Electronic Colloquium on Computational Complexity, Report No. 92 (2006)

Worst Case and Probabilistic Analysis of the 2-Opt Algorithm for the TSP*

Matthias Englert Heiko Röglin Berthold Vöcking

Department of Computer Science
RWTH Aachen, Germany
{englert,roeglin,voecking}@cs.rwth-aachen.de

July 5, 2006

Abstract

2-Opt is probably the most basic and widely used local search heuristic for the TSP. This heuristic achieves amazingly good results on "real world" Euclidean instances both with respect to running time and approximation ratio. There are numerous experimental studies on the performance of 2-Opt. However, the theoretical knowledge about this heuristic is still very limited. Not even its worst case running time on Euclidean instances was known so far. In this paper, we clarify this issue by presenting a family of Euclidean instances on which 2-Opt can take an exponential number of steps.

Previous probabilistic analyses were restricted to instances in which n points are placed uniformly at random in the unit square $[0,1]^2$, where it was shown that the expected number of steps is bounded by $\tilde{O}(n^{10})$ for Euclidean instances. We consider a more advanced model of probabilistic instances in which the points can be placed according to general distributions on $[0,1]^2$. In particular, we allow different distributions for different points. We study the expected running time in terms of the number n of points and the maximal density ϕ of the probability distributions. We show an upper bound on the expected length of any 2-Opt improvement path of $\tilde{O}(n^{4+1/3} \cdot \phi^{8/3})$. When starting with an initial tour computed by an insertion heuristic, the upper bound on the expected number of steps improves even to $\tilde{O}(n^{3+5/6} \cdot \phi^{8/3})$. If the distances are measured according to the Manhattan metric, then the expected number of steps is bounded by $\tilde{O}(n^{3.5} \cdot \phi)$. In addition, we prove an upper bound of $O(\sqrt{\phi})$ on the expected approximation factor w. r. t. both of these metrics.

Let us remark that our probabilistic analysis covers as special cases the uniform input model with $\phi = 1$ and a smoothed analysis with Gaussian perturbations of standard deviation σ with $\phi \sim 1/\sigma^2$. Besides random metric instances, we also consider an alternative random input model in which an adversary specifies a graph and distributions for the edge lengths in this graph. In this model, we achieve even better results on the expected running time of 2-Opt.

*The first author is supported by DFG grant WE 2842/1. The second and the third author are supported by DFG grant VO 889/2.

ISSN 1433-8092

Contents

1 Introduction

In the *traveling salesperson problem (TSP)*, we are given a set $\{v_1, v_2, \ldots, v_n\}$ of *vertices* (*cities*) and for each pair $\{v_i, v_j\}$ of distinct vertices a distance $d(v_i, v_j)$. The goal is to find a tour of minimal length visiting each vertex exactly once and returning to the initial vertex at the end, that is, the goal is to compute a permutation π minimizing

$$\sum_{i=1}^{n-1} d(v_{\pi(i)}, v_{\pi(i+1)}) + d(v_{\pi(n)}, v_{\pi(1)}) \enspace .$$

Despite many theoretical analyses and experimental evaluations of the TSP, there is still a considerable gap between the theoretical results and the experimental observations. The Euclidean TSP, for example, is known to be NP-hard in the strong sense [Pap77]. In a breakthrough result, Arora has shown that the Euclidean TSP admits a polynomial time approximation scheme (PTAS) and, hence, can be approximated arbitrarily well in polynomial time [Aro98]. Arora's PTAS is based on dynamic programming. However, the most successful algorithms on practical instances rely on the principle of local search and very little is known about their complexity.

The *2-Opt* algorithm is probably the most basic and widely used local search heuristic for the TSP. 2-Opt starts with an arbitrary initial tour and incrementally improves on this tour by making successive improvements that exchange two of the edges in the tour with two other edges. More precisely, in each *improving step* the 2-Opt algorithm selects two edges $\{u_1, u_2\}$ and $\{v_1, v_2\}$ from the tour such that u_1, u_2, v_1, v_2 are distinct and appear in this order in the tour, and the algorithm replaces these edges by the edges $\{u_1, v_1\}$ and $\{u_2, v_2\}$, provided that this change decreases the length of the tour. The algorithm terminates in a local optimum in which no further improving step is possible. We use the term *2-change* to denote a local improvement made by 2-Opt. This simple heuristic performs amazingly well on "real-life" Euclidean instances like, e. g., the ones in the well-known TSPLIB [Rei91]. Usually the 2-Opt heuristic needs a clearly subquadratic number of improving steps until it reaches a local optimum and the computed solution lies within a few percentage points of the global optimum [JM97].

There are numerous experimental studies on the performance of 2-Opt. However, the theoretical knowledge about this heuristic is still very limited. Let us first discuss the number of local improvement steps made by 2-Opt before it finds a locally optimal solution. When talking about the number of local improvements, it is convenient to consider the *state graph*. The vertices in this graph correspond to the possible tours and an arc from a vertex v to a vertex u is contained if u is obtained from v by performing one improving 2-Opt step. On the positive side, van Leeuwen and Schoone consider a 2-Opt variant for the Euclidean plane in which only steps are allowed that remove a crossing from the tour. Observe that such steps can introduce new crossings. However, van Leeuwen and Schoone show that after $O(n^3)$ steps, 2-Opt has found a tour without any crossing [vLS80]. On the negative side, Lueker constructs TSP instances whose state graphs contain exponentially long paths, that is, 2-Opt can take an exponential number of steps before it finds a locally optimal solution [Lue75]. His result is generalized to k-Opt for arbitrary $k \geq 2$ by Chandra, Karloff, and Tovey [CKT99]. These results, however, use arbitrary graphs whose edge lengths do not satisfy the triangle inequality. Hence they leave open the question about the worst case complexity of 2-Opt on metric TSP instances. In particular, Chandra, Karloff, and Tovey ask whether it is possible to construct Euclidean TSP instances on which 2-Opt can take an exponential number of steps. In this paper, we settle this question. We construct Euclidean instances whose state graphs contain exponentially longs paths, that is, Euclidean instances on which 2-Opt can take an exponential number of steps before finding a locally optimal solution. In chip design applications, often TSP instances arise in which the distances are measured according to the Manhattan metric. Also for this metric, we construct instances with exponentially long paths in the 2-Opt state graph.

Theorem 1.

a) *For every $n \in \mathbb{N}$, there is a graph in the Euclidean plane with $8n$ vertices whose corresponding state graph contains a path of length $2^{n+3} - 14$.*

b) *For every $n \in \mathbb{N}$, there is a graph in the plane with Manhattan metric with $16n$ vertices whose corresponding state graph contains a path of length $2^{n+4} - 22$.*

For Euclidean instances in which n points are placed uniformly at random in the unit square, Kern shows that the length of the longest path in the state graph is bounded by $O(n^{16})$ with probability $1 - c/n$ for some constant c [Ker89]. Chandra, Karloff, and Tovey improve this result by bounding the expected length of the longest path in the state graph by $O(n^{10} \log n)$ [CKT99]. That is, independent of the initial tour and the choice of the local improvements, the expected running time of 2-Opt is bounded by $O(n^{10} \log n)$. For instances in which n points are placed uniformly at random in the unit square and the distances are measured according to the Manhattan metric, Chandra, Karloff, and Tovey show that the expected length of the longest path in the state graph is bounded by $O(n^6 \log n)$.

We consider a more general probabilistic input model and improve the previously known bounds. The probabilistic model underlying our analysis allows that different vertices are placed in the plane using different continuous probability distributions. The distribution of vertex v_i is defined by a density function $f_i : [0,1]^2 \to [0, \phi]$ for some given $\phi \geq 1$. Our upper bounds depend on the number n of vertices and the upper bound ϕ on the density. We denote instances created by this input model as *ϕ-perturbed Euclidean* or *Manhattan instances* depending on the underlying metric. The parameter ϕ can be seen as a parameter specifying how close the analysis is to a worst case analysis since the larger ϕ is, the better worst case instances can be approximated by the distributions. For $\phi = 1$, every point has a uniform distribution over the unit square and hence the input model equals the uniform model analyzed before. Our results narrow the gap between the subquadratic number of improving steps observed in experiments [JM97] and the upper bounds from the probabilistic analysis. With slight modifications, this model also covers a smoothed analysis, in which first an adversary specifies the positions of the points and after that each position is slightly perturbed by adding a Gaussian random variable with small standard deviation σ. In this case, one has to set $\phi \sim 1/\sigma^2$.

We also consider a model in which an arbitrary graph $G = (V, E)$ is given and for each edge $e \in E$, a probability distribution according to which the edge length $d(e)$ is chosen independently of the other edge lengths. Again, we restrict the choice of distributions to distributions which can be specified by density functions $f : [0,1] \to [0, \phi]$ with maximal density at most ϕ for a given $\phi \geq 1$. We denote inputs created by this input model as *ϕ-perturbed graphs*. Observe that in this input model only the distances are perturbed whereas the graph structure is not touched by the randomization. This can be useful if one wants to explicitely prohibit certain edges. However, if the graph G is not complete, one has to initialize 2-Opt with a Hamiltonian cycle to start with.

We prove the following theorem about the expected length of the longest path in the 2-Opt state graph for the three probabilistic input models discussed above.

Theorem 2. *The expected length of the longest path in the 2-Opt state graph*

a) is $O(n^4 \cdot \phi)$ for ϕ-perturbed Manhattan instances with n points.

b) is $O(n^{4+1/3} \cdot \log(n\phi) \cdot \phi^{8/3})$ for ϕ-perturbed Euclidean instances with n points.

c) is $O(m \cdot n^{1+o(1)} \cdot \phi)$ for ϕ-perturbed graphs with n vertices and m edges.

The experimental study in [JM97] suggests to use an insertion heuristic to compute the initial tour to improve both the approximation ratio and running time. We show that also from a theoretical point of view, using such an insertion heuristic yields a significant improvement since the initial tour 2-Opt starts with is much shorter than the longest possible tour. In the following theorem, we summarize our results on the expected number of local improvements.

Theorem 3. *The expected number of steps performed by 2-Opt*

a) is $O(n^{3.5} \cdot \log n \cdot \phi)$ on ϕ-perturbed Manhattan instances with n points when one starts with a tour obtained by an arbitrary insertion heuristic.

b) is $O(n^{3+5/6} \cdot \log^2(n\phi) \cdot \phi^{8/3})$ on ϕ-perturbed Euclidean instances with n points when one starts with a tour obtained by an arbitrary insertion heuristic.

c) is $O(m \cdot n^{1+o(1)} \cdot \phi)$ on ϕ-perturbed graphs with n vertices and m edges.

In fact, our analysis shows not only that the expected running time is polynomially bounded but it also shows that the second moment and hence the variance is bounded polynomially for ϕ-perturbed Manhattan and graph instances. For the Euclidean metric, we cannot bound the variance but the 3/2-th moment polynomially.

Similar to the running time, the good approximation ratios obtained by 2-Opt on practical instances cannot be explained by a worst-case analysis. In fact, there are quite negative results on the worst-case behavior of 2-Opt. For example, Chandra, Karloff, and Tovey show that there are Euclidean instances for which 2-Opt has local optima whose costs are $\Omega\left(\frac{\log n}{\log \log n}\right)$ times larger than the optimal costs [CKT99]. However, the same authors also show that the expected approximation ratio for instances with n points drawn uniformly at random from the unit square is bounded from above by a constant. We generalize their result to our input model in which different points can have different distributions with bounded density ϕ. For both Euclidean and Manhattan instances, we obtain the following theorem.

Theorem 4. *For ϕ-perturbed Manhattan and Euclidean instances, the expected approximation ratio of the worst tour that is locally optimal for 2-Opt is bounded by $O(\sqrt{\phi})$.*

The remainder of the paper is organized as follows. We start by stating some basic definitions and notations in Section 2. In Section 3, we present the lower bounds. In Section 4, we analyze the expected running time and prove Theorems 2 and 3. Finally, in Sections 5 and 6, we prove Theorem 4 about the expected approximation factor and we present some extensions of our analysis.

2 Preliminaries

We begin by stating some basic definitions and notations. First of all, an instance of the TSP consists of a set $V = \{v_1, \dots, v_n\}$ of *vertices* (depending on the context, synonymously referred to as *points*) and a symmetric *distance function* $d : V \times V \to \mathbb{R}_{\geq 0}$ which associates with each pair $\{v_i, v_j\}$ of distinct vertices a distance $d(v_i, v_j) = d(v_j, v_i)$. The goal is to find a tour of minimal length visiting each vertex exactly once and returning to the initial vertex at the end.

A pair (V, d) of a nonempty set V and a function $d : V \times V \to \mathbb{R}_{\geq 0}$ is called a *metric space* if for all $x, y, z \in V$ the following properties are satisfied:

(a) $d(x, y) = 0$ if and only if $x = y$ *(reflexivity)*,

(b) $d(x, y) = d(y, x)$ *(symmetry)*,

(c) $d(x, z) \leq d(x, y) + d(y, z)$ *(triangle inequality)*.

If (V, d) is a metric space, then d is called a *metric on* V. A TSP instance with vertices V and distance function d is called *metric TSP instance* if (V, d) is a metric space.

