max Cut problem [14]. Here, we are given a graph  $G = (V, E)$  and our goal is to find a set  $S \subseteq V$  such that  $|\delta(S, \bar{S})|$  is maximized, where  $\delta(A, B)$  is the set of edges in the cut separating  $A$  from  $B$ .

**Theorem 3.** *A priori TSP is NP-complete when  $|S_j| \leq 4$  for all  $j$ .*

*Proof.* We are given an instance of Max Cut. Create an instance of *a priori* TSP by making a complete graph  $G'$  on  $V \cup \{s, t\}$ . All edges with  $s$  or  $t$  as endpoint, except edge  $(s, t)$ , have length 1 and all other edges have length 2 (see Fig. 1). For every edge  $(a, b)$  in  $E$ , we create a scenario  $\{a, b, s, t\}$ . All scenarios have equal probability. Note that a scenario can only have a contribution to the objective value of 4 or 6. We say that a scenario is satisfied if its resulting tour has length 4. Hence, minimizing the expected length is equivalent to maximizing the number of satisfied scenarios.

Suppose there is a cut of size at least  $k$  in  $G$ , say  $(Q_1, Q_2)$ . First, visit the vertices of  $Q_1$  in arbitrary order. After that, we visit  $s$ . Finally, we visit the vertices of  $Q_2$  in arbitrary order followed by  $t$ . It is easy to see that every scenario corresponding to an edge in the cut has length 4, whereas other scenarios have length 6. Hence, there is a tour satisfying at least  $k$  scenarios.

On the other hand, suppose that we have a tour in  $G'$  satisfying at least  $k$  scenarios. Without loss of generality, the tour can be written as  $sR_1tR_2$ , where  $R_1$  and  $R_2$  are sequences of vertices. The only way to satisfy a scenario is by putting one vertex in  $R_1$  and one vertex in  $R_2$ . Hence, the  $k$  satisfied scenarios correspond to edges in the cut  $(R_1, R_2)$  which has size at least  $k$ .  $\square$



**Fig. 1.** Graph  $G'$  as in the proof of Theorem 3.

By adjusting the proof of Theorem 3, we can prove that the master tour problem with scenarios is NP-complete when  $|S_j| \leq 5$ . This is done by reducing from Set Splitting instead of Max Cut. The fact above follows because 3-Set

Splitting is NP-complete [15]. The master tour problem with scenarios is still open for  $|S_j| \leq 4$ .

Note that the graph we used in the proof above can be obtained by taking the metric completion of  $K_{2,n}$ . This graph is planar, bipartite and it has treewidth and pathwidth equal to 2. Deterministic TSP would be polynomially solvable on such a graph with bounded treewidth. Furthermore, there is a PTAS for deterministic TSP in planar graphs [16]. The next theorem shows that this is not the case for *a priori* TSP (since the proof uses the same graph as before, a metric completion of  $K_{2,n}$ ). This theorem relies on the fact that Max Cut cannot be approximated within the Goemans-Williamson [17] constant, i.e. approximately 0.878567, unless the Unique Games Conjecture (UGC) fails [18], and it cannot be approximated within a factor  $\frac{16}{17}$ , unless  $P = NP$  [19].

**Theorem 4.** *There is no 1.0117-approximation for a priori TSP with  $|S_j| \leq 4$ , unless  $P = NP$ , and no 1.0242-approximation under UGC.*

*Proof.* Consider the reduction from the proof of Theorem 3. Let  $\text{OPT}_{\text{TSP}}$  and  $\text{OPT}_{\text{CUT}}$  be the optimal values of *a priori* TSP in the created instance and of Max Cut in the original instance respectively. We have  $\text{OPT}_{\text{TSP}} = 6|E| - 2\text{OPT}_{\text{CUT}}$ . If we have an  $(1 + \alpha)$ -approximation algorithm, we get a tour with total length at most  $(1 + \alpha)(6|E| - 2\text{OPT}_{\text{CUT}})$ . This implies that there are at least  $(1 + \alpha)\text{OPT}_{\text{CUT}} - 3\alpha|E|$  satisfied scenarios. These correspond to edges in the cut, hence we have

