Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 89 (2000)

Approximation Hardness of TSP with Bounded Metrics

Lars Engebretsen[1,⋆] and Marek Karpinski[2,⋆⋆]

[1] Department of Numerical Analysis and Computer Science
Royal Institute of Technology
SE-100 44 Stockholm
SWEDEN
E-mail: enge@nada.kth.se

[2] Department of Computer Science
University of Bonn
53117 Bonn
GERMANY
E-mail: marek@cs.uni-bonn.de

Abstract. The general asymmetric TSP with triangle inequality is known to be approximable only within an $O(\log n)$ factor, and is also known to be approximable within a constant factor as soon as the metric is bounded. In this paper we study the asymmetric and symmetric TSP problems with bounded metrics, i.e., metrics where the distances are integers between one and some upper bound B. We first prove approximation lower bounds of 321/320 and 741/740 for the asymmetric and symmetric TSP with distances one and two, improving over the previous best lower bounds of 2805/2804 and 5381/5380. Then we consider the TSP with triangle inequality and distances that are integers between one and eight and prove approximation lower bounds of 131/130 for the asymmetric and 405/404 for the symmetric, respectively, version of that problem, improving over the previous best lower bounds of 2805/2804 and 3813/3812 by an order of magnitude.

Key words. Approximation Ratios; Lower Bounds; Metric TSP; Bounded Metric.

1 Introduction

A common special case of the Traveling Salesman Problem (TSP) is the metric TSP, where the distances between the cities satisfy the triangle inequality. The decision version of this special case was shown to be **NP**-complete by

⋆This research was performed while the author was visiting MIT with support from the Marcus Wallenberg Foundation and the Royal Swedish Academy of Sciences.
⋆⋆Supported in part by DFG grant, DIMACS, and IST grant 14036 (RAND-APX).

ISSN 1433-8092

Karp [13], which means that we have little hope of computing exact solutions in polynomial time. Christofides [7] has constructed an elegant algorithm approximating the metric TSP within 3/2, i.e., an algorithm that always produces a tour whose weight is at most a factor 3/2 from the weight of the optimal tour. For the case when the distance function may be asymmetric, the best known algorithm approximates the solution within $O(\log n)$, where n is the number of cities [10], although a constant factor approximation algorithm has recently been conjectured [6]. As for lower bounds, Papadimitriou and Yannakakis [15] have shown that there exists some constant, see also [1], such that it is **NP**-hard to approximate the TSP where the distances are constrained to be either one or two—note that such a distance function always satisfies the triangle inequality—within that constant. This lower bound was improved by Engebretsen [8] to $2805/2804 - \epsilon$ for the asymmetric and $5381/5380 - \epsilon$ for the symmetric, respectively, TSP with distances one and two. Böckenhauer et. al [4, 5] considered the symmetric TSP with distances one, two and three, and were able to prove a lower bound of $3813/3812 - \epsilon$. (For a discussion of bounded metric TSP, see also Trevisan [16].) It appears that the metric TSP lacks the good definability properties which seem to be needed for proving strong inapproximability results. Therefore, any new insights into explicit lower bounds here seem to be of a considerable interest.

Papadimitriou and Vempala [14] recently announced lower bounds of $42/41 - \epsilon$ and $129/128 - \epsilon$, respectively, for the asymmetric and symmetric versions, respectively, of the TSP with graph metric, but left the question of the approximability for the case with bounded metric open. However, their proof contained an error influencing the explicit constants. A new proof with the new constants of $98/97 - \epsilon$ and $234/233 - \epsilon$, respectively, has been communicated to us by Papadimitriou and Vempala (April 2001). Apart from being an interesting question on its own, it is conceivable that the special cases with bounded metric are easier to approximate than the cases when the distance between two points can grow with the number of cities in the instance. Indeed, the asymmetric TSP with distances bounded by B can be approximated within B by just picking any tour as the solution and the asymmetric TSP with distances one and two can be approximated within 4/3 [3]. The symmetric version of the latter problem can be approximated within 7/6 [15].

Definition 1.1. *The* Asymmetric Traveling Salesman Problem (ATSP) *is the following minimization problem: Given a collection of cities and a matrix whose entries are interpreted as the distance from a city to another, find the shortest tour starting and ending in the same city and visiting every city* *exactly once.*

Definition 1.2. (1,B)-ATSP *is the special case of ATSP where the entries in the distance matrix obey the triangle inequality and the off-diagonal entries in the distance matrix are integers between 1 and B.* (1,B)-TSP *is the special case of (1,B)-ATSP where the distance matrix is symmetric.*

In this paper, we first prove that it is **NP**-hard to approximate the asymmetric TSP with distances one and two within $321/320 - \epsilon$, for any constant $\epsilon > 0$. For the symmetric version of the latter problem we show a lower bound of $741/740 - \epsilon$. The previously best known bounds for this case are $2805/2804 - \epsilon$ and $5381/5380 - \epsilon$, respectively [8]. Our proofs contain explicit reductions from certain bounded dependency instances of linear equations satisfiability. The main idea is to represent every equation by a gadget and then link these gadgets together in a circle.

We then consider the case when the metric contains only integer distances between one and eight and prove, by generalizing the recent construction of Papadimitriou and Vempala, lower bound of $131/130 - \epsilon$ for the asymmetric case and $405/404 - \epsilon$ for the symmetric case.* This is an improvement of an order of magnitude compared to the previous best bounds of $2805/2804 - \epsilon$ and $3813/3812 - \epsilon$ for this case, respectively [4, 5, 8]. This requires us to establish the existence of a 7-regular bipartite graph with certain expansion properties—a result that may be of independent interest.

2 The hardness of (1,2)-ATSP

To prove a lower bound for (1,2)-ATSP we reduce from the problem *Hybrid*, introduced by Berman and Karpinski [2] to prove hardness results for instances of several combinatorial optimization problems where the number of occurrences of every variable is bounded by some constant.

Definition 2.1. Hybrid *is the following maximization problem: Given a system of linear equations mod 2 containing n variables, m_2 equations with exactly two unknowns, and m_3 equations exactly with three unknowns, find an assignment to the variables that satisfies as many equations as possible.*

Theorem 2.1 [2]. *There exists instances of Hybrid with 42ν variables, 60ν equations with two variables, and 2ν equations with three variables such that:*

*In a preliminary version of this paper [9] we erroneously claimed slightly better bounds for these two problems.

1. *Each variable occurs exactly three times.*

2. *For any constant $\epsilon' > 0$, it is **NP**-hard to decide if at most $\epsilon'\nu$ or at least $(1-\epsilon')\nu$ equations are left unsatisfied.*

Since we adopt the construction of Berman and Karpinski [2], we can partly rely on their main technical lemmas, which simplifies our proof of correctness.

In fact, every instance of Hybrid produced by the Berman-Karpinski construction has a very special structure. It can be viewed as a collection of wheels where the nodes correspond to variables and edges to equations. Figure 1 gives an example of one such wheel. The (hyper)edges connecting different wheels represent equations containing three unknowns. They are of the form $x + y + z = \{0, 1\}$; the number of such equations with right-hand side 0 is equal to the number of such equations with right-hand side 1. The edges within a wheel all represent equations of the form $x_i + x_j = 0$. To get exactly one negated occurrence and exactly two unnegated occurrences of each variable, we rewrite the equations corresponding to the perimeter of the wheel as $x_1 + \bar{x}_2 = 1$, $x_2 + \bar{x}_3 = 1$, $\ldots$, $x_{k-1} + \bar{x}_k = 1$, and $x_k + \bar{x}_1 = 1$.

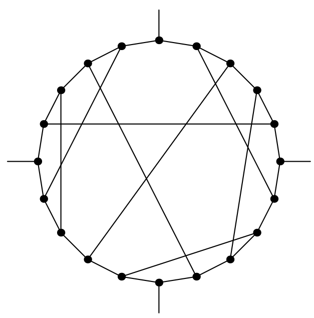


Figure 1. The Hybrid instance produced by the Berman-Karpinski construction can be viewed as a collection of wheels where the nodes correspond to variables and edges to equations.

Corollary 2.1. *There are instances of Hybrid with 42ν variables, 42ν equations of the form $x+\bar{y} = 1 \bmod 2$, 18ν equations of the form $x+y = 0 \bmod 2$, ν equations of the form $x + y + z = 0 \bmod 2$, and ν equations of the form $x + y + z = 1 \bmod 2$ such that:*

1. *Each variable occurs exactly three times, one time negatively and two times positively.*

2. For any constant $\epsilon' > 0$, it is **NP***-hard to decide if at most $\epsilon'\nu$ or at least $(1-\epsilon')\nu$ equations are left unsatisfied.*

To prove our hardness result for (1,2)-ATSP, we reduce instances of Hybrid of the form described in Corollary 2.1 to instances of (1,2)-ATSP and prove that, given a tour in the (1,2)-ATSP instance, it is possible to construct an assignment to the variables in the original Hybrid instance with the property that the number of unsatisfied equations in the Hybrid instance is related to the length of the tour in the (1,2)-ATSP instance.

Theorem 2.2. *Suppose that we are given an arbitrary instance of Hybrid with n variables, $m_{2,0}$ equations of the form $x+y=0 \bmod 2$, $m_{2,1}$ equations of the form $x+\bar{y}=1 \bmod 2$, $m_{3,0}$ equations of the form $x+y+z=0 \bmod 2$, and $m_{3,1}$ equations of the form $x+y+z=1 \bmod 2$ such that $m_{3,0}>0$ and each variable occurs exactly three times, two times positively and one time negatively.*

Then we can construct an instance of (1,2)-ATSP with the property that a tour of length $6n+m_{2,0}+m_{2,1}+4m_{3,0}+4m_{3,1}+u$ corresponds to an assignment satisfying all but u of the equations in the Hybrid instance.

Corollary 2.2. *For any constant $\epsilon > 0$, it is* **NP***-hard to approximate (1,2)-ATSP within $321/320-\epsilon$.*

Proof. Select $\epsilon' > 0$ such that $(321-\epsilon')/(320+\epsilon') \geq 321/320-\epsilon$. Consider an instance of Hybrid with the structure described in Corollary 2.1. By Theorem 2.2 we can construct an instance of (1,2)-ATSP with the property that a tour of length

$$6n+m_{2,0}+m_{2,1}+4m_{3,0}+4m_{3,1}+u = 6\cdot 42\nu+42\nu+18\nu+4\nu+4\nu+u = 320\nu+u$$

corresponds to an assignment satisfying all but u of the equations in the Hybrid instance. By Corollary 2.1 it is **NP**-hard to distinguish the cases $u \leq \epsilon'$ and $u \geq 1-\epsilon'$; therefore it is **NP**-hard to approximate (1,2)-ATSP within $(321-\epsilon')/(320+\epsilon') \geq 321/320-\epsilon$. ∎

The rest of this section is devoted to the proof of Theorem 2.2. On a high level, the (1,2)-ATSP instance in our reduction consists of a circle formed by *equation gadgets* representing the equations occurring in the corresponding instance of Hybrid. These equation gadgets are also connected through *consistency checkers* to ensure that it is possible to recover a consistent assignment from a tour.

To describe a (1,2)-TSP instance, it is enough to specify the edges of weight one. We do this by constructing a graph G and then let the (1,2)-TSP instance have the nodes of G as cities. The distance between two cities

u and v is defined to be one if (u, v) is an edge in G and two otherwise. To compute the weight of a tour, it is enough to study the parts of the tour traversing edges of G. In the asymmetric case G is a directed graph.

Definition 2.2. *We call a node where the tour leaves or enters* G *an* endpoint. *A node with the property that the tour both enters and leaves* G *in that particular node is called a* double endpoint *and counts as two endpoints.*

If c is the number of cities and $2e$ is the total number of endpoints, the weight of the tour is $c + e$ since every edge of weight two corresponds to two endpoints.

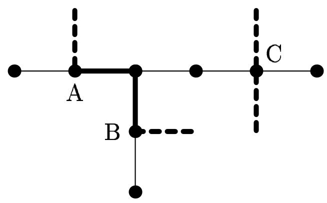


Figure 2. The above figure contains two partial tours—one entering the graph at A and leaving at B, and one both entering and leaving at C. The vertices A and B are endpoints and C is a double endpoint. The dashed parts of the tour denotes parts where the tour traverses edges with weight two.

2.1 The gadgets

The equation gadgets for equations of the form $x + y + z = \{0, 1\}$ are shown in Fig. 3. For homogeneous equations, the gadget is syntactically the same as the one used by Papadimitriou and Vempala [14]. The ticked edges are syntactic sugar for a construction ensuring consistency among the three occurrences of each variable. As we noted above, exactly one of the occurrences of a variable is negated. The construction in Fig. 4 ensures that the occurrences are given consistent values, i.e., that either $x = 0$ and $\bar{x} = 1$, or $x = 1$ and $\bar{x} = 0$. We say that the consistency checker has an *upper level*, corresponding to the negative occurrence of the variable, and a *lower level*, corresponding to the two positive occurrences. We use the term *occurrence* also to denote the path inside the consistency checker from an incoming connecting edge to the corresponding outgoing one. We say that an occurrence of a literal is *traversed* if both of its connecting edges are traversed, *untraversed* if none of its connecting edges are traversed, and *semitraversed* otherwise.

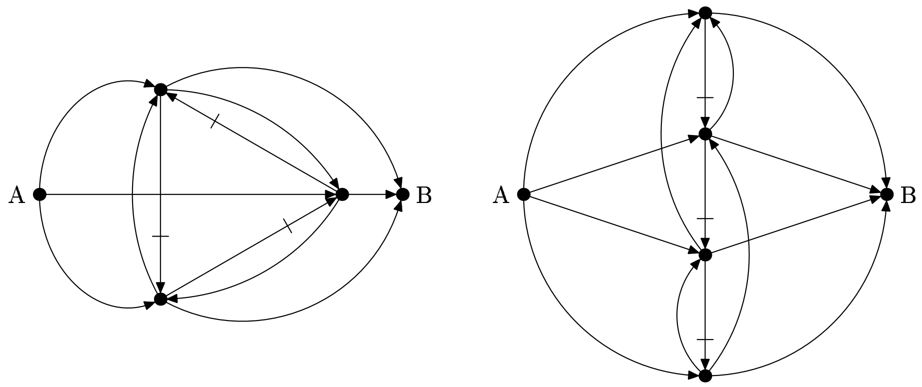


Figure 3. The gadget for equations of the form $x + y + z = 0$ (left) and $x + y + z = 1$ (right). There is a path of length 4 from A to B in the left gadget only if an even number of ticked edges is traversed and a path of length 5 in the right gadget only if an odd number of the ticked edges is traversed. All other traversals have an extra cost of at least 1.

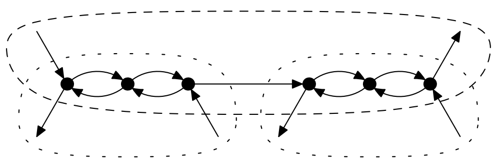


Figure 4. The gadget ensuring consistency for a variable. The ticked edges corresponding to the two positive occurrences are represented by the parts enclosed in the dotted curves and the ticked edge corresponding to the negative occurrence is represented by the part enclosed in the dashed curve.