A well-known class of metrics on $\mathbb{R}^2$ is the class of L_p *metrics*. For $p \geq 1$, the distance $d_{L_p}(P_1, P_2)$ of two points $P_1 = (x_1, y_1) \in \mathbb{R}^2$ and $P_2 = (x_2, y_2) \in \mathbb{R}^2$ with respect to the L_p metric is given by $d_{L_p}(P_1, P_2) = \sqrt[p]{|x_1 - x_2|^p + |y_1 - y_2|^p}$. The L_1 metric is often called *Manhattan metric*, and the L_2 metric is well-known as *Euclidean metric*. A TSP instance (V, d) with $V \subseteq \mathbb{R}^2$ in which d equals d_{L_p} restricted to V is called an L_p *instance*. For $p = 2$, we also use the term *Euclidean instance*.

A *tour construction heuristic* for the TSP incrementally constructs a tour and stops as soon as a valid tour is created. Usually, a tour constructed by such a heuristic is used as the initial solution 2-Opt starts with. The most successful tour construction heuristics for metric TSP instances are so-called *insertion heuristics*. These heuristics insert the vertices into the tour one after another, and every vertex is inserted between two consecutive vertices in the current tour where it fits best. To make this more precise, let T_i denote a subtour on a subset S_i of i vertices, and suppose $v \notin S_i$ is the next vertex to be inserted. If (x, y) denotes an edge in T_i that minimizes $d(x, v) + d(v, y) - d(x, y)$, then the new tour T_{i+1} is obtained from T_i by deleting the edge (x, y) and adding the edges (x, v) and (v, y). Depending on the order in which the vertices are inserted into the tour, one distinguishes between several different insertion heuristics. Rosenkrantz et al. show an upper bound of $\lceil \log n \rceil + 1$ on the approximation factor of any insertion heuristic on metric TSP instances [RSI77]. Furthermore, they show that two variants which they call *nearest insertion* and *cheapest insertion* achieve an approximation ratio of 2 for metric TSP instances. The nearest insertion heuristic always inserts the vertex with the smallest distance to the current tour, and the cheapest insertion heuristic always inserts the vertex whose insertion leads to the cheapest tour T_{i+1}.

3 L_1 and L_2 Instances with Exponentially Long Sequences of Improving 2-Changes

In 1975 Lueker constructed a family of TSP instances for which there exist exponentially long sequences of improving 2-changes [Lue75]. Later his construction was generalized to k-changes with arbitrary $k \geq 2$ by Chandra, Karloff, and Tovey [CKT99]. The instances in these constructions are general TSP instances that do not satisfy the triangle inequality, and hence they leave open the question whether there exist metric TSP instances with exponentially long sequences of improving 2-changes. In particular, Chandra, Karloff, and Tovey ask whether it is possible to construct such instances in the Euclidean plane. We answer this question affirmatively by presenting families of L_1 and L_2 instances with exponentially long sequences of improving 2-changes.

In Lueker's construction many of the 2-changes remove two edges which are far apart in the current tour in the sense that many vertices are visited between them, no matter of how the direction of the tour is chosen. This is crucial to the construction and also to its generalization to k-changes. Our construction differs significantly from the previous ones as the 2-changes affect the tour only locally. The instances we construct are composed of gadgets of constant size. Each of these gadgets has a *zero state* and a *one state*, and there exists a sequence of improving 2-changes starting in the zero state and eventually leading to the one state. Let $G_0, \dots, G_{n-1}$ denote these gadgets. If gadget G_i with $i > 0$ has reached state one, then it can be reset to its zero state by gadget G_{i-1}. The crucial property of our construction is that whenever a gadget G_{i-1} changes its state from zero to one, it resets gadget G_i twice. Hence, if in the initial tour, gadget G_0 is in its zero state and every other gadget is in state one, then for every i with $0 \leq i \leq n-1$, gadget G_i performs 2^i state changes from zero to one as for $i > 0$, gadget G_i is reset 2^i times.

In Section 3.1, we describe the structure of the instances that we construct for the Euclidean plane and we present the exponentially long sequence of 2-changes. Each of these 2-changes imposes an inequality on the distances of the vertices as it has to decrease the length of the tour strictly. In Section 3.2, we construct a set of points in the plane whose L_2 distances satisfy all of these inequalities. In Section 3.3 we describe how one can adapt the structure of the instance for the L_1 metric, and in Section 3.4 we give points in the plane whose L_1 distances satisfy the necessary inequalities.

3.1 Exponentially Long Sequences of 2-Changes

Every gadget is composed of 2 subgadgets which we refer to as *blocks*. Each of these blocks consists of 4 vertices that are consecutively visited in the tour. Let $\mathcal{B}_1^i$ and $\mathcal{B}_2^i$ denote the blocks of gadget G_i and let A_j^i, B_j^i, C_j^i, and D_j^i denote the four points the j-th block of the i-th gadget consist of. If one ignores certain intermediate configurations that arise when one gadgets resets another one, our construction ensures the following properties: The points are always consecutive in the tour, the edge between B_j^i and C_j^i is contained in every tour, and B_j^i and C_j^i are always the inner points of the block. That is, if one excludes the intermediate configurations, only the configurations $A_j^i B_j^i C_j^i D_j^i$ and $A_j^i C_j^i B_j^i D_j^i$ occur during the sequence of 2-changes. Observe that the change from one of these configurations to the other corresponds to a single 2-change in which the edges $A_j^i B_j^i$ and $C_j^i D_j^i$ are replaced by the edges $A_j^i C_j^i$ and $B_j^i D_j^i$, or vice versa. In the following, we assume that the sum $d(A_j^i, B_j^i) + d(C_j^i, D_j^i)$ of the distances between A_j^i and B_j^i and between C_j^i and D_j^i is strictly smaller than the sum $d(A_j^i, C_j^i) + d(B_j^i, D_j^i)$, and we refer to the configuration $A_j^i B_j^i C_j^i D_j^i$ as the *short state* of the block and to the configuration $A_j^i C_j^i B_j^i D_j^i$ as the *long state*. Another property of our construction is that neither the order in which the blocks are visited nor the order of the gadgets is changed during the sequence of 2-changes. Again with the exception of the intermediate configurations, the order in which the blocks are visited is $\mathcal{B}_1^0 \mathcal{B}_2^0 \mathcal{B}_1^1 \mathcal{B}_2^1 \dots \mathcal{B}_1^{n-1} \mathcal{B}_2^{n-1}$. See Figure 1 for an illustration.

Due to the aforementioned properties, we can describe every non-intermediate tour that occurs during the sequence of 2-changes completely by specifying for every block if it is in its short state or in its long state. In the following, we denote the state of a gadget G_i by a vector (x_1, x_2) with $x_j \in \{S, L\}$, meaning that block $\mathcal{B}_j^i$ is in its short state if and only if $x_j = S$. Since every gadget consists of two blocks, there are four possible states for each gadget. However, only three of them appear in the sequence of 2-changes, namely (L, L), (S, L), and (S, S). We call state (L, L) the *zero state* and state (S, S) the *one state*. In order to guarantee the existence of an exponentially long sequence of 2-changes, the gadgets we construct possess the following properties:

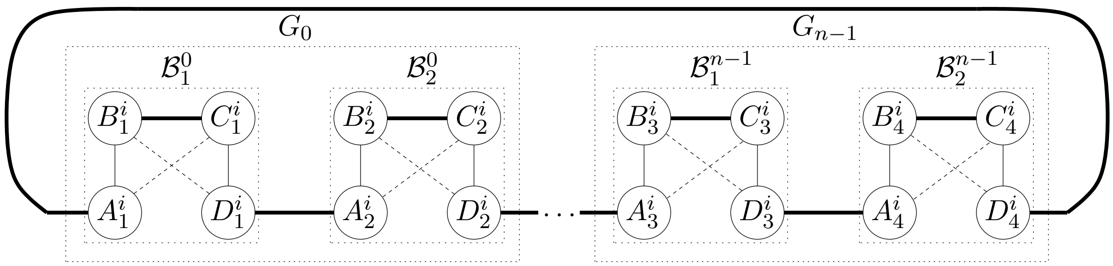

Figure 1: Every tour that occurs in the sequence of 2-changes contains the thick edges. For each block either both solid or both dotted edges are contained. In the former case the block is in its short state; in the latter case the block is in its long state.

1. If gadget G_i with $0 \le i \le n-2$ is in state (L, L) and gadget G_{i+1} is in state (S, S), then there exists a sequence of 7 consecutive 2-changes involving only edges of and between the gadgets G_i and G_{i+1} terminating with gadget G_i being in state (S, L) and gadget G_{i+1} in state (L, L).
2. If gadget G_i with $0 \le i \le n-2$ is in state (S, L) and gadget G_{i+1} is in state (S, S), then there exists a sequence of 7 consecutive 2-changes involving only edges of and between the gadgets G_i and G_{i+1} terminating with gadget G_i being in state (S, S) and gadget G_{i+1} in state (L, L).

If these properties are satisfied and if in the initial tour gadget G_0 is in its zero state (L, L) and every other gadget is in its one state (S, S), then there exists an exponentially long sequence of 2-changes in which gadget G_i changes 2^i times from state zero to state one. In order to see this, we prove the following lemma.

Lemma 5. *If gadget G_i with $0 \le i \le n-1$ is in the zero state (L, L) and all gadgets G_j with $j > i$ are in the one state (S, S), then there exists a sequence of $2^{n+3-i} - 14$ 2-changes in which only edges of and between the gadgets G_j with $j \ge i$ are involved and that terminates in a state in which all gadgets G_j with $j \ge i$ are in the one state.*

Proof. We prove the lemma by induction on i. If gadget G_{n-1} is in state (L, L), then it can change its state with two 2-changes to (S, S) without affecting the other gadgets. Hence, the lemma is true for $i = n-1$. Now assume that the lemma is true for $i+1$ and consider a state in which gadget G_i is in state (L, L) and all gadgets G_j with $j > i$ are in state (S, S). Due to the first property, there exists a sequence of 7 consecutive 2-changes in which only edges of and between G_i and G_{i+1} are involved terminating with G_i being in state (S, L) and G_{i+1} being in state (L, L). By the induction hypothesis there exists a sequence of $2^{n+2-i} - 14$ 2-changes after which all gadgets G_j with $j > i$ are in state (S, S). Then, due to the second property, there exists a sequence of 7 consecutive 2-changes in which only G_i changes its state from (S, L) to (S, S) while resetting gadget G_{i+1} again from (S, S) to (L, L). Hence, we can apply the induction hypothesis again, yielding that after another $2^{n+2-i} - 14$ 2-changes all gadgets G_j with $j \ge i$ are in state (S, S). This concludes the proof as the number of 2-changes performed is $14 + 2(2^{n+2-i} - 14) = 2^{n+3-i} - 14$. □

We still need to show how the aforementioned properties are achieved, that is, how the sequence of 2-changes exactly looks like. We first present a sequence of 2-changes that satisfies the second property. Observe that the initial situation in the second property is as follows: There are three consecutive blocks, namely $\mathcal{B}_2^i$, $\mathcal{B}_1^{i+1}$, and $\mathcal{B}_2^{i+1}$, the leftmost one is in its long state, and the other blocks are in their short states. We need to present a sequence of 2-changes in which only edges of and between these three blocks are involved and after which the first block is in its short state and the other blocks are in their long states. Remember that when the edges $\{u_1, u_2\}$ and $\{v_1, v_2\}$ are removed from the tour and the vertices appear in the order u_1, u_2, v_1, v_2 in the current tour, then the edges $\{u_1, v_1\}$ and $\{u_2, v_2\}$ are added to the tour and the subtour between u_1 and v_2 is visited in reverse order. If, e. g., the current tour corresponds to the permutation $(1, 2, 3, 4, 5, 6, 7)$ and the edges $\{1, 2\}$ and $\{5, 6\}$ are removed, then the new tour is $(1, 5, 4, 3, 2, 6, 7)$. The following sequence of seven 2-changes satisfies the second property; brackets indicate the edges that are removed from the tour.

$$
\begin{array}{rcccc|cccc|cccc}
1) & [A_2^i & C_2^i] & B_2^i & D_2^i & A_1^{i+1} & B_1^{i+1} & C_1^{i+1} & D_1^{i+1} & A_2^{i+1} & B_2^{i+1} & [C_2^{i+1} & D_2^{i+1}] \\
2) & A_2^i & C_2^{i+1} & [B_2^{i+1} & A_2^{i+1}] & D_1^{i+1} & C_1^{i+1} & B_1^{i+1} & A_1^{i+1} & [D_2^i & B_2^i] & C_2^i & D_2^{i+1} \\
3) & A_2^i & C_2^{i+1} & [B_2^{i+1} & D_2^i] & A_1^{i+1} & B_1^{i+1} & [C_1^{i+1} & D_1^{i+1}] & A_2^{i+1} & B_2^i & C_2^i & D_2^{i+1} \\
4) & A_2^i & C_2^{i+1} & B_2^{i+1} & C_1^{i+1} & [B_1^{i+1} & A_1^{i+1}] & D_2^i & D_1^{i+1} & A_2^{i+1} & B_2^i & [C_2^i & D_2^{i+1}] \\
5) & [A_2^i & C_2^{i+1}] & B_2^{i+1} & C_1^{i+1} & B_1^{i+1} & C_2^i & [B_2^i & A_2^{i+1}] & D_1^{i+1} & D_2^i & A_1^{i+1} & D_2^{i+1} \\
6) & A_2^i & B_2^i & C_2^i & B_1^{i+1} & [C_1^{i+1} & B_2^{i+1}] & C_2^{i+1} & A_2^{i+1} & D_1^{i+1} & D_2^i & [A_1^{i+1} & D_2^{i+1}] \\
7) & A_2^i & B_2^i & [C_2^i & B_1^{i+1}] & C_1^{i+1} & A_1^{i+1} & [D_2^i & D_1^{i+1}] & A_2^{i+1} & C_2^{i+1} & B_2^{i+1} & D_2^{i+1} \\
 & A_2^i & B_2^i & C_2^i & D_2^i & A_1^{i+1} & C_1^{i+1} & B_1^{i+1} & D_1^{i+1} & A_2^{i+1} & C_2^{i+1} & B_2^{i+1} & D_2^{i+1}
\end{array}
$$

A sequence of steps that satisfies the first property can be constructed analogously. Additionally, one has to take into account that the three involved blocks $\mathcal{B}_1^i$, $\mathcal{B}_1^{i+1}$, and $\mathcal{B}_2^{i+1}$ are not consecutive in the tour but that block $\mathcal{B}_2^i$ lies between them. However, one can easily verify that this block is not affected by the sequence of 2-changes, that is, after the seven 2-changes have been performed, the block is in the same state and at the same position as before.

