FULL LENGTH PAPER

Hard to solve instances of the Euclidean Traveling Salesman Problem

Stefan Hougardy[1] · Xianghui Zhong[1]

Abstract

The well known 4/3 conjecture states that the integrality ratio of the subtour LP is at most 4/3 for metric Traveling Salesman instances. We present a family of Euclidean Traveling Salesman instances for which we prove that the integrality ratio of the subtour LP converges to 4/3. These instances (using the rounded Euclidean norm) turn out to be hard to solve exactly with `Concorde`, the fastest existing exact TSP solver. For a 200 vertex instance from our family of Euclidean Traveling Salesman instances `Concorde` needs several days of CPU time. This is more than 1,000,000 times more runtime than for a TSPLIB instance of similar size. Thus our new family of Euclidean Traveling Salesman instances may serve as new benchmark instances for TSP algorithms.

1 Introduction

The traveling salesman problem (TSP) is probably the most well-known problem in discrete optimization. An instance is given by n vertices and their pairwise distances. The task is to find a shortest tour visiting each vertex exactly once. This problem is known to be NP-hard [12].

If the distances satisfy the triangle inequality, we obtain an important special case called METRIC TSP. For this problem, no better algorithm than the $\frac{3}{2}$-approximation algorithm proposed by Christofides in 1976 [4] (and independently developed by Serdjukov [17]) is known. A well studied special case of the METRIC TSP is the

✉ Stefan Hougardy
   hougardy@or.uni-bonn.de

[1]   Research Institute for Discrete Mathematics, University of Bonn, Lennéstr. 2, 53113 Bonn, Germany

EUCLIDEAN TSP. Here an instance consists of points in the Euclidean plane and distances are defined by the l_2 norm. The EUCLIDEAN TSP is still NP-hard [8,14] but is in some sense easier than the METRIC TSP: For the EUCLIDEAN TSP there exists a PTAS [3] while for the METRIC TSP there cannot exist a $\frac{123}{122}$-approximation algorithm unless $P = NP$ [13].

The *subtour LP* is a relaxation of a well known integer linear program for the TSP [5]. If K_n is a complete graph with non-negative edge costs c_e for all $e \in E(K_n)$ the subtour LP is given by:

$$\min \sum_{e \in E(K_n)} c_e x_e$$

$$0 \le x_e \le 1 \qquad\qquad \text{for all } e \in E(K_n)$$

$$\sum_{e \in \delta(v)} x_e = 2 \qquad\qquad \text{for all } v \in V(K_n)$$

$$\sum_{e \in E(K_n[X])} x_e \le |X| - 1 \qquad\qquad \text{for all } X \subset V(K_n).$$

Although this LP has exponentially many inequalities, the separation problem and hence the LP itself can be solved in polynomial time [9].

The *integrality ratio* of a TSP instance is the ratio of the length of an optimum tour to the length of an optimum solution to the subtour LP. The integrality ratio of a TSP variant is the supremum over the integrality ratios of all instances of this TSP variant. The exact integrality ratio of the METRIC TSP is not yet known but it must lie between 4/3 [19] and 3/2 [20]. It is conjectured that the exact value is 4/3 [19, page 35] and this conjecture is known under the name 4/3-*Conjecture*. For the EUCLIDEAN TSP we also know only that the integrality ratio must lie between 4/3 [10] and 3/2 [20]. The lower bound of 4/3 was proven in [10] by showing that for a certain family of EUCLIDEAN TSP instances the integrality ratio converges to 4/3. In these instances all vertices lie on three parallel lines whose distances depend on the number of vertices.

New results In this paper we present a new family of instances of the EUCLIDEAN TSP which we call *tetrahedron instances* as they arise as certain subdivisions of a 2-dimensional projection of the edges of a tetrahedron. For these tetrahedron instances we prove that the integrality ratio of the subtour LP converges to 4/3. The rate of convergence is faster than for the instances constructed in [10]. Moreover, knowing structurally different families of instances for the METRIC TSP with integrality ratio converging to 4/3 may be useful for attacking the 4/3-Conjecture. Finding optimum solutions for the tetrahedron instances turns out to be much more difficult than for any known metric TSP instances of similar sizes: When using Concorde [2], the fastest known exact TSP solver, we observe that on instances with about 200 vertices, Concorde is more than 1,000,000 times slower than on TSPLIB instances [15] of similar size. Therefore, our tetrahedron instances may serve as new benchmark instances for the TSP and we provide them for download in TSPLIB format [11].

Outline of the paper In Sect. 2 we present the construction of the tetrahedron instances and introduce a modification of the tetrahedron instances which results in instances

with (up to symmetry) unique optimum tours. In Sect. 3 we prove some structural results of the optimum tours in these modified tetrahedron instances which allow us to bound the length of an optimum TSP tour in these instances. We also compute a bound for an optimum solution of the subtour LP for the modified tetrahedron instances. By combining these two results we can prove that for certain families of the modified tetrahedron instances the integrality ratio converges to 4/3. We then show how to carry over this result to the (unmodified) tetrahedron instances.

In Sect. 4 we present runtime experiments with `Concorde` [2] on the tetrahedron instances. We compare the runtimes with the runtimes on the instances proposed in [10] and on instances of comparable size from the TSPLIB [15].

2 The tetrahedron instances and their structural properties

In this section we first define the tetrahedron instances and the modified tetrahedron instances. Then, we prove some general properties of the optimal tour and geometrical properties of these instances. With this preparation we show structure theorems for the optimal tour that determine it uniquely up to certain symmetries.

2.1 Construction of the tetrahedron instances $T_{n,m}$

Now, we construct the instance $T_{n,m}$. Figure 1 shows as an example the instance $T_{9,5}$. Denote the Euclidean distance between two points x and y in the plane by $\mathrm{dist}(x, y)$. Let A, B, C be the vertices of an equilateral triangle with center M. The three sides of the triangle are called *base sides*, the closed line segments connecting A, B, or C to M are called *internal segments*. The vertex M belongs to all three internal segments. Denote the base side opposite to A, B respectively C by a, b respectively c and the internal segments connecting A, B respectively C with M by e, f respectively g.

Given such an equilateral triangle ABC we define the *tetrahedron instance $T_{n,m}$* for $n, m \in \mathbb{N}$ as follows. We refine each of the base sides a, b, and c by $n - 1$ equidistant vertices $a_1, \ldots, a_{n-1}$, $b_1, \ldots, b_{n-1}$, and $c_1, \ldots, c_{n-1}$. Moreover we define $a_0 := B$, $a_n := C$, $b_0 := C$, $b_n := A$, $c_0 := A$, and $c_n := B$. For $i \in \{0, \ldots, n\}$ the vertices a_i, b_i, and c_i are called *base vertices*.

Similarly, we refine each of the internal segments e, f, and g by $m - 1$ equidistant vertices $e_1, \ldots, e_{m-1}$, $f_1, \ldots, f_{m-1}$, and $g_1, \ldots, g_{m-1}$ numbered in ascending order from $A =: e_0$, $B =: f_0$, respectively $C =: g_0$ to the center $M =: e_m = f_m = g_m$. For $i \in \{1, \ldots, m\}$ the vertices e_i, f_i, and g_i are called *internal vertices*.

Finally, we rotate and scale all coordinates such that the side c is parallel to the x-axis, the vertex C is above the side c, and the distance between two consecutive base vertices is 1:

$$\mathrm{dist}(a_i, a_{i+1}) = \mathrm{dist}(b_i, b_{i+1}) = \mathrm{dist}(c_i, c_{i+1}) = 1 \quad \text{for } 0 \leq i < n. \quad (1)$$

This implies $\mathrm{dist}(A, M) = \mathrm{dist}(B, M) = \mathrm{dist}(C, M) = \dfrac{n}{\sqrt{3}}$ and therefore

$$\mathrm{dist}(e_i, e_{i+1}) = \mathrm{dist}(f_i, f_{i+1}) = \mathrm{dist}(g_i, g_{i+1}) = \dfrac{n}{\sqrt{3} \cdot m} \quad \text{for } 0 \leq i < m. \quad (2)$$
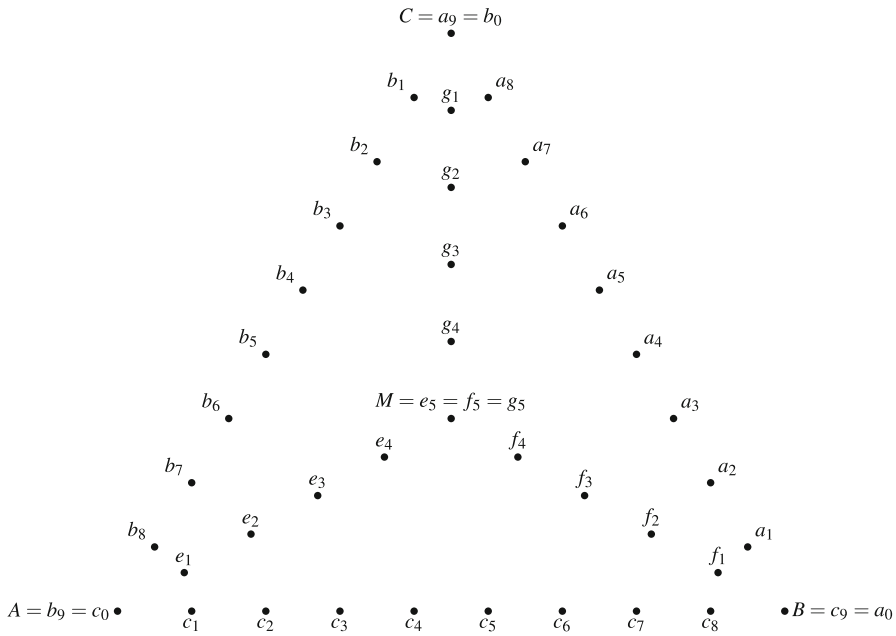
Fig. 1 The tetrahedron instance $T_{9,5}$