The equation gadgets for equations of the form $x + y = \{0, 1\}$ are shown in Fig. 5. Note that there is no vertex between the two ticked edges in the gadget corresponding to equations of the form $x + y = 0$. Instead, the edge leaving the consistency checker corresponding to the first ticked edge is merged with the edge entering the consistency checker corresponding to the second ticked edge as shown in Fig. 6. This implies that the number of traversed edges in a satisfying traversal of the gadget—a traversal where either both or none of the occurrences are traversed—is 1, not counting the 6 edges that are attributed to each consistency checker.

The equation gadgets are hooked together in a circle in such a way that the vertex B in each gadget is identified with the vertex A in another gadget. The order of the gadgets is as follows: first all gadgets for equations of the form $x+y+z = 1$, then the gadgets for equations of the form $x+y+z = 0$, and finally the gadgets for equations with arity two. To make the amortized cost of the gadgets for equations of the form $x + y + z = 1$ lower—to match the cost of the gadgets for equations of the form $x + y + z = 0$—we combine the three outgoing edges for one gadget with the incoming edges for the next gadget as shown in Fig. 7.

The tour is intended to traverse the consistency checkers as shown in Fig. 11. This makes, for every variable x, the ticked edges corresponding to x and $\bar{x}$, respectively, traversed in a consistent way. If we let a traversal encode that the corresponding occurrence should be 1, it is easy to see that there will be two endpoints in the equation gadgets corresponding to unsatisfied equations and no endpoints anywhere else. A slight technicality arises here since the three occurrences in a gadget corresponding to equations of the form $x + y + z = 0$ cannot be simultaneously traversed—that would result in a short cycle. Therefore, we allow the tour to leave the consistency checker corresponding to one of the occurrences just after the bridge, see Figs. 12a–d, and then jump directly to B in Fig. 3. Similarly, we allow gadgets corresponding to equations of the form $x + y = 0$ to have one untraversed and one semitraversed consistency checker. In both of the above cases we still have two endpoints, one in the consistency checker and one in the equation gadget.

Definition 2.3. *We say that a tour is* normal *if all occurrences are either traversed or untraversed, except for the case when there are two traversed and one semitraversed occurrence in an equation gadget for equations of the form $x + y + z = 0$ or one untraversed and one semitraversed occurrence in an equation gadget for equations of the form $x + y = 0$.*

Definition 2.4. *A tour is* strictly normal*, if it is normal and it is impossible*

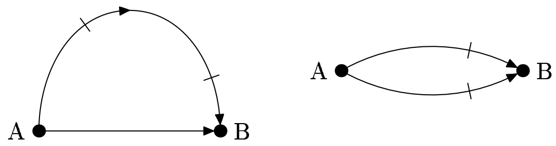


Figure 5. The gadget for equations of the form $x+y=0$ (left) and $x+y=1$ (right). There is a Hamiltonian path from A to B only if an even (left) or odd (right) number of the ticked edges is traversed.

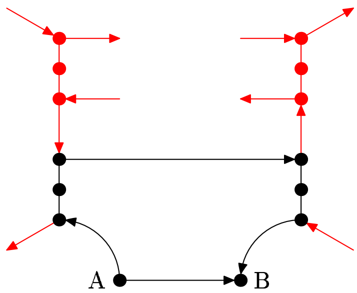


Figure 6. A more detailed view of the gadget for equations of the form $x+y=0$. In this figure the ticked edges have been expanded to show the consistency checkers. The black edges correspond to the gadget shown in Fig. 5

Figure 7. The cost of the gadgets for equations of the form $x+y+z=1$ is lowered by the above transformation. The last equation gadget for equations of the form $x+y+z=1$ is joined with the first equation gadget for equations of the form $x+y+z=0$ in a similar way.

to change any traversal of the form shown in Fig. 9 into the form shown in Fig. 11 without increasing the length of the tour.

Lemma 2.1. *Suppose that we are given an instance of Hybrid and construct from that instance an instance of (1,2)-ATSP as described above. Then it is possible to obtain from a strictly normal tour in this (1,2)-ATSP instance an assignment to the variables in the Hybrid instance such that there are two endpoints in the tour for every equation that is left unsatisfied and no other endpoints.*

Proof. The variables are given assignments as follows: Variables whose consistency checkers are traversed according to Figs. 9c and 11a are assigned 0; variables whose consistency checkers are traversed according to Figs. 9a–b, 11b and 12a–d are assigned 1. Since the tour is assumed to be strictly normal, and therefore normal, this covers all possible cases.

The only equations that are unsatisfied by this assignment are the ones where there are two endpoints within the corresponding equation gadget. Moreover there are no endpoints in other equation gadgets. ∎

2.2 Normalizing a tour

We want to prove that every unsatisfied equation has an extra cost of one associated with it. At first, it would seem that this is very easy—the gadgets in Fig. 3 is traversed by a path of length four if the equation is satisfied and a path of length at least five otherwise; the gadgets in Fig. 5 are traversed by a path of length one if the equation is satisfied and a path of length at least two otherwise; and the construction in Fig. 4 ensures consistency and is traversed by a tour of length six, not counting the edges that were accounted for above. Unfortunately, things are slightly more complicated than this. Due to the consistency checkers, the tour can leave a ticked edge when it is half-way through it, which forces us to be more careful in our analysis. Using a chain of local alterations of the tour, we prove that any tour can be *normalized*, i.e., converted into a strictly normal tour, of equal or lower cost. We can then recover an assignment to the variables in the Hybrid instance from this tour as described in Lemma 2.1. The normalization proceeds in four phases.

2.2.1 Make all bridges traversed

In the first phase, we first make all *bridges*, i.e., all pairs of undirected edges in the consistency checkers, traversed. By Lemma 2.2, this does not increase the length of the tour. Then we observe that some of the resulting traversals are suboptimal and can be transformed into other traversals according to Lemma 2.3.

Lemma 2.2. *Any tour can be modified to traverse both bridges in every consistency checker. Moreover, this transformation does not increase the length of the tour.*

Proof. By case analysis on the traversed connection edges. The lemma is clearly true if either the upper two or the lower four connection edges are traversed—then it is locally optimal to traverse the gadget as shown in Fig. 11. The case when none of the upper (but a subset of the lower) edges are traversed, and the case when none of the lower (but a subset of the upper) edges are traversed are treated in the same way. We now cover the remaining cases by an argument involving each bridge separately. When at most one of the four attaching edges are traversed by the tour, it is clearly locally optimal to traverse the bridge. The remaining cases are shown in Fig. 8. ∎

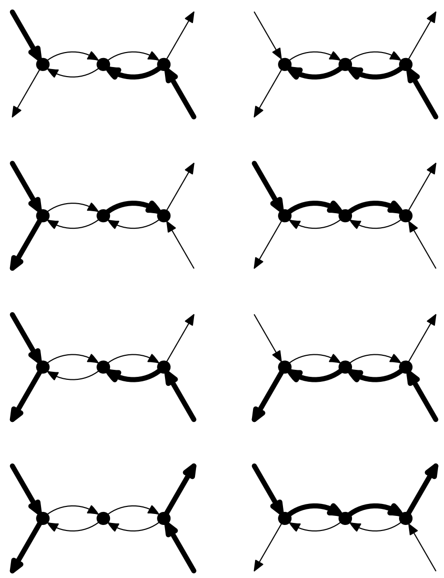


Figure 8. It is possible to change the traversals in the left column into the traversals in the right column without increasing the total number of endpoints in the graph.

Lemma 2.3. *It is locally optimal to traverse the consistency checker as shown in Figs. 9, 10a–d, 11, and 12a–h.*

Proof. By Lemma 2.2 it is locally optimal to traverse both bridges in the consistency checker. Therefore, an exhaustive list of possible traversals is shown in Figs. 9, 10, 11, and 12. Without increasing the number of endpoints in the tour, we can replace the traversals shown in Figs. 10e, g and i with the traversal shown in Fig. 10a; the ones shown in Figs. 10f, h and j with the one shown in Fig. 10b; and the ones shown in Figs. 12i–l with the one shown in Fig. 9c. ∎

2.2.2 Remove most of the semitraversals

In the second phase we remove most of the semitraversals on the lower level of the consistency checkers by repeatedly using the following observation:

Lemma 2.4. *A semitraversed occurrence on the lower level of the consistency gadget can be made traversed without affecting the other occurrences in the gadget.*

Proof. By Lemma 2.2, we can assume that the semitraversed occurrence has only one of its connecting edges traversed (see also Fig. 12). Therefore, we can make the semitraversed occurrence traversed without changing the tour on edges of the gadget outside that occurrence. ∎

Note that lower semitraversals always correspond to positive occurrences. We now transform the equation gadgets in such a way that most of the lower semitraversals disappear and the length of tour does not increase. Remember that we allow an equation gadget to have one semitraversed occurrence in certain cases according to Definition 2.3. We treat each type of equation gadget separately. After the transformations, we can assume that the tour has the following structure: The consistency checkers are traversed as shown in Figs. 9, 10a–d, 11, and 12a–h. The only semitraversed occurrences that remain appear in equations of the form $x+\bar{y}=1$ together with an untraversed occurrence, in equations of the form $x+y=0$ together with an untraversed occurrence and in equations of the form $x+y+z=0$ together with two traversed occurrences. Moreover, the semitraversed occurrences in equations of the form $x+\bar{y}=1$ always correspond to the negated variable.

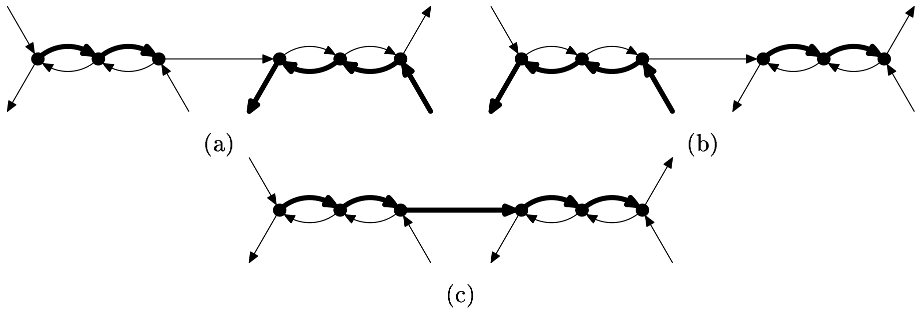


Figure 9. If there are no semitraversed occurrences in the consistency checker but the occurrences are still inconsistent, the checker has to be traversed as shown above.

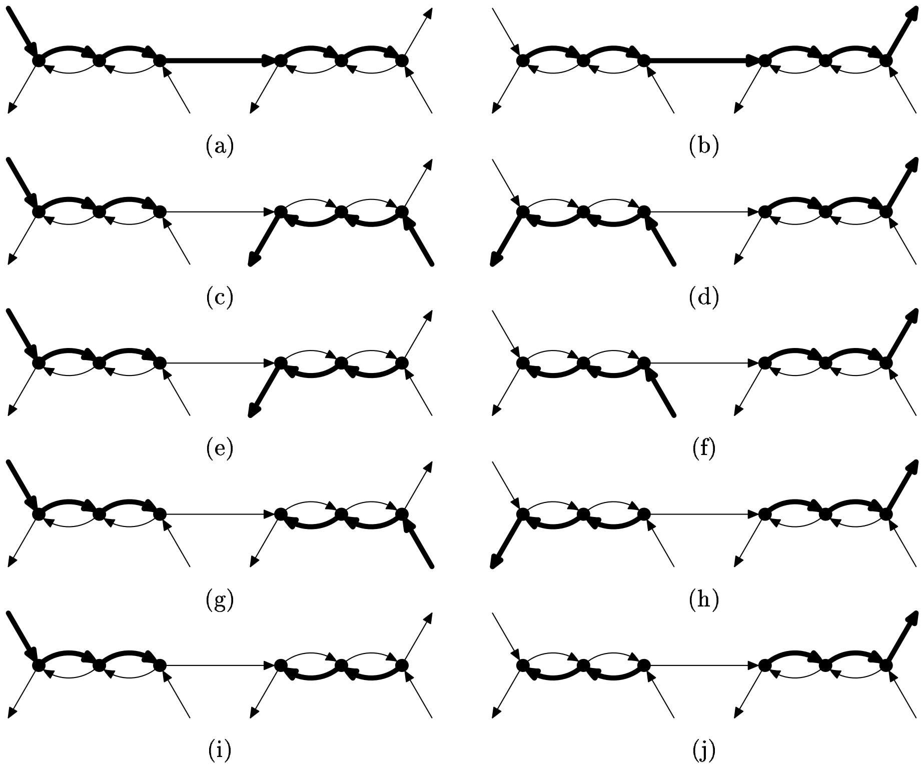


Figure 10. If the negative occurrence in the consistency checker is semitraversed, the checker has to be traversed as shown above.

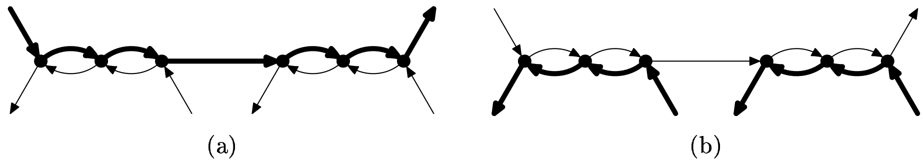


Figure 11. If either the upper two or the lower four connection edges are traversed in the consistency gadget, it is locally optimal to traverse the gadget as shown above.

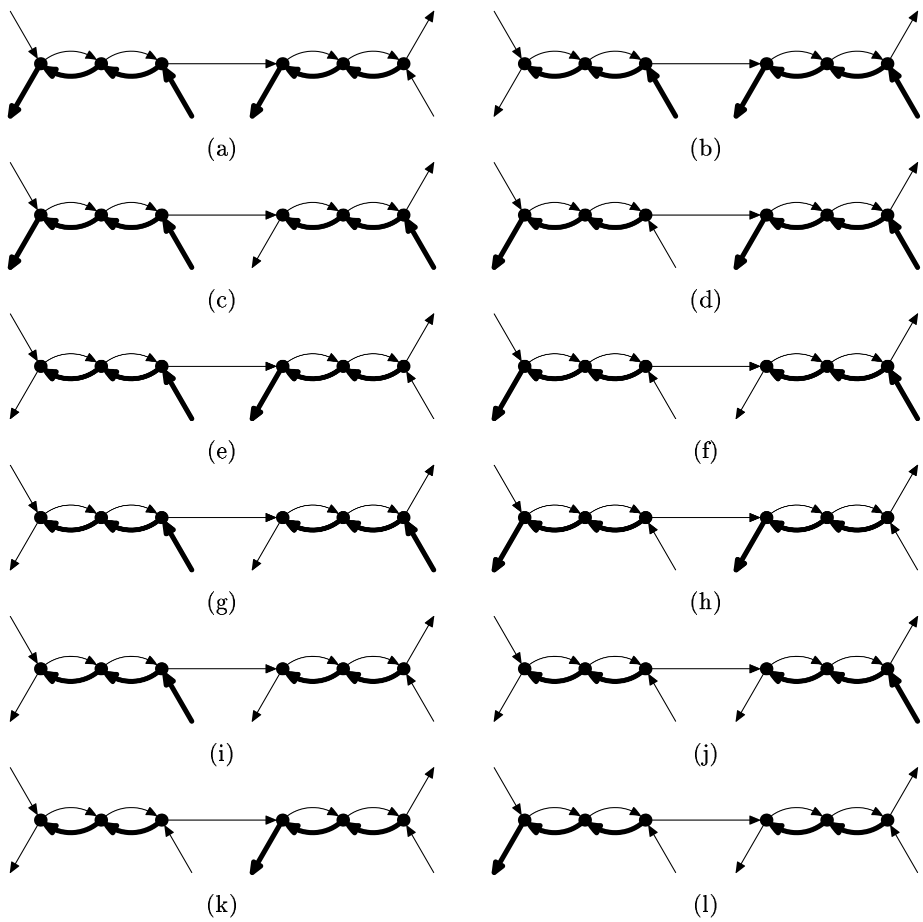


Figure 12. If there is at least one semitraversed occurrence in the consistency checker but the upper level is untraversed, the checker has to be traversed as shown above.