3.2 Embedding the Construction into the Euclidean Plane

The only missing step in the proof of Theorem 1 a) is to find points in the Euclidean plane such that all of the 2-changes that we described in the previous section are improving. We specify the positions of the points of gadget G_{n-1} and give a rule how the points of gadget G_i can be derived when all points of the gadgets G_j with $j > i$ have already been placed. In our construction it happens that different points have exactly the same coordinates. This is only for ease of notation; if one wants to obtain a TSP instance in which distinct points have distinct coordinates, one can slightly move these points without affecting the property that all 2-changes are improving. We choose the following coordinates for the points in both blocks $\mathcal{B}_1^{n-1}$ and $\mathcal{B}_2^{n-1}$ of gadget G_{n-1}.

	A_j^{n-1}	B_j^{n-1}	C_j^{n-1}	D_j^{n-1}
x-coordinate	0	1	-0.1	-1.1
y-coordinate	0	0	1.4	4.8

First we show that $A_j^{n-1}B_j^{n-1}C_j^{n-1}D_j^{n-1}$ is the short state and that $A_j^{n-1}C_j^{n-1}B_j^{n-1}D_j^{n-1}$ is the long state. This follows since

$$4.54\ldots = d(A_j^{n-1}, B_j^{n-1}) + d(C_j^{n-1}, D_j^{n-1}) < d(A_j^{n-1}, C_j^{n-1}) + d(B_j^{n-1}, D_j^{n-1}) = 6.64\ldots \;.$$

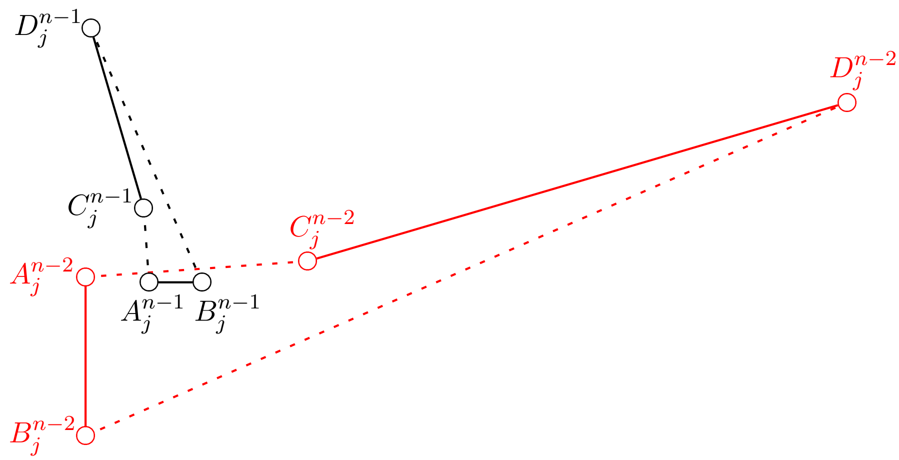

Figure 2: This illustration shows the points of the gadgets G_{n-1} and G_{n-2}. One can see that G_{n-2} is a copy of G_{n-1} rotated by $-\pi/2$, scaled by 3, and translated by $(-1.2, 0.1)$.

In order to describe how the points of gadget G_i are chosen when the points of the gadgets G_j with $j > i$ are already determined, we use the fact that if one has a set of points in the Euclidean plane that admit certain improving 2-changes, then these 2-changes are still improving if one scales, rotates, and translates all points in the same fashion. In fact, our construction has the property that each gadget is just a scaled, rotated, and translated copy of gadget G_{n-1}. Therefore, one can assume without loss of generality that after appropriate scaling, rotating, and translating the points of gadget G_{i+1} have exactly the coordinates which we specified above for gadget G_{n-1}. Under this assumption, we place the points of gadget G_i as follows (see Figure 2):

1. Start with the coordinates that we specified for the points of gadget G_{n-1}.
2. Rotate these points around the origin by $3\pi/2$.
3. Scale each coordinate with a factor of 3.
4. Translate the points by the vector $(-1.2, 0.1)$.

Hence, if the points of the gadgets G_j with $j > i$ have been scaled, rotated, and translated such that the points of G_{i+1} are at those positions which we specified for gadget G_{n-1}, we obtain the following coordinates for the points of gadget G_i. Observe that due to the construction, the points in the two blocks $\mathcal{B}_1^i$ and $\mathcal{B}_2^i$ again have the same coordinates.

	A_j^i	B_j^i	C_j^i	D_j^i
x-coordinate	-1.2	-1.2	3	13.2
y-coordinate	0.1	-2.9	0.4	3.4

Now it only remains to show that the sequences of 2-changes which reset gadget G_{i+1} from (S, S) to (L, L) consist of improving 2-changes only. There are two such sequences; in the first one, gadget G_i changes its state from (L, L) to (S, L), in the second one, gadget G_i changes its state from (S, L) to (S, S). Since the coordinates in both blocks of gadget G_i are the same, the inequalities for both sequences are also the same. Hence, it suffices to consider only the first sequence.

$$
\begin{array}{rcl}
7.75\ldots = d(A_1^i, C_1^i) + d(C_2^{i+1}, D_2^{i+1}) & > & d(A_1^i, C_2^{i+1}) + d(C_1^i, D_2^{i+1}) = 7.71\ldots \\
16.71\ldots = d(B_2^{i+1}, A_2^{i+1}) + d(D_1^i, B_1^i) & > & d(B_2^{i+1}, D_1^i) + d(A_2^{i+1}, B_1^i) = 15.80\ldots \\
16.20\ldots = d(B_2^{i+1}, D_1^i) + d(C_1^{i+1}, D_1^{i+1}) & > & d(B_2^{i+1}, C_1^{i+1}) + d(D_1^i, D_1^{i+1}) = 16.14\ldots \\
7.01\ldots = d(B_1^{i+1}, A_1^{i+1}) + d(C_1^i, D_2^{i+1}) & > & d(B_1^{i+1}, C_1^i) + d(A_1^{i+1}, D_2^{i+1}) = 6.96\ldots \\
4.84\ldots = d(A_1^i, C_2^{i+1}) + d(B_1^i, A_2^{i+1}) & > & d(A_1^i, B_1^i) + d(C_2^{i+1}, A_2^{i+1}) = 4.40\ldots \\
6.70\ldots = d(C_1^{i+1}, B_2^{i+1}) + d(A_1^{i+1}, D_2^{i+1}) & > & d(C_1^{i+1}, A_1^{i+1}) + d(B_2^{i+1}, D_2^{i+1}) = 6.64\ldots \\
16.40\ldots = d(C_1^i, B_1^{i+1}) + d(D_1^i, D_1^{i+1}) & > & d(C_1^i, D_1^i) + d(B_1^{i+1}, D_1^{i+1}) = 15.87\ldots
\end{array}
$$

This concludes the proof of Theorem 1 a) as it shows that all 2-changes in Lemma 5 are improving.

3.3 Modifying the Construction for the L_1 Metric

One important difference between the L_1 and the L_2 metric is that the L_1 metric is not invariant under rotations of the coordinate system. It is, however, invariant under rotations of multiples of $\pi/2$. Hence, one might think that points in the plane whose L_1 distances satisfy all inequalities can be constructed like the points in the Euclidean plane from the previous section. However, we were not able to find such a construction for the plane with L_1 metric, and, in fact, we believe that such a construction does not exist for the L_1 metric.

Hence, we have to modify the construction of the gadgets and the sequence of 2-changes that we described in Section 3.1. Our construction for the L_1 metric is based on the construction for the Euclidean plane, but it does not possess the property that every gadget resets its neighboring gadget twice. Instead this property is only true for half of the gadgets. To be more precise, we construct two different types of gadgets which we call *reset gadgets* and *propagation gadgets*. Reset gadgets perform the same sequence of 2-changes as the gadgets that we constructed for the Euclidean plane. Propagation gadgets also have the same structure as the gadgets for the Euclidean plane, but when such a gadget changes its state from (L, L) to (S, S), it resets its neighboring gadget only once. Due to this relaxed requirements it is possible to find points in the plane with L_1 metric whose distances satisfy all necessary inequalities. Instead of n gadgets our constructions consists of $2n$ gadgets, namely n propagation gadgets G_i^P and n reset gadgets G_i^R. The order in which these gadgets appear in the tour is $G_0^P G_0^R G_1^P G_1^R \ldots G_{n-1}^P G_{n-1}^R$.

As before, every gadget consists of two blocks and the order in which the blocks and the gadgets are visited does not change during the sequence of 2-changes. Consider a reset gadget G_i^R and its neighboring propagation gadget G_{i+1}^P. Then the two properties on which our previous construction is based are still satisfied, that is, if G_i^R is in state (L, L) (respectively (S, L)) and G_{i+1}^P is in state (S, S), then there exists a sequence of 7 consecutive 2-changes resetting gadget G_{i+1}^P to state (L, L) and leaving gadget G_i^R in state (S, L) (respectively (S, S)). The situation is different for a propagation gadget G_i^P and its neighboring reset gadget G_i^R. In this case, if G_i^P is in state (L, L), it first changes its state with a single 2-change to (S, L). After that, gadget G_i^P changes its state to (S, S) while resetting gadget G_i^R from state (S, S) to state (L, L) by a sequence of 7 consecutive 2-changes. In both cases, the sequences of 2-changes in which one block changes from its long to its short state while resetting two blocks from their short to their long states are chosen analogously to our previous construction for the Euclidean plane.

In the initial tour, only gadget G_0^P is in state (L, L) and every other gadget is in state (S, S). With similar arguments as for the Euclidean plane one can show that gadget G_i^R is reset from its one state (L, L) to its zero state (S, S) 2^i times and that the total number of steps is $2^{n+4} - 22$.

3.4 Embedding the Construction into the Plane with L_1 Metric

Similar to the construction in the Euclidean plane, the points in both blocks of a reset gadget G_i^R have the same coordinates. Also in this case one can slightly move all the points without affecting the inequalities if one wants distinct coordinates for distinct points. Again, we first choose points for the gadgets G_{n-1}^P and G_{n-1}^R and later describe how the points of the gadgets G_i^P and G_i^R can be chosen when all points of gadgets G_j^P and G_j^R with $j > i$ are already chosen. We choose the following points for G_{n-1}^R.

	$A_{R,j}^{n-1}$	$B_{R,j}^{n-1}$	$C_{R,j}^{n-1}$	$D_{R,j}^{n-1}$
x-coordinate	0	0	-0.7	-1.2
y-coordinate	1	0	0.1	0.08

Furthermore, we choose the following points for G_{n-1}^P.