The smallest possible distance between any two different internal vertices will be denoted by γ, i.e., we have

$$\gamma = \frac{n}{\sqrt{3} \cdot m}. \tag{3}$$

In total, the instance $T_{n,m}$ has $3(n + m) - 2$ vertices and a possible way to assign explicit coordinates to all these vertices satisfying the above conditions is to assign for $i = 0, \ldots, n$ and for $j = 0, \ldots, m$:

$$\left.\begin{aligned}
a_i &:= \left(n - i/2, \qquad i \cdot \sqrt{3}/2 \right) \\
b_i &:= \left(n/2 - i/2, \qquad (n - i) \cdot \sqrt{3}/2 \right) \\
c_i &:= \left(i, \qquad 0 \right) \\
e_j &:= \left(j \cdot n/(2m), \qquad j \cdot n/(2\sqrt{3}m) \right) \\
f_j &:= \left(n - j \cdot n/(2m), \quad j \cdot n/(2\sqrt{3}m) \right) \\
g_j &:= \left(n/2, \qquad n \cdot \sqrt{3}/2 - j \cdot n/(\sqrt{3}m) \right)
\end{aligned}\right\} \tag{4}$$

Whenever we use the words left, right, above or below to express the relative position between two points in an instance $T_{n,m}$, we always consider this instance to be embedded in such a way that c is parallel to the x-axis and A, B, and C are oriented counterclockwise.
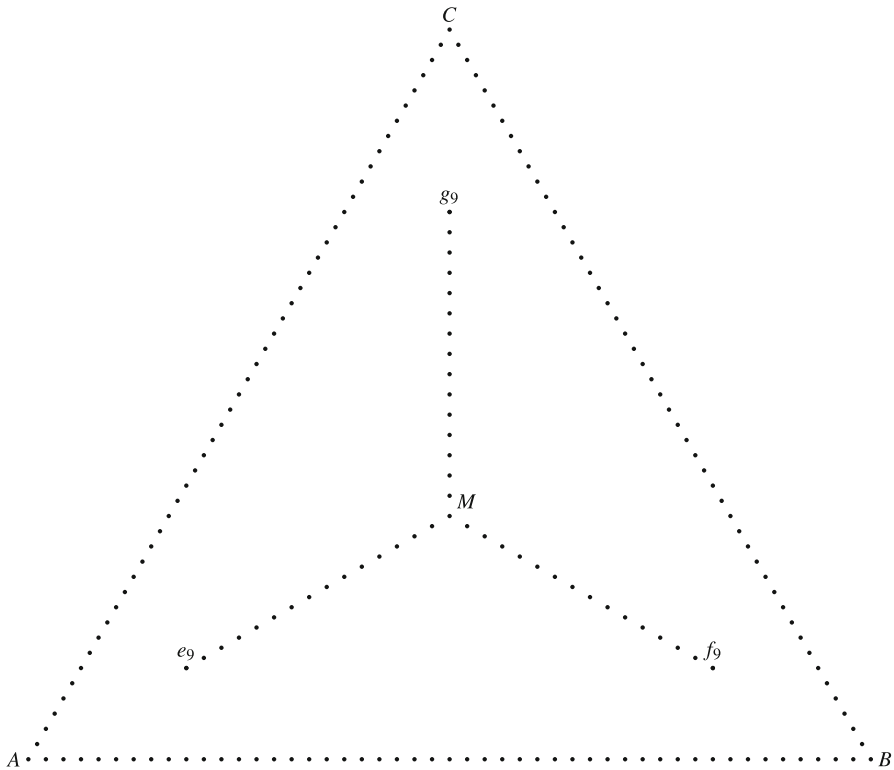
Fig. 2 The modified tetrahedron instance $T'_{48,24}$

2.2 The modified tetrahedron instances $T'_{n,m}$

To analyze the integrality ratio of the instances $T_{n,m}$ it turns out to be useful to introduce slightly modified instances $T'_{n,m}$. The instance $T'_{n,m}$ is obtained from $T_{n,m}$ by removing all internal vertices whose distance to one of the vertices A, B, or C is less than $\max\{10, 4 + 4\gamma\}$. Figure 2 shows the instance $T'_{48,24}$.

We have:

$$\text{dist}(A, e_j) \geq \max\{10, 4 + 4\gamma\} \text{ for all } e_j \in T'_{n,m}. \tag{5}$$

As $\text{dist}(c_i, e_j) \geq \text{dist}(A, e_j) \cdot \sin(30°) = \frac{1}{2} \cdot \text{dist}(A, e_j)$ we get

$$\text{dist}(c_i, e_j) \geq 5 \text{ for all } c_i, e_j \in T'_{n,m}. \tag{6}$$

We will need that an instance $T'_{n,m}$ contains at least four internal vertices. We therefore require that

$$n \geq 40 \text{ and } m \geq 22. \tag{7}$$

Using (2) this implies

$$\text{dist}(A, e_{m-1}) = \frac{(m-1) \cdot n}{\sqrt{3} \cdot m} \geq \frac{21}{22\sqrt{3}} \cdot n > \frac{n}{2} \geq 10$$

and using (3) this implies

$$\text{dist}(A, e_{m-1}) \ = \ \frac{(m-5) \cdot n}{\sqrt{3} \cdot m} + \frac{4 \cdot n}{\sqrt{3} \cdot m} \geq \frac{17}{22\sqrt{3}} \cdot n + 4\gamma \ > \ 4 + 4\gamma$$

and therefore $T'_{n,m}$ contains at least the four internal vertices e_{m-1}, f_{m-1}, g_{m-1}, and M if $n \geq 40$ and $m \geq 22$.

Note that as long as the ratio n/m is constant, the instances $T_{n,m}$ and $T'_{n,m}$ differ by a constant number of vertices only. Therefore, as we will see later, $T_{n,m}$ and $T'_{n,m}$ have asymptotically the same integrality ratio, if n/m is constant.

3 Structural properties of optimal tours

In this section we show that the shape of the tour shown in Fig. 7 is the unique optimal tour for $T'_{n,m}$ up to symmetry. To achieve this, we first summarize some previous results for optimal tours. Then, we introduce the concept of a pseudo-tour from which we derive structural properties of optimal tours in Theorem 1. After that, we make further geometrical observations culminating in Theorem 2 describing the unique optimal tour up to symmetry.

A *tour* for a Euclidean Traveling Salesman instance is a polygon that contains all the given vertices. A polygon is called *simple* if no two of its edges intersect except for the common vertex of two consecutive edges. As an immediate consequence from the triangle inequality we have

Lemma 1 (Flood 1956 [7]) *Unless all vertices lie on one line, an optimal tour of a Euclidean Traveling Salesman instance is a simple polygon.* □

An important consequence of Lemma 1 is the following result:

Lemma 2 ([6], page 142) *An optimal tour of a Euclidean Traveling Salesman instance visits the vertices on the boundary of the convex hull of all vertices in their cyclic order.* □

In case of the tetrahedron instances, the vertices on the boundary of the convex hull are exactly the base vertices. From now on we fix the orientation of an optimal TSP tour of the tetrahedron instance such that the base vertices are visited in counterclockwise order. We use the notation (x, y) for an edge of an oriented optimum tour.

A *subpath* of an oriented tour consists of vertices $v_1, \ldots, v_l$, such that v_{i+1} is visited by the tour immediately after v_i for all $i = 1, \ldots, l-1$. A subpath of a tour starting and ending at base vertices and containing no other base vertices is called a *trip* if it contains at least one internal vertex. The first and the last internal vertex of a trip (which may coincide) are called the *connection vertices* of the trip. By Lemma 2 the base vertices are visited counterclockwise. Therefore, the two end vertices of a trip must be consecutive base vertices belonging to the same side of the triangle ABC. We call this side the *main side* of the trip.

Each optimum tour of a tetrahedron instance can be decomposed into a set of trips and a set of edges connecting consecutive base vertices such that all internal vertices
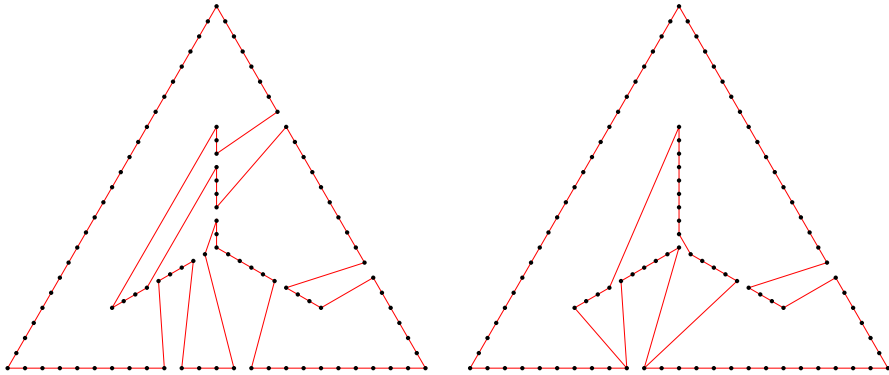
Fig. 3 A tour with four trips in the instance $T'_{24,18}$ (left). A pseudo-tour in the instance $T'_{24,18}$ (right)

are contained in some trip and two different trips intersect in at most one base vertex (see Fig. 3, left). A set of edges between consecutive base vertices together with a set of trips that contain all internal vertices is called a *pseudo-tour* if it is not a tour and satisfies the following property: for any two consecutive base vertices there exists at least one trip having these two vertices as end vertices if and only if these two vertices are not connected by an edge (see Fig. 3, right).