$\boldsymbol{x+y+z=0}$**:** If there are three semitraversed occurrences we make two of them traversed—that removes two endpoints. If there are two semitraversed occurrences and one traversed we make one of the semitraversed occurrences traversed—that keeps the number of endpoints constant. If there are two semitraversed occurrences and one untraversed we make both semitraversed occurrences traversed—that removes two endpoints. If there is one semitraversed occurrence and two untraversed ones we make the semitraversed occurrence traversed—that keeps the number of endpoints constant. If there is one semitraversed occurrence, one traversed and one untraversed we make the semitraversed one traversed—that removes two endpoints. The remaining cases—either no semitraversals or one semitraversal together with two traversals—are left unchanged.

$\boldsymbol{x+y+z=1}$**:** We can make any semitraversed occurrence traversed and then adjust the tour on the gadget in such a way that the total number of endpoints does not increase: If there are initially three semitraversed occurrences we remove at least two endpoints. If there are initially two semitraversed occurrences and one traversed, we remove two endpoints. If there are initially two semitraversed occurrences and one untraversed, we keep the number of endpoints constant. If there is initially one semitraversed occurrence and either two traversed or two untraversed, we remove two endpoints. Finally, if there is initially one semitraversed, one traversed and one untraversed occurrence, we keep the number of endpoints constant.

$\boldsymbol{x+y=0}$**:** If both occurrences are semitraversed we make them traversed—that removes two endpoints. If one occurrence is traversed and one is semitraversed we make the semitraversed one traversed—that removes two endpoints. It is impossible to have one traversed and one untraversed occurrence due to the construction of the equation gadget (see Fig. 6). The remaining two cases are left unchanged.

$\boldsymbol{x+\bar{y}=1}$**:** If both occurrences are semitraversed we make x traversed and $\bar{y}$ untraversed—that keeps the number of endpoints constant but removes the two semitraversals. If x is semitraversed and $\bar{y}$ is untraversed we make x traversed—that removes two endpoints. It is impossible to have two traversed occurrences since that results in a vertex which the alleged tour leaves in two different directions. Similarly, it is impossible two have one traversed and one semitraversed occurrence. The remaining case, x untraversed and $\bar{y}$ semitraversed, is left unchanged.

2.2.3 Remove the upper semitraversals

In the third phase we remove the upper semitraversals. This is actually the most non-trivial transformation, since it relies on the fact that the tour already has a certain structure. Note, that if the occurrence is traversed as shown in Fig. 12a–b it can be made traversed without affecting the lower occurrences, otherwise a more careful argument is needed. After this phase, the tour is normal. We repeat the following procedure for every semitraversed upper occurrence $\bar{y}$:

If $\bar{y}$ appears together with an untraversed occurrence x in an equation of the form $x + \bar{y} = 1$ and $\bar{y}$ is traversed as shown in Fig. 12a–b, we make $\bar{y}$ traversed—that saves two endpoints.

If $\bar{y}$ appears together with an untraversed occurrence x in an equation of the form $x + \bar{y} = 1$ and $\bar{y}$ is traversed as shown in Fig. 12c–d, we also make $\bar{y}$ traversed. This makes the lower traversed occurrence untraversed and adds two endpoints in the equation gadget corresponding to that occurrence, but this cost is set off against two saved endpoints in the gadget corresponding to $x + \bar{y} = 1$. Hence, the total number of endpoints is unchanged by the transformation.

2.2.4 Remove some inconsistent non-semitraversed checkers

In the fourth and final phase, we iterate through all consistency checkers that are traversed according to Fig. 9. If it is possible to make them traversed according to Fig. 11 without increasing the number of endpoints, we do so. After this phase, the tour is strictly normal.

2.2.5 Construct the assignment

Since the tour is now strictly normal, we can construct an assignment to the variables in the obvious way. Moreover, this assignment has the property that there are two endpoints for every unsatisfied equation and no other endpoints. Theorem 2.2 follows.

Proof of Theorem 2.2. Given an instance of Hybrid with the structure described in Corollary 2.1, the instance of (1,2)-ATSP is constructed as described above. By the normalization we can assume that the tour is strictly normal. We can thus construct an assignment to the variables as described in the proof of Lemma 2.1; this assignment leaves u equations unsatisfied when the length of the tour is $6n + m_{2,0} + m_{2,1} + 4m_{3,0} + 4m_{3,1} + u$. ∎

3 The hardness of (1,2)-TSP

It is possible to adapt the above construction for (1,2)-ATSP to prove a lower bound also for (1,2)-TSP, yielding the following result:

Theorem 3.1. *Suppose that we are given an arbitrary instance of Hybrid with n variables, $m_{2,0}$ equations of the form $x + y = 0 \bmod 2$, $m_{2,1}$ equations of the form $x + \bar{y} = 1 \bmod 2$, $m_{3,0}$ equations of the form $x + y + z = 0 \bmod 2$, and $m_{3,1}$ equations of the form $x + y + z = 1 \bmod 2$ such that $m_{3,0} > 0$ and each variable occurs exactly three times, two times positively and one time negatively.*

Then we can construct an instance of (1,2)-TSP with the property that a tour of length $16n + m_{2,0} + m_{2,1} + 4m_{3,0} + 4m_{3,1} + u$ corresponds to an assignment satisfying all but u of the equations in the Hybrid instance.

Corollary 3.1. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,2)-TSP within $741/740 - \epsilon$.*

Proof. Select $\epsilon' > 0$ such that $(741 - \epsilon')/(740 + \epsilon') \geq 741/740 - \epsilon$. Consider an instance of Hybrid with the structure described in Corollary 2.1. By Theorem 3.1 we can construct an instance of (1,2)-TSP with the property that a tour of length

$$16 \cdot 42\nu + 42\nu + 18\nu + 4\nu + 4\nu + u = 740\nu + u$$

corresponds to an assignment satisfying all but u of the equations in the Hybrid instance. By Corollary 2.1 it is **NP**-hard to distinguish the cases $u \leq \epsilon'$ and $u \geq 1 - \epsilon'$; therefore it is **NP**-hard to approximate (1,2)-TSP within $(741 - \epsilon')/(740 + \epsilon') \geq 741/740 - \epsilon$. ∎

We now describe how to construct the instances of (1,2)-TSP described in Theorem 3.1. The equation gadgets for the symmetric case are shown in Figs. 13 and 14. As in the asymmetric case the ticked edges are syntactic sugar for a construction ensuring consistency among the three occurrences of each variable. The construction is shown in Fig. 15 and expanded versions of the gadgets for equations of the form $x + y + z = 0$ and $x + y = 0$, respectively, are shown in Figs. 16 and 17, respectively. In the same way as in the asymmetric case, the equation gadgets are connected in a circle. We also use the same trick as in the asymmetric case to lower the cost for the equation gadgets for equations of the form $x + y + z = 1$ by one, to make it the same as the cost for the equation gadgets for equations of the form $x + y + z = 0$.

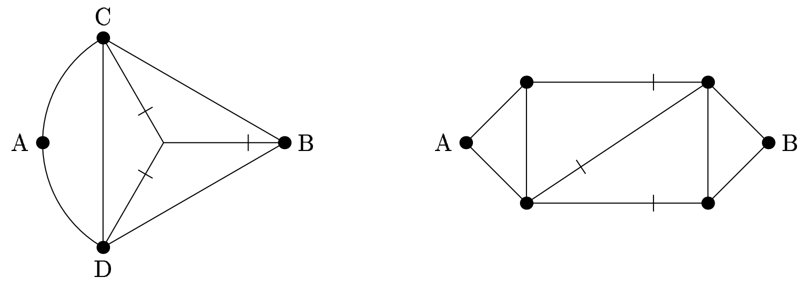


Figure 13. The gadget for equations of the form $x + y + z = 0$ (left) and $x + y + z = 1$ (right). There is a path of length 4 from A to B in the left gadget only if an even number of ticked edges is traversed and a path of length 5 in the right gadget only if an odd number of the ticked edges is traversed. All other traversals have an extra cost of at least 1.

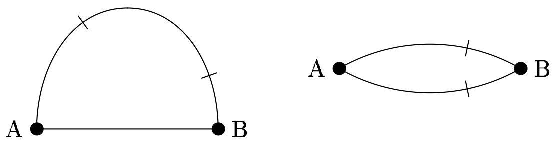


Figure 14. The gadget for equations of the form $x + y = 0$ (left) and $x + y = 1$ (right). There is a Hamiltonian path from A to B only if an even (left) or odd (right) number of the ticked edges is traversed.

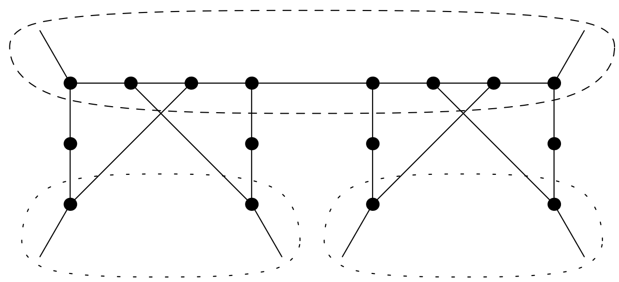


Figure 15. The gadget ensuring consistency for a variable. The ticked edges corresponding to the two positive occurrences are represented by the parts enclosed in the dotted curves and the ticked edge corresponding to the negative occurrence is represented by the part enclosed in the dashed curve.

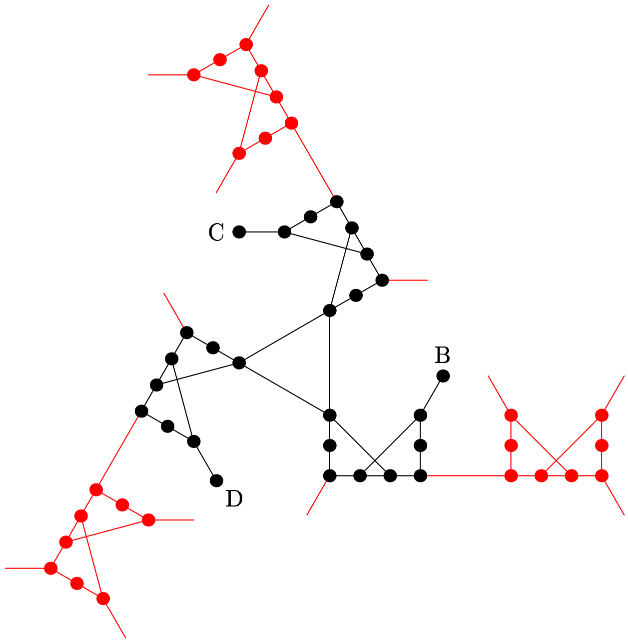


Figure 16. A more detailed view of the gadget for equations of the form $x + y + z = 0$. The figure shows how the three variable gadgets meet in the center of the gadget. The black edges above correspond to the ticked edges in Fig. 13 and the three labeled vertices above are the same as the corresponding vertices in Fig. 13.

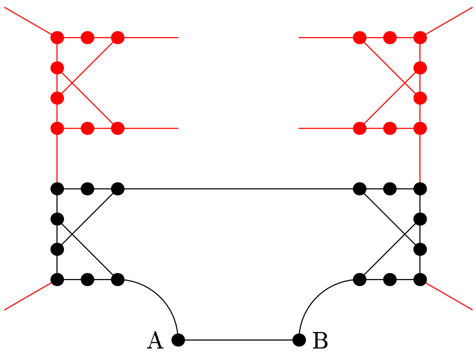


Figure 17. A more detailed view of the gadget for equations of the form $x + y = 0$. In this figure the ticked edges have been expanded to show the consistency checkers. The black edges correspond to the gadget shown in Fig. 14

The tour is intended to traverse the consistency checkers as shown in Fig. 18. This makes, for every variable x, the ticked edges corresponding to x and $\bar{x}$, respectively, traversed in a consistent way. If we let a traversal encode that the corresponding occurrence should be 1, it is easy to see that there will be two endpoints in the equation gadgets corresponding to unsatisfied equations and no endpoints anywhere else. Again, a slight technicality arises here since the three occurrences in a gadget corresponding to equations of the form $x + y + z = 0$ cannot be simultaneously traversed—this technicality is resolved in the same way as in the asymmetric case. Similarly, we allow gadgets corresponding to equations of the form $x + y = 0$ to have one untraversed and one semitraversed consistency checker. In both of the above cases we still have two endpoints, one in the consistency checker and one in the equation gadget.

Lemma 3.1. *Suppose that we are given an instance of Hybrid and construct from that instance an instance of (1,2)-TSP as described above. Then it is possible to obtain from a strictly normal tour in this (1,2)-TSP instance an assignment to the variables in the Hybrid instance such that there are two endpoints in the tour for every equation that is left unsatisfied and no other endpoints.*

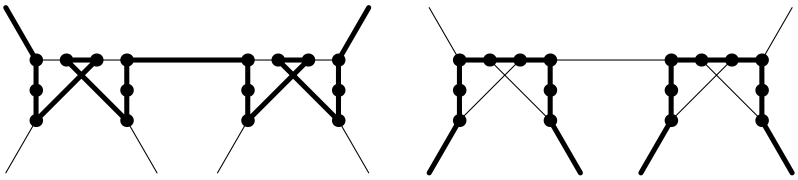
Proof. The variables are given assignments as follows: Variables whose consistency checkers are traversed according to Figs. 18a and 20c are assigned 0; variables whose consistency checkers are traversed according to Figs. 18b, 20a–b and 21 are assigned 1. Since the tour is assumed to be strictly normal, and therefore normal, this covers all possible cases.

The only equations that are unsatisfied by this assignment are the ones where there are two endpoints within the corresponding equation gadget. Moreover there are no endpoints in other equation gadgets. ∎

3.1 Normalizing a tour

As in the asymmetric case, we need to prove that every unsatisfied equation has an extra cost of one associated with it. The steps in the normalization are exactly the same in the symmetric and the asymmetric cases; therefore we only give the proof of the analogue of Lemma 2.2—the lemma establishing that it is locally optimal to traverse the bridges in the consistency checkers.

Lemma 3.2. *Any tour can be modified to traverse both bridges in every consistency checker. Moreover, this transformation does not increase the length of the tour.*



(a) (b)

Figure 18. If either the upper two or the lower four connection edges are traversed in the consistency gadget, it is locally optimal to traverse the gadget as shown above.