$$\begin{aligned} \text{Size of cut} &\geq (1 + \alpha)\text{OPT}_{\text{CUT}} - 3\alpha|E| \\ &\geq (1 + \alpha)\text{OPT}_{\text{CUT}} - 6\alpha\text{OPT}_{\text{CUT}} \\ &= (1 - 5\alpha)\text{OPT}_{\text{CUT}}, \end{aligned}$$

where the second inequality follows from  $\text{OPT}_{\text{CUT}} \geq |E|/2$ . Assuming  $P \neq NP$  or the Unique Games Conjecture, this means that there is no  $(1 + \alpha)$ -approximation for our problem for  $\alpha$ 's with  $1 - 5\alpha \geq \frac{16}{17}$  or  $1 - 5\alpha \geq 0.878567$  respectively. These inequalities are tight for  $\alpha \approx 0.0117$  and  $\alpha \approx 0.0242$  respectively.  $\square$

One could also consider the path-version of *a priori* TSP. In fact, the application on photolithography is modeled as the path-version. It is easy to see that this problem is trivial when  $|S_j| \leq 2$  for all  $j$ . If we delete  $t$  from the graph created in the reduction of Theorem 3, we can use this graph and the same reduction to show that the path-version of *a priori* TSP is NP-complete when  $|S_j| \leq 3$ . It is easy to see that this graph can be obtained by taking the metric completion of the star graph. Since the star has pathwidth 2, the problem is NP-complete on graphs with this property. On the other hand, the problem is trivially solvable on path graphs. Note that we can also adjust Theorem 4 to the path-version which will give the same inapproximability result, i.e. there is no 1.0117-approximation, unless  $P = NP$ , and there is no 1.0242-approximation under UGC.

We can strengthen the inapproximability of *a priori* TSP by using strong results on Permutation CSP's [20]. The problem that we need will be called 4-Undirected Cyclic Ordering (4-UCO). In this problem, we are given a ground

set  $T$  and a set of 4-tuples  $\Delta$  using elements from  $T$ . Our goal is to construct an ordering on  $T$  that maximizes the number of satisfied 4-tuples. We say that 4-tuple  $(a, b, c, d)$  is satisfied if one of the following sequences is a subsequence of the total ordering:  $(a, b, c, d)$ ,  $(b, c, d, a)$ ,  $(c, d, a, b)$ ,  $(d, a, b, c)$ ,  $(d, c, b, a)$ ,  $(c, b, a, d)$ ,  $(b, a, d, c)$ ,  $(a, d, c, b)$ . In other words, we get a collection of cycles and we want to find an ordering maximizing the number of cycles that can be embedded in it. To the best of our knowledge, the problem has never been considered. For completeness, we first show that the problem is NP-complete by using a reduction from Cyclic Ordering. In this problem, we are given a set of ordered triples  $\Delta$  of ground set  $T$ . The question is whether there exists a cyclic ordering on all elements such that each triple is ordered in the right direction. This problem is NP-complete [21].

**Theorem 5.** *4-Undirected Cyclic Ordering is NP-complete.*

In [20], it is shown that every Permutation CSP of constant arity is approximation resistant. This means that, under the Unique Games Conjecture, the best we can do is constructing a random ordering. Classical problems like Cyclic Ordering and Betweenness are in this class of problems. It is easy to see that 4-UCO is also in this class. A corollary of the work of Guruswami et al. is that there is no approximation algorithm with guarantee greater than  $\frac{1}{3}$ , assuming the Unique Games Conjecture is true. The natural generalization of 4-UCO is 5-UCO. For this problem, there is no algorithm having a guarantee greater than  $\frac{1}{12}$ . This gives the following results.

**Theorem 6.** *Under UGC, there is no  $\alpha$ -approximation for a priori TSP with*

- (a)  $\alpha < \frac{10}{9}$  when  $|S_j| \leq 6$ ,
- (b)  $\alpha < \frac{7}{6}$  when  $|S_j| \leq 8$ ,
- (c)  $\alpha < \frac{71}{60}$  when  $|S_j| \leq 10$ .