The following result will be useful to prove that certain tours of the tetrahedron instances are not optimum.

Lemma 3 *A pseudo-tour in $T'_{n,m}$ is more than 1 longer than an optimum tour in $T'_{n,m}$.*

Proof Let T be a pseudo-tour of minimum length in $T'_{n,m}$. There exist two consecutive base vertices that are the end vertices of at least two trips. Wlog we may assume that the two base vertices are c_i and c_{i+1}. Let $c_i, x_1, \ldots, x_k, c_{i+1}$ be the vertices of the first trip and $c_i, y_1, \ldots, y_l, c_{i+1}$ be the vertices of a second trip with $k, l \geq 1$. By (6) we know that the distance from c_i to x_1 or y_1 and from c_{i+1} to x_k or y_l is at least 5. Moreover, we have $\text{dist}(c_i, c_{i+1}) = 1$ and T is intersection free. If there is no $u \in \{x_k, y_l\}$ and $v \in \{x_1, y_1\}$ such that the rays $\overrightarrow{c_i u}$ and $\overrightarrow{c_{i+1} v}$ intersect, we have $\angle x_1 c_i y_1 + \angle x_k c_{i+1} y_l \leq 180°$. Otherwise, consider the intersection P of two of these rays such that no other intersection point lies on or in the triangle $c_i c_{i+1} P$ and let Q be the intersection of the angle bisector of $\angle c_i P c_{i+1}$ with c. Then,

$$\angle x_1 c_i y_1 + \angle x_k c_{i+1} y_l \leq 180° - \angle c_{i+1} c_i P + 180° - \angle P c_{i+1} c_i$$
$$= 360° - (180° - \angle c_i P c_{i+1})$$
$$= 180° + \angle c_i P c_{i+1}.$$

On the other hand either $\text{dist}(c_i, Q) \leq \frac{1}{2}$ or $\text{dist}(Q, c_{i+1}) \leq \frac{1}{2}$ so wlog $\text{dist}(c_i, Q) \leq \frac{1}{2}$. Using the law of sines we get:

$$\sin\left(\frac{\angle c_i P c_{i+1}}{2}\right) = \sin(\angle c_i P Q) = \frac{\text{dist}(c_i, Q)}{\text{dist}(c_i, P)} \sin(\angle P Q c_i) \le \frac{\frac{1}{2}}{5} < \frac{1}{5}.$$

Therefore, the total sum of the two angles $x_1 c_i y_1$ and $x_k c_{i+1} y_l$ is at most $180° + 2\alpha$ with $\sin\alpha < 1/5$.

Wlog we may assume that the angle $x_1 c_i y_1$ is at most $90° + \alpha$. Set $x := \text{dist}(c_i, x_1)$, $y := \text{dist}(c_i, y_1)$, and $z := \text{dist}(x_1, y_1)$. As by (6) $x, y \ge 5$ we get $-4x + xy \ge xy/5$ and $-4y + xy \ge xy/5$ which implies

$$4 - 4x - 4y + 2xy > 2xy/5.$$

By adding $x^2 + y^2$ to both sides of this inequality and using $\frac{1}{5} > \sin(\alpha) = -\cos(90° + \alpha)$ we get with the law of cosines

$$(x + y - 2)^2 > x^2 + y^2 + 2xy/5 > x^2 + y^2 - 2xy\cos(90° + \alpha) \ge z^2.$$

Therefore, we have

$$z + 2 < x + y.$$

We now replace the edges (c_i, x_1) and (c_i, y_1) by the edges (c_i, c_{i+1}) and (x_1, y_1) and shortcut the two edges (c_i, c_{i+1}) and (c_{i+1}, y_l) by the edge (c_i, y_l). This yields either a pseudo-tour that is shorter than T, contradicting the choice of T. Or it yields a tour that is by more than 1 shorter than T which proves the result. $\qquad\square$

Lemma 4 *In an optimum tour for $T'_{n,m}$ a trip with end vertices c_i and c_{i+1} cannot contain the edge (c_i, g_j) for $1 \le j \le m - 1$.*

Proof Suppose there exists a trip with end vertices c_i and c_{i+1} that contains the edge (c_i, g_j) for some $j \in \{1, \ldots, m - 1\}$. Let (x, c_{i+1}) be the last edge of the trip.

If $x \in \{f_1, \ldots, f_m, g_1, \ldots, g_{m-1}\}$ then we can replace the edges (c_i, g_j) and (x, c_{i+1}) by (a_{n-i}, g_j) and $(x, a_{n-(i+1)})$ and add the edge (c_i, c_{i+1}). Note that $\text{dist}(c_i, g_j) > \text{dist}(a_{n-i}, g_j)$ and $\text{dist}(x, c_{i+1}) \ge \text{dist}(x, a_{n-(i+1)})$. If $(a_{n-(i+1)}, a_{n-i})$ is an edge in the tour, then we can remove this edge and get a shorter tour, contradiction. Otherwise, we got a pseudo-tour that is at most 1 longer than an optimum tour. This contradicts Lemma 3.

If $x \in \{e_1, \ldots, e_{m-1}\}$ we can replace the edges (c_i, g_j) and (x, c_{i+1}) by (b_{n-i}, g_j) and $(x, b_{n-(i+1)})$ and use an analogous argument. $\qquad\square$

From now on we denote by x' the reflection of x across c.

Lemma 5 *Let (c_i, x) and (y, c_{i+1}) be the first and last edge of a trip in an optimum tour for $T'_{n,m}$. Then $\text{dist}(x, y') + 1 \ge \text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) > \text{dist}(x, y') - 1$.*

Proof By symmetry and the triangle inequality, we get

$$\text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) > \text{dist}(c_i, x) + \text{dist}(y, c_i) - 1$$
$$= \text{dist}(c_i, x) + \text{dist}(y', c_i) - 1$$
$$\geq \text{dist}(x, y') - 1.$$

On the other hand, consider the intersection P of xy' with c and let c_j be the next vertex left of P. Since the trip belongs to an optimal tour we have $\text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) \leq \text{dist}(c_j, x) + \text{dist}(y, c_{j+1})$, otherwise we can replace (c_i, x), (y, c_{i+1}) by (c_j, x), (y, c_{j+1}) and (c_i, c_{i+1}) to get a pseudo-tour that is at most 1 longer than an optimum tour contradicting Lemma 3 or remove in addition (c_j, c_{j+1}) to get a shorter tour. Hence

$$\text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) \leq \text{dist}(c_j, x) + \text{dist}(y, c_{j+1})$$
$$\leq \text{dist}(x, P) + \text{dist}(c_j, P) + \text{dist}(P, y)$$
$$+ \text{dist}(P, c_{j+1})$$
$$= \text{dist}(x, y') + 1.$$

$\square$

Lemma 6 *For x and y lying on the same internal segment we have:*

$$\text{dist}(x, y') \geq \text{dist}(x, y) + 2 + 2\gamma.$$

Proof Wlog we may assume that x, y are on e and x is to the left of y. Let P be the perpendicular foot of x on AM' (Fig. 4). For reasons of symmetry, we get $\text{dist}(A, x) = \text{dist}(A, x')$ and $\angle x'Ax = 2 \cdot \angle BAM = \angle BAC = 60°$. Therefore, the triangle $Ax'x$ is equilateral and we have $\text{dist}(P, x') = \frac{1}{2}\text{dist}(A, x)$. Together with the triangle inequality and as the triangle $Py'x$ has the hypotenuse xy' and (5) we get:

$$\text{dist}(x, y') > \text{dist}(P, y') = \text{dist}(P, x') + \text{dist}(x', y') = \text{dist}(x, y) + \frac{1}{2}\text{dist}(A, x)$$

$$\geq \text{dist}(x, y) + \frac{1}{2}(4 + 4\gamma) = \text{dist}(x, y) + 2 + 2\gamma.$$

$\square$

Theorem 1 *An optimum tour for $T'_{n,m}$ does not contain a trip where the two connection vertices lie on the same internal segment.*

Proof Assume such a trip t exists. Wlog let t have c as main side, starting at c_i and ending at c_{i+1}. Let x resp. y be the first resp. last connection vertex of the trip. By Lemma 4 the vertices x and y cannot belong to $\{g_1, \ldots, g_{m-1}\}$. We may assume that x and y lie wlog on e.