	$A_{P,1}^{n-1}$	$B_{P,1}^{n-1}$	$C_{P,1}^{n-1}$	$D_{P,1}^{n-1}$	$A_{P,2}^{n-1}$	$B_{P,2}^{n-1}$	$C_{P,2}^{n-1}$	$D_{P,2}^{n-1}$
x	-2	-3.3	-1.3	1.5	-0.7	-1.5	1.9	-0.8
y	1.8	2.8	1.4	0.9	1.6	1.2	-1.5	-1.1

Before we describe how the points of the other gadgets are chosen, we first show that the 2-changes within and between the gadgets G_{n-1}^P and G_{n-1}^R are improving. Therefore, consider first the 2-change in which G_{n-1}^P changes its state from (L, L) to (S, L). This 2-change is improving since

$$5.6 = d(A_{P,1}^{n-1}, B_{P,1}^{n-1}) + d(C_{P,1}^{n-1}, D_{P,1}^{n-1}) < d(A_{P,1}^{n-1}, C_{P,1}^{n-1}) + d(B_{P,1}^{n-1}, D_{P,1}^{n-1}) = 7.8 \ .$$

Next we show that the sequence of 2-changes in which G_{n-1}^P changes its state from (S, L) to (S, S) while resetting G_{n-1}^R consists of improving steps only.

$$\begin{array}{rcl}
6.22 = d(A_{P,2}^{n-1}, C_{P,2}^{n-1}) + d(C_{R,2}^{n-1}, D_{R,2}^{n-1}) & > & d(A_{P,2}^{n-1}, C_{R,2}^{n-1}) + d(C_{P,2}^{n-1}, D_{R,2}^{n-1}) = 6.18 \\
4.0 = d(B_{R,2}^{n-1}, A_{R,2}^{n-1}) + d(D_{P,2}^{n-1}, B_{P,2}^{n-1}) & > & d(B_{R,2}^{n-1}, D_{P,2}^{n-1}) + d(A_{R,2}^{n-1}, B_{P,2}^{n-1}) = 3.6 \\
2.42 = d(B_{R,2}^{n-1}, D_{P,2}^{n-1}) + d(C_{R,1}^{n-1}, D_{R,1}^{n-1}) & > & d(B_{R,2}^{n-1}, C_{R,1}^{n-1}) + d(D_{P,2}^{n-1}, D_{R,1}^{n-1}) = 2.38 \\
5.68 = d(B_{R,1}^{n-1}, A_{R,1}^{n-1}) + d(C_{P,2}^{n-1}, D_{R,2}^{n-1}) & > & d(B_{R,1}^{n-1}, C_{P,2}^{n-1}) + d(A_{R,1}^{n-1}, D_{R,2}^{n-1}) = 5.52 \\
3.2 = d(A_{P,2}^{n-1}, C_{R,2}^{n-1}) + d(B_{P,2}^{n-1}, A_{R,2}^{n-1}) & > & d(A_{P,2}^{n-1}, B_{P,2}^{n-1}) + d(C_{R,2}^{n-1}, A_{R,2}^{n-1}) = 2.8 \\
2.92 = d(C_{R,1}^{n-1}, B_{R,2}^{n-1}) + d(A_{R,1}^{n-1}, D_{R,2}^{n-1}) & > & d(C_{R,1}^{n-1}, A_{R,1}^{n-1}) + d(B_{R,2}^{n-1}, D_{R,2}^{n-1}) = 2.88 \\
4.98 = d(C_{P,2}^{n-1}, B_{R,1}^{n-1}) + d(D_{P,2}^{n-1}, D_{R,1}^{n-1}) & > & d(C_{P,2}^{n-1}, D_{P,2}^{n-1}) + d(B_{R,1}^{n-1}, D_{R,1}^{n-1}) = 4.38
\end{array}$$

Again, our construction possesses the property that each pair of gadgets G_i^P and G_i^R is just a scaled and translated version of the pair G_{n-1}^P and G_{n-1}^R. Since we have relaxed the requirements for the gadgets, we do not even need rotations here. Hence, when we choose the coordinates for the points of the gadgets G_i^P and G_i^R, we can after appropriate scaling and translating, assume without loss of generality that the gadgets G_{i+1}^P and G_{i+1}^R lie at those positions that we have specified above for the gadgets G_{n-1}^P and G_{n-1}^R. Under this assumption, we place the points of G_i^P and G_i^R as follows:

1. Start with the coordinates that we specified for the points of gadgets G_{n-1}^P and G_{n-1}^R.
2. Scale each coordinate with a factor of 7.7.
3. Translate the points by the vector $(1.93, 0.3)$.

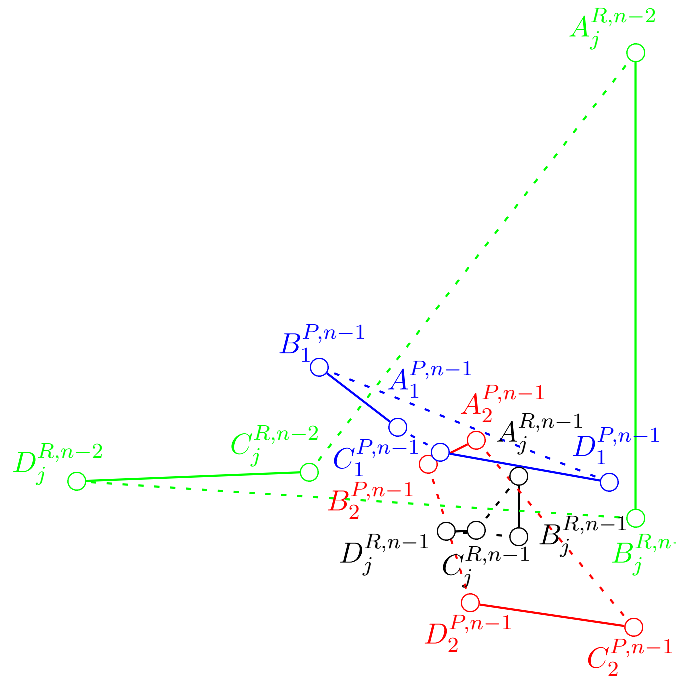

Figure 3: This illustration shows the points of the gadgets G^P_{n-1}, G^R_{n-1}, and G^R_{n-2}. One can see that G^R_{n-2} is a copy of G^R_{n-1}, scaled by 7.7 and translated by $(1.93, 0.3)$.

Figure 3 shows the points of the gadgets G^P_{n-1}, G^R_{n-1}, and G^R_{n-2}.

Now it only remains to show that the sequences of 2-changes in which one block of G^R_i changes from its long to its short state while resetting G^P_{i+1} from (S,S) to (L,L) are improving. As the coordinates of the points in the two blocks of gadget G^R_i are identical, it suffices to show the inequalities for the sequence in which the first block of G^R_i changes from its long to its short state. The coordinates of the points of gadget G^R_i are as follows:

	$A^i_{R,j}$	$B^i_{R,j}$	$C^i_{R,j}$	$D^i_{R,j}$
x-coordinate	-2	-3.3	-1.3	1.5
y-coordinate	1.8	2.8	1.4	0.9

This yields the following inequalities.

$$\begin{aligned}
15.42 = d(A^i_{R,1}, C^i_{R,1}) + d(C^{i+1}_{P,2}, D^{i+1}_{P,2}) &> d(A^i_{R,1}, C^{i+1}_{P,2}) + d(C^i_{R,1}, D^{i+1}_{P,2}) = 14.36\\
11.056 = d(B^{i+1}_{P,2}, A^{i+1}_{P,2}) + d(D^i_{R,1}, B^i_{R,1}) &> d(B^{i+1}_{P,2}, D^i_{R,1}) + d(A^{i+1}_{P,2}, B^i_{R,1}) = 10.024\\
9.394 = d(B^{i+1}_{P,2}, D^i_{R,1}) + d(C^{i+1}_{P,1}, D^{i+1}_{P,1}) &> d(B^{i+1}_{P,2}, C^{i+1}_{P,1}) + d(D^i_{R,1}, D^{i+1}_{P,1}) = 9.226\\
7.13 = d(B^{i+1}_{P,1}, A^{i+1}_{P,1}) + d(C^i_{R,1}, D^{i+1}_{P,2}) &> d(B^{i+1}_{P,1}, C^i_{R,1}) + d(A^{i+1}_{P,1}, D^{i+1}_{P,2}) = 5.99\\
13.46 = d(A^i_{R,1}, C^{i+1}_{P,2}) + d(B^i_{R,1}, A^{i+1}_{P,2}) &> d(A^i_{R,1}, B^i_{R,1}) + d(C^{i+1}_{P,2}, A^{i+1}_{P,2}) = 13.4\\
4.5 = d(C^{i+1}_{P,1}, B^{i+1}_{P,2}) + d(A^{i+1}_{P,1}, D^{i+1}_{P,2}) &> d(C^{i+1}_{P,1}, A^{i+1}_{P,1}) + d(B^{i+1}_{P,2}, D^{i+1}_{P,2}) = 4.1\\
10.716 = d(C^i_{R,1}, B^{i+1}_{P,1}) + d(D^i_{R,1}, D^{i+1}_{P,1}) &> d(C^i_{R,1}, D^i_{R,1}) + d(B^{i+1}_{P,1}, D^{i+1}_{P,1}) = 10.704
\end{aligned}$$

This concludes the proof of Theorem 1 b) as it shows that all 2-changes are improving.

4 The Expected Number of 2-Changes

In Sections 4.1, 4.2, and 4.3 we analyze the expected number of 2-changes on random L_1, L_2, and general TSP instances, respectively. Basically the previous results on the expected number of 2-changes are obtained by analyzing the improvement made by the smallest improving 2-change. If the smallest improvement is not too small, then the number of improvements cannot be large. In our analyses for the L_1 and the L_2 metric, we consider not only a single step but certain pairs of steps. We show that the smallest improvement made by any such pair is typically much larger than the improvement made by a single step which yields our improved bounds. Our approach is not restricted to pairs of steps. One could also consider sequences of steps of length k for any small enough k. In fact, for general ϕ-perturbed graphs with m edges, we consider sequences of length $\sqrt{\log m}$. The reason why we can analyze longer sequences for general graphs is that these inputs possess more randomness than ϕ-perturbed L_1 and L_2 instances since every edge length is a random variable that is independent of the other edge lengths. Hence, the analysis for general ϕ-perturbed graphs demonstrates the limits of our approach under optimal conditions. For L_1 and L_2 instances, the gain of considering longer sequences is small due to the dependencies between the edge lengths.

4.1 The L_1 Metric

In this section, we analyze the expected number of 2-changes on ϕ-perturbed L_1 instances. First we prove a weaker bound than the one in Theorem 2. The proof of this weaker bound illustrates our approach and reveals the problems one has to tackle in order to improve the upper bounds. It is solely based on an analysis of the smallest improvement made by any of the possible 2-Opt steps. If with high probability every 2-Opt step decreases the tour length by a polynomially large amount, then with high probability only polynomially many 2-Opt steps are possible before a local optimum is reached.

Theorem 6. *Starting with an arbitrary tour, the expected number of steps performed by 2-Opt on ϕ-perturbed L_1 instances is $O(n^6 \cdot \log n \cdot \phi)$.*

Proof. In order to prove the desired bound on the expected convergence time, we only need two simple observations. First, the initial tour can have length at most $2n$ as the number of edges is n and every edge has length at most 2. And second, every 2-Opt step decreases the weight of the tour by a polynomially large amount. The latter can be shown by a union bound over all possible 2-Opt steps. Consider a fixed 2-Opt step S, let e_1 and e_2 denote the edges removed from the tour in step S, and let e_3 and e_4 denote the edges added to the tour. Then the improvement $\Delta(S)$ of step S can be written as

$$\Delta(S) = d(e_1) + d(e_2) - d(e_3) - d(e_4) \ . \tag{1}$$

Without loss of generality let $e_1 = (v_1, v_2)$ be the edge between the vertices v_1 and v_2, and let $e_2 = (v_3, v_4)$, $e_3 = (v_1, v_3)$, and $e_4 = (v_2, v_4)$. Furthermore, let x_i denote the x-coordinate of vertex v_i and let y_i denote its y-coordinate. Then the improvement $\Delta(S)$ of step S can be written as

$$\begin{aligned}\Delta(S) =& |x_1 - x_2| + |x_3 - x_4| - |x_1 - x_3| - |x_2 - x_4| \\ &+ |y_1 - y_2| + |y_3 - y_4| - |y_1 - y_3| - |y_2 - y_4| \ .\end{aligned}$$

Depending on the order of the x- and y-coordinates, $\Delta(S)$ can be written as linear combination of the coordinates, e. g., for $x_1 \geq x_2 \geq x_3 \geq x_4$ and $y_1 \geq y_2 \geq y_3 \geq y_4$ the improvement $\Delta(S)$

can be written as $-2x_2 + 2x_3 - 2y_2 + 2y_3$. There are $(4!)^2$ such orders and each one gives rise to a linear combination of the x- and y-coordinates with integer coefficients. For each of these linear combinations, the probability that it takes a value in the interval $(0, \varepsilon]$ is bounded by $\varepsilon\phi$. In order to see this, assume that all variables that appear in the linear combination, i. e., whose coefficients are not zero, except one of them have already been determined by an adversary. No matter of how the adversary has chosen these variables, the linear combination can only take a value in $(0, \varepsilon]$ if the last variable takes a value in some interval of length at most ε. This interval is determined by the decisions of the adversary. However, for every such interval the probability that the last variables takes a value in it is bounded by $\varepsilon\phi$. Since $\Delta(S)$ can only take a value in the interval $(0, \varepsilon]$ if one of the linear combinations takes a value in this interval, the probability of the event $\Delta(S) \in (0, \varepsilon]$ can be upper bounded by $(4!)^2\varepsilon\phi$.