For the path-version, we can strengthen previous results by using the maximization version of Betweenness. In this problem, we are given a set of triples  $\Delta$  from elements of  $T$ . The triple  $(a, b, c)$  is satisfied if  $(a, b, c)$  or  $(c, b, a)$  is a subsequence of the total ordering. The goal is to find an ordering on  $T$  maximizing the number of satisfied triples. By [20], the best approximation ratio is  $\frac{1}{3}$ , unless the Unique Games Conjecture fails. Under the assumption that  $P \neq NP$ , there is no approximation for Max Betweenness with a factor better than  $\frac{1}{2}$  [22].

**Theorem 7.** *There is no  $\frac{9}{8}$ -approximation for a priori path-TSP with  $|S_j| \leq 5$ , unless  $P = NP$ , and no  $\frac{7}{6}$ -approximation under UGC.*

Finally, we note that by using twice the diameter of a scenario as a lower bound, we can show that an arbitrary tour is a  $c/2$ -approximation when  $|S_j| \leq c$ . A random tour gives a value of at most  $(c^2 - 3c + 4/2c - 2)$  times the optimal value in expectation. This factor approaches  $c/2$  for  $c$  large. Similar results hold for the path-version.

## 4 Big Scenarios

In this section, we investigate the special case of big scenarios, i.e. the case when each scenario has size at least  $n - c$ , for small  $c$ . One would expect that the optimal tour on the entire instance would perform well on these instances. Here, we analyze this option. Let us denote  $\text{OPT}(S)$  for the optimal value of a tour on  $S$ . Further, let  $\text{OPT}(S)|_T$  denote the value of the optimal tour on  $S$  shortcutted to  $T$ . As before, let  $D_S$  denote the diameter of the graph restricted to  $S$ .

**Lemma 1.** *For  $S \subset V$  and  $1 \leq c \leq \lceil n/2 \rceil$  such that  $|S| = n - c$ , we have*

$$\text{OPT}(V)|_S \leq \text{OPT}(S) + cD_S.$$

*Proof.* Suppose  $S = V \setminus \{a_1, \dots, a_c\}$ . Let  $\Delta_S^{a_i} = \min_{u \in S} d(u, a_i)$  for  $i = 1, \dots, c$ . Since we can extend our tour on  $S$  to  $V$  by going back and forth to each  $a_i$ , we have  $\text{OPT}(V) \leq \text{OPT}(S) + 2 \sum_{i=1}^c \Delta_S^{a_i}$ . We want to show that  $\text{OPT}(V)|_S \leq \text{OPT}(S) + cD_S$ . Suppose this is not the case, i.e.  $\text{OPT}(V)|_S > \text{OPT}(S) + cD_S$ . Furthermore, suppose w.l.o.g. that  $b_i$  and  $d_i$  are the two nodes in  $S$  that are visited before and after  $a_i$  in the optimal tour of  $V$ . If two consecutive vertices on the tour are not in  $S$ , then one can reconstruct the tour accordingly without increasing the length of the tour restricted to  $S$ . Then

$$\begin{aligned} \text{OPT}(V) &= \text{OPT}(V)|_S + \sum_{i=1}^c (d(b_i, a_i) + d(a_i, d_i) - d(b_i, d_i)) \\ &\geq \text{OPT}(V)|_S + \sum_{i=1}^c (2\Delta_S^{a_i} - d(b_i, d_i)) \\ &\geq \text{OPT}(V)|_S - cD_S + 2 \sum_{i=1}^c \Delta_S^{a_i} \\ &> \text{OPT}(S) + 2 \sum_{i=1}^c \Delta_S^{a_i} \end{aligned}$$

But this is a contradiction to our previous observation. Hence  $\text{OPT}(V)|_S \leq \text{OPT}(S) + cD_S$ .  $\square$

**Theorem 8.** *The optimal solution on  $V$  is a  $(1 + \frac{c}{2})$ -approximation for a priori TSP with  $|S_i| \geq n - c$ , where  $1 \leq c \leq \lceil \frac{n}{2} \rceil$ .*

## 5 Nested Scenarios

Let us now consider the case of nested scenarios, i.e.  $S_1 \subseteq S_2 \subseteq \dots \subseteq S_m$ . Here, the following algorithm gives a constant factor approximation. First, compute an 1.5-approximate tour  $T_j$  for scenario  $S_j$  for all  $j$ . Let  $\alpha_1 = 1$ . Next, for  $h = 2, 3, \dots$  let  $\alpha_h$  be the largest number  $k > \alpha_{h-1}$  for which  $T_k \leq 2T_{\alpha_{h-1}}$ . If no such  $k$  exists then let  $\alpha_h = \alpha_{h-1} + 1$ . The first-stage tour is obtained by visiting vertices in the order  $T_{\alpha_1}, T_{\alpha_2}, \dots$ .