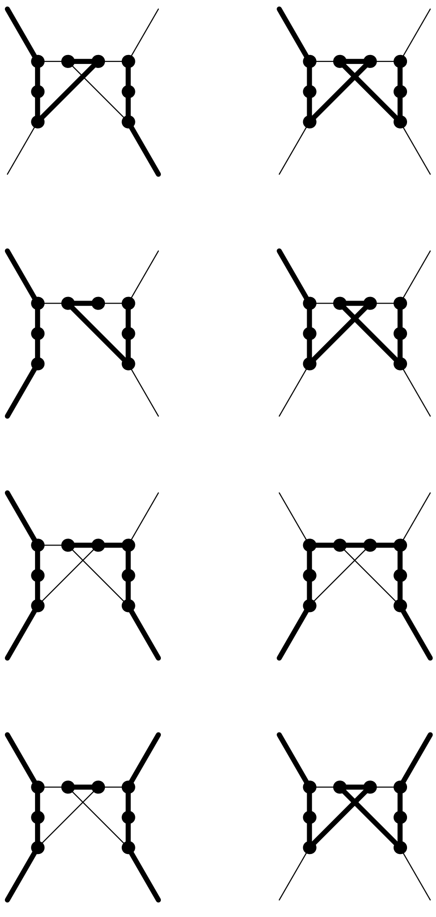


Figure 19. It is possible to change the traversals in the left column into the traversals in the right column without increasing the total number of endpoints in the graph.

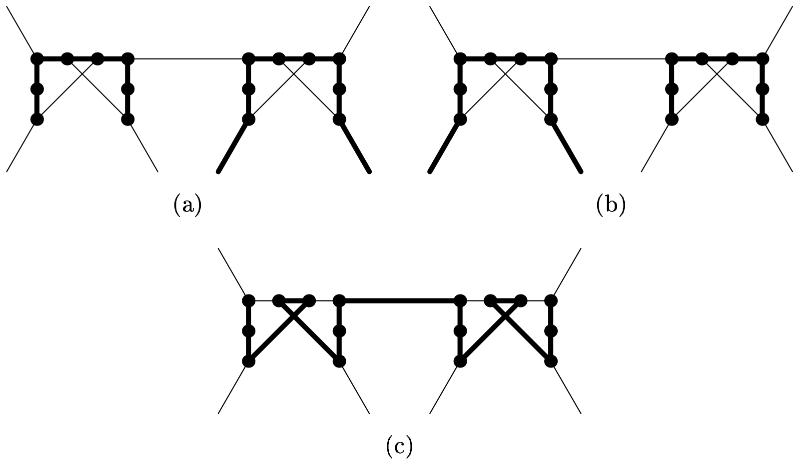


Figure 20. If there are no semitraversed occurrences in the consistency checker but the occurrences are still inconsistent, the checker has to be traversed as shown above.

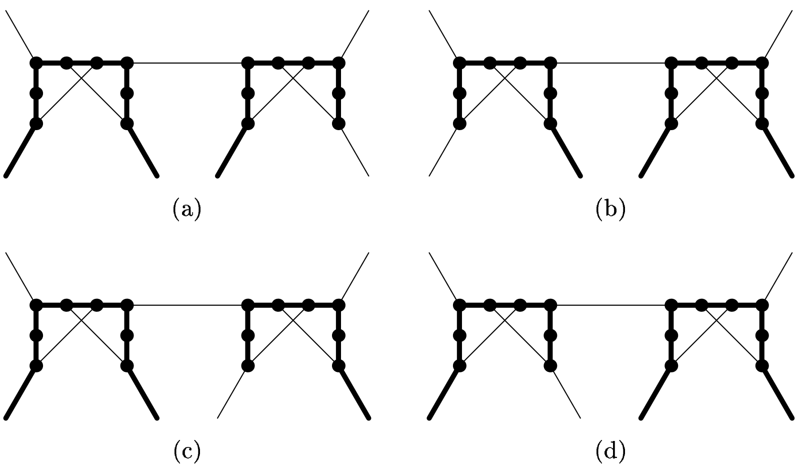


Figure 21. The semitraversals shown above are the only ones that can occur in a strictly normal tour.

Proof. The proof is very similar to the proof for the asymmetric case. The only additional complication is that some occurrences actually have two connection edges on one side due to the construction of the gadget for equations of the form $x+y+z=0$ (Fig. 16). However, it can never be locally optimal to traverse both of these connection edges since that introduces an endpoint in the consistency checker. Figure 22 shows one example of this—the other cases are treated similarly.

If either the upper two or the lower four connection edges are traversed, the lemma clearly holds—then it is locally optimal to traverse the gadget as shown in Fig. 18. The case when none of the upper (but a subset of the lower) edges are traversed, and the case when none of the lower (but a subset of the upper) edges are traversed are treated in the same way. We now cover the remaining cases by an argument involving each bridge separately. When at most one of the four attaching edges are traversed by the tour, it is clearly locally optimal to traverse the bridge. The remaining cases are shown in Fig. 19. ∎

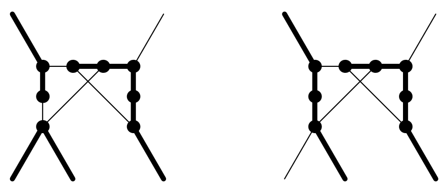


Figure 22. Some consistency checkers have double connection edges at one point, see also Fig. 16. By the above local transformation we can assume that at most one of the double edges are traversed.

The other steps in the normalization are identical to the corresponding steps for the asymmetric case—we omit the details. As in the asymmetric case, the theorem follows.

Proof of Theorem 3.1. Given an instance of Hybrid with the structure described in Corollary 2.1, the instance of (1,2)-ATSP is constructed as described above. By the normalization we can assume that the tour is strictly normal. We can thus construct an assignment to the variables as described in the proof of Lemma 3.1; this assignment leaves u equations unsatisfied when the length of the tour is $16n + m_{2,0} + m_{2,1} + 4m_{3,0} + 4m_{3,1} + u$. ∎

4 The hardness of (1,$\boldsymbol{B}$)-ATSP

In this section, we establish that the construction introduced by Papadimitriou and Vempala [14] can be adapted to bounded metrics. We reduce, similarly to Papadimitriou and Vempala [14], from Håstad's lower bound for E3-Lin mod 2 [12]. In fact, our gadgets for the (1,B)-ATSP case are syntactically identical to those of Papadimitriou and Vempala [14] but we use a slightly different accounting method. The construction consists of a circle of *equation gadgets* testing odd parity. This is no restriction since we can easily transform a test for even parity into a test for odd parity by flipping a literal. Three of the edges in the equation gadget correspond to the variables involved in the parity check. These edges are in fact gadgets, so called *edge gadgets*, themselves. Edge gadgets from different equation gadgets are connected to ensure consistency among the edges representing a literal. This requires the number of negative occurrences of a variable to be equal to the number of positive occurrences. This is no restriction since we can duplicate every equation a constant number of times and flip literals to reach this property.

Definition 4.1. E3-Lin mod 2 *is the following problem: Given an instance of* n *variables and* m *equations over* $\boldsymbol{Z}_2$ *with exactly three unknowns in each equation, find an assignment to the variables that satisfies as many equations as possible.*

Theorem 4.1 [12]. *For any constant* $\epsilon > 0$*, there exists instances of E3-Lin mod 2 with* $2m$ *equations such that it is* $\boldsymbol{NP}$*-hard to decide if at most* ϵm *or at least* $(1-\epsilon)m$ *equations are left unsatisfied by the optimal assignment. Each variable in the instance occurs a constant number of times, half of them negated and half of them unnegated.*

We describe our instance of (1,B)-ATSP by constructing a weighted directed graph and then let the (1,B)-ATSP instance have the nodes of this graph as cities. The distance between two cities u and v is the (1,B)-ATSP instance is then defined to be $\min\{B, \ell(u,v)\}$, where $\ell(u,v)$ be the length of the shortest path from u to v in the graph.

4.1 The gadgets

The gadgets are parametrized by the parameters a, b and d; they will be specified later. The equation gadget for equations of the form $x+y+z=0$ is shown in Fig. 23. The key property of this gadget is that there is a Hamiltonian path through the gadget only if zero or two of the ticked edges are traversed. To form the circle of equation gadgets, vertex A in one gadget coincides with vertex B in another gadget.

The ticked edges in Fig. 23 are gadgets themselves. This gadget is shown in Fig. 24. Each of the bridges is shared between two different edge gadgets, one corresponding to a positive occurrence of the literal and one corresponding to a negative occurrence. The precise coupling is provided by a perfect matching in a d-regular bipartite multigraph $(V_1 \cup V_2, E)$ on $2k$ vertices with the following property: For any partition of V_1 into subsets S_1, U_1 and T_1 and any partition of V_2 into subsets S_2, U_2 and T_2 such that there are no edges from T_1 to T_2 and no edges from U_1 to U_2,

$$\begin{aligned}(|S_1| + |S_2|) \min\{a/2, b, a/2 + b/2 - 1, a/2 + b/4 - 1/2\} \geq \\ \min\{k, |U_1| + |T_2| + |S_1| + |S_2|, |U_2| + |T_1| + |S_1| + |S_2|\}.\end{aligned}$$

The purpose of this construction is to ensure that it is always optimal for the tour to traverse the graph in such a way that all variables are given consistent values. The edge gadget gives an assignment to an occurrence of a variable by the way it is traversed.

Definition 4.2. *We call an edge gadget where all bridges are traversed from left to right in Fig. 24* traversed *and an edge gadget where all bridges are traversed from right to left* untraversed. *All other edge gadgets are called* semitraversed.

4.2 The lower bound

If we assume that the tour behaves nicely, i.e., that the edge gadgets are either traversed or untraversed, it is straightforward to establish a correspondence between the length of the tour and the number of unsatisfied equations.

Lemma 4.1. *The only way to traverse the equation gadget in Fig. 23 with a tour of length* 4*—if the edge gadgets count as length one for the moment—is to traverse an odd number of edge gadgets. All other locally optimal traversals have length* 5.

Proof. It is easy to see that any tour traversing two ticked edges and leaving the third one untraversed has length 4. Any tour traversing one ticked edge and leaving the other two ticked edges untraversed has length at least 5. Strictly speaking, it is impossible to have three traversals since this does not result in a tour. However, we can regard the case when the tour leaves the

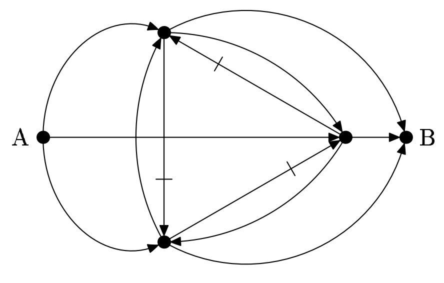


Figure 23. The gadget for equations of the form $x + y + z = 0$. There is a Hamiltonian path from A to B only if zero or two of the ticked edges, which are actually gadgets themselves (Fig. 24), are traversed. The non-ticked edges have weight 1.

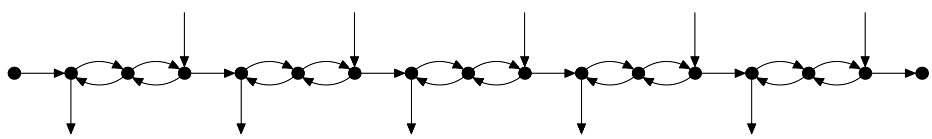


Figure 24. The edge gadget consists of d bridges. Each of the bridges are shared between two different edge gadgets and consist of a undirected edges of weight 1 each. The leftmost directed edge above has weight $b/2$, the rightmost has weight $b/2 + 1$, all other horizontal edges entering or leaving a bridge have weight b.

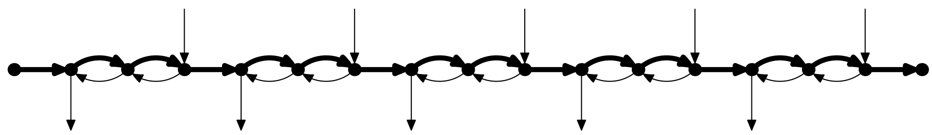


Figure 25. A traversed edge gadget represents the value 1.

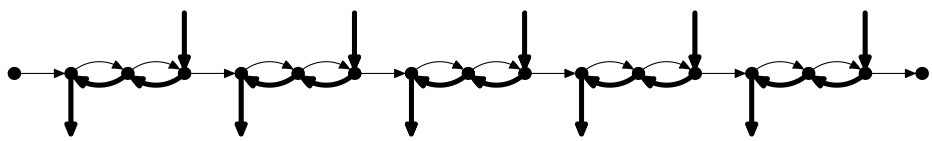


Figure 26. An untraversed edge gadget represents the value 0.

edge gadget by jumping directly to the exit node of the equation gadget as a tour with three traversals; such a tour gives a cost of 5. ∎

Lemma 4.2. *In addition to the length* 1 *attributed to the edge gadget above, the length of a tour traversing an edge gadget in the intended way is* $d(a+b)$.

Proof. Each bridge has length a, and every bridge must have one of the incoming edge traversed. Thus, the total cost is $d(a+b)$. ∎

Lemma 4.3. *Suppose that there are* $2m$ *equations in the E3-Lin mod 2 instance. If the tour is shaped in the intended way, i.e., every edge gadget is either traversed or untraversed, the length of the tour is* $3md(a+b)+4m+u$*, where* u *is the number of unsatisfied equations resulting from the assignment represented by the tour.*

Proof. The length of the tour on an edge gadgets is $d(a+b)$. There are three edge gadgets corresponding to every equation and every bridge in the edge gadget is shared between two equation gadgets. Thus, the length of the tour on the edge gadgets is $2m \cdot 3d(a+b)/2 = 3md(a+b)$ The length of the tour on an equation gadget is 4 if the equation is satisfied and 5 otherwise. Thus, the total length is $3md(a+b) + 4m + u$. ∎

The main challenge now is to prove that the above correspondence between the length of the optimum tour and the number of unsatisfied equation holds also when we drop the assumption that the tour is shaped in the intended way. Our proof uses the following technical lemma (we provide a proof in the appendix):

Lemma A.1. *For every large enough constant* k*, there exists an 7-regular bipartite multigraph on* $2k$ *vertices such that for every partition of the left vertices into sets* T_1*,* U_1 *and* S_1 *and every partition of the right vertices into sets* T_2*,* U_2 *and* S_2 *such that there are no edges from* T_1 *to* T_2*, and there are no edges from* U_1 *to* U_2*,*

$$2(|S_1|+|S_2|) \geq \min\{k, |U_1|+|T_2|+|S_1|+|S_2|, |U_2|+|T_1|+|S_1|+|S_2|\}$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ *or* $S_1 = S_2 = T_1 = U_2 = \emptyset$.