Suppose there are internal vertices p and q with $\text{dist}(p, q) = \gamma$ and p is visited by the trip t, but q is not. In this case, we replace (c_i, x) and (y, c_{i+1}) by $(p, q), (q, p)$,
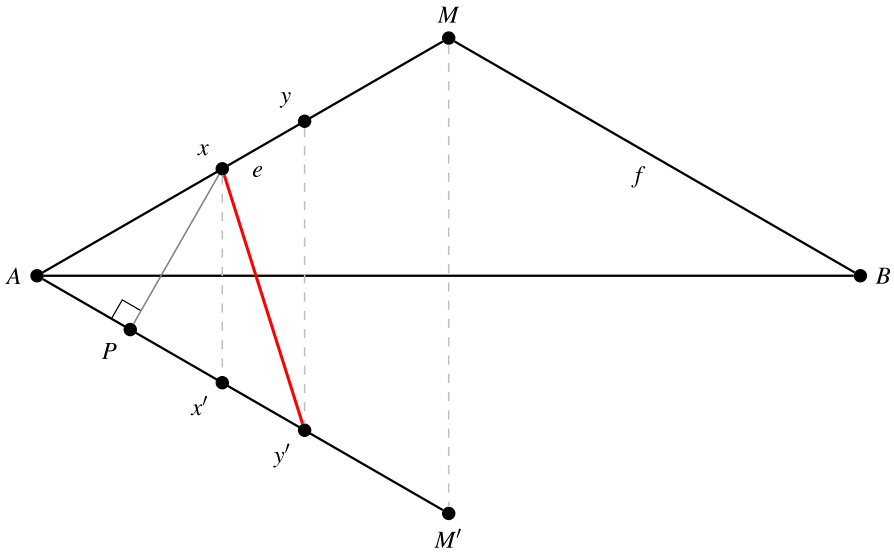
Fig. 4 The situation in the proof of Lemma 6

(y, x), and (c_i, c_{i+1}) (Fig. 5). Since q is not visited by the trip t, we get an upper bound for the length of an optimal tour. By Lemmas 5 and 6 we have

$$\text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) > \text{dist}(x, y) + 1 + 2\gamma$$
$$= \text{dist}(x, y) + \text{dist}(c_i, c_{i+1})$$
$$+ \text{dist}(p, q) + \text{dist}(q, p).$$

Thus, the modification makes the tour shorter, contradiction. Hence, there is no such vertex p visited by the trip with a neighbor q at distance γ not visited by the trip. It follows that the trip visits all internal vertices and it is the only trip.

Since the trip visits all internal vertices, there is an edge (u, v) leaving the internal segment e the first time where u lies to the left of y. Furthermore v cannot lie on the internal segment f, since otherwise (u, v) would intersect the edge (y, c_{i+1}) Therefore, v lies on the internal segment g. Consider b_{n-i}, the reflection of c_i with respect to e. Because there is only one trip, b_{n-i} is not a starting vertex of a trip and the edge $(b_{n-i}, b_{n-(i+1)})$ is part of the tour. We replace the edges (c_i, x), (y, c_{i+1}) and $(b_{n-i}, b_{n-(i+1)})$ by (b_{n-i}, x), $(y, b_{n-(i+1)})$ and (c_i, c_{i+1}). We have by symmetry:

$$\text{dist}(c_i, x) + \text{dist}(y, c_{i+1}) + \text{dist}(b_{n-i}, b_{n-(i+1)})$$
$$= \text{dist}(b_{n-i}, x) + \text{dist}(y, b_{n-(i+1)}) + \text{dist}(c_i, c_{i+1}).$$

Therefore, the new tour is optimum as well. But $(y, b_{n-(i+1)})$ intersects (u, v), contradicting the optimality. □
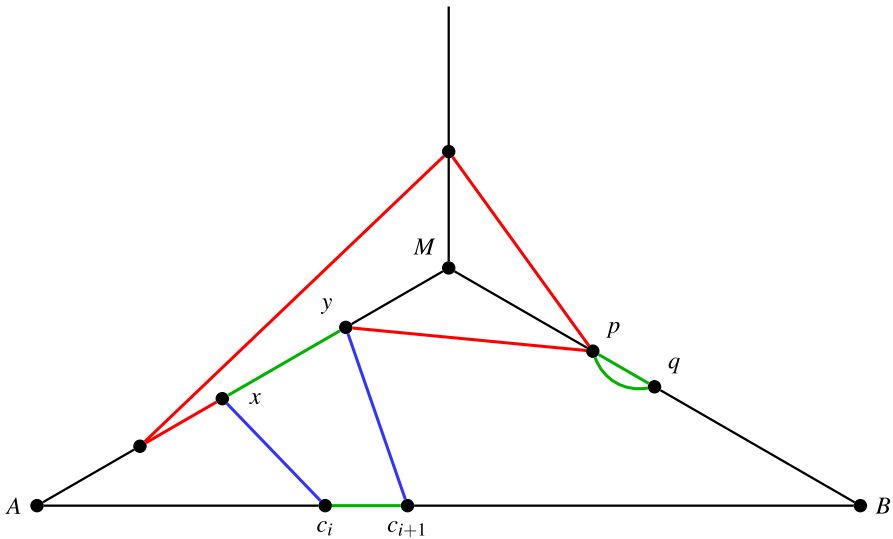
Fig. 5 Illustration of the proof of Theorem 1. The trip consists of the red and blue edges. The blue edges are replaced by the green edges to get an upper bound on the length of the trip (color figure online)

Corollary 1 *Let t be a trip with main side c in an optimum tour for $T'_{n,m}$. Then the first connection vertex of t lies on e and the second connection vertex lies on f.*

Proof By Theorem 1, the two connection vertices lie on different internal segments. By Lemma 4 the first connection vertex has to lie on e and the second connection vertex on f. □

The next step is to show that the optimal tour consists of only one trip of a certain shape. Let i_0 be the smallest integer such that e_{i_0} is a vertex in $T'_{n,m}$.

Lemma 7 *We have*

$$\frac{\text{dist}(A, e_{i_0})}{\text{dist}(A, M)} < \frac{1}{2}.$$

Proof By definition of the instances $T'_{n,m}$ we have $\text{dist}(A, e_{i_0}) \leq \max\{10, 4+4\gamma\} + \gamma$ and $\text{dist}(A, M) = \frac{n}{\sqrt{3}} = m \cdot \gamma$. Moreover $\frac{10+\gamma}{m \cdot \gamma} = \frac{10 \cdot \sqrt{3}}{n} + \frac{1}{m} < \frac{1}{2}$ and $\frac{4+5\gamma}{m \cdot \gamma} = \frac{4 \cdot \sqrt{3}}{n} + \frac{5}{m} < \frac{1}{2}$ as by assumption (7) we have $n \geq 40$ and $m \geq 22$. □

Let $\mathcal{T}$ be the set of trips that occur in any optimum tour. For $t \in \mathcal{T}$ let $g(t)$ be the sum of the lengths of the first and last edge of t, and $\delta := \min_{t \in \mathcal{T}} (g(t))$. Let M_c be the center of c.

Lemma 8 *Let t be a trip in an optimum tour of $T'_{n,m}$ such that t contains at least one of $\{e_{i_0}, f_{i_0}, g_{i_0}\}$ as a connection vertex. Then the total length of the first and last edge in t is at least $\text{dist}(e_{i_0}, f_{m-1}) + 1$.*
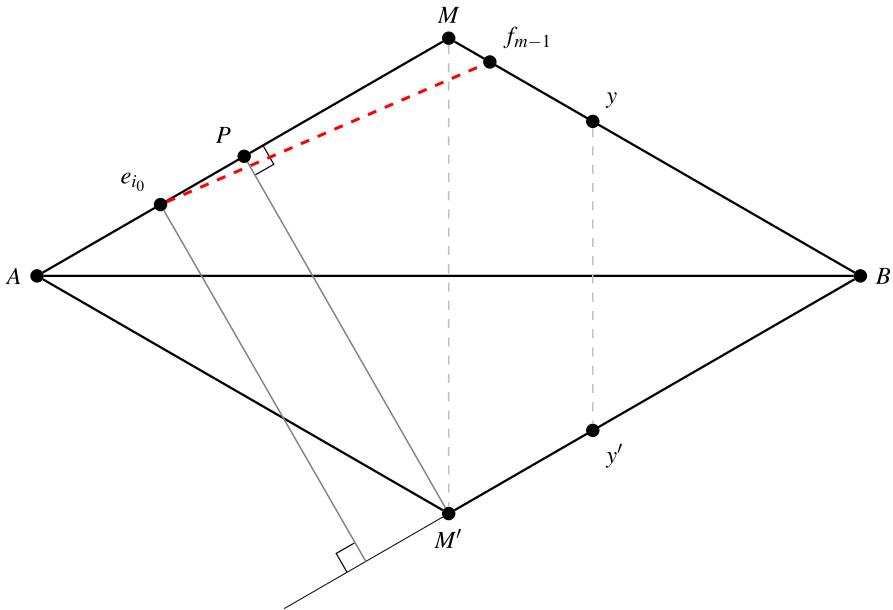
Fig. 6 The situation in the proof of Lemma 8

Proof Let wlog (c_i, e_{i_0}) and (y, c_{i+1}) be the first and last edge of such a trip and P be the perpendicular foot of M' to AM (Fig. 6). By Corollary 1 we know that y lies on f. Since the triangle $AM'M$ is equilateral, P is the center of AM. Therefore by Lemma 7 e_{i_0} lies left of P. Hence the perpendicular foot of e_{i_0} to $M'B$ lies left of M' and we have $\mathrm{dist}(e_{i_0}, y') \geq \mathrm{dist}(e_{i_0}, M')$. By Lemmas 5 and 6 we get

$$\mathrm{dist}(c_i, e_{i_0}) + \mathrm{dist}(y, c_{i+1}) \geq \mathrm{dist}(e_{i_0}, y') - 1 \geq \mathrm{dist}(e_{i_0}, M') - 1$$
$$\geq \mathrm{dist}(e_{i_0}, M) + 1 + 2\gamma > \mathrm{dist}(e_{i_0}, f_{m-1}) + 1.$$

$\square$

Lemma 9 *We have* $\delta \geq \mathrm{dist}(P, M') - 1$ *where* P *is the perpendicular foot of* M' *to* AM.