**Theorem 9.** *The algorithm above is a 9-approximation for nested scenarios.*

*Proof.* Consider scenario  $S_j$ . The last vertices of this scenario will be visited on the tour  $T_{\alpha_h}$ , where  $h$  is the smallest index such that  $\alpha_h \geq j$ . Note that for any  $h \geq 2$ , we have  $T_{\alpha_h} > 2T_{\alpha_{h-2}}$ . Hence, we can decompose the concatenated tour up to  $T_{\alpha_h}$  into two parts which correspond to even and odd  $h$  respectively, such that both parts have geometrically increasing tour lengths. The length of the concatenated tour up to  $T_{\alpha_h}$  is therefore at most

$$2T_{\alpha_{h-1}} + 2T_{\alpha_h}.$$

If  $\alpha_h = j$  then the length of the tour is at most  $2T_{\alpha_{h-1}} + 2T_{\alpha_h} \leq 4T_{\alpha_h} = 4T_j \leq 6T_j^*$ .

If  $\alpha_h > j$  then we must have  $T_{\alpha_h} \leq 2T_{\alpha_{h-1}}$  so the length of the tour is at most  $2T_{\alpha_{h-1}} + 2T_{\alpha_h} \leq 6T_{\alpha_{h-1}} \leq 6T_j \leq 9T_j^*$ .  $\square$

The problem is still open for laminar scenarios, i.e. when for each  $i, j$ , either  $S_i \cap S_j = \emptyset$  or  $S_i \subseteq S_j$  or  $S_j \subseteq S_i$ . It is even open in the case when the scenarios have the following structure.

$$S_i \cap S_j = \emptyset \text{ for } i \neq j, i, j = 1, \dots, m-1, \text{ and } S_m = \bigcup_{j=1}^{m-1} S_j.$$

It would be interesting if one could get a constant factor approximation for these “starlike” (the inclusion graph is a star) instances.

## 6 Relation with Minimum Spanning Tree Problems

It would be nice to have a similar relation between *a priori* TSP and *a priori* MST as in the deterministic setting. We consider two versions of *a priori* MST. The first one is defined by Bertsimas [23], who called it *a priori* MST, while it seems more natural to call it *a priori* Steiner Tree. The second problem is defined by Boria et al. [24], who called it Probabilistic MST under Closest Ancestor. In both problems, we have a graph  $G = (V, E)$  and a probability distribution over subsets of vertices. The second problem also has a root  $r$  that is always active. This is optional in the first problem. The goal is to construct a tree on the entire vertex set in the first stage. A subset  $S$  of the vertices, drawn according to the probability distribution, is revealed in the second stage. In the *a priori* MST, the second-stage tree will be obtained by deleting inactive vertices, provided that the remaining tree stays connected. In the Probabilistic MST under Closest Ancestor, the second-stage tree only contains active vertices. This is done by taking an edge between an active vertex and its closest active ancestor in the rooted first-stage tree. In both problems, the goal is to construct a first-stage tour that minimizes the expected length of the second-stage tree.

Unfortunately, it turns out that the expected length of the optimal *a priori* MST defined by Bertsimas is not smaller than the optimal *a priori* TSP in general. The gap between the optimal values of a *a priori* MST and a *a priori* TSP can be arbitrarily large.

**Theorem 10.** *The optimal value of the a priori MST can be arbitrarily greater than the optimal value of the a priori TSP.*

However, the Probabilistic MST under Closest Ancestor can be used as a lower bound for a priori TSP. In fact, we only lose a factor 2. Note that this only works for the rooted case, since Probabilistic MST under Closest Ancestor is defined with a root vertex.