Given the above lemma, the following argument gives a lower bound on the extra cost, not counting the "normal" cost of $d(a+b)$ per edge gadget and 4 per equation gadget, that results from a non-standard behavior of the tour. We have already seen that an unsatisfied equation adds an extra cost of 1. Edge gadgets that are either traversed or untraversed do not add any extra cost. Note that traversed edge gadgets never can share the same bridge, neither can untraversed edge gadgets. The following lemma gives a lower bound on the additional length of the tour due to semitraversed edge gadgets (we provide a proof in § 4.3):

Lemma 4.4. *Consider an instance of (1,B)-ATSP constructed as described above from an instance of E3-Lin mod 2 with equations of the form $x + y + z = 0$ where x, y and z are variables or negated variables and each variable occurs equally many times positively and negatively. For any tour in such an instance, it is possible to associate a cost of at least* $\min\{a/2, b, a/2 + b/2 - 1, a/2 + b/4 - 1/2\}$ *with every semitraversed edge gadget given that* $B \geq \min\{3b, a + b, 2a + b - 2\}$.

By combining the above two lemmas we can prove that it is never optimal to have inconsistent traversals. The parameters a, b and d are chosen to give a lower bound that is as good as possible. By Lemma A.1 we must have $d = 7$ and by choosing $a = 2b$ in Lemma 4.4 we get a cost of at least $\min\{b, 3b/2 - 1, 5b/4 - 1/2\}$. This cost must be at least 2 for Lemma A.1 to apply; therefore we select $b = 2$ which in turn implies that $a = 4$ and $B = 8$.

Lemma 4.5. *For $a = 4$, $b = 2$, $d = 7$, and $B = 8$, there exists a coupling of the equation gadgets with the property that it can never be advantageous to have inconsistently traversed equation gadgets.*

Proof. Repeat the following argument for every variable x:

Let k be the number of occurrences of x (and also the number of occurrences of $\bar{x}$). Pick a bipartite multigraph on $2k$ vertices such that for every partition of the left vertices into sets T_1, U_1 and S_1 and every partition of the right vertices into sets T_2, U_2 and S_2 such that there are no edges from T_1 to T_2, and there are no edges from U_1 to U_2,

$$2(|S_1| + |S_2|) \geq \min\{k, |U_1| + |T_2| + |S_1| + |S_2|, |U_2| + |T_1| + |S_1| + |S_2|\}$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ or $S_1 = S_2 = T_1 = U_2 = \emptyset$. We know by Lemma A.1 that such a graph exists—since the graph has constant size, we can try all possible graphs in constant time.

Put occurrences of x at one side and occurrences of $\bar{x}$ on the other side of the bipartite graph. Each vertex in the graph can be labeled as T, U or S, depending on whether it is traversed, untraversed or semitraversed. Let T_1 be the set of traversed positive occurrences and T_2 be the set of traversed negative occurrences. Define U_1, U_2, S_1, and S_2 similarly. We can assume that $|U_1| + |T_2| \leq |U_2| + |T_1|$—otherwise we just change the indexing convention.

We now consider a modified tour where the positive occurrences are traversed and the negative occurrences are untraversed. This decreases the cost of tour by at least $2(|S_1|+|S_2|)$ and increases it by $\min\{k, |S_1|+|S_2|+|U_1|+|T_2|\}$. But the bipartite graph has the property that

$$2(|S_1| + |S_2|) \geq \min\{k, |U_1| + |T_2| + |S_1| + |S_2|\}$$

which implies that the cost of tour decreases by this transformation. Thus, we can assume that x is given a consistent assignment by the tour. ∎

Theorem 4.2. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,8)-ATSP within $131/130 - \epsilon$.*

Proof. Given an instance of E3-Lin mod 2 with $2m$ equations where every variable occurs a constant number of times, we construct the corresponding instance of (1,8)-ATSP with $a = 4$, $b = 2$ and $d = 7$. This can be done in polynomial time. By the above lemma, we can assume that all edge gadgets are traversed consistently in this instance. The assignment obtained from this traversal satisfies $2m - u$ equations if the length of the tour is $3md(a+b) + 4m + u$. If we could decide if the length of the optimum tour is at most $(3d(a+b)+4+\epsilon_1)m$ or at least $(3d(a+b)+5-\epsilon_2)m$, we could decide if at most $\epsilon_1 m$ or at least $(1-\epsilon_2)m$ of the equations are left unsatisfied by the corresponding assignment. But to decide this is **NP**-hard by Theorem 4.1. Therefore it is **NP**-hard to approximate (1,8)-ATSP within

$$\frac{3d(a+b)+5-\epsilon_2}{3d(a+b)+4+\epsilon_1} \geq \frac{131}{130} - \epsilon.$$ ∎

4.3 Proof of Lemma 4.4

The theorem follows from a direct combination of the following lemmas. We first note that we can assume that the tour has a certain structure inside the bridges.

Lemma 4.6. *Let (u, v) be an edge of the tour and suppose that u and v both belong to the same bridge. Then u and v are neighbors in the graph defining the (1,B)-ATSP instance if $B \geq a$.*

Let u and v be neighbors on the same bridge and assume that there is no edge between u and v in the tour. Let (u, u') and (v, v') be edges of the tour and assume that $d(u, u') < B$ and that $d(v, v') < B$. Then we can assume that the shortest path from u to u' does not intersect the shortest path from v to v'.

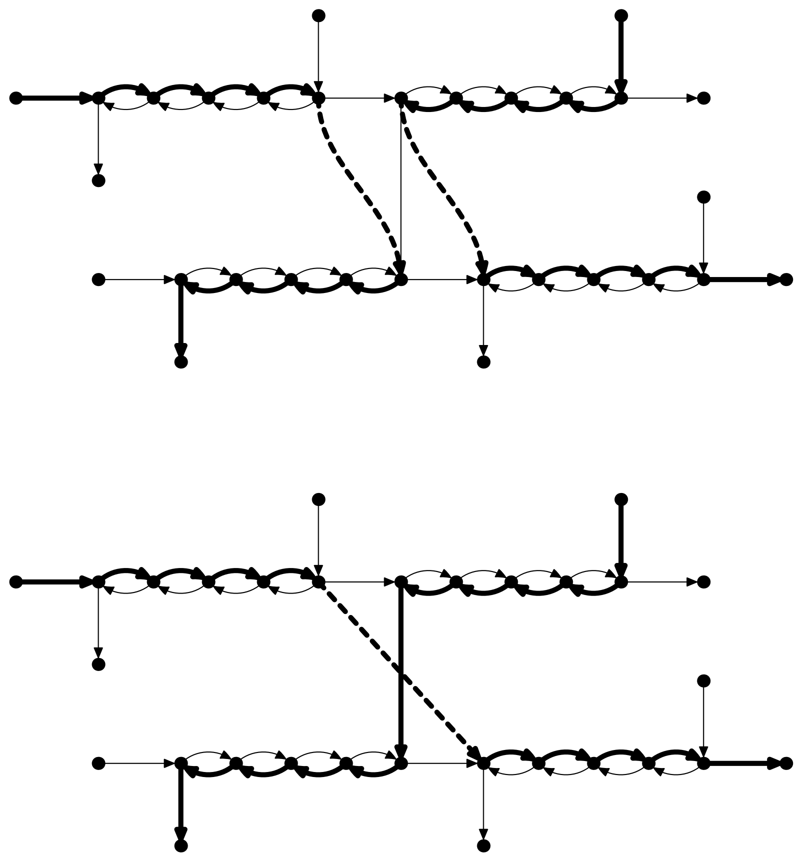


Figure 27. Switching from traversing an edge gadget representing an occurrence of x to traversing another edge gadget representing an occurrence of x gives an extra cost of at least b. The dotted edge above has length $3b$; that gives an extra cost of $2b$ which is then shared evenly among the two semitraversed edge gadgets.

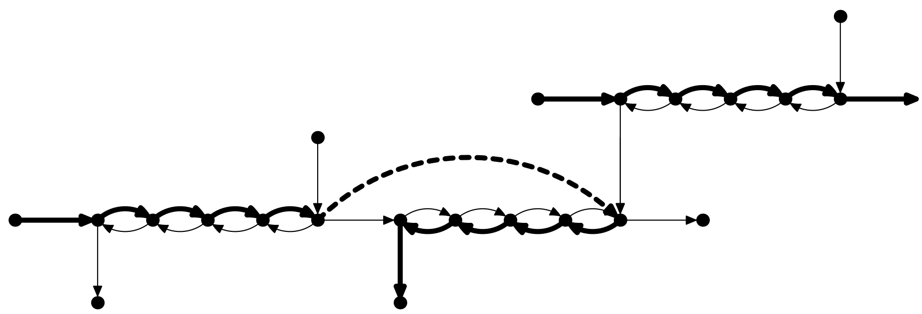


Figure 28. Switching from traversing an edge gadget representing an occurrence of x to traversing an edge gadget representing an occurrence of $\bar{x}$ gives an extra cost of at least $a/2$. The dashed edges above has length $a + b$; that gives an extra cost of a which is then shared evenly among the two semitraversed edge gadgets.

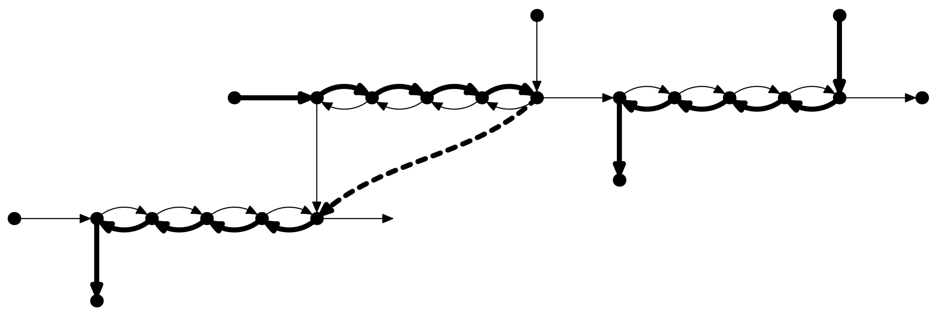


Figure 29. Switching from traversing an edge gadget representing an occurrence of x to traversing an edge gadget representing an occurrence of $\bar{x}$ gives an extra cost of at least $a/2$. The dashed edges above has length $a + b$; that gives an extra cost of a which is then shared evenly among the two semitraversed edge gadgets.

Proof. Suppose that u and v are not neighbors. Then we can produce another tour with equal or shorter cost as follows: Let the tour follow the shortest path from u to v in the graph defining the instance instead of jumping directly from u to v. Since $B \geq a$ this does not increase the length of the tour. This change will make the tour pass through some cities—the cities that are on the shortest path from u to v in the graph—twice. For all such cities w, do the following: Let w' be the city visited immediately before w and w'' be the city visited immediately after w. Then replace the edges (w', w) and (w, w'') by the single edge (w', w'') in the tour. By triangle inequality this procedure does not increase the length of the tour.

Let the tour follow the shortest path from u to u' in the graph defining the instance instead of jumping directly from u to u'. This change will make the tour pass through some cities—the cities that are on the shortest path from u to u' in the graph—twice. For all such cities w, do the following: Let w' be the city visited immediately before w and w'' be the city visited immediately after w. Then replace the edges (w', w) and (w, w'') by the single edge (w', w'') in the tour. By triangle inequality this procedure does not increase the length of the tour. ∎

Consider a semitraversed edge gadget. We now argue by case analysis that it introduces an extra cost in addition to the "standard" cost of $a + b$ per bridge. When analyzing the extra cost due to semitraversals, we use the following convention: Suppose that the tour leaves some bridge following an edge of length ℓ. If $\ell < B$, there is a corresponding path of length ℓ in the graph defining the (1,B)-ATSP instance. Then penalties corresponding to the part of the path that is inside bridge, including $b/2$ of the penalty corresponding to traversed edges connecting the bridge to other bridges, is attributed to the bridge that the tour is leaving. if $\ell \geq B$, a cost of $B/2$ is attributed to both of the involved bridges. We use a similar convention for tours entering a bridge following a long edge; only the length "inside" the bridge is attributed to the cost of that bridge.

Lemma 4.7. *Given that $B \geq \min\{3b, a+b\}$ it is possible to associate a cost of at least $\min\{a/2, b\}$ with every semitraversed edge gadget where no bridge has an undefined traversal.*

Proof. We first consider the case when the metric is not bounded; we will show later how to extend the argument to cover also bounded metrics. In the unbounded case, the distance between two vertices u and v is exactly the length of the shortest path from u to v in the graph defining the instance.

If no bridge has an undefined traversal, there must be two adjacent bridges that are traversed in different directions. Suppose that the left of these bridges is traversed from left to right and that the right one is traversed from right to left, i.e., that two parts of the tour "collide". Consider the tour leaving the left bridge. Since its natural way to escape the bridge is blocked by the tour on the right bridge, it has to make a jump. There are three subcases.

1. **The tour goes down (Fig. 27).** The earliest available free city is a distance of $2b$ away; that blocks the tour leaving the right bridge, forcing it to also make a jump of at least $2b$. The next available free city is a distance of $3b$ away. Both these cases give a total extra cost of $2b$, which is split evenly between the two involved semitraversed edges. Note that the above argument is valid also when the right bridge is close to the vertex connecting the edge gadget with the equation gadget—this corresponds to replacing the bottom right bridge with the final vertex of the edge gadget in Fig. 27.

2. **The tour goes forwards (Fig. 28).** The earliest available free city is a distance of $a + b$ away, giving a total extra cost of a. This cost is split evenly between the two involved semitraversed edges.

3. **The tour goes backwards (Fig. 29).** The earliest available free city is a distance of $a + b$ away, giving a total extra cost of a. This cost is split evenly between the two involved semitraversed edges.

Now suppose that the left of these bridges is traversed from right to left and that the right one is traversed from left to right, i.e., that the tours "depart". Consider the tour entering the left bridge. As in the "colliding case" above, we get three subcases—each of them corresponds to a subcase of the "colliding" case.

1. **The tour comes from above (Fig. 27).** The earliest available free city is a distance of $2b$ away, but that blocks the tour entering the right bridge, forcing it to also make a jump of at least $2b$. The next available free city is a distance of $3b$ away. Both these cases give a total extra cost of $2b$, which is split evenly between the two involved semitraversed edges. Note that the above argument is valid also when the right bridge is close to the vertex connecting the edge gadget with the equation gadget—this corresponds to replacing the top right bridge with the final vertex of the edge gadget in Fig. 27.

2. **The tour comes from the front (Fig. 28).** The earliest available free city is a distance of $a+b$ away, giving a total extra cost of a. This cost is split evenly between the two involved semitraversed edges.

3. **The tour comes from behind (Fig. 29).** The earliest available free city is a distance of $a+b$ away, giving a total extra cost of a. This cost is split evenly between the two involved semitraversed edges.

If the tour makes a larger jump than the shortest possible jumps stated above, the additional cost can never decrease, thanks to the triangle inequality. In all cases, the extra cost is split evenly between the two semitraversed edges that are involved—the one where the jump starts and the one where the jump ends. Therefore, each semitraversed edge that does not have an undefined traversal has an extra cost of at least $\min\{a/2, b\}$.

Finally, since the distance of a shortest jump has to be at least $\max\{3b, a+b\}$ for the penalty to be high enough, the argument generalizes to metrics with integer distances between 1 and B as soon as $B \geq \max\{3b, a+b\}$. ∎

Lemma 4.8. *Given that* $B \geq 2a + b - 2$ *it is possible to associate a cost of at least* $\min\{a+b-2, a+b/2-1\}$ *with every bridge containing an undefined traversal.*

Proof. Since the bridge has an undefined traversal, there must be two adjacent cities u and v that are not neighbors in the tour. Consider the edges (u, u') and (v, v') in the tour—thanks to Lemma 4.6 we can assume that neither u' nor v' belong to the bridge.