Let $\Delta_{\min}$ denote the improvement of the smallest improving 2-Opt step S, i. e., $\Delta_{\min} = \min\{\Delta(S) | \Delta(S) > 0\}$. We can estimate $\Delta_{\min}$ by a union bound, yielding

$$\mathbf{Pr}\left[\Delta_{\min} \leq \varepsilon\right] \leq (4!)^2 \varepsilon n^4 \phi$$

as there are at most n^4 different 2-Opt steps. Let T denote the random variable describing the number of 2-Opt steps before a local optimum is reached. Observe that T can only exceed a given number t if the smallest improvement $\Delta_{\min}$ is less than $2n/t$, and hence

$$\mathbf{Pr}\left[T \geq t\right] \leq \mathbf{Pr}\left[\Delta_{\min} \leq \frac{2n}{t}\right] \leq \frac{2(4!)^2 n^5 \phi}{t} \ .$$

Since there are at most $(n!)$ different TSP tours and none of these tours can appear twice during the local search, T is always bounded by $(n!)$. Altogether, we can bound the expected value of T by

$$\mathbf{E}\left[T\right] = \sum_{t=1}^{n!} \mathbf{Pr}\left[T \geq t\right] \leq \sum_{t=1}^{n!} \frac{2(4!)^2 n^5 \phi}{t} \ .$$

Bounding the n-th harmonic number by $\ln(n) + 1$ and using $\ln(n!) = O(n \log n)$ yields

$$\mathbf{E}\left[T\right] = O(n^6 \cdot \log n \cdot \phi) \ .$$

□

The bound in Theorem 6 is only based on the smallest improvement $\Delta_{\min}$ made by any of the 2-Opt steps. Intuitively, this is too pessimistic since most of the steps performed by 2-Opt yield a larger improvement than $\Delta_{\min}$. In particular, two consecutive steps yield an improvement of at least $\Delta_{\min}$ plus the improvement $\Delta_{\min}^{(2)}$ of the second smallest step. This observation alone, however, does not suffice to improve the bound substantially. Instead, we regroup the 2-changes to pairs such that each pair of 2-changes is *linked* by an edge, i. e., one edge added to the tour in the first 2-change is removed from the tour in the second 2-change, and we analyze the smallest improvement made by any pair of linked 2-Opt steps. Obviously, this improvement is at least $\Delta_{\min} + \Delta_{\min}^{(2)}$ but one can hope that it is in fact much larger since it is unlikely that the 2-change that yields the smallest improvement and the 2-change that yields the second smallest improvement form a pair of linked steps. We show that this is indeed the case and use this result to prove the bound on the expected length of the longest path in the state graph of 2-Opt on ϕ-perturbed L_1 instances claimed in Theorem 2.

4.1.1 Construction of Pairs of Linked 2-Changes

Consider an arbitrary sequence of consecutive 2-changes of length t. The following lemma guarantees that the number of disjoint and linked pairs of 2-changes in every such sequence increases linearly with the length t.

Lemma 7. *In every sequence of t consecutive 2-changes the number of disjoint pairs of 2-changes that are linked by an edge, i. e., pairs such that there exists an edge added to the tour in the first 2-change of the pair and removed from the tour in the second 2-change of the pair, is at least $t/3 - n(n-1)/12$.*

Proof. Let $S_1, \dots, S_t$ denote an arbitrary sequence of consecutive 2-changes. The sequence is processed step by step and a list $\mathcal{L}$ of linked pairs of 2-changes is created. However, these pairs are not necessarily disjoint. Hence, after the list has been created, pairs have to be removed from the list until there are no non-disjoint pairs left. Assume that the 2-changes $S_1, \dots, S_{i-1}$ have already been processed and that now 2-change S_i has to be processed. Assume further that in step S_i the edges e_1 and e_2 are exchanged with the edges e_3 and e_4. Let j denote the smallest index with $j > i$ such that edge e_3 is removed from the tour in step S_j if such a step exists. In this case, the pair (S_i, S_j) is added to the constructed list $\mathcal{L}$. Analogously, let j' denote the smallest index with $j' > i$ such that edge e_4 is removed from the tour in step $S_{j'}$ if such a step exists. In this case, also the pair $(S_i, S_{j'})$ is added to the list $\mathcal{L}$.

After the sequence has been processed completely, each pair in $\mathcal{L}$ is linked by an edge but we still have to identify a subset $\mathcal{L}'$ of $\mathcal{L}$ consisting only of pairwise disjoint pairs. This subset is constructed in a greedy fashion. We process the list $\mathcal{L}$ step by step, starting with an empty list $\mathcal{L}'$. For each pair in $\mathcal{L}$, we check whether it is disjoint from all pairs which have already been inserted into $\mathcal{L}'$ or not. In the former case, the current pair is inserted into $\mathcal{L}'$. This way, we obtain a list $\mathcal{L}'$ of disjoint pairs such that each pair is linked by an edge. The number of pairs in $\mathcal{L}$ is at least $2t - n(n-1)/2$ since each of the t steps gives rise to 2 different pairs, unless an edge is added to the tour which is never removed again. Each 2-change occurs in at most 4 different pairs in $\mathcal{L}$, hence, each pair in $\mathcal{L}$ is non-disjoint from at most 6 other pairs in $\mathcal{L}$. This implies that $\mathcal{L}$ contains at most 6 times as many pairs as $\mathcal{L}'$ which concludes the proof. ∎

Consider a fixed pair of 2-changes linked by an edge. Without loss of generality assume that in the first step the edges $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are exchanged with the edges $\{v_1, v_3\}$ and $\{v_2, v_4\}$, for distinct vertices $v_1, \dots, v_4$. Also without loss of generality assume that in the second step the edges $\{v_1, v_3\}$ and $\{v_5, v_6\}$ are exchanged with the edges $\{v_1, v_5\}$ and $\{v_3, v_6\}$. However, note that the vertices v_5 and v_6 are not necessarily distinct from the vertices v_2 and v_4. We distinguish between three different types of pairs.

1. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 0$. We can assume w. l. o. g. that in the second step the edges $\{v_1, v_5\}$ and $\{v_3, v_6\}$ are added to the tour. See Figure 4.
2. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 1$. We can assume w. l. o. g. that $v_2 \in \{v_5, v_6\}$. We have to distinguish between two subcases: a) The edges $\{v_1, v_5\}$ and $\{v_2, v_3\}$ are added to the tour in the second step. b) The edges $\{v_1, v_2\}$ and $\{v_3, v_5\}$ are added to the tour in the second step. These cases are illustrated in Figure 5.
3. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 2$. The case $v_2 = v_5$ and $v_4 = v_6$ cannot appear, as it would imply that the tour is not changed by performing the considered pair of steps. Hence, for pairs of this type, we must have $v_2 = v_6$ and $v_4 = v_5$.

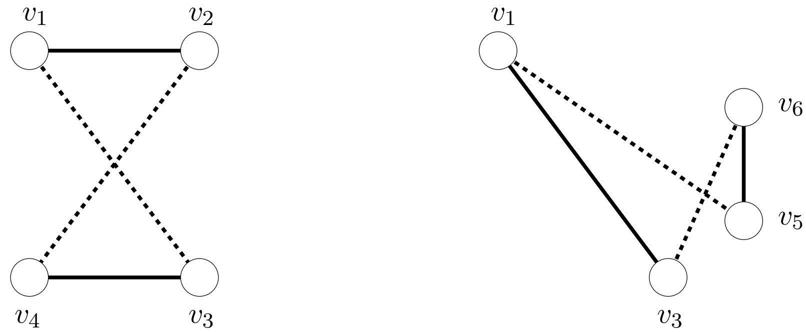

Figure 4: A pair of type 1: First the edges $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are exchanged with the edges $\{v_1, v_3\}$ and $\{v_2, v_4\}$, then $\{v_1, v_3\}$ and $\{v_5, v_6\}$ are exchanged with $\{v_1, v_5\}$ and $\{v_3, v_6\}$.

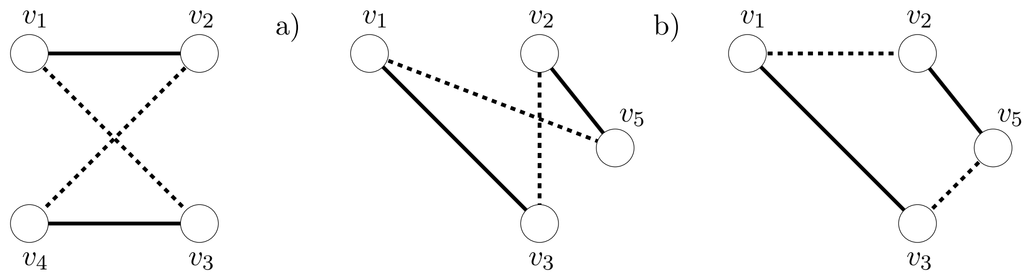

Figure 5: A pair of type 2: First the edges $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are exchanged with the edges $\{v_1, v_3\}$ and $\{v_2, v_4\}$, then $\{v_1, v_3\}$ and $\{v_2, v_5\}$ are exchanged with either $\{v_1, v_5\}$ and $\{v_2, v_3\}$ or $\{v_1, v_2\}$ and $\{v_3, v_5\}$.

When distances are measured according to the L_2 metric, pairs of type 3 result in vast dependencies and hence the probability that there exists a pair of this type in which both steps are improvements by at most ε w.r.t. the L_2 metric cannot be bounded appropriately. In order to reduce the number of cases we have to consider and in order to prepare the analysis of ϕ-perturbed L_2 instances, we exclude pairs of type 3 from our probabilistic analysis by leaving out all pairs of type 3 when constructing the list $\mathcal{L}$ in the proof of Lemma 7.

We only need to show that there are always enough pairs of type 1 or 2. Therefore consider two steps S_i and S_j with $i < j$ that form a pair of type 3. Assume that in step S_i the edges $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are replaced by the edges $\{v_1, v_3\}$ and $\{v_2, v_4\}$, and that in step S_j these edges are replaced by the edges $\{v_1, v_4\}$ and $\{v_2, v_3\}$. Now consider the next step S_l with $l > j$ in which the edge $\{v_1, v_4\}$ is removed from the tour if such a step exists and the next step $S_{l'}$ with $l' > j$ in which the edge $\{v_2, v_3\}$ is removed from the tour if such a step exists. Observe that (S_j, S_l) and $(S_j, S_{l'})$ cannot be pairs of type 3, that is $l \neq l'$, since otherwise the improvement of one of the steps S_i, S_j, or S_l must be negative. Hence, with each pair (S_i, S_j) of type 3 we can associate two pairs (S_j, S_l) and $(S_j, S_{l'})$ of type 1 or 2, unless one of the edges added to the tour in S_j is never removed again. Observe that each pair of type 1 or 2 can be associated with at most 2 different pairs of type 3. Hence, if x denotes the number of type 3 pairs, then the total number of pairs of type 1 or 2 must be at least $x - n(n-1)/4$. Since the total number of pairs is at most $2t$, we have $2x - n(n-1)/4 \leq 2t$ and hence $x \leq t + n(n-1)/8$. Since the total number of pairs is at least $2t - n(n-1)/2$, the total number of pairs of type 1 or 2 is at least $t - 5n(n-1)/8$. Now the list $\mathcal{L}'$ is constructed exactly as before, yielding the following lemma.

Lemma 8. *In every sequence of t consecutive 2-changes the number of disjoint pairs of 2-changes of type 1 or 2 is at least $t/6 - 5n(n-1)/48$.* □

4.1.2 Analysis of Pairs of Linked 2-Changes

We prove the following lemmas about pairs of type 1 and 2.

Lemma 9. *In a ϕ-perturbed L_1 instance with n vertices, the probability that there exists a pair of type 1 in which both 2-changes are improvements by at most ε is bounded by $O(n^6 \cdot \varepsilon^2 \cdot \phi^2)$.*

Lemma 10. *In a ϕ-perturbed L_1 instance with n vertices, the probability that there exists a pair of type 2 in which both 2-changes are improvements by at most ε is bounded by $O(n^5 \cdot \varepsilon^2 \cdot \phi^2)$.*

Proof of Lemma 9. We use the notations from Figure 4, i. e., we assume that in the first step the edges $\{v_1, v_2\}$ and $\{v_3, v_4\}$ are replaced by the edges $\{v_1, v_3\}$ and $\{v_2, v_4\}$ and that in the second step the edges $\{v_1, v_3\}$ and $\{v_5, v_6\}$ are replaced by the edges $\{v_1, v_5\}$ and $\{v_3, v_6\}$. For $i \in \{1, \dots, 6\}$, let x_i denote the x-coordinate of vertex v_i and let y_i denote its y-coordinate. Furthermore, let Δ_1 denote the (possibly negative) improvement of the first step and let Δ_2 denote the (possibly negative) improvement of the second step. Δ_1 and Δ_2 can be written as

$$\begin{aligned}\Delta_1 =& |x_1 - x_2| + |x_3 - x_4| - |x_1 - x_3| - |x_2 - x_4| \\ &+ |y_1 - y_2| + |y_3 - y_4| - |y_1 - y_3| - |y_2 - y_4|\end{aligned}$$

and

$$\begin{aligned}\Delta_2 =& |x_1 - x_3| + |x_5 - x_6| - |x_1 - x_5| - |x_3 - x_6| \\ &+ |y_1 - y_3| + |y_5 - y_6| - |y_1 - y_5| - |y_3 - y_6| \ .\end{aligned}$$

For any fixed order of the x-coordinates and the y-coordinates, Δ_1 and Δ_2 can be expressed as linear combinations of the coordinates with integer coefficients. For fixed orders σ_x and σ_y of the x- and y-coordinates, let $\Delta_1^{\sigma_x,\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y}$ denote the corresponding linear combinations. Let $\mathcal{A}$ denote the event that both Δ_1 and Δ_2 take values in the interval $(0, \varepsilon]$, and for fixed orders σ_x and σ_y, let $\mathcal{A}^{\sigma_x,\sigma_y}$ denote the event that both linear combinations $\Delta_1^{\sigma_x,\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y}$ take values in the interval $(0, \varepsilon]$. Obviously $\mathcal{A}$ can only occur if for at least one pair of σ_x and σ_y, the event $\mathcal{A}^{\sigma_x,\sigma_y}$ occurs. Hence, we obtain the following estimate

$$\mathbf{Pr}[\mathcal{A}] \le \sum_{\sigma_x,\sigma_y} \mathbf{Pr}[\mathcal{A}^{\sigma_x,\sigma_y}] \ .$$

Since there are only a constant number of different orders σ_x and σ_y, we only need to show that for every pair of orders σ_x and σ_y, the probability of the event $\mathcal{A}^{\sigma_x,\sigma_y}$ is bounded from above by $O(\varepsilon^2\phi^2)$. Then a union bound over all possible pairs of linked 2-changes of type 1 yields the lemma.