**Theorem 11.** *If there is an  $\alpha$ -approximation for the Probabilistic MST under Closest Ancestor, then there is a  $2\alpha$ -approximation for the a priori TSP, and vice versa.*

*Proof.* We show that the following inequalities are valid, where  $\text{OPT}_{\text{MST}}$  and  $\text{OPT}_{\text{TSP}}$  denote the optimal values of Closest Ancestor and a priori TSP respectively.

$$\text{OPT}_{\text{MST}} \leq \text{OPT}_{\text{TSP}} \leq 2\text{OPT}_{\text{MST}} \leq 2\text{OPT}_{\text{TSP}}.$$

The first inequality can be proven by taking the optimal a priori TSP-tour and deleting one edge. This gives a spanning tree on  $V$ , called  $T$ . If we look at a specific active set  $S$ , then the optimal a priori TSP-tour restricted to  $S$  will have exactly one edge less than before. Namely, if we delete edge  $(a, b)$  from tour  $(1, \dots, a, b, \dots, n)$ , only edge  $(\max\{k \in S : k \leq a\}, \min\{k \in S : k \geq b\})$  will disappear from the restricted tour on  $S$ . Note that for active set  $S$ , the tour without this edge is the same as  $T$  shortcutted to  $S$ . Hence, this is a feasible solution for Probabilistic MST under Closest Ancestor with cost no larger than the optimal value of a priori TSP, and the first inequality has been proven.

The second inequality is proven by doubling the optimal tree and shortcutting the obtained Eulerian tour. In each scenario, the cost of the edges is at most twice the cost of the edges in the tree restricted to the scenario. The third inequality follows from the first inequality.  $\square$

**Corollary 1.** *There is a randomized 8-approximation and a deterministic 13-approximation for Probabilistic MST under Closest Ancestor in the independent decision model. There is also a  $O(\log n)$ -approximation in the black-box model.*

Unfortunately, this does not imply a 2-approximation for a priori TSP, since we can prove that Probabilistic MST under Closest Ancestor is NP-complete in the scenario model. For this, we need the following lemma.

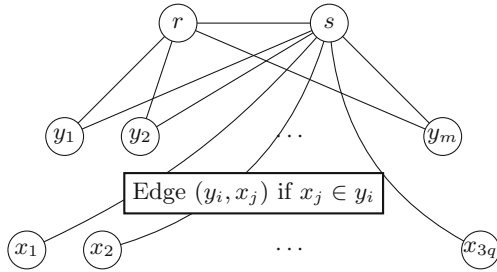
**Lemma 2.** *If Probabilistic MST under Closest Ancestor is NP-complete in the non-metric case, then it is NP-complete in the metric case.*

Boria et al. [24] showed that Probabilistic MST under Closest Ancestor is NP-complete in the independent decision model, but only for the non-metric case. Using Lemma 2, we obtain the following corollary.

**Corollary 2.** *Probabilistic MST under Closest Ancestor is NP-complete in the independent decision model, even if the triangle inequality is satisfied.*

**Theorem 12.** *Probabilistic MST under Closest Ancestor in the scenario model is NP-complete.*

*Proof.* We reduce the problem from the NP-complete problem Exact Cover by 3-Sets [14]. In this problem, we are given  $3q$  elements,  $x_1, \dots, x_{3q}$ , and  $m$  sets,  $y_1, \dots, y_m$ , containing three elements. The problem asks whether there are  $q$  sets that together cover all elements. Create the graph as in Fig. 2. Non-present edges have weight equal to  $M$ , where  $M$  is a large number. There are  $m$  scenarios with probability  $1/m$ . In each scenario, all  $x_j$ 's,  $r$  and  $s$  are active as is one of the  $y_i$ 's.



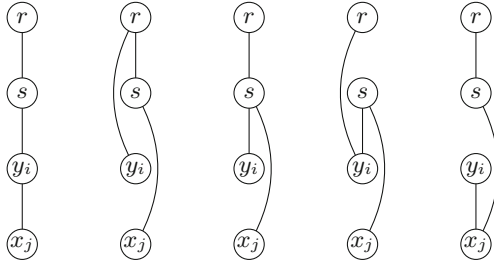
**Fig. 2.** Graph used in proof of Theorem 12. Edges  $(r, s)$  and  $(r, y_i)$  have length 0. Edges  $(s, y_i)$  and  $(y_i, x_j)$  have length 1. Edges  $(s, x_j)$  have length 2. Non-drawn edges have length  $M$ .