1. $\boldsymbol{d(u, u') = B}$**.** By our convention, we can attribute half of that cost to the bridge. The tour must visit the other cities on the bridge with one incoming and one outgoing edge of length at least $\min\{b/2, 1\}$ each—note that the edge leaving one city could potentially be the edge entering another city an the bridge. If v is the first or last city of the bridge, this gives a total cost of at least $B/2+3b/2+a-1 \geq 2a+2b-2$, i.e., an extra cost of $a+b-2$ in addition to the normal cost of $a+b$ for the bridge. If v is an interior city, but u is the first or last city, it gives a total cost of at least $B/2+b+a \geq 2a+3b/2-1$, i.e., an extra cost of $a+b/2-1$. If both u and v are interior cities, it gives a total cost of at least $B/2+b+a$, i.e., an extra cost of $a+b/2-1$.

2. $\boldsymbol{d(v, v') = B}$**.** This case is identical to the case when (u, u') has length B.

3. $\boldsymbol{d(u,u') < B \wedge d(v,v') < B}$. In this case, the edges (u, u') and (v, v') correspond to paths on the edges of the graph defining the (1,B)-ATSP instance. By Lemma 4.6 we can assume that the path from u to u' and the path from v to v' do not intersect. This gives a total cost of at least $a + b - 1 + c(u, u') + c(v, v')$, i.e., an extra cost of $c(u, u') + c(v, v') - 1$, where $c(\cdot, \cdot)$ denotes the part of the cost associated with the bridge according to our convention. Since the path from u to u' and the path from v to v' do not intersect, $c(u, u') + c(v, v') \geq a + b - 1$, therefore the extra cost is at least $a + b - 2$.

To sum up, we have established that an extra cost of $\min\{a+b-2, a+b/2-1\}$ can be associated with each bridge that has an undefined traversal. ∎

Corollary 4.1. *Given that* $B \geq 2a + b - 2$ *it is possible to associate a cost of at least* $\min\{a/2 + b/2 - 1, a/2 + b/4 - 1/2\}$ *with every semitraversed edge gadget containing an undefined traversal.*

Proof. Every bridge is shared between two edge gadgets. ∎

To complete the proof of the theorem, we note that the fact that a jump may start from a semitraversed gadget with no undefined traversal and end in an undefined traversal, and vice versa, does not void the above analysis.

Proof of Lemma 4.4. Suppose that $B \geq \max\{3b, a + b, 2a + b - 2\}$. Consider a semitraversed edge gadget. If it has no undefined traversals, it is possible to associate a cost of at least $\min\{a/2, b\}$ to the edge gadget according to Lemma 4.7. If it has an undefined traversal, it is possible to associate a cost of at least $(a + b)/2 - 1$ with it according to Corollary 4.1.

The analysis in Lemmas 4.7 and 4.8 is valid also if the tour jumps from a semitraversed edge gadget with no undefined traversals to a semitraversed edge gadget with an undefined traversal by the convention we use when attributing the cost of long edges in the tour to the involved bridges. ∎

5 The hardness of (1,B)-TSP

To adapt the construction from the previous section for the symmetric case we need to change the gadgets. The equation gadget is replaced with the gadget in Fig. 31—note that this gadget tests odd parity instead of even parity—and the edge gadget is changed according to Fig. 32. If we assume that the tour behaves nicely, it is straightforward to prove a correspondence between the length of a tour and the number of equations left unsatisfied by the corresponding assignment.

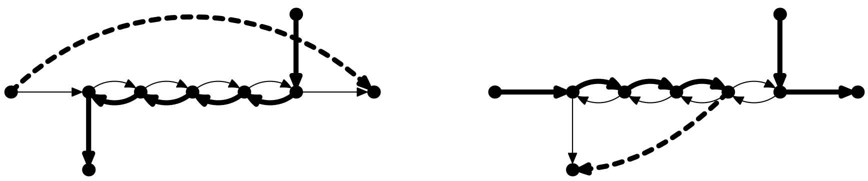


Figure 30. We can assume that traversals shown in the left figure above never occur since they can be transformed into the traversal shown in the right figure without increasing the length of the tour. A bridge with a traversal of that form gives an extra cost of at least $\min\{a+b-2, a+b/2-1\}$ if $B \geq 2a+b-2$.

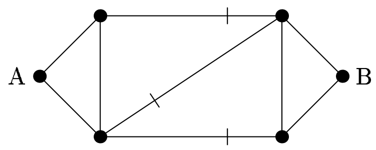


Figure 31. The symmetric gadget for equations of the form $x+y+z=1$. There is a Hamiltonian path from A to B only if zero or two of the ticked edges are traversed.

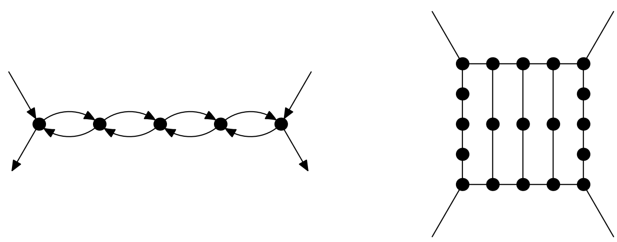


Figure 32. To transform the edge gadget from Fig. 24 into a gadget that can be used in the symmetric case, all occurrences of the structure to the left above are replaced with the structure to the right above. All edges in the right figure have weight 1.

Lemma 5.1. *The only way to traverse the equation gadget in Fig. 31 with a tour of length* 5*—if the edge gadgets count as length one for the moment—is to traverse an odd number of edge gadgets. All other locally optimal traversals have length* 6*.*

Proof. It is easy to see that any tour traversing either zero or two of the ticked edges and leaving the third one untraversed has length 5. Any tour traversing an odd number of ticked edges gets stuck in the center node and needs an extra cost of at least one to get out to a corner of the enclosing triangle. ∎

Lemma 5.2. *In addition to the length* 1 *attributed to the edge gadget above, the length of a tour traversing an edge gadget in the intended way is* 119*.*

Proof. The total cost is $7 \cdot 19 = 133$. ∎

Lemma 5.3. *Suppose that there are* $2m$ *equations in the E3-Lin instance. If the tour is shaped in the intended way, i.e., every edge gadget is either traversed or untraversed, the length of the tour is* $404m + u$*, where* u *is the number of unsatisfied equations resulting from the assignment represented by the tour.*

Proof. There are three edge gadgets corresponding to every equation and every bridge in the edge gadget is shared between two equation gadgets. Thus, the length of the tour on the edge gadgets is $2m \cdot 3 \cdot 133/2 = 399m$. The length of the tour in the equation gadgets is 5 if the equation is satisfied and 6 otherwise. Thus, the total length is $404m + u$. ∎

In the same way as in the asymmetric case, it can now be shown that the tour can be assumed to behave in the intended way. This gives the following lemmas (we omit the proof):

Lemma 5.4. *Suppose that* $B \geq 8$*. Then every semitraversed edge gadget adds an extra cost of at least* 2 *to the length of the tour.*

Lemma 5.5. *There exists a coupling of the edge gadgets with the property that there can never be advantageous to have inconsistently traversed edge gadgets.*

Given the above lemmas, the main theorem follows in exactly the same way as in the asymmetric case.

Theorem 5.1. *For any constant* $\epsilon > 0$*, it is* ***NP****-hard to approximate (1,8)-TSP within* $405/404 - \epsilon$*.*

Proof. Given an instance of E3-Lin mod 2 with $2m$ equations where every variable occurs a constant number of times, we construct the corresponding instance of (1,8)-TSP. This can be done in polynomial time. By the above lemma, we can assume that all edge gadgets are traversed consistently in this instance. The assignment obtained from this traversal satisfies $2m - u$ equations if the length of the tour is $404m + u$. If we could decide if the length of the optimum tour is at most $(404 + \epsilon_1)m$ or at least $(405 - \epsilon_2)m$, we could decide if at most $\epsilon_1 m$ or at most $(1 - \epsilon_2)m$ of the equations are let unsatisfied by the corresponding assignment. But to decide this is **NP**-hard by Theorem 4.1. ∎

6 Conclusions

It should be possible to improve the reduction by eliminating the vertices that connect the equation gadgets for $x + y + z = \{0, 1\}$ with each other. This reduces the cost of those equation gadgets by one, which improves our bounds—but only by a miniscule amount. The big bottleneck, especially in the (1,2) case, is the consistency gadgets. If, for the asymmetric case, we were able to decrease the cost of them to four instead of six, we would improve the bound to $237/236 - \epsilon$; if we could decrease the cost to three, the bound would become $195/194 - \epsilon$. We conjecture that some improvement for the $(1, 2)$ case is still possible along these lines.

Acknowledgments

We thank Santosh Vempala for many clarifying discussions on the subject of this paper.

References

1. Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Márió Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.
2. Piotr Berman and Marek Karpinski. On some tighter inapproximability results. In *Proceedings of 26th International Colloquium on Automata, Languages and Programming*, volume 1644 of *Lecture Notes in Computer Science*, pages 200–209, 1999.
3. Markus Bläser and Bodo Siebert. Computing cycle covers without short cycles. In *Proceedings of the 9th Annual European Symposium on Algorithms*, 2001. To appear.
4. Hans-Joachim Böckenhauer, Juraj Hromkovič, Ralf Klasing, Sebastian Seibert, and Walter Unger. An improved lower bound on the approximability of metric TSP and

approximation algorithms for the TSP with sharpened triangle inequality. In *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1770 of *Lectures Notes in Computer Science*, pages 382–391, 2000.

5. Hans-Joachim Böckenhauer and Sebastian Seibert. Improved lower bounds on the approximability of the traveling salesman problem. *RAIRO Theoretical Informatics and Applications*, 34(3):213–255, 2000.
6. Robert D. Carr and Santosh Vempala. Towards a 4/3 approximation for the asymmetric traveling salesman problem. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 116–125, 2000.
7. Nicos Christofides. Worst-case analysis of a new heuristic for the traveling salesman problem. Technical Report CS-93-13, G.S.I.A., Carnegie Mellon University, Pittsburgh, 1976.
8. Lars Engebretsen. An explicit lower bound for TSP with distances one and two. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1563 of *Lecture Notes in Computer Science*, pages 373–382, 1999.
9. Lars Engebretsen and Marek Karpinski. Approximation hardness of TSP with bounded metrics. In *Proceedings of 28th International Colloquium on Automata, Languages and Programming*, volume 2076 of *Lecture Notes in Computer Science*, pages 201–212, 2001.
10. Alan Frieze, Giulia Galbiati, and Francesco Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. *Networks*, 12(1):23–39, 1982.
11. Ramon E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.
12. Johan Håstad. Some optimal inapproximability results. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 1–10, 1997. Accepted for publication in *Journal of the ACM*.
13. Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.
14. Christos H. Papadimitriou and Santosh Vempala. On the approximability of the traveling salesman problem. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 126–133, 2000.
15. Christos H. Papadimitriou and Mihalis Yannakakis. The traveling salesman problem with distances one and two. *Mathematics of Operations Research*, 18(1):1–11, 1993.
16. Luca Trevisan. When Hamming meets Euclid: The approximability of geometric TSP and MST. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 21–29, 1997.

A The bipartite graph

This section is devoted to the proof of the following technical lemma:

Lemma A.1. *For every large enough constant k, there exists an 7-regular bipartite multigraph on $2k$ vertices such that for every partition of the left vertices into sets T_1, U_1 and S_1 and every partition of the right vertices into sets T_2, U_2 and S_2 such that there are no edges from T_1 to T_2, and there are no edges from U_1 to U_2,*

$$2(|S_1|+|S_2|) \geq \min\{k, |U_1|+|T_2|+|S_1|+|S_2|, |U_2|+|T_1|+|S_1|+|S_2|\} \quad (1)$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ or $S_1 = S_2 = T_1 = U_2 = \emptyset$.

The proof uses the same main idea as the proof of a similar expansion theorem for 8-regular graphs communicated to us by Papadimitriou and Vempala in April 2001: It uses a lemma that bounds the size of neighbor sets in 7-regular bipartite graphs.

Lemma A.2. *For every large enough constant k, there exists a 7-regular bipartite multigraph on $2k$ vertices such that every subset U of vertices contained entirely in V_1 or V_2 has a set $N(U)$ of neighbors satisfying the following constraints:*

$$\begin{aligned}
0 < |U| \leq k/10 &\implies |N(U)| > 29|U|/10, && (2)\\
k/10 \leq |U| \leq 3k/10 &\implies |N(U)| > 13k/100 + 8|U|/5, && (3)\\
3k/10 \leq |U| \leq 39k/100 &\implies |N(U)| > 31k/100 + |U|, && (4)\\
39k/100 \leq |U| \leq 62k/100 &\implies |N(U)| > k/2 + |U|/2, && (5)\\
62k/100 \leq |U| < k &\implies |N(U)| > \max\{85k/100, k/2 + |U|/2\}. && (6)
\end{aligned}$$

Proof. We select a d-regular bipartite graph on $2k$ vertices by selecting d perfect matchings independently and uniformly at random. Let $A_{U,N}$ be the event that the set U has neighbors only inside the set N and let Ω be the subset of $\{0, 1, 2, \ldots, k\} \times \{0, 1, 2, \ldots, k\}$ such that if $(a, b) \in \Omega$,

$$\begin{aligned}
0 < a \leq k/10 &\implies b \leq 29a/10,\\
k/10 \leq a \leq 3k/10 &\implies b \leq 13k/100 + 8a/5,\\
3k/10 \leq a \leq 39k/100 &\implies b \leq 31k/100 + a,\\
39k/100 \leq a \leq 62k/100 &\implies b \leq k/2 + a/2,
\end{aligned}$$

$$62k/100 \leq a < k \implies b \leq \max\{85/100, k/2 + a/2\}.$$

Denote the vertex set of the bipartite graph by $V_1 \cup V_2$. We need to prove that

$$\Pr\Big[\bigcup_{i=1}^{2} \bigcup_{(a,b)\in\Omega} \bigcup_{\substack{U\subset V_i\\|U|=a}} \bigcup_{\substack{N\subset V_{2-i}\\|N|=b}} A_{U,N}\Big] < 1$$

and we do this by using the union bound, i.e., we prove that

$$\sum_{i=1}^{2} \sum_{(a,b)\in\Omega} \sum_{\substack{U\in V_i\\|U|=a}} \sum_{\substack{N\in V_{2-i}\\|N|=b}} \Pr[A_{U,N}] < 1.$$

First note that $\Pr[A_{U,N}] = 0$ when $|U| > |N|$, therefore it suffices to consider only (a,b) such that $a \leq b$. Since Ω contains less than k^2 pairs and

$$\Pr[A_{U,N}] = \binom{d|N|}{d|U|} \frac{(d|U|)!(dk - d|U|)!}{(dk)!} = \frac{\binom{d|N|}{d|U|}}{\binom{dk}{d|U|}}$$

when $a \leq b$ it is enough to prove that

$$2k^2 \max_{\substack{(a,b)\in\Omega\\a\leq b}} \binom{k}{a}\binom{k}{b}\frac{\binom{db}{da}}{\binom{dk}{da}} = \max_{\substack{(a,b)\in\Omega\\a\leq b}} P(a,b) < 1.$$

We prove this inequality by case analysis. When a can be written as αk and b can be written as βk where $10^{-5} \leq \alpha \leq \beta \leq 1 - 10^{-5}$ we expand the above expression using Stirling's formula. The cases when either a or b are very close to either 0 or k are dealt with separately.