Proof Let x and y be the first and last connection vertex of a trip contained in an optimum tour of $T'_{n,m}$. By Lemma 5 the total length of the first and last edge of this trip is at least $\mathrm{dist}(x, y') - 1$. By Corollary 1 $\mathrm{dist}(x, y') \geq \mathrm{dist}(P, M')$, since $\mathrm{dist}(P, M')$ is the distance between the parallel segments AM and $M'B$. $\square$

Lemma 10 *We have* $\mathrm{dist}(g_{i_0}, f_{m-1}) + \delta > \min_i(\mathrm{dist}(c_i, e_{i_0}) + \mathrm{dist}(f_{i_0}, c_{i+1})) + \gamma$

Proof By Lemma 9 it suffices to prove that

$$\mathrm{dist}(g_{i_0}, f_{m-1}) + \mathrm{dist}(P, M') - 1 > \min_i(\mathrm{dist}(c_i, e_{i_0}) + \mathrm{dist}(f_{i_0}, c_{i+1})) + \gamma. \quad (8)$$

As by the triangle inequality

$$\min_i(\text{dist}(c_i, e_{i_0}) + \text{dist}(f_{i_0}, c_{i+1})) = \min_i(\text{dist}(c_i, e_{i_0}) + \text{dist}(f_{i_0}, c_{i+1})$$
$$+ \text{dist}(c_i, c_{i+1})) - 1$$
$$< \text{dist}(e_{i_0}, f'_{i_0}) + 1$$

the inequality (8) follows from the inequality $\text{dist}(g_{i_0}, M) + \text{dist}(P, M') - 2 - \gamma > \text{dist}(e_{i_0}, f'_{i_0})$ which we are now going to prove.

Consider the point on S on AM with $\text{dist}(M, S) = \text{dist}(P, M') - 2 - \gamma$ and that lies on the opposite side of M than A. It is enough to show that $\frac{\text{dist}(e_{i_0}, f'_{i_0})}{\text{dist}(e_{i_0}, S)} < 1$. By symmetry, we have $2 \cdot \text{dist}(e_{i_0}, M_c) = \text{dist}(e_{i_0}, f'_{i_0})$. By the sine law in the triangle $e_{i_0} M_c S$, we have $\frac{\text{dist}(e_{i_0}, M_c)}{\text{dist}(e_{i_0}, S)} = \frac{\sin(\angle e_{i_0} S M_c)}{\sin(\angle S M_c e_{i_0})}$. If we move e_{i_0} towards A, $\angle e_{i_0} S M_c$ is fixed and $\angle S M_c e_{i_0}$ is monotonically increasing. Since the sine is concave in $[0, \pi]$, it is enough to verify the statement for the cases where e_{i_0} is as far and as near as possible to A. By Lemma 7 e_{i_0} lies between A and P. For $e_{i_0} = A$, we have using (7):

$$\text{dist}(e_{i_0}, S) = \text{dist}(e_{i_0}, M) + \text{dist}(M, S) = \frac{\sqrt{3}}{3}n + \frac{1}{2}n - 2 - \gamma$$
$$\geq \frac{\sqrt{3}}{3}n + \frac{1}{2}n - 2 - \frac{\sqrt{3}}{66}n > n + \frac{1}{20}n - 2 \geq n > \text{dist}(e_{i_0}, f'_{i_0}).$$

For $e_{i_0} = P$, we have

$$\text{dist}(e_{i_0}, S) = \text{dist}(e_{i_0}, M) + \text{dist}(M, S)$$
$$\geq \frac{1}{2\sqrt{3}}n + \frac{1}{2}n - 2 - \gamma$$
$$\geq \frac{1}{\sqrt{3}}n + \frac{\sqrt{3}-1}{2\sqrt{3}}n - 2 - \frac{\sqrt{3}}{66}n$$
$$> \frac{1}{\sqrt{3}}n = \text{dist}(A, M) = \text{dist}(e_{i_0}, f'_{i_0}).$$

$\square$

Lemma 11 $c_{\lfloor \frac{n-1}{2} \rfloor}$ *resp.* $c_{\lfloor \frac{n-1}{2} \rfloor + 1}$ *and* $c_{\lceil \frac{n-1}{2} \rceil}$ *resp.* $c_{\lceil \frac{n-1}{2} \rceil + 1}$ *are the optimal starting resp. ending vertices of a trip with main side c and connection vertices e_{i_0} and f_{i_0}.*

Proof We have to show $\text{dist}(c_i, e_{i_0}) + \text{dist}(f_{i_0}, c_{i+1})$ is minimal for $i \in \{\lfloor \frac{n-1}{2} \rfloor, \lceil \frac{n-1}{2} \rceil\}$. Let $\overline{f}_{i_0}$ be the point obtained by shifting f_{i_0} to the left by 1. Let $\overline{M}$ be the point of intersection of the perpendicular bisector of $e_{i_0}\overline{f}_{i_0}$ and c. We have $\text{dist}(c_i, e_{i_0}) + \text{dist}(f_{i_0}, c_{i+1}) = \text{dist}(c_i, e_{i_0}) + \text{dist}(\overline{f}_{i_0}, c_i)$. The trace of points X satisfying $\text{dist}(X, e_{i_0}) + \text{dist}(\overline{f}_{i_0}, X) = k$ for a fixed k is an ellipsoid with focal points e_{i_0} and $\overline{f}_{i_0}$. The size of the ellipsoid is strictly monotonically decreasing with k. Consider the ellipsoid through c_i with focal points e_{i_0} and $\overline{f}_{i_0}$. The size of the ellipsoid

is strictly monotonically decreasing with the distance of c_i to $\overline{M}$. Thus, the sum $\text{dist}(c_i, e_{i_0}) + \text{dist}(\overline{f}_{i_0}, c_i)$ is strictly monotonically decreasing with the distance of c_i to $\overline{M}$. For odd n, $\overline{M}$ is $c_{\frac{n-1}{2}}$, for even n, it is the midpoint of $c_{\lfloor \frac{n-1}{2} \rfloor}$ and $c_{\lceil \frac{n-1}{2} \rceil}$. This proves the statement. $\qquad\square$

Lemma 12 *Let x resp. y be vertices on e resp. f, c_i be any base vertex. If the edge (c_i, x) is part of an optimum tour in $T'_{n,m}$, then there exists no z on e and to the left of x such that (z, g_k) is part of the tour for any vertex g_k. Similarly if (y, c_i) is part of the tour, then there exists no z on f and to the right of y such that (g_k, z) is part of the tour for any vertex g_k.*

Proof Assume there is a leaving edge (z, g_k) with z to the left of x. Then, (c_i, z) is not in the tour, since this would result in a pseudo-tour and (x, g_k) is not in the tour since otherwise the edge (z, g_k) would be oriented (g_k, z). Hence, we can replace the edges (c_i, x) and (z, g_k) by (c_i, z) and (x, g_k). If the old tour was $(c_i, x)X(z, g_k)Y$ where X, Y are subpaths, we get the new tour $(c_i, z)\overline{X}(x, g_k)Y$. Here $\overline{X}$ denotes the reversed subpath of X. Since the segments $c_i x$ and $f_k z$ intersect, $\text{dist}(c_i, z) + \text{dist}(f_k, x) < \text{dist}(c_i, x) + \text{dist}(f_k, z)$ follows from the (strict) triangle inequality applied to the triangles $zc_i s$ and $sf_k x$, where s is the point of intersection of $c_i x$ and $f_k z$. By symmetry we have $\text{dist}(f_k, x) = \text{dist}(g_k, x)$ and $\text{dist}(f_k, z) = \text{dist}(g_k, z)$ and thus $\text{dist}(c_i, z) + \text{dist}(g_k, x) < \text{dist}(c_i, x) + \text{dist}(g_k, z)$. Hence, the new tour is shorter, which is a contradiction to the optimality. Analogously we can prove the second statement. $\qquad\square$

Lemma 13 *For all internal vertices, the distance to the nearest vertex is γ, for all other vertices it is 1.*

Proof By definition the smallest distance from an internal vertex to another internal vertex is γ, and the smallest distance from a noninternal vertex to another noninternal vertex is 1. On the other hand, the smallest distance between an internal vertex and a noninternal vertex is at least $\sin(30°) \cdot \text{dist}(A, e_{i_0}) \geq \frac{1}{2} \max\{10, 4 + 4\gamma\}$ which is larger than 1 and γ. $\qquad\square$

Consider the tour T^* that only contains one trip which visits e_{i_0} first and f_{i_0} last and visits adjacent internal vertices except for the edge (g_{i_0}, f_{m-1}). Moreover, the first and last vertex of the trip in T^* are the vertices $c_{\lfloor \frac{n-1}{2} \rfloor}$ and $c_{\lfloor \frac{n-1}{2} \rfloor + 1}$ (Fig. 7). We show that up to rotations and reflections the tour T^* is the only optimum tour in $T'_{n,m}$.