We divide the set of possible pairs of linear combinations $(\Delta_1^{\sigma_x,\sigma_y}, \Delta_2^{\sigma_x,\sigma_y})$ into three classes. We say that a pair of linear combinations belongs to class A if at least one of the linear combinations equals 0, we say that it belongs to class B if $\Delta_1^{\sigma_x,\sigma_y} = -\Delta_2^{\sigma_x,\sigma_y}$, and we say that it belongs to class C if $\Delta_1^{\sigma_x,\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y}$ are linearly independent. For orders σ_x and σ_y that yield pairs from class A or

B, the event $\mathcal{A}^{\sigma_x,\sigma_y}$ can never occur since in both cases the value of at least one linear combination is at most 0. For orders σ_x and σ_y that yield pairs from class C, we can apply Lemma 31 from Appendix A which yields that the probability of the event $\mathcal{A}^{\sigma_x,\sigma_y}$ is bounded from above by $\varepsilon^2\phi^2$. Hence, we only need to show that every pair $(\Delta_1^{\sigma_x,\sigma_y}, \Delta_2^{\sigma_x,\sigma_y})$ of linear combinations belongs either to class A, B, or C.

Consider fixed orders σ_x and σ_y. We split $\Delta_1^{\sigma_x,\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y}$ into the x and the y parts, i. e., $\Delta_1^{\sigma_x,\sigma_y} = X_1^{\sigma_x} + Y_1^{\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y} = X_2^{\sigma_x} + Y_2^{\sigma_y}$, and we show that the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ belongs either to class A, B, or C. For symmetry reasons the same must then be true for the pair $(Y_1^{\sigma_y}, Y_2^{\sigma_y})$. One can easily see that this directly implies that also $(\Delta_1^{\sigma_x,\sigma_y}, \Delta_2^{\sigma_x,\sigma_y})$ must belong to one of the aforementioned classes.

Now assume for contradiction that the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ is linearly dependent for the fixed order σ_x but does not belong to class A or B. Observe that this can only happen if $X_1^{\sigma_x}$ does not contain x_2 and x_4 and if $X_2^{\sigma_x}$ does not contain x_5 and x_6. The former can only happen if either $x_3 \geq x_4$, $x_2 \geq x_4$, and $x_2 \geq x_1$ or if $x_3 \leq x_4$, $x_2 \leq x_4$, and $x_2 \leq x_1$. The latter can only happen if either $x_5 \geq x_6$, $x_3 \geq x_6$, and $x_5 \geq x_1$ or if $x_5 \leq x_6$, $x_3 \leq x_6$, and $x_5 \leq x_1$.

In the case $x_3 \geq x_4$, $x_2 \geq x_4$, and $x_2 \geq x_1$, we can write $X_1^{\sigma_x}$ as

$$X_1^{\sigma_x} = -x_1 + x_3 - |x_1 - x_3| = \begin{cases} -2x_1 + 2x_3 & \text{if } x_1 \geq x_3 \\ 0 & \text{if } x_3 \geq x_1 \end{cases} .$$

In the case $x_3 \leq x_4$, $x_2 \leq x_4$, and $x_2 \leq x_1$, we can write $X_1^{\sigma_x}$ as

$$X_1^{\sigma_x} = x_1 - x_3 - |x_1 - x_3| = \begin{cases} 0 & \text{if } x_1 \geq x_3 \\ 2x_1 - 2x_3 & \text{if } x_3 \geq x_1 \end{cases} .$$

In the case $x_5 \geq x_6$, $x_3 \geq x_6$, and $x_5 \geq x_1$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = x_1 - x_3 + |x_1 - x_3| = \begin{cases} 2x_1 - 2x_3 & \text{if } x_1 \geq x_3 \\ 0 & \text{if } x_3 \geq x_1 \end{cases} .$$

In the case $x_5 \leq x_6$, $x_3 \leq x_6$, and $x_5 \leq x_1$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = -x_1 + x_3 + |x_1 - x_3| = \begin{cases} 0 & \text{if } x_1 \geq x_3 \\ -2x_1 + 2x_3 & \text{if } x_3 \geq x_1 \end{cases} .$$

Hence, if one chooses the order such that x_2, x_4, x_5, and x_6 cancel out, then in the case $x_1 \geq x_3$, we have

$$X_1^{\sigma_x} \in \{0, -2x_1 + 2x_3\}$$

and

$$X_2^{\sigma_x} \in \{0, 2x_1 - 2x_3\} .$$

Hence, the pair of resulting linear combinations in the case $x_1 \geq x_3$ belongs either to class A or to class B. Analogously, in the case $x_3 \geq x_1$ we have

$$X_1^{\sigma_x} \in \{0, 2x_1 - 2x_3\}$$

and

$$X_2^{\sigma_x} \in \{0, -2x_1 + 2x_3\} .$$

Hence, also in this case, the pair of resulting linear combinations belongs either to class A or to class B. $\square$

With similar arguments we prove Lemma 10.

Proof of Lemma 10. We first prove the lemma for pairs of type 2 a). Using the same notations as in the proof of Lemma 9, we can write the improvement Δ_2 now as

$$\begin{aligned}\Delta_2 =& |x_1 - x_3| + |x_2 - x_5| - |x_1 - x_5| - |x_2 - x_3| \\ &+ |y_1 - y_3| + |y_2 - y_5| - |y_1 - y_5| - |y_2 - y_3| \ .\end{aligned}$$

Again we show that the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ belongs either to class A, B, or C, which implies the same for the pair $(\Delta_1^{\sigma_x,\sigma_y}, \Delta_2^{\sigma_x,\sigma_y})$ we are interested in. Assume for contradiction that the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ is linearly dependent for the fixed order σ_x but does not belong to class A or B. Observe that this can only happen if $X_1^{\sigma_x}$ does not contain x_4 and if $X_2^{\sigma_x}$ does not contain x_5. The former can only happen if either $x_3 \geq x_4$ and $x_2 \geq x_4$ or if $x_3 \leq x_4$ and $x_2 \leq x_4$. The latter can only happen if either $x_2 \geq x_5$ and $x_1 \geq x_5$ or if $x_2 \leq x_5$ and $x_1 \leq x_5$.

In the case $x_3 \geq x_4$ and $x_2 \geq x_4$, we can write $X_1^{\sigma_x}$ as

$$X_1^{\sigma_x} = |x_1 - x_2| - |x_1 - x_3| - x_2 + x_3 = \begin{cases} -2x_2 + 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_1 \geq x_2 \\ -2x_1 + 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_1 \leq x_2 \\ 2x_1 - 2x_2 & \text{if } x_1 \leq x_3 \text{ and } x_1 \geq x_2 \\ 0 & \text{if } x_1 \leq x_3 \text{ and } x_1 \leq x_2 \end{cases} .$$

In the case $x_3 \leq x_4$ and $x_2 \leq x_4$, we can write $X_1^{\sigma_x}$ as

$$X_1^{\sigma_x} = |x_1 - x_2| - |x_1 - x_3| + x_2 - x_3 = \begin{cases} 0 & \text{if } x_1 \geq x_3 \text{ and } x_1 \geq x_2 \\ -2x_1 + 2x_2 & \text{if } x_1 \geq x_3 \text{ and } x_1 \leq x_2 \\ 2x_1 - 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_1 \geq x_2 \\ 2x_2 - 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_1 \leq x_2 \end{cases} .$$

In the case $x_2 \geq x_5$ and $x_1 \geq x_5$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = |x_1 - x_3| - |x_2 - x_3| - x_1 + x_2 = \begin{cases} 0 & \text{if } x_1 \geq x_3 \text{ and } x_2 \geq x_3 \\ 2x_2 - 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_2 \leq x_3 \\ -2x_1 + 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_2 \geq x_3 \\ -2x_1 + 2x_2 & \text{if } x_1 \leq x_3 \text{ and } x_2 \leq x_3 \end{cases} .$$

In the case $x_2 \leq x_5$ and $x_1 \leq x_5$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = |x_1 - x_3| - |x_2 - x_3| + x_1 - x_2 = \begin{cases} 2x_1 - 2x_2 & \text{if } x_1 \geq x_3 \text{ and } x_2 \geq x_3 \\ 2x_1 - 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_2 \leq x_3 \\ -2x_2 + 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_2 \geq x_3 \\ 0 & \text{if } x_1 \leq x_3 \text{ and } x_2 \leq x_3 \end{cases} .$$

Hence, if one chooses the order such that x_4 and x_5 cancel out, then in the case $x_1 \geq x_3$, we have

$$X_1^{\sigma_x} \in \{0, -2x_1 + 2x_2, -2x_1 + 2x_3, -2x_2 + 2x_3\}$$

and

$$X_2^{\sigma_x} \in \{0, 2x_1 - 2x_2, 2x_1 - 2x_3, 2x_2 - 2x_3\} \ .$$

Hence, the pair of resulting linear combinations in the case $x_1 \geq x_3$ belongs either to class A, B, or C. Analogously, in the case $x_3 \geq x_1$ we have

$$X_1^{\sigma_x} \in \{0, 2x_1 - 2x_2, 2x_1 - 2x_3, 2x_2 - 2x_3\}$$

and

$$X_2^{\sigma_x} \in \{0, -2x_1 + 2x_2, -2x_1 + 2x_3, -2x_2 + 2x_3\} \ .$$

Hence, also in this case, the pair of resulting linear combinations belongs either to class A, B, or C.

It remains to consider pairs of type 2 b). For these pairs, we can write Δ_2 as

$$\begin{aligned}\Delta_2 =& |x_1 - x_3| + |x_2 - x_5| - |x_1 - x_2| - |x_3 - x_5| \\ &+ |y_1 - y_3| + |y_2 - y_5| - |y_1 - y_2| - |y_3 - y_5| \ .\end{aligned}$$

Assume that the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ is linearly dependent for the fixed order σ_x but does not belong to class A or B. Observe that this can only happen if $X_1^{\sigma_x}$ does not contain x_4 and if $X_2^{\sigma_x}$ does not contain x_5. As we have already seen for pairs of type 2 a), the former can only happen if either $x_3 \geq x_4$ and $x_2 \geq x_4$ or if $x_3 \leq x_4$ and $x_2 \leq x_4$. The latter can only happen if either $x_2 \geq x_5$ and $x_3 \geq x_5$ or if $x_2 \leq x_5$ and $x_3 \leq x_5$.

In the case $x_2 \geq x_5$ and $x_3 \geq x_5$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = -|x_1 - x_2| + |x_1 - x_3| + x_2 - x_3 = \begin{cases} 2x_2 - 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_1 \geq x_2 \\ 2x_1 - 2x_3 & \text{if } x_1 \geq x_3 \text{ and } x_1 \leq x_2 \\ -2x_1 + 2x_2 & \text{if } x_1 \leq x_3 \text{ and } x_1 \geq x_2 \\ 0 & \text{if } x_1 \leq x_3 \text{ and } x_1 \leq x_2 \end{cases} .$$

In the case $x_2 \leq x_5$ and $x_3 \leq x_5$, we can write $X_2^{\sigma_x}$ as

$$X_2^{\sigma_x} = -|x_1 - x_2| + |x_1 - x_3| - x_2 + x_3 = \begin{cases} 0 & \text{if } x_1 \geq x_3 \text{ and } x_1 \geq x_2 \\ 2x_1 - 2x_2 & \text{if } x_1 \geq x_3 \text{ and } x_1 \leq x_2 \\ -2x_1 + 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_1 \geq x_2 \\ -2x_2 + 2x_3 & \text{if } x_1 \leq x_3 \text{ and } x_1 \leq x_2 \end{cases} .$$

Hence, if one chooses the order such that x_4 and x_5 cancel out, then in the case $x_1 \geq x_3$ and $x_1 \geq 2$, we have

$$X_1^{\sigma_x} \in \{0, -2x_2 + 2x_3\}$$

and

$$X_2^{\sigma_x} \in \{0, 2x_2 - 2x_3\} \ .$$

Hence, the pair of resulting linear combinations in this case belongs either to class A or to class B. One can easily see that also in the other three cases only pairs of class A, B, or C can occur. □

4.1.3 The Expected Number of 2-Changes on ϕ-perturbed L_1 instances

Based on Lemmas 8, 9, and 10, we are now able to prove part a) of Theorem 2.