Case I: $10^{-5}k \leq a \leq b \leq (1 - 10^{-5})k$. By Stirling's formula

$$\binom{k}{\alpha k} = (\alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)})^k \operatorname{poly}(k).$$

Now write $a = \alpha k$ and $b = \beta k$ and apply Stirling's formula to the expression we want to bound. This gives us the equality

$$P(\alpha k, \beta k) = \left(\frac{(1-\alpha)^{(d-1)(1-\alpha)}\beta^{(d-1)\beta}}{\alpha^{\alpha}(1-\beta)^{(1-\beta)}(\beta-\alpha)^{d(\beta-\alpha)}}\right)^k \operatorname{poly}(k)$$

This expression is certainly strictly less than 1 for all (α, β) such that $\alpha \leq \beta$ and $(\alpha k, \beta k) \in \Omega$ as soon as there exists a universal constant $c < 1$, strictly bounded away from 1, such that

$$Q(\alpha, \beta) = \frac{(1-\alpha)^{(d-1)(1-\alpha)}\beta^{(d-1)\beta}}{\alpha^{\alpha}(1-\beta)^{(1-\beta)}(\beta-\alpha)^{d(\beta-\alpha)}} \leq c$$

for all (α, β) such that $\alpha \leq \beta$ and $(\alpha k, \beta k) \in \Omega$. The validity of the latter inequality is established in Lemma B.1.

Case II: $0 < a \leq 10^{-5}k$. For every fixed a in that range and every b such that $a \leq b \leq 10a$, $P(a, b)$ is increasing with b. Therefore it suffices to prove that $P(a, b) < 1$ when $b = 3a$; that implies (2). Let us first note that

$$P(1,3) = 2k^2 \binom{k}{1}\binom{k}{3}\frac{\binom{21}{7}}{\binom{7k}{7}} < \frac{240k^6}{(k-1)^7},$$

therefore $P(1, 3) < 1$ when $k > 250$. We now show that $P(a, 3a)/P(a + 1, 3(a + 1)) > 1$ when $0 < a/k < 10^{-5}$ and $k > 10^5$, thereby establishing that (2) holds in that region. Since

$$P(a, 3a) = 2k^2 \binom{k}{a}\binom{k}{3a}\frac{\binom{21a}{7a}}{\binom{7k}{7a}}$$

we need to bound quotients of the following forms:

$$\binom{k}{a} \Big/ \binom{k}{a+1} = \frac{a+1}{k-1},$$

$$\binom{k}{3a} \Big/ \binom{k}{3a+3} = \frac{(3a+3)!(k-3a-3)!}{(3a)!(k-3a)!} > \left(\frac{3a+1}{k-3a}\right)^3,$$

$$\binom{7k}{7a+7} \Big/ \binom{7k}{7a} = \frac{(7a)!(k-7a)!}{(7a+7)!(k-7a-7)!} > \left(\frac{k-a-1}{a+1}\right)^7,$$

$$\binom{21a}{7a} \Big/ \binom{21a+21}{7a+7} = \frac{(21a)!(7a+7)!(14a+14)!}{(21a+21)!(7a)!(14a)!} > \frac{1}{54^7}.$$

The above bounds imply that when $0 < a \leq \delta k$, where $\delta = 10^{-5}$ and $k > 10^5$,

$$\frac{P(a, 3a)}{P(a+1, 3(a+1))} > \frac{a+1}{k-1}\left(\frac{3a+1}{k-3a}\right)^3\left(\frac{a+1}{k-a-1}\right)^7\frac{1}{54^7}$$
$$> \frac{(k-\delta k-1)^7}{k^4(\delta k+1)^3 54^7} = \frac{(1-\delta-1/k)^7}{(\delta+1/k)^3 54^7}$$

$$
\begin{aligned}
&> \frac{1 - 14 \cdot 10^{-5}}{8 \cdot 10^{-15} \cdot 54^7} > \frac{9900}{8 \cdot 10^{-10} \cdot 54^7} = \frac{99 \cdot 10^{12}}{8 \cdot 9^7 \cdot 6^7} \\
&> \frac{11 \cdot 10^6}{8 \cdot 6^7} = \frac{11 \cdot 5^6}{48 \cdot 3^6} \\
&> \frac{5^5}{3^6} = \frac{3125}{729} \\
&> 1.
\end{aligned}
$$

Case III: $(1 - 10^{-5})k \leq b < k$. Note that $P(a, b) = P(k - b, k - a)$ since

$$
\begin{aligned}
\binom{d(k-a)}{d(k-b)} \Big/ \binom{dk}{d(k-b)} &= \frac{(dk - da)!}{(dk - db)!(db - da)!} \cdot \frac{(dk - db)!(db)!}{(dk)!} \\
&= \frac{(dk - da)!(db)!}{(dk)!(db - da)!} \\
&= \frac{(dk - da)!(da)!}{(dk)!} \cdot \frac{(db)!}{(db - da)!(da)!} \\
&= \binom{db}{da} \Big/ \binom{dk}{da}.
\end{aligned}
$$

Therefore, (6) in the region $1 - 10^{-5} \leq b/k \leq 1$ follows by a direct application of Case II. ∎

Proof of Lemma A.1. We use the shorthands $|T_1| = t_1$, $|U_1| = u_1$, $|S_1| = s_1$, $|T_2| = t_2$, $|U_2| = u_2$, and $|S_2| = s_2$. We can assume without loss of generality that $u_1 \leq u_2$. This implies that $u_1 < 7k/20$; otherwise $t_2 + s_2 \geq |N(U_1)| > k - u_1$ which is equivalent to $u_2 = k - t_2 - s_2 < u_1$, a contradiction. The proof now proceeds by case analysis on t_2 and u_1.

Case I a: $0 < t_2 \leq 3k/10$ and $u_1 \leq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 2t_2$ and $t_2 + s_2 \geq |N(U_1)| \geq 2u_1$ by (2) and (3)—we use non-strict inequality to also cover the case when $u_1 = 0$. Adding these two inequalities gives the inequality $s_1 + s_2 > t_2 + u_1$ which is equivalent to $2(s_1 + s_2) > u_1 + t_2 + s_1 + s_2$; therefore (1) holds.

Case I b: $0 < t_2 \leq k/10$ and $u_1 \geq 3k/10$. Then $t_2 + s_2 \geq |N(U_1)| > 31k/100 + u_1$ by (4) which is equivalent to $s_2 > 31k/100 + u_1 - t_2 > k/2$; therefore (1) holds.

Case I c: $k/10 \leq t_2 \leq 3k/10$ and $u_1 \geq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 13k/100 + 8t_2/5$ by (3) and $t_2 + s_2 \geq |N(U_1)| > 31k/100 + u_1$ by (4). Adding these inequalities gives the inequality $s_1 + s_2 > 44k/100 + 3t_2/5 \geq k/2$; therefore (1) holds.

Case II a: $3k/10 \leq t_2 \leq 39k/100$ and $u_1 \leq k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 31k/100 + t_2$ by (4) which is equivalent to $s_1 > 31k/100 + t_2 - u_1 > k/2$, therefore (1) holds.

Case II b: $3k/10 \leq t_2 \leq 39k/100$ and $u_1 \geq k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 31k/100 + t_2$ by (4) which is equivalent to $s_1 + s_2 > 31k/100 + t_2 + s_2 - u_1$, and $s_2 + t_2 \geq |N(U_1)| > 13k/100 + 8u_1/5$ by (3). Therefore $s_1 + s_2 > 31k/100 + t_2 + s_2 - u_1 > 44k/100 + 3u_1/5 > k/2$ and (1) holds.

Case III a: $39k/100 \leq t_2 < k$ and $u_1 \leq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > k/2 + t_2/2$ by (5) and (6), which is equivalent to $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1$, and $t_2 + s_2 \geq |N(U_1)| > 2u_1$ by (2) and (3); therefore $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1 > k/2 + s_2 + t_2/2 - (s_2 + t_2)/2 \geq k/2$ and (1) holds.

Case III b: $39k/100 \leq t_2 \leq 62k/100$ and $u_1 \geq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > k/2 + t_2/2$ by (5), which is equivalent to $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1$, and $t_2 + s_2 \geq |N(U_1)| > 31k/100 + u_1$ by (4); therefore $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1 > 81k/100 - t_2/2 \geq k/2$ and (1) holds.

Case III c: $62k/100 \leq t_2 < k$ and $u_1 \geq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 85k/100$ by (6), which is equivalent to $s_1 > 85k/100 - u_1 \geq k/2$ where the last inequality follows since $u_1 \leq 7k/20$; therefore (1) holds.

Case IV: $t_2 = 0$. Since $u_1 \leq u_2$ and $t_2 = 0$, $u_1 + t_2 \leq u_2 + t_1$, therefore it suffices to show that $2(s_1 + s_2) > \min\{u_1 + s_1 + s_2, k\}$. But this always holds if $u_1 > 0$ since then $s_2 \geq |N(U_1)| > u_1$. And if $u_1 = 0$, the inequality holds trivially as soon as either s_1 or s_2 are non-zero. Therefore, (1) holds when $t_2 = 0$.

Case V: $t_2 = k$. Since vertices in T_1 are not connected to vertices in T_2, $t_2 = k$ implies that $t_1 = 0$. Moreover, since $u_1 \leq u_2 = 0$, also $u_1 = 0$. Therefore, $s_1 = k$, which implies that (1) holds when $t_2 = k$. ∎

B The function $Q(\alpha, \beta)$

The proof of Lemma A.2 relies on the fact that a certain inequality holds in a certain region. In this section we prove this fact using a blend of analytic techniques and computer-assisted verification using interval arithmetic.

Lemma B.1. *Let*

$$Q(\alpha, \beta) = \frac{(1-\alpha)^{(d-1)(1-\alpha)}\beta^{(d-1)\beta}}{\alpha^{\alpha}(1-\beta)^{(1-\beta)}(\beta-\alpha)^{d(\beta-\alpha)}}.$$

Then $Q(\alpha,\beta) < 1-2.8\cdot 10^{-4}$ *for all* (α,β) *such that* $10^{-5} \le \alpha \le \beta \le 1-10^{-5}$ *and*

$$\begin{aligned}
&10^{-5} \le \alpha \le 1/10 \implies \alpha \le \beta \le 29\alpha/10,\\
&1/10 \le \alpha \le 3/10 \implies \alpha \le \beta \le 13/100 + 8\alpha/5,\\
&3/10 \le \alpha \le 39/100 \implies \alpha \le \beta \le 31/100 + \alpha,\\
&39/100 \le \alpha \le 62/100 \implies \alpha \le \beta \le 1/2 + \alpha/2,\\
&62/100 \le \alpha \le 1-10^{-5} \implies \alpha \le \beta \le \max\{85/100, 1/2 + \alpha/2\}.
\end{aligned}$$

if $(\beta-\alpha)^{d(\beta-\alpha)}$ *is defined to be* 1 *when* $\alpha = \beta$.

The fact that we define $(\beta-\alpha)^{d(\beta-\alpha)}$ to be 1 when $\alpha = \beta$ can be motivated by a continuity argument and it also gives the right interpretation to the probabilistic experiment that Q is describing. If we in addition define $Q(0,0) = Q(0,1) = Q(1,1) = 1$, Q is continuous in the region $\{(\alpha,\beta) : 0 \le \alpha \le 1 \wedge \alpha \le \beta \le 1\}$

The overall structure of the proof of the above lemma is as follows: We first rotate the coordinate system to be able to exploit that $Q(\alpha,\beta)$ is symmetric along the line $\alpha+\beta = 1$. We then prove that maximizing Q over a closed set with a certain structure is equivalent to maximizing Q over the border of that set. Finally, we prove that the region we are interested in can be extended to a convex set and then establish that Q is strictly less than $1-2.8\cdot 10^{-4}$ on the border of that set.

We first make the substitutions $\beta+\alpha = x+1$ and $\beta-\alpha = y$, which is equivalent to $\alpha = (1+x-y)/2$ and $\beta = (1+x+y)/2$. That changes Q to

$$Q(x,y) = 2^{-(d-2)}(2y)^{-dy} \times \sqrt{\frac{(1+x+y)^{(d-1)(1+x+y)}(1-x+y)^{(d-1)(1-x+y)}}{(1+x-y)^{(1+x-y)}(1-x-y)^{(1-x-y)}}} \tag{7}$$

where y^{-dy} is defined to be 1 when $y = 0$ and $Q(-1,0) = Q(0,1) = Q(1,0) = 1$. Again, this function is continuous in the region $\{(x,y) : -1 \le x \le 1 \wedge 0 \le$

$y \le \min\{1-x, 1+x\}\}$. The following lemma relates the area of interest in the (x, y) plane to the area of interest in the (α, β) plane.

Lemma B.2. *Let A be the convex hull of the points*

$$\begin{aligned}
(x_1, y_1) &= (1 - 2 \cdot 10^{-5}, 0),\\
(x_2, y_2) &= (1 - 3.9 \cdot 10^{-5}, 1.9 \cdot 10^{-5}),\\
(x_3, y_3) &= (0.61, 0.19),\\
(x_4, y_4) &= (0.47, 0.23),\\
(x_5, y_5) &= (0.09, 0.31),\\
(x_6, y_6) &= (-0.09, 0.31),\\
(x_7, y_7) &= (-0.47, 0.23),\\
(x_8, y_8) &= (-0.61, 0.19),\\
(x_9, y_9) &= (-1 + 3.9 \cdot 10^{-5}, 1.9 \cdot 10^{-5}),\\
(x_{10}, y_{10}) &= (-1 + 2 \cdot 10^{-5}, 0).
\end{aligned}$$

If $Q(x, y) < 1 - 2.8 \cdot 10^{-4}$ in A, Lemma B.1 holds.

Proof. We transform the (x_i, y_i) pairs into (α_i, β_i) pairs using the transformations $\alpha = (1 + x - y)/2$ and $\beta = (1 + x + y)/2$:

$$\begin{aligned}
(\alpha_1, \beta_1) &= (1 - 10^{-5}, 1 - 10^{-5}),\\
(\alpha_2, \beta_2) &= (1 - 2.9 \cdot 10^{-5}, 1 - 10^{-5}),\\
(\alpha_3, \beta_3) &= (0.71, 0.90),\\
(\alpha_4, \beta_4) &= (0.62, 0.85),\\
(\alpha_5, \beta_5) &= (0.39, 0.70),\\
(\alpha_6, \beta_6) &= (0.30, 0.61),\\
(\alpha_7, \beta_7) &= (0.15, 0.38),\\
(\alpha_8, \beta_8) &= (0.10, 0.29),\\
(\alpha_9, \beta_9) &= (10^{-5}, 2.9 \cdot 10^{-5}),\\
(\alpha_{10}, \beta_{10}) &= (10^{-5}, 10^{-5}).
\end{aligned}$$

The convex hull of these points contains the region defined in Lemma B.1, therefore Lemma B.1 holds as soon as $Q(x, y) < 1 - 2.8 \cdot 10^{-4}$ in A. ∎

We now prove that the function $f(x) = \ln Q(x;y)$ is convex on a certain interval, depending on y. This in fact implies that also Q is convex for fixed y since

$$f'' = \frac{Q''_{xx}Q - (Q'_x)^2}{Q^2} \geq 0 \implies Q''_{xx} \geq \frac{(Q'_x)^2}{Q} > 0$$

since Q is non-negative, but we do not need this in our proof.