Let $K := 3n - 1 + (3(m - i_0) - 1)\gamma$. This is the maximum total length of all pairwise point distances that are 1 or γ that can be contained in a tour in $T'_{n,m}$. Note that the tour T^* has length $\text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0}) + \text{dist}(f_{i_0}, c_{\lfloor \frac{n-1}{2} \rfloor + 1}) + \text{dist}(g_{i_0}, f_{m-1}) + K$.

Let the *long edges* of a TSP tour be the edges that are incident to one of $\{e_{i_0}, f_{i_0}, g_{i_0}\}$ and have length larger than γ. A long edge is called *internal* if it connects two internal vertices, it is called *externalizing* if it is the starting or ending edge of a trip.

Theorem 2 *Up to symmetry the tour T^* is the unique optimum tour for $T'_{n,m}$.*

Proof Let T be an optimum tour. First we prove that there exist at least three long edges in T. Consider the neighbors of e_{i_0} in T. As the tour T is a simple polygon by
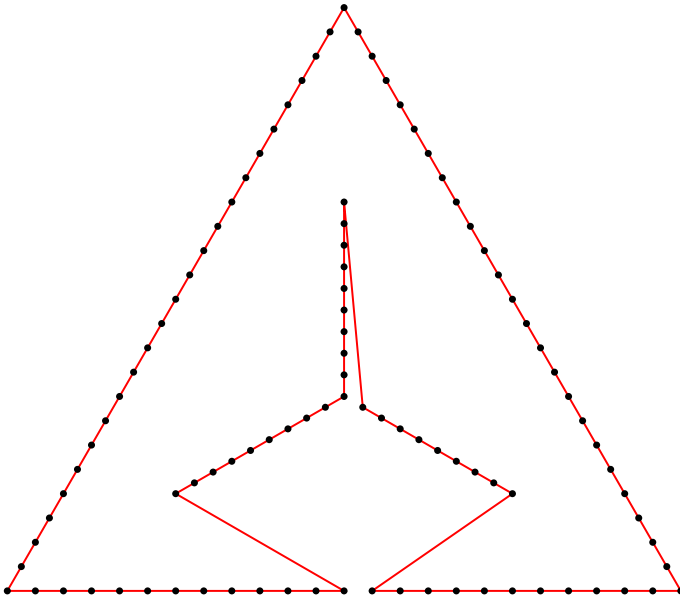
Fig. 7 An optimum tour for the instance $T'_{24,18}$

Lemma 1 one of the two edges incident to e_{i_0} is either an internal or externalizing long edge. With the same argument for f_{i_0} and g_{i_0} we get at least two long edges incident to $\{e_{i_0}, f_{i_0}, g_{i_0}\}$. Assume that we get exactly two long edges. Then, we have at least one long edge connecting two of $\{e_{i_0}, f_{i_0}, g_{i_0}\}$, wlog the edge $\{e_{i_0}, g_{i_0}\}$. In this case e_{i_0} or g_{i_0} cannot be connection vertices, since otherwise we would get three long edges. Hence, the trip containing the edge $\{e_{i_0}, g_{i_0}\}$ cannot have b as main side, since this would intersect the edge. By symmetry, assume that (c_i, x) and (y, c_{i+1}) are the first and last edge of the trip, respectively. By Corollary 1 x lies on e. If the orientation of the edge $\{e_{i_0}, g_{i_0}\}$ is (e_{i_0}, g_{i_0}) then (c_i, x) and (e_{i_0}, g_{i_0}) contradict Lemma 12, since e_{i_0} and g_{i_0} are not connection vertices. If the orientation is (g_{i_0}, e_{i_0}), the part of the tour connecting c_{i+1} and e_{i_0} has to intersect that part connecting c_i and g_{i_0}. This contradicts the optimality of the tour. Hence, we get at least three long edges.

Given a set of long edges, we can get a lower bound of the tour by bounding the length of the remaining edges by the distance of the vertices to the closest neighbors. If two of them are externalizing long edges sharing a trip and the third is internal, then the tour has at least length $K + \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0}) + \text{dist}(f_{i_0}, c_{\lfloor \frac{n-1}{2} \rfloor + 1}) + \text{dist}(g_{i_0}, f_{m-1})$ which is the length of T^*. If the third edge is an externalizing edge, the tour is longer than $K - 1 + \text{dist}(c_i, e_{i_0}) + \text{dist}(f_{i_0}, c_{i+1}) + \text{dist}(e_{i_0}, f_{m-1}) + 1$ by Lemma 8. This is more than the length of T^*. It remains the case where no two of these three edges belong to the same trip. If there is one externalizing and two internal edges the tour has length at least $K - \gamma + 2\,\text{dist}(g_{i_0}, f_{m-1}) + \delta$. By Lemma 10 this is larger than the length of T^*. If all three long edges are internal edges, we get an even higher lower bound for the length of T of value $K - 2\gamma + 3\,\text{dist}(g_{i_0}, f_{m-1}) + \delta$ since every tour has at least one trip. Tours where at least two of the three long edges are externalizing

Fig. 8 An optimum solution for the instance $T_{17,17}$

edges have at least length $\min(K - 1 - \gamma + \text{dist}(g_{i_0}, f_{m-1}) + 2\delta, K - 2 - \gamma + 3\delta)$ which is larger than the length of T^* by Lemmas 8 and 10. Together with Lemma 11 the tour T^* is unique up to symmetry. □

Optimum tours in $T_{n,m}$ are not unique and may contain several trips. An example of an optimum tour in $T_{n,m}$ is shown in Fig. 8.

3.1 Asymptotic value of the integrality ratio

To get a lower bound on the integrality ratio of the instances $T_{n,m}$ we first compute a lower bound on the length of an optimum TSP tour of the instance $T'_{n,m}$ and an upper bound on an optimum solution to the subtour LP for the instance $T'_{n,m}$.

Theorem 3 *For $n \leq 3/2 \cdot m$ an optimum TSP tour of the instance $T'_{n,m}$ has length at least $4n + 4n/\sqrt{3} - 69$ and at most $4n + 4n/\sqrt{3} - 17$.*

Proof From the proof of Theorem 2 it follows that an optimum TSP tour for $T'_{n,m}$ has length

$$3n - 1 + 3 \cdot \text{dist}(M, e_{i_0}) - \gamma + \text{dist}(g_{i_0}, f_{m-1}) + \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0})$$
$$+ \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, f_{i_0}),$$

where i_0 is the smallest index such that e_{i_0} is a vertex of $T'_{n,m}$. For $n \leq 3/2 \cdot m$ we have $\gamma = \frac{n}{\sqrt{3} \cdot m} \leq \frac{\sqrt{3}}{2}$ and therefore $\max\{10, 4 + 4\gamma\} = 10$. This implies $10 \leq$

$\text{dist}(A, e_{i_0}) < 10 + \gamma$ and using the triangle inequality we get

$$n/\sqrt{3} - 10 + \gamma > \text{dist}(g_{i_0}, f_{m-1}) > \text{dist}(M, e_{i_0}) > n/\sqrt{3} - 10 - \gamma.$$

Applying the triangle inequality once more we have:

$$\text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) + (10 + \gamma) > \text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) + \text{dist}(A, e_{i_0}) > \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0})$$

and

$$\text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0}) \geq \text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) - \text{dist}(A, e_{i_0}) \geq \text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) - (10 + \gamma).$$

Combining these inequalities with similar inequalities for $\text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, f_{i_0})$ we get:

$$
\begin{aligned}
\text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0}) &+ \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, f_{i_0}) \\
&\geq \text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) + \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, B) - 2(10 + \gamma) \\
&= \text{dist}(A, B) - 21 - 2\gamma \\
&= n - 21 - 2\gamma
\end{aligned}
$$

and

$$
\begin{aligned}
\text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor}, e_{i_0}) &+ \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, f_{i_0}) \\
&\leq \text{dist}(A, c_{\lfloor \frac{n-1}{2} \rfloor}) + \text{dist}(c_{\lfloor \frac{n-1}{2} \rfloor + 1}, B) + 2(10 + \gamma) \\
&= \text{dist}(A, B) + 19 + 2\gamma \\
&= n + 19 + 2\gamma.
\end{aligned}
$$

Combining all these inequalities and using $\gamma < 1$ for $n \leq 3/2 \cdot m$ we get that

$$4n + 4n/\sqrt{3} - 69$$

is a lower bound and

$$4n + 4n/\sqrt{3} - 17$$

is an upper bound on the length of an optimum TSP tour in $T'_{n,m}$. $\quad\square$

Theorem 4 *For $n \leq 3/2 \cdot m$ an optimum solution to the subtour LP of the instance $T'_{n,m}$ has length at most $3n + 3n/\sqrt{3}$ and at least $3n + 3n/\sqrt{3} - 33$.*