If there is an exact cover, then construct the following solution. If set  $y_i$  is chosen in the cover, then use edge  $(s, y_i)$  and the edges from vertex  $y_i$  to the corresponding elements of  $y_i$ . If set  $y_i$  is not in the cover, then use edge  $(r, y_i)$ . Finally, use edge  $(r, s)$ . This solution has expected value equal to  $q(1/m \cdot 4 + (m - 1)/m \cdot 6) = q(6 - 2/m)$ .

Note that an optimal tree will never use edges with weight  $M$  or a combination of edges that enforce using an edge of weight  $M$  in the shortcut solution. This leaves five ways of connecting a specific set vertex  $y_i$  and element vertex  $x_j$ , where  $j$  is in set  $i$ , to  $r$  and  $s$ . The five subtrees are depicted in Fig. 3.

Tree  $T_3$  is dominated by  $T_1$ , since  $T_1$  only has cost 2 for connecting  $x_j$  when  $y_i$  is inactive while  $T_3$  always has cost 2. Similarly,  $T_4$  is dominated by  $T_2$  and  $T_5$  is dominated by  $T_1$ . So, an optimal tree is a combination of  $T_1$  and  $T_2$ . Suppose that the tree connects  $k$  set vertices to  $s$  which connect  $\ell$  elements vertices. The other set vertices are connected to  $r$  whereas the other element vertices are connected to  $s$ . Number the  $k$  set vertices connected to  $s$  as  $1, \dots, k$  and say that set vertex  $i$  connects  $\ell_i$  element vertices. This tree has an expected value of

$$\frac{1}{m} \sum_{i=1}^k ((\ell_i + 1) - 2(3q - \ell_i)) + \frac{m - k}{m} 6q = 6q + \frac{1}{m}(k - \ell),$$



**Fig. 3.** Subtrees  $T_1$  up to  $T_5$ .

which is equal to  $q(6 - 2/m)$  if and only if  $k = q$  and  $\ell = 3q$ . Hence, there is a tree with expected value at most  $q(6 - 2/m)$  if and only if there is an exact cover. Using Lemma 2 completes the proof.  $\square$

## 7 Conclusion

In this paper, we showed how to get constant factor approximation for some well-structured problem instances. An interesting question that remains unanswered is whether there exists a constant factor approximation for laminar scenarios. More specifically, it is still open whether we can do this on “starlike” scenarios. One could also consider instances with restricted metrics. In Sect. 3 we showed that there is no PTAS for planar graphs. We do not have such results in the Euclidean plane. It would be interesting to settle the approximability of the problem in this metric. It is easy to construct examples where the optimal solution crosses itself and hence the non-crossing property does not hold. This property was a crucial ingredient of the PTAS by Arora [25] for the deterministic problem. So far, we have not been able to show any lower bound or improve the upper bound for this special case.

We did not succeed in improving the  $O(\log n)$ -approximation for the general problem. Next to the master tour lower bound, we investigated minimum spanning tree and linear programming approaches. However, preliminary results suggest that these approaches might not help us to break the barrier. In fact, we conjecture that there is no  $o(\log n)$ -approximation algorithm for *a priori* TSP in the scenario model in the general case.

**Acknowledgments.** We would like to thank Karen Aardal, Jan Driessen and Neil Olver for useful discussions. A part of the work by Teun Janssen has been performed in the project INTEGRATE “Integrated Solutions for Agile Manufacturing in High-mix Semiconductor Fabs”, co-funded by grants from France, Italy, Ireland, The Netherlands and the ECSEL Joint Undertaking. Martijn van Ee and René Sitters are supported by the NWO Grant 612.001.215. Leo van Iersel was partially supported by NWO and partially by the 4TU Applied Mathematics Institute.