Proof of Theorem 2 a). Let T denote the random variable that describes the length of the longest path in the state graph. If $T \geq t$, then there must exist a sequence $S_1, \ldots, S_t$ of t consecutive 2-changes in the state graph. We start by identifying a set of linked pairs of type 1 and 2 in this sequence. Due to Lemma 8, we know that we can find at least $t/6 - 5n(n-1)/48$ such pairs. For $i \in \{1,2\}$, let $\Delta_{\min}^{(i)}$ denote the smallest improvement made by any pair of improving 2-Opt steps of type i. For $t > n^2$, we have $t/6 - 5n(n-1)/48 > t/16$ and hence due to Lemmas 9 and 10,

$$\begin{aligned} \mathbf{Pr}\left[T \geq t\right] &\leq \mathbf{Pr}\left[\Delta_{\min}^{(1)} \leq \frac{32n}{t}\right] + \mathbf{Pr}\left[\Delta_{\min}^{(2)} \leq \frac{32n}{t}\right] \\ &= O\left(\min\left\{\frac{n^8\phi^2}{t^2}, 1\right\}\right) + O\left(\min\left\{\frac{n^7\phi^2}{t^2}, 1\right\}\right) \\ &= O\left(\min\left\{\frac{n^8\phi^2}{t^2}, 1\right\}\right) \quad . \end{aligned}$$

Since T is at most $n!$, this implies the following bound on the expected running time

$$\mathbf{E}\left[T\right] = n^2 + \sum_{t=1}^{n!} O\left(\min\left\{\frac{n^8\phi^2}{t^2}, 1\right\}\right) = O(n^4 \cdot \phi) \quad .$$

This concluses the proof of part a) of the theorem. □

It is well-known that for an arbitrary set of n points in the unit square and for an arbitrary metric on $\mathbb{R}^2$ the optimal tour visiting all n points has length $O(\sqrt{n})$ (see, e. g., [CKT99]). Furthermore, every insertion heuristic finds an $O(\log n)$-approximation [RSI77]. Hence, if one starts with a solution calculated by an insertion heuristic, the initial tour has length $O(\sqrt{n} \cdot \log n)$. Using this observation yields part a) of Theorem 3.

Proof of Theorem 3 b). Since the initial tour has length $O(\sqrt{n} \cdot \log n)$, we obtain for an appropriate constant c and $t > n^2$,

$$\begin{aligned} \mathbf{Pr}\left[T \geq t\right] &\leq \mathbf{Pr}\left[\Delta_{\min}^{(1)} \leq \frac{c \cdot \sqrt{n} \cdot \log n}{t}\right] + \mathbf{Pr}\left[\Delta_{\min}^{(2)} \leq \frac{c \cdot \sqrt{n} \cdot \log n}{t}\right] \\ &= O\left(\min\left\{\frac{n^7 \cdot \log^2 n \cdot \phi^2}{t^2}, 1\right\}\right) \quad . \end{aligned}$$

This yields

$$\mathbf{E}\left[T\right] = n^2 + \sum_{t=1}^{n!} O\left(\min\left\{\frac{n^7 \cdot \log^2 n \cdot \phi^2}{t^2}, 1\right\}\right) = O(n^{3.5} \cdot \log n \cdot \phi) \quad .$$

□

4.2 The L_2 Metric

Now we consider ϕ-perturbed L_2 instances. The structure of the analysis is very similar to the analysis of ϕ-perturbed L_1 instances that we presented in the previous section. In fact, we only exchange Lemmas 9 and 10 with their equivalent versions for the L_2 metric.

Lemma 11. *For ϕ-perturbed L_2 instances, the probability that there exists a pair of type 1 in which both 2-changes are improvements by at most ε is bounded by $O(n^6 \cdot \varepsilon^2 \cdot (\log^2(1/\varepsilon) + 1) \cdot \phi^5)$.*

Lemma 12. *For ϕ-perturbed L_2 instances, the probability that there exists a pair of type 2 in which both 2-changes are improvements by at most ε is bounded by $O(n^5 \cdot \varepsilon^{3/2} \cdot (\log(1/\varepsilon) + 1) \cdot \phi^4)$.*

The bounds which these lemmas provide are slightly weaker than their L_1 counterparts, and hence, also the bound on the expected running time is slightly worse for L_2 instances. The proof of these lemmas is based on a careful probabilistic analysis of a single 2-change which is then used to yield bounds on the improvement of pairs of linked 2-changes. We like to mention that already the analysis of a single 2-change when used as in the proof of Theorem 6 yields a bound of $O(n^7 \cdot \log^2 n \cdot \phi^3)$ for the expected number of 2-changes. For L_2 instances in which all points are distributed uniformly over the unit square, this bound improves already the previous known bound of $O(n^{10} \cdot \log n)$.

4.2.1 Analysis of a Single 2-Change

Before we can prove Lemmas 11 and 12, we have to understand the random variable that describes the improvement of a single 2-change. In this section, we analyze this variable under certain conditions. If, e. g., we would like to analyze a pair of linked 2-changes that share an edge e, it is helpful to know the densities of the random variables that describe the improvements of the first and the second 2-change under the condition that the length of e is given.

We analyze a 2-change in which the edges $\{O, Q_1\}$ and $\{P, Q_2\}$ are exchanged with the edges $\{O, Q_2\}$ and $\{P, Q_1\}$ for some vertices O, P, Q_1, and Q_2. In the considered input model, each of these points has a probability distribution over the unit square according to which it is chosen. We consider a simplified random experiment in which O is chosen to be the origin and P, Q_1, and Q_2 are chosen independently and uniformly from the interior of a circle with radius $\sqrt{2}$ around the origin. Due to the rotational symmetry of this model, we assume further that that P lies at position $(0, T)$, where T denotes the distance between O and P. In the next section, we explain how the analysis of this simple random experiment helps us to analyze the actual random experiment that occurs in the probabilistic input model.

Let Z_1 denote the difference $d(O, Q_1) - d(P, Q_1)$ and let Z_2 denote the difference $d(O, Q_2) - d(P, Q_2)$. Then the improvement Δ of the 2-Opt step can be expressed as $Z_1 - Z_2$. Before we analyze Δ, we consider the random variables Z_1 and Z_2 and prove the following lemma about their densities.

Lemma 13. *Let $i \in \{1, 2\}$, let $Q = Q_i$, and let $R = d(O, Q)$ denote the distance between the origin O and the point Q. Let Z denote the random variable $d(O, Q) - d(P, Q)$, i. e., $Z = Z_i$. For $z \in [-\tau, \min\{\tau, 2r - \tau\}]$, the conditional density $f_{Z|T=\tau,R=r}$ of the random variable Z, given $T = \tau$ and $R = r$ with $0 \le r, \tau \le \sqrt{2}$, can be upper bounded by*

$$f_{Z|T=\tau,R=r}(z) \le \begin{cases} \sqrt{\frac{2}{\tau^2 - z^2}} & \text{if} \quad r \ge \tau \\ \sqrt{\frac{2}{(\tau+z)(2r-\tau-z)}} & \text{if} \quad r \le \tau \end{cases}.$$

For $z \notin [-\tau, \min\{\tau, 2r-\tau\}]$, the density $f_{Z|T=\tau,R=r}(z)$ is 0.

The proof of this Lemma can be found in Appendix B.1.

Based on Lemma 13, we prove the following lemma which shows an upper bound on the conditional density of the improvement Δ of the 2-Opt step given the distance T and the distances $R_1 = d(O, Q_1)$ and $R_2 = d(O, Q_2)$. Observe that the distance T affects both Z_1 and Z_2 while the distances R_1 and R_2 affect only Z_1 or Z_2, respectively.

Lemma 14. *Let τ, r_1, and r_2 be distances with $r_1 \le r_2$ and $0 \le \tau, r_1, r_2 \le \sqrt{2}$. Furthermore, let Z_1 and Z_2 be independent random variables drawn according to the densities $f_{Z|T=\tau,R=r_1}(z)$ and $f_{Z|T=\tau,R=r_2}(z)$, respectively, and let $\Delta = Z_1 - Z_2$. For a sufficiently large constant κ, the conditional density of Δ for $\delta \ge 0$, given τ, r_1, and r_2, is bounded by*

$$f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \quad \le \quad \begin{cases} \frac{\kappa}{\tau}\left(\ln\left(\frac{1}{\delta}\right)+1\right) & \text{if } \tau \le r_1 \text{ and } \tau \le r_2 \\ \frac{\kappa}{\sqrt{r_1 r_2}}\left(\ln\left(\frac{1}{\delta}\right)+\ln\left(\frac{1}{|2(r_1-r_2)-\delta|}\right)+1\right) & \text{if } r_1 \le \tau \text{ and } r_2 \le \tau \\ \frac{\kappa}{\sqrt{r_1\tau}}\left(\ln\left(\frac{1}{\delta}\right)+1\right) & \text{if } r_1 \le \tau \le r_2 \\ \frac{\kappa}{\sqrt{r_2\tau}}\left(\ln\left(\frac{1}{\delta}\right)+\ln\left(\frac{1}{|2(\tau-r_2)-\delta|}\right)+1\right) & \text{if } r_2 \le \tau \le r_1 \end{cases} .$$

In order to prove this lemma, one basically just needs to calculate the convolution of the density given in Lemma 13 with itself. We present these calculations in Appendix B.2. The Lemmas 13 and 14 contain most of the work. Now the necessary bounds on the density of Δ and Z_i that are summarized in the next four lemmas follow from these lemmas simply by integrating over some of the conditions. The proofs of these lemmas can be found in Appendix B.3.

The next lemma gives an upper bound on the conditional density of the random variable Δ under the condition that only one of the distances R_1 and R_2 is given. It follows from Lemma 14 by integrating over all values of the other distance and of T.

Lemma 15. *Let r be an arbitrary distance with $0 \le r \le \sqrt{2}$. For a sufficiently large constant κ and for $i \in \{1,2\}$, the conditional density $f_{\Delta|R_i=r}(\delta)$ of Δ for $\delta \ge 0$ under the condition $d(O, Q_i) = r$ can be bounded by*

$$f_{\Delta|R_i=r}(\delta) \le \frac{\kappa}{\sqrt{r}} \cdot \left(\ln\left(\frac{1}{\delta}\right)+1\right) .$$

If we integrate the conditional density given in Lemma 14 over all values which the distances R_1 and R_2 can take, we obtain the following lemma.

Lemma 16. *Let τ be an arbitrary distance with $0 \le \tau \le \sqrt{2}$. For a sufficiently large constant κ, the conditional density $f_{\Delta|T=\tau}(\delta)$ of Δ for $\delta \ge 0$ under the condition $T = \tau$ can be bounded by*

$$f_{\Delta|T=\tau}(\delta) \le \frac{\kappa}{\tau} \cdot \left(\ln\left(\frac{1}{\delta}\right)+1\right) .$$

The following lemma gives a bound on the conditional density of the random variable Z_i, given the distance T. It follows by integrating the conditional density given in Lemma 13 over all values which the distance R can take.

Lemma 17. *Let τ be an arbitrary distance with $0 \le \tau \le \sqrt{2}$. For a sufficiently large constant κ and for $i \in \{1,2\}$, the conditional density $f_{Z_i|T=\tau}$ of Z_i under the condition $T = \tau$ can be bounded by*

$$f_{Z_i|T=\tau}(z) \le \frac{\kappa}{\sqrt{\tau^2 - z^2}}$$

if $|z| \leq \tau$. Since Z_i takes only values in the interval $[-\tau, \tau]$, the conditional density $f_{Z_i|T=\tau}(z)$ is 0 *for $z \notin [-\tau, \tau]$.*

If we consider only a single 2-change, then it is not necessary to consider the conditional density of Δ. In order to get rid of the last condition, we can take the conditional density from Lemma 16 and integrate over all values T can take. This yields the following lemma.

Lemma 18. *For a sufficiently large constant κ, the density $f_\Delta(\delta)$ of Δ for $\delta \geq 0$ can be bounded by*

$$f_\Delta(\delta) \leq \kappa \left(\ln\left(\frac{1}{\delta}\right) + 1 \right) .$$

4.2.2 Simplified Random Experiments

In the previous section we did not analyze the random experiment that really takes place. Instead of choosing the points according to the given density functions, we simplified their distributions by placing point O in the origin and by giving the other points P, Q_1, and Q_2 uniform distributions centered around the origin. In our input model, however, each of these points is described by a density function over the unit square. We consider the probability of the event $\Delta \in [0, \varepsilon]$ in both the original input model as well as in the simplified random experiment. In the following, we denote this event by $\mathcal{E}$. We claim that the simplified random experiment that we analyze is only slightly dominated by the original random experiment, in the sense that the probability of the event $\mathcal{E}$ in the simplified random experiment is smaller by at most some factor depending on ϕ.