Proof A feasible solution to the subtour LP is shown in Fig. 9. Its total length is at most $3n + 3n/\sqrt{3}$. This proves the upper bound.
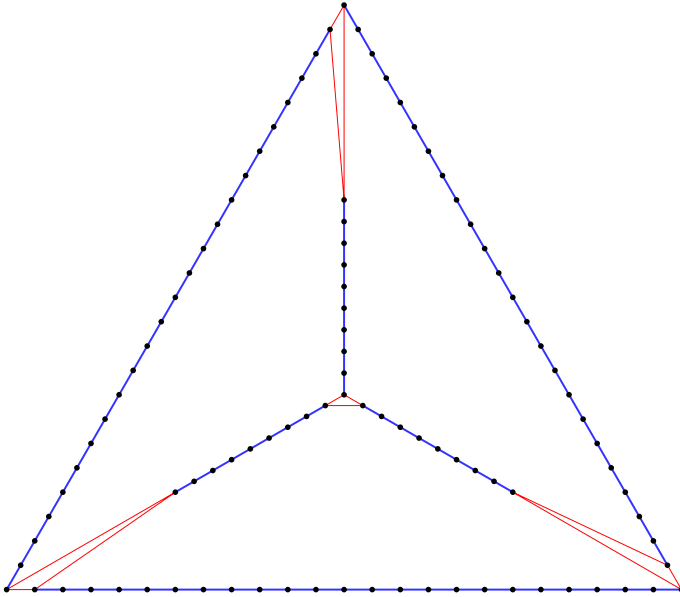
Fig. 9 A feasible solution for the subtour LP of the instance $T'_{24,18}$. The red lines correspond to variables with value $1/2$ while the blue lines correspond to variables with value 1 (color figure online)

For the lower bound we observe that the nearest neighbor of a base vertex has distance 1 while a nearest neighbor of an internal vertex has distance γ. Therefore, the total length of a feasible solution to the subtour LP must be at least

$$3n + 3(n/\sqrt{3} - 10 - \gamma) > 3n + 3n/\sqrt{3} - 33.$$

$\square$

Theorem 5 *For $n \leq 3/2 \cdot m$ the integrality ratio of the instances $T'_{n,m}$ converges to $4/3$ for $n \to \infty$.*

Proof Using Theorems 3 and 4 we conclude that the integrality ratio of $T'_{n,m}$ is at most $\frac{4n+4n/\sqrt{3}-17}{3n+3n/\sqrt{3}-33}$ and at least $\frac{4n+4n/\sqrt{3}-69}{3n+3n/\sqrt{3}}$. Both these values converge to $4/3$ for $n \to \infty$. $\square$

The length of an optimum tour in $T_{n,m}$ is clearly at least as long as an optimum tour in $T'_{n,m}$ and at most by some constant value larger if $n \leq 3/2 \cdot m$. The bounds for a feasible solution to the subtour LP of $T'_{n,m}$ carry over to $T_{n,m}$. Therefore we get:

Corollary 2 *For $n \leq 3/2 \cdot m$ the integrality ratio of the instances $T_{n,m}$ converges to $4/3$ for $n \to \infty$.*

By comparing the rate of convergence obtained in the proof of Theorem 5 with Theorem 13 in [10] one easily sees that for a given number of vertices the integrality

ratio of the instances $T'_{n,m}$ converges much faster to 4/3 than the instances constructed in [10]. Theorem 5 yields the integrality ratio $\frac{4}{3} - \Theta(N^{-1})$ for an N-vertex instance while the instances constructed in [10] have integrality ratio $\frac{4}{3} - \Theta(N^{-2/3})$.

4 Experimental results

All runtime experiments described in this section were performed using `Concorde` version 03.12.19 [2]. This is the fastest known code to solve large TSP instances exactly. The source code of `Concorde` can be downloaded at [2]. We used `gcc 4.8.5` to compile `Concorde` and used `CPLEX 12.6` as the LP solver. All experiments were performed on a 2.20GHz Intel Xeon E5-2699 v4 using a single core for each job. Up to 16 jobs were run in parallel on the machine. We have switched off the turbo boost mode of the machine to reduce the variation in runtimes.

Repeated runs of `Concorde` on the same instance may vary a lot in their runtime if different random seeds are used for `Concorde`. On some instances we observed a runtime difference of more than a factor of 10 between runs with different random seeds. In our experiments we therefore took the average runtime of k runs where k was chosen to be 10 or 100. We used the `Concorde` option `-s #` to set the random seed to i in the ith of the k runs.

4.1 TSPLIB results

The TSPLIB is a very well known collection of TSP instances and can be downloaded at [18]. It contains 111 instances with sizes between 14 and 85,900. For all instances optimum solutions are known [1]. For our experiments we used all TSPLIB instances with at most 2000 vertices except the instance `linhp318.tsp` which contains a fixed edge. Thus, our TSPLIB testbed contains 93 instances with sizes between 16 and 1889. In Fig. 10 we show for each of the 93 instances the minimum, average and maximum runtime taken over 100 runs of `Concorde`.

It can be seen that `Concorde` solves each TSPLIB instance with less than 1000 vertices within a minute. The slowest run took 56,834 s and was on the instance `u1817`.

4.2 Results for points on three parallel lines

In [10] a family of Euclidean TSP instances is proposed that arises from three sets of n equidistant points placed on three parallel lines. It was shown that this family of instances has an integrality ratio converging to 4/3 if the distances of the three lines are chosen appropriately. We denote by $P_{n,d}$ the instance with n points at distance 1 on each of the three parallel lines which have distance d. We have generated these instances for all n with $34 \leq n \leq 64$ and the distances d ranging between 0.1 and 10.0 in increments of 0.1. All point coordinates were scaled by 10,000. Thus in total we generated 3300 instances. For each instance we measured the average runtime of 100 runs of `Concorde`. From these results we chose for each n the instance having the largest average runtime.
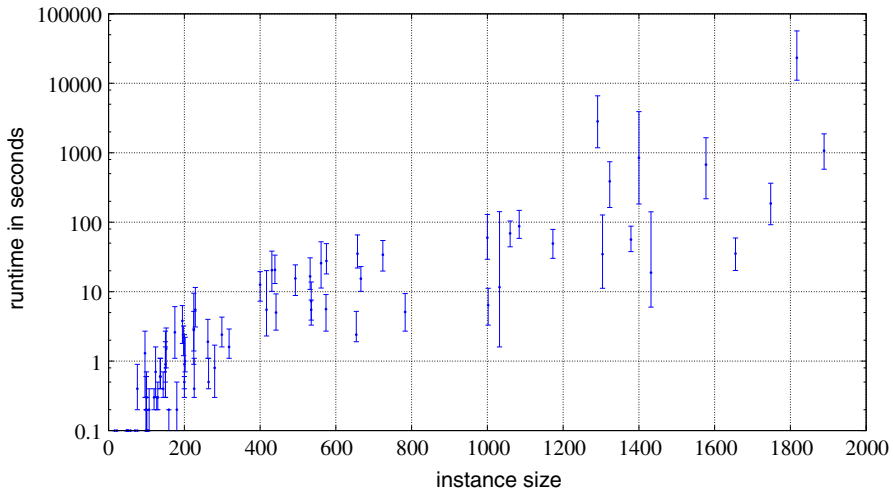
Fig. 10 Runtimes of `Concorde` on 93 TSPLIB instances with sizes between 16 and 1889. The x-axis shows the instance size while the y-axis gives the log scaled runtime. For each instance a vertical line is drawn indicating the minimal and maximal runtime seen over 100 runs using different random seeds. A dot on this line marks the average runtime of these 100 runs

Figure 11 shows the minimum, average and maximum runtime for all these chosen instances. The red line shown in this figure is a least-squares fit of the average runtimes.

4.3 Results for the instances $T_{n,m}$

The instances $T_{n,m}$ have $N := 3(n + m) - 2$ vertices. We created these instances in TSPLIB format by scaling the point coordinates by 10,000 and rounding to the nearest integer. The distance between two points is defined as the rounded Euclidean distance called EUC_2D in the TSPLIB format. All these instances are available for download at [11]. We measured the runtime of `Concorde` on the instances $T_{n,m}$ for all $5 \leq n \leq 33$ and $5 \leq m \leq 33$ to get some idea for which choices of n and m the largest runtimes appear if the number of vertices of the instance is fixed. Our conclusion from that experiment is that for a given number N of vertices with $N \equiv 1 \mod 3$ and $N \geq 50$ the following choices for n and m lead to high runtimes of `Concorde`:

$$n := \lfloor \frac{3N - 40}{10} \rfloor \quad \text{and} \quad m := \frac{N + 2}{3} - n \tag{9}$$