## References

1. Driessen, J., Janssen, T.: Minimizing the blading in lithography machines: an application of the *a priori* TSP problem. Unpublished manuscript (2016)
2. Shmoys, D., Talwar, K.: A constant approximation algorithm for the *a priori* traveling salesman problem. In: Lodi, A., Panconesi, A., Rinaldi, G. (eds.) IPCO 2008. LNCS, vol. 5035, pp. 331–343. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-68891-4\\_23](https://doi.org/10.1007/978-3-540-68891-4_23)
3. Zuylen, A.: Deterministic sampling algorithms for network design. *Algorithmica* **60**, 110–151 (2011)
4. Schalekamp, F., Shmoys, D.B.: Algorithms for the universal and *a priori* TSP. *Oper. Res. Lett.* **36**, 1–3 (2008)
5. Hajiaghayi, M.T., Kleinberg, R., Leighton, T.: Improved lower and upper bounds for universal TSP in planar metrics. In: Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 649–658 (2006)
6. Gorodezky, I., Kleinberg, R.D., Shmoys, D.B., Spencer, G.: Improved lower bounds for the universal and *a priori* TSP. In: Serna, M., Shaltiel, R., Jansen, K., Rolim, J. (eds.) APPROX/RANDOM 2010. LNCS, vol. 6302, pp. 178–191. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15369-3\\_14](https://doi.org/10.1007/978-3-642-15369-3_14)
7. Immorlica, N., Karger, D.R., Minkoff, M., Mirrokni, V.S.: On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems. In: Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 691–700 (2004)
8. Ravi, R., Sinha, A.: Hedging uncertainty: approximation algorithms for stochastic optimization. *Math. Program.* **108**, 97–114 (2006)
9. Chen, L., Megow, N., Rischke, R., Stougie, L.: Stochastic and robust scheduling in the cloud. In: Proceedings of the 18th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, pp. 175–186 (2015)
10. Feuerstein, E., Marchetti-Spaccamela, A., Schalekamp, F., Sitters, R., Ster, S., Stougie, L., Zuylen, A.: Scheduling over scenarios on two machines. In: Cai, Z., Zelikovsky, A., Bourgeois, A. (eds.) COCOON 2014. LNCS, vol. 8591, pp. 559–571. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-08783-2\\_48](https://doi.org/10.1007/978-3-319-08783-2_48)
11. Christofides, N.: Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, DTIC Document (1976)
12. Deineko, V.G., Rudolf, R., Woeginger, G.J.: Sometimes travelling is easy: the master tour problem. *SIAM J. Discrete Math.* **11**, 81–93 (1998)
13. Ee, M., Sitters, R.: On the complexity of master problems. In: Italiano, G.F., Pighizzini, G., Sannella, D.T. (eds.) MFCS 2015. LNCS, vol. 9235, pp. 567–576. Springer, Heidelberg (2015). doi:[10.1007/978-3-662-48054-0\\_47](https://doi.org/10.1007/978-3-662-48054-0_47)
14. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds.) Complexity of Computer Computations, pp. 85–103. Springer, Heidelberg (1972)
15. Lovász, L.: Coverings and colorings of hypergraphs. In: Proceedings of the 4th Southeastern Conference on Combinatorics, Graph Theory and Computing, pp. 3–12 (1973)
16. Arora, S., Grigni, M., Karger, D.R., Klein, P.N., Woloszyn, A.: A polynomial-time approximation scheme for weighted planar graph TSP. In: Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 33–41 (1998)
17. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM (JACM)* **42**, 1115–1145 (1995)

18. Khot, S., Kindler, G., Mossel, E., O’Donnell, R.: Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.* **37**, 319–357 (2007)
19. Håstad, J.: Some optimal inapproximability results. *J. ACM* **48**, 798–859 (2001)
20. Guruswami, V., Håstad, J., Manokaran, R., Raghavendra, P., Charikar, M.: Beating the random ordering is hard: every ordering CSP is approximation resistant. *SIAM J. Comput.* **40**, 878–914 (2011)
21. Galil, Z., Megiddo, N.: Cyclic ordering is NP-complete. *Theor. Comput. Sci.* **5**, 179–182 (1977)
22. Austrin, P., Manokaran, R., Wenner, C.: On the NP-hardness of approximating ordering-constraint satisfaction problems. *Theor. Comput.* **11**, 257–283 (2015)
23. Bertsimas, D.: Probabilistic combinatorial optimization problems. Ph.D. thesis, Massachusetts Institute of Technology (1988)
24. Boria, N., Murat, C., Paschos, V.: On the probabilistic min spanning tree problem. *J. Math. Model. Algorithms* **11**, 45–76 (2012)
25. Arora, S.: Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM (JACM)* **45**, 753–782 (1998)