Lemma B.3. *Let $f(x;y) = \ln Q(x,y)$ be a function of x. Then f is convex in the interval*

$$|x| \leq \sqrt{\frac{d(1-y)-2}{d(1+y)-2}(1-y^2)}$$

for every fixed $y \in [0,(d-2)/d]$.

Proof. We can write f as $f(x;y) = C(y) + \frac{1}{2}g(x;y)$ where

$$\begin{aligned} C(y) \quad &= -dy(1+\ln y) - (d-2) && (8)\\ g(x;y) &= (d-1)(1+x+y)\ln(1+x+y) \\ &\quad + (d-1)(1-x+y)\ln(1-x+y) \\ &\quad - (1+x-y)\ln(1+x-y) - (1-x-y)\ln(1-x-y). && (9) \end{aligned}$$

To establish that f is convex in some region, it is enough to prove that g is convex in that region. To this end we take the first and second derivatives of g:

$$\begin{aligned} g'(x;y) \; &= (d-1)((1+x+y)\ln(1+x+y) - \ln(1-x+y)) \\ &\quad - (\ln(1+x-y) - \ln(1-x-y)), \\ g''(x;y) &= \frac{d-1}{1+x+y} + \frac{d-1}{1-x+y} - \frac{1}{1+x-y} - \frac{1}{1-x-y} \end{aligned}$$

We now rewrite the second derivative as

$$\begin{aligned} g''(x;y) &= \frac{d-1}{1+x+y} + \frac{d-1}{1-x+y} - \frac{1}{1-x-y} - \frac{1}{1+x-y} \\ &= \frac{(d-1)(2+2y)}{(1+x+y)(1-x+y)} - \frac{2-2y}{(1-x-y)(1+x-y)} \\ &= 2\frac{(d-2) - dy - (d-2)x^2 - (d-2)y^2 - dx^2y + dy^3}{(1+x+y)(1-x+y)(1-x-y)(1+x-y)} \end{aligned}$$

$$= 2\frac{(d-2)(1-x^2-y^2) - dy(1+x^2-y^2)}{(1+x+y)(1-x+y)(1-x-y)(1+x-y)}$$

and obtain that $g''(x;y)$ is non-negative when

$$(d-2)(1-x^2-y^2) \geq dy(1+x^2-y^2)$$

which is equivalent to

$$x^2 \leq \frac{d(1-y)-2}{d(1+y)-2}(1-y^2).$$

∎

The above lemma implies that along a horizontal line, the function Q is maximized at one of the endpoints. Let us state this more formally:

Lemma B.4. *Let x_0 and x_1 be functions satisfying*

$$-\sqrt{\frac{d(1-y)-2}{d(1+y)-2}(1-y^2)} \leq x_0(y) \leq x_1(y) \leq \sqrt{\frac{d(1-y)-2}{d(1+y)-2}(1-y^2)}$$

on the interval $[y_0, y_1] \subseteq [0, (d-2)/2]$ and define

$$A = \{(x,y) : y_0 \leq y \leq y_1 \wedge x_0(y) \leq x \leq x_1(y)\}.$$

Then $\max_{(x,y)\in A} Q(x,y) = \max_{(x,y)\in \partial A} Q(x,y)$ where ∂A denotes the border of A.

Proof. Suppose that Q attains it maximum at (x^*, y^*). If (x^*, y^*) is an interior point of A, then either $f(x_0(y^*); y^*) \geq f(x^*, y^*)$ or $f(x_1(y^*); y^*) \geq f(x^*, y^*)$ since f is convex for every fixed $y \in [0, (d-2)/2]$ in the interval

$$|x| \leq \sqrt{\frac{d(1-y)-2}{d(1+y)-2}(1-y^2)}$$

by Lemma B.3. But since $f(x;y) = \ln Q(x,y)$ and the logarithm is a one-to-one mapping, this implies that either $Q(x_0(y^*), y^*) \geq Q(x^*, y^*)$ or $Q(x_1(y^*), y^*) \geq f(x^*, y^*)$, therefore the maximum is also attained at a point in ∂A. ∎

Lemma B.5. *Let A be the convex hull of the points*

$$(x_1, y_1) = (1 - 2 \cdot 10^{-5}, 0),$$
$$(x_2, y_2) = (1 - 3.9 \cdot 10^{-5}, 1.9 \cdot 10^{-5}),$$
$$(x_3, y_3) = (0.61, 0.19),$$
$$(x_4, y_4) = (0.47, 0.23),$$
$$(x_5, y_5) = (0.09, 0.31),$$

$$
\begin{aligned}
(x_6, y_6) &= (-0.09, 0.31),\\
(x_7, y_7) &= (-0.47, 0.23),\\
(x_8, y_8) &= (-0.61, 0.19),\\
(x_9, y_9) &= (-1 + 3.9 \cdot 10^{-5}, 1.9 \cdot 10^{-5}),\\
(x_{10}, y_{10}) &= (-1 + 2 \cdot 10^{-5}, 0).
\end{aligned}
$$

Then $Q(x, y) < 1 - 2.8 \cdot 10^{-4}$ *in* A.

Proof. Let $y_0 = 0$, $y_1 = 0.31$, $x_0 = -x_1$, and let x_1 be the piecewise linear function defined by $x_1(0) = 1 - 2 \cdot 10^{-5}$, $x_1(1.9 \cdot 10^{-5}) = 1 - 3.9 \cdot 10^{-5}$, $x_1(0.19) = 0.61$, $x_1(0.23) = 0.47$, and $x(0.31) = 0.09$. These choices fulfill the requirements in Lemma B.4; therefore it is enough to show that $Q(x, y) < 1$ along the curves $x_0(y)$ and $x_1(y)$ when $y \in [y_0, y_1]$ in order to complete the proof. Furthermore, it is enough to consider $Q(x, y)$ along the curve $x_1(y)$ since $Q(x, y) = Q(-x, y)$.

Case I: $\mathbf{1.9 \cdot 10^{-5} \leq y \leq 0.31}$. Substituting $d = 7$ in Eq. (7) gives the expression

$$
Q(x, y) = \frac{(2y)^{-7y}(1 + x + y)^{3(1+x+y)}(1 - x + y)^{3(1-x+y)}}{32(1 + x - y)^{(1+x-y)/2}(1 - x - y)^{(1-x-y)/2}} \tag{10}
$$

We now need to verify that this expression is less than $1 - 2.8 \cdot 10^{-4}$ along the lines

$$
\begin{aligned}
&x = 1 - 39y/19 \qquad \text{when } y \in [1.9 \cdot 10^{-5}, 0.19],\\
&x = 0.61 - 3.5(y - 0.19) \qquad \text{when } y \in [0.19, 0.23],\\
&x = 0.47 - 4.75(y - 0.23) \qquad \text{when } y \in [0.23, 0.31].
\end{aligned}
$$

Substitution these expressions into (10) gives the expressions

$$
\begin{aligned}
Q_1(y) &= \frac{(2y)^{-7y}(2 - 20y/19)^{6-60y/19}(58y/19)^{174y/19}}{32(2 - 58y/19)^{1-29y/19}(20y/19)^{10y/19}}\\
Q_2(y) &= \frac{(2y)^{-7y}(2.275 - 2.5y)^{6.825-7.5y}(4.5y - 0.275)^{13.5y-0.825}}{32(2.275 - 4.5y)^{1.1375-2.25y}(2.5y - 0.275)^{1.25y-0.1375}}\\
Q_3(y) &= \frac{(2y)^{-7y}(2.5625 - 3.75y)^{7.6875-11.25y}(5.75y - 0.5625)^{17.25y-1.6875}}{32(2.5625 - 5.75y)^{1.28125-2.875y}(3.75y - 0.5625)^{1.875y-0.28125}}
\end{aligned}
$$

To verify that these functions are less than $1 - 2.8 \cdot 10^{-4}$ in the intervals $[1.9 \cdot 10^{-5}, 0.19]$, $[0.19, 0.23]$ and $[0.23, 0.31]$, respectively, we first prove that the functions assume their maximum value in one of the endpoints of the intervals where they are defined. To do this, we use the same approach as we did earlier in this section: We prove that $\ln Q_i$ is convex. Using the identity

$$\frac{d^2(\ln(ay+b)^{c(ay+b)})}{dy^2} = \frac{a^2c}{ay+b}$$

we obtain that

$$\begin{aligned}
\frac{d^2(\ln Q_1)}{dy^2} &= -\frac{7}{y} + \frac{3\cdot 20^2/19^2}{2-20y/19} + \frac{174}{19y} - \frac{58^2/(2\cdot 19^2)}{2-58y/19} - \frac{10}{19y} \\
&= \frac{31}{19y} + \frac{600}{19^2-190y} - \frac{29^2}{19^2-29\cdot 19y} \\
&> \frac{31}{19y} - \frac{29^2}{19^2-29\cdot 19y}
\end{aligned}$$

where the inequality follows since $600/(19^2-190y) > 0$ in $1.9\cdot 10^{-5}, 0.19]$. The last expression above is minimized when $y = 0.19$; therefore

$$\frac{d^2(\ln Q_1)}{dy^2} > \frac{3100}{19^2} - \frac{29^2}{19^2 - 29\cdot 19^2/100} > \frac{3100-2\cdot 29^2}{19^2} = \frac{1418}{19^2} > 0.$$

In a similar way we obtain the bounds

$$\begin{aligned}
\frac{d^2(\ln Q_2)}{dy^2} &= -\frac{7}{y} + \frac{750}{91-100y} + \frac{2430}{180y-11} - \frac{405}{91-180y} - \frac{125}{100y-11} \\
&> -\frac{7}{0.19} + \frac{750}{72} + \frac{2430}{34.2} - \frac{405}{49.60} - \frac{125}{8} \\
&> -40+10+69-9-16 = 14 > 0, \\
\frac{d^2(\ln Q_3)}{dy^2} &= -\frac{7}{y} + \frac{675}{41-60y} + \frac{1587}{92y-9} - \frac{529}{82-184y} - \frac{75}{40y-6} \\
&> \frac{7}{0.23} + \frac{675}{27.2} + \frac{1587}{19.52} - \frac{529}{24.96} - \frac{75}{3.20} \\
&> -\frac{7}{0.2} + \frac{675}{30} + \frac{1580}{20} - \frac{530}{20} - \frac{75}{3} \\
&= -35+22.5+79-26.5-25 = 15 > 0.
\end{aligned}$$

Therefore, it is enough to show that the function Q is less than $1-2.8\cdot 10^{-4}$ at the points $(1-3.9\cdot 10^{-5}, 1.9\cdot 10^{-5})$, $(0.61, 0.19)$, $(0.47, 0.23)$, and $(0.09, 0.31)$. To do this, we computed an upper bound on the values of the function Q at the above four points using interval arithmetic [11]. The source code for the program computing the bounds is presented in Appendix C.

Case II: $0 \leq y \leq 1.9 \cdot 10^{-5}$. In this interval $x = 1 - 2 \cdot 10^{-5} - y$. With this substitution, equation (9) becomes

$$\begin{aligned} g(x;y) = (d-1)(2 - 2\cdot 10^{-5}) \ln(2 - 2\cdot 10^{-5}) \\ + (d-1)(2\cdot 10^{-5} + 2y) \ln(2\cdot 10^{-5} + 2y) \\ - (2 - 2\cdot 10^{-5} - 2y) \ln(2 - 2\cdot 10^{-5} - 2y) \\ - 2\cdot 10^{-5} \ln(2\cdot 10^{-5}); \end{aligned}$$

therefore we can write $h(y) = \ln Q(x_1(y), y)$ as

$$\begin{aligned} h(y) = -dy(1 + \ln y) + (d-1)(10^{-5} + y) \ln(2\cdot 10^{-5} + 2y) \\ - (1 - 10^{-5} - y) \ln(2 - 2\cdot 10^{-5} - 2y) + C. \end{aligned}$$

Taking the derivative of h we obtain

$$\begin{aligned} h'(y) &= -d \ln y + (d-1) \ln(2\cdot 10^{-5} + 2y) + \ln(2 - 2\cdot 10^{-5} - 2y) \\ &> -d \ln y + (d-1) \ln(2\cdot 10^{-5} + 2y) \\ &> -d \ln(2\cdot 10^{-5}) + (d-1) \ln(2\cdot 10^{-5}) \\ &= 4 \ln 2 + 5 \ln 5 > 0, \end{aligned}$$

where the first inequality follows since $2 - 2\cdot 10^{-5} - 2y > 0$ in the interval and the second one follows since $-d \ln y$ is decreasing in the interval $[0, 2\cdot 10^{-5}]$ and $(d-1)\ln(2\cdot 10^{-5} + 2y)$ is increasing in the same interval. Therefore, the maximum value of $Q(x_1(y), y)$ in this region is $Q(1 - 3.9\cdot 10^{-5}, 1.9\cdot 10^{-5})$ which, by Case I, is less than $1 - 2.8\cdot 10^{-4}$. ∎

C Source code

This section contains the Fortran 90 source code for the program that verifies Case I in the proof of Lemma B.5 using interval arithmetic. It needs Sun Forte 6.0, update 1, to compile. The only option given to the complier was `-xia`, which enables interval arithmetic.

```
PROGRAM QPOINTS
IMPLICIT NONE
INTERVAL Q, X, Y

Q(X,Y) = (2.0*Y)**(-7*Y) / 32.0 &
       * (1.0+X+Y)**(3.0*(1.0+X+Y)) * (1.0-X+Y)**(3.0*(1.0-X+Y)) &
       * (1.0+X-Y)**((-1.0-X+Y)/2.0) * (1.0-X-Y)**((-1.0+X+Y)/2.0)

PRINT *, "Q(0.999961,0.000019) IS IN", Q([0.999961], [0.000019])
```

```
PRINT *, "Q(0.61,0.19) IS IN", Q([0.61], [0.19])
PRINT *, "Q(0.47,0.23) IS IN", Q([0.47], [0.23])
PRINT *, "Q(0.09,0.31) IS IN", Q([0.09], [0.31])

END PROGRAM QPOINTS
```

When the above source code was compiled with the command `f90 -xia` on a Sun Ultra Sparc running SunOS 5.6 and the resulting program was executed on the same computer it produced the following output:

```
Q(0.999961,0.000019) IS IN [0.99971203376169426,0.99971203376170371]
Q(0.61,0.19) IS IN [0.96016606004041482,0.96016606004042105]
Q(0.47,0.23) IS IN [0.91593524118552094,0.9159352411855265]
Q(0.09,0.31) IS IN [0.96341302921105986,0.96341302921106676]
```


ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'