Figure 12 shows the minimum, average and maximum runtime for all instances $T_{n,m}$ with at least 50 and at most 200 vertices and with n and m defined by Eq. (9). The red line shown in this figure is a least-squares fit of the average runtimes. It corresponds to the function
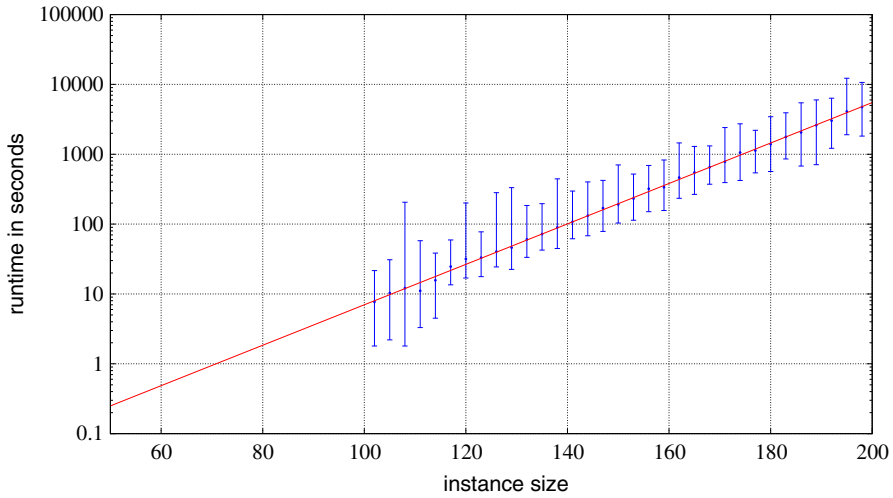
Fig. 11 Runtimes of Concorde on the $P_{n,d}$ instances with at least 100 and at most 200 vertices. The distance d was chosen to maximize the largest average runtime. The x-axis shows the instance size while the y-axis gives the log scaled runtime. For each instance a vertical line is drawn indicating the minimal and maximal runtime seen over 100 runs using different random seeds. A dot on this line marks the average runtime of these 100 runs. The red line is a least-squares fit of the average runtimes (color figure online)
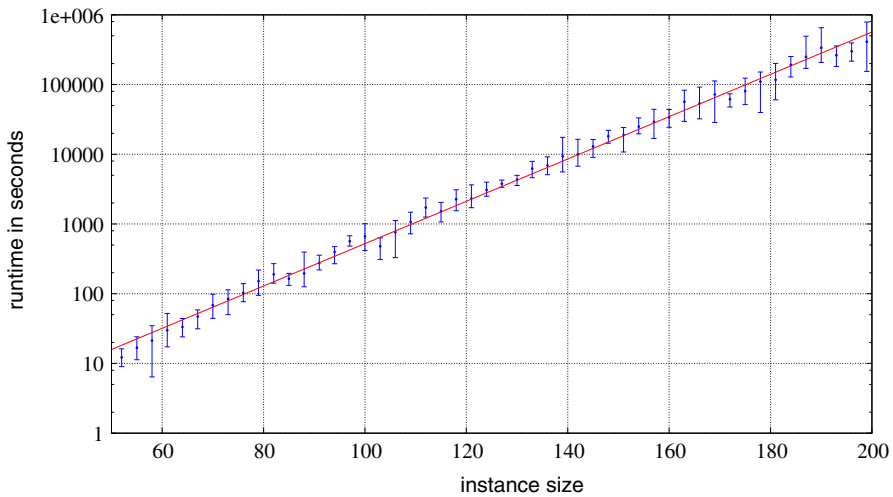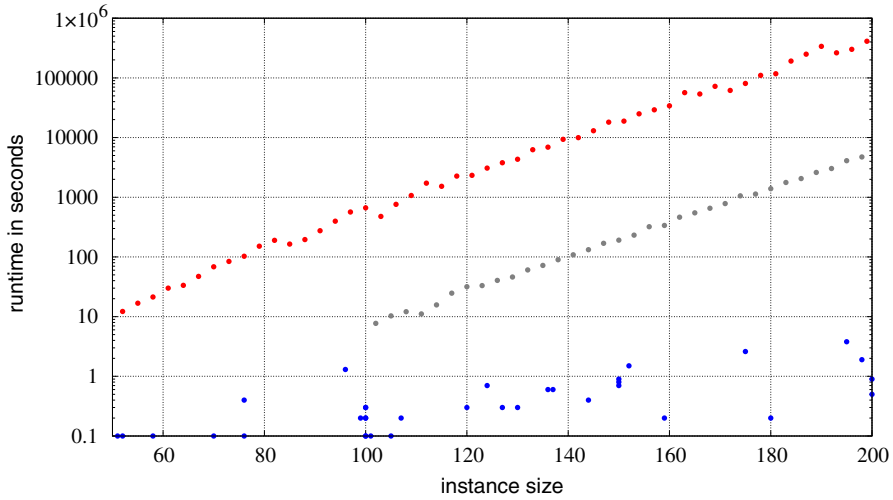

Fig. 12 Runtimes of Concorde on the $T_{n,m}$ instances with at least 52 and at most 200 vertices. The values n and m were selected according to Eq. (9). The x-axis shows the instance size while the y-axis gives the log scaled runtime. For each instance a vertical line is drawn indicating the minimal and maximal runtime seen over 10 runs using different random seeds. A dot on this line marks the average runtime of these 10 runs. The red line is a least-squares fit of the average runtimes (color figure online)

Fig. 13 Comparison of the runtimes of `Concorde` on the $T_{n,m}$ instances (red dots), on the $P_{n,d}$ instances (gray dots), and on the TSPLIB instances (blue dots). The instances have sizes between 51 and at most 200 vertices. The values n and m for the $T_{n,m}$ instances were chosen according to Eq. (9). The x-axis shows the instance size while the y-axis gives the log scaled runtime in seconds. The runtime shown for the $T_{n,m}$ instances is the average taken over 10 independent runs of `Concorde`. The runtime shown for the $P_{n,d}$ instances and the TSPLIB instances is the average taken over 100 independent runs of `Concorde` (color figure online)

$$\text{runtime in seconds} = 0.480 \cdot 1.0724^N \tag{10}$$

Thus, this function estimates the runtime in seconds needed by `Concorde` for the instances $T_{n,m}$ with $N = 3(n + m) - 2$ and n and m defined by Eq. (9). From (10) we get for example the following very rough runtime estimates:

- $n = 60$ and $m = 12$ implies $N = 214$ and runtime estimate 17 days.
- $n = 71$ and $m = 13$ implies $N = 250$ and runtime estimate 216 days.
- $n = 296$ and $m = 38$ implies $N = 1000$ and runtime estimate 3×10^{22} years.

The largest runtime that we have measured for `Concorde` on a TSPLIB instance with at most 1000 vertices was 129.2 s on the instance `dsj1000`. According to the above runtime estimates `Concorde` would need for the 1000 vertex instance $T_{296,38}$ more than 10^{27} times as long.

4.4 Comparison of runtime results

In Fig. 13 we compare the runtimes of all TSPLIB instances with up to 200 vertices with the runtimes for the $T_{n,m}$ instances with n and m chosen according to (9) and the $P_{n,d}$ instances with up to 200 points. As one can see already for quite small instances the runtimes of `Concorde` on the $T_{n,m}$ instances are by several orders of magnitude larger than on the TSPLIB instances. Moreover, `Concorde`'s runtime on the $T_{n,m}$ instances is two orders of magnitude slower than for the $P_{n,d}$ instances. Therefore, the $T_{n,m}$ instances with n and m chosen according to (9) may be useful benchmark

instances for TSP algorithms. All these instances and C++ code for generating them are available for download at [11]. It should be noted that there exist polynomial time algorithms for the $P_{n,d}$ instances and the $T_{n,m}$ instances if their special structure is exploited [16].

References

1. Applegate, D., Bixby, R., Chvátal, V., Cook, W.: The Traveling Salesman Problem. A Computational Study. Princeton University Press, Princeton (2006)
2. Applegate, D., Bixby, R., Chvátal, V., Cook, W.: Concorde-03.12.19. http://www.math.uwaterloo.ca/tsp/concorde/downloads/downloads.htm. Accessed 26 Feb 2020
3. Arora, S.: Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. J. ACM **45**(5), 753–782 (1998)
4. Christofides, N.: Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, Carnegie-Mellon University (1976)
5. Dantzig, G., Fulkerson, R., Johnson, S.: Solution of a large-scale traveling-salesman problem. J. Oper. Res. Soc. Am. **2**(4), 393–410 (1954)
6. Deineko, V.G., van Dal, R., Rote, G.: The convex-hull-and-line traveling salesman problem: a solvable case. Inf. Process. Lett. **51**, 141–148 (1994)
7. Flood, M.M.: The traveling-salesman problem. Oper. Res. **4**(1), 61–75 (1956)
8. Garey, M.R., Graham, R.L., Johnson, D.S.: Some NP-complete geometric problems. In: Proceedings of the Eighth Annual ACM Symposium on Theory of Computing (STOC '76), pp. 10–22 (1976)
9. Grötschel, M., Lovász, L., Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. Combinatorica **1**, 169–197 (1981)
10. Hougardy, S.: On the integrality ratio of the subtour LP for Euclidean TSP. Oper. Res. Lett. **42**(8), 495–499 (2014)
11. Hougardy, S., Zhong, X.: The $T_{n,m}$ instances in TSPLIB format. https://doi.org/10.5281/zenodo.3688446; http://www.or.uni-bonn.de/~hougardy/HardTSPInstances.html. Accessed 26 Feb 2020 (2018)
12. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W. (eds.) Complexity of Computer Computations. Plenum Press, New York (1972)
13. Karpinski, M., Lampis, M., Schmied, R.: New inapproximability bounds for TSP. J. Comput. Syst. Sci. **81**, 1665–1677 (2015)
14. Papadimitriou, C.H.: The euclidean traveling salesman problem is NP-complete. Theor. Comput. Sci. **4**(3), 237–244 (1977)
15. Reinelt, G.: TSPLIB 95. Technical report, Universität Heidelberg. http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/tsp95.pdf. Accessed 26 Feb 2020 (1995)
16. Rubinstein, J.H., Thomas, D.A., Wormald, N.C.: A polynomial algorithm for a constrained traveling salesman problem. Networks **38**, 68–75 (2001)
17. Serdjukov, A.I.: Some extremal bypasses in graphs. Upravlyaemye Sistemy **17**, 76–79 (1978). [**in Russian**]
18. TSPLIB: http://softlib.rice.edu/pub/tsplib/. Accessed 26 Feb 2020

19. Williamson, D.P.: Analysis of the Held–Karp heuristic for the traveling salesman problem. Master's thesis, Massachusetts Institute of Technology (1990)
20. Wolsey, L.A.: Heuristic analysis, linear programming and branch and bound. Math. Program. Study **13**, 121–134 (1980)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.