In order to compare the probabilities in the original and in the simplified random experiment, consider the original experiment and assume that the point O has position $(x, y) \in [0,1]^2$. Then one can identify a region $\mathcal{R}_{(x,y)} \subseteq \mathbb{R}^6$ with the property that the event $\mathcal{E}$ occurs if and only if the random vector (P, Q_1, Q_2) lies in $\mathcal{R}_{(x,y)}$. No matter of how the position (x, y) of O is chosen, this region always has the same shape, only its position is shifted. Let $\mathcal{V} = \sup_{(x,y)\in[0,1]^2} \mathrm{vol}(\mathcal{R}_{(x,y)} \cap [0,1]^6)$. Then the probability of $\mathcal{E}$ can be bounded from above by $\phi^3 \cdot \mathcal{V}$ in the original random experiment. One can easily see that

$$\mathcal{R}_{(x,y)} \cap [0,1]^6 = \mathcal{R}_{(0,0)} \cap ([-x, 1-x] \times [-y, 1-y])^3 \subseteq \mathcal{R}_{(0,0)} \cap [-1,1]^6 .$$

Hence for $\mathcal{V}' = \mathrm{vol}(\mathcal{R}_{(0,0)} \cap [-1,1]^6)$ we have $\mathcal{V} \leq \mathcal{V}'$. Observe that the probability of $\mathcal{E}$ in the simplified random experiment can be bounded from below by $(1/(2\pi))^3 \cdot \mathcal{V}'$ since the circle centered around the origin with radius $\sqrt{2}$ contains the square $[-1,1]^2$ completely. Hence, the probability of $\mathcal{E}$ in the simplified random experiment is smaller by at most a factor of $(2\pi\phi)^3$ compared to the original random experiment.

Taking into account this factor and using Lemma 18 and a union bound over all possible 2-changes yields the following lemma about the improvement of a single 2-change.

Lemma 19. *The probability that there exists an improving 2-change whose improvement is at most ε is bounded from above by*

$$O\left(n^4 \cdot \varepsilon \cdot \left(\log\left(\frac{1}{\varepsilon}\right) + 1 \right) \cdot \phi^3 \right) .$$

Using similar arguments as in the proof of Theorem 6, yields the following upper bound on the expected number of 2-changes.

Theorem 20. *Starting with an arbitrary tour, the expected number of steps performed by 2-Opt on ϕ-perturbed L_2 instances is $O(n^7 \cdot \log^2 n \cdot \phi^3)$.*

Pairs of Type 1. In order to improve upon Theorem 20, we consider pairs of linked 2-changes as in the analysis of ϕ-perturbed L_1 instances. Since our analysis of pairs of linked 2-changes is based on the analysis of a single 2-change that we presented in the previous section, we also have to consider simplified random experiments when analyzing pairs of 2-changes. For a fixed pair of type 1, we assume that point v_3 is chosen to be the origin and the other points v_1, v_2, v_4, v_5, and v_6 are chosen uniformly from a circle with radius $\sqrt{2}$ centered at v_3. Let $\mathcal{E}$ denote the event that both Δ_1 and Δ_2 lie in the interval $[0, \varepsilon]$, for some given ε. With the same arguments as above, one can see that the probability of $\mathcal{E}$ in the simplified random experiment is smaller compared to the original experiment by at most a factor of $(2\pi\phi)^5$.

Pairs of Type 2. For a fixed pair of type 2, we consider the simplified random experiment in which v_2 is placed in the origin and the other points v_1, v_3, v_4, and v_5 are chosen uniformly from a circle with radius $\sqrt{2}$ centered at v_2. In this case the probability in the simplified random experiment is smaller by at most a factor of $(2\pi\phi)^4$.

4.2.3 Analysis of Pairs of Linked 2-Changes

Finally, we can prove Lemmas 11 and 12.

Proof of Lemma 11. We consider the simplified random experiment in which v_3 is chosen to be the origin and the other points are drawn uniformly at random from a circle with radius $\sqrt{2}$ centered at v_3. If the position of the point v_1 is fixed, then the events $\Delta_1 \in [0, \varepsilon]$ and $\Delta_2 \in [0, \varepsilon]$ are independent as only the vertices v_1 and v_3 appear in both the first and the second step. In fact, because the densities of the points v_2, v_4, v_5, and v_6 are rotationally symmetric, the concrete position of v_1 is not important in our simplified random experiment anymore, but only the distance R between v_1 and v_3 is of interest.

For $i = 1$ and $i = 2$, we determine the conditional probability of the event $\Delta_i \in [0, \varepsilon]$ under the condition that the distance $d(v_1, v_3)$ is fixed with the help of Lemma 15 and obtain, for a sufficiently large constant κ,

$$\begin{aligned} \mathbf{Pr}\left[\Delta_i \in [0,\varepsilon] \mid d(v_1,v_3) = r\right] &= \int_{\delta=0}^{\varepsilon} f_{\Delta_i \mid d(v_1,v_3)=r}(\delta)\, d\delta \\ &\leq \frac{\kappa}{\sqrt{r}} \cdot \varepsilon \cdot \left(\ln\left(\frac{1}{\varepsilon}\right) + 1\right) \quad . \end{aligned}$$

Since for fixed distance $d(v_1, v_3)$ the random variables Δ_1 and Δ_2 are independent, we obtain

$$\begin{aligned} \mathbf{Pr}\left[\Delta_1, \Delta_2 \in [0,\varepsilon] \mid d(v_1,v_3) = r\right] &\leq \frac{\kappa^2}{r} \cdot \varepsilon^2 \cdot \left(\ln\left(\frac{1}{\varepsilon}\right) + 1\right)^2 \\ &\leq \frac{\kappa'}{r} \cdot \varepsilon^2 \cdot \left(\ln^2\left(\frac{1}{\varepsilon}\right) + 1\right) \qquad (2) \end{aligned}$$

for a sufficiently large constant κ'. For $r \in [0, \sqrt{2}]$, the density $f_{d(v_1,v_3)}$ of the random variable $d(v_1, v_3)$ in the simplified random experiment is $f_{d(v_1,v_3)}(r) = r$. Combining this observation with the bound given in (2) yields

$$\begin{aligned}\mathbf{Pr}\left[\Delta_1, \Delta_2 \in [0,\varepsilon]\right] &= \int_0^{\sqrt{2}} r \cdot \mathbf{Pr}\left[\Delta_1, \Delta_2 \in [0,\varepsilon] \mid d(v_1, v_3) = r\right] \, dr \\ &\leq \sqrt{2}\kappa' \cdot \varepsilon^2 \cdot \left(\ln^2\left(\frac{1}{\varepsilon}\right) + 1\right) \quad .\end{aligned}$$

There are $O(n^6)$ different pairs of type 1, hence a union bound over all of them concludes the proof of the lemma when taking into account the factor $(2\pi\phi)^5$ that results from considering the simplified random experiment. □

Proof of Lemma 12. We consider the simplified random experiment in which v_2 is chosen to be the origin and the other points are drawn uniformly at random from a circle with radius $\sqrt{2}$ centered at v_2. In contrast to pairs of type 1, pairs of type 2 exhibit larger dependencies as only 5 different vertices are involved in these pairs. Fix one pair of type 2. The two 2-changes share the whole triangle consisting of v_1, v_2 and v_3. In the second step, there is only one new vertex, namely v_5. Hence, there is not enough randomness contained in a pair of type 2 such that Δ_1 and Δ_2 are nearly independent as for pairs of type 1.

We start by analyzing pairs of type 2 a). First we analyze the probability that Δ_1 lies in the interval $[0, \varepsilon]$. After that we analyze the probability that Δ_2 lies in the interval $[0, \varepsilon]$ under the condition that the points v_1, v_2, v_3 and v_4 have already been chosen. In the analysis of the second step we cannot make use of the fact that the distances $d(v_1, v_3)$ and $d(v_2, v_3)$ are random variables anymore since we exploited their randomness already in the analysis of the first step. The only distances whose randomness we can exploit are the distances $d(v_1, v_5)$ and $d(v_2, v_5)$. We pessimistically assume that the distances $d(v_1, v_3)$ and $d(v_2, v_3)$ have been chosen by an adversary. This means the adversary can determine an interval of length ε in which the random variable $d(v_2, v_5) - d(v_1, v_5)$ must lie in order for Δ_2 to lie in $[0, \varepsilon]$.

Due to Lemma 15, the probability of the event $\Delta_1 \in [0, \varepsilon]$ under the condition $d(v_1, v_2) = r$ can be bounded by

$$\mathbf{Pr}\left[\Delta_1 \in [0,\varepsilon] \mid d(v_1, v_2) = r\right] \leq \frac{\kappa}{\sqrt{r}} \cdot \varepsilon \cdot \left(\ln\left(\frac{1}{\varepsilon}\right) + 1\right) \tag{3}$$

for a sufficiently large constant κ. Due to Lemma 17, the conditional density of the random variable $Z = d(v_2, v_5) - d(v_1, v_5)$ under the condition $d(v_1, v_2) = r$ can be bounded by

$$f_{Z \mid d(v_1,v_2)=r}(z) \leq \frac{\kappa}{\sqrt{r^2 - z^2}}$$

for $|z| \leq r$.

The intervals the adversary can specify which have the highest probability of Z falling into them are $[-r, -r+\varepsilon]$ and $[r-\varepsilon, r]$. Hence, the conditional probability of the event $\Delta_2 \in [0, \varepsilon]$ under the condition $d(v_1, v_2) = r$ and for fixed points v_3 and v_4 is bounded from above by

$$\int_{z=\max\{r-\varepsilon,-r\}}^{r} \frac{\kappa}{\sqrt{r^2 - z^2}} \, dz \leq \frac{\kappa}{\sqrt{r}} \cdot \int_{z=\max\{r-\varepsilon,-r\}}^{r} \frac{1}{\sqrt{r - |z|}} \, dz \leq \frac{\kappa' \sqrt{\varepsilon}}{\sqrt{r}} \quad ,$$

for a sufficiently large constant κ'. Combining this inequality with (3) yields

$$\mathbf{Pr}\left[\Delta_1, \Delta_2 \in [0,\varepsilon] \mid d(v_1, v_2) = r\right] \leq \frac{\kappa \cdot \kappa'}{r} \cdot \varepsilon^{3/2} \cdot \left(\ln\left(\frac{1}{\varepsilon}\right) + 1\right) \quad .$$

In order to get rid of the condition $d(v_1,v_2)=r$, we integrate over all possible values the random variable $d(v_1,v_2)$ can take. For $0 \le r \le \sqrt{2}$, the density of $f_{d(v_1,v_2)}(r)$ equals r. Hence, we get

$$\mathbf{Pr}\left[\Delta_1,\Delta_2 \in [0,\varepsilon]\right] \le \int_{r=0}^{\sqrt{2}} r \cdot \frac{\kappa \cdot \kappa'}{r} \cdot \varepsilon^{3/2} \cdot \left(\ln\left(\frac{1}{\varepsilon}\right)+1\right)\, dr = O\left(\varepsilon^{3/2}\left(\ln\left(\frac{1}{\varepsilon}\right)+1\right)\right) \quad .$$

Applying a union bound over all $O(n^5)$ possible pairs of type 2 concludes the proof when one takes into account the factor $(2\pi\phi)^4$ due to considering the simplified random experiment.

For pairs of type 2 b), the situation looks somewhat similar. We analyze the first step and in the second step, we can only exploit the randomness of the distances $d(v_2,v_5)$ and $d(v_3,v_5)$. Due to Lemma 16, the probability of the event $\Delta_1 \in [0,\varepsilon]$ under the condition $d(v_2,v_3)=\tau$ can be bounded by

$$\mathbf{Pr}\left[\Delta_1 \in [0,\varepsilon] \mid d(v_2,v_3)=\tau\right] \le \frac{\kappa}{\tau} \cdot \varepsilon \cdot \left(\ln\left(\frac{1}{\varepsilon}\right)+1\right) \tag{4}$$

for a sufficiently large constant κ.

Due to Lemma 17, the conditional density of the random variable $Z = d(v_2,v_5) - d(v_3,v_5)$ under the condition $d(v_2,v_3)=\tau$ can be bounded by

$$f_{Z|d(v_2,v_3)=\tau}(z) \le \frac{\kappa}{\sqrt{\tau^2 - z^2}}$$

for $|z| \le \tau$.

Similar to pairs of type 2 a), the conditional probability of the event $\Delta_2 \in [0,\varepsilon]$ under the condition $d(v_2,v_3)=\tau$ and for fixed points v_1 and v_3 is bounded from above by

$$\int_{z=\max\{\tau-\varepsilon,-\tau\}}^{\tau} \frac{\kappa}{\sqrt{\tau^2-z^2}}\, dz \le \frac{\kappa}{\sqrt{\tau}} \cdot \int_{z=\max\{\tau-\varepsilon,-\tau\}}^{\tau} \frac{1}{\sqrt{\tau-|z|}}\, dz \le \frac{\kappa'\sqrt{\varepsilon}}{\sqrt{\tau}}$$

for a sufficiently large constant κ'. Combining this inequality with (4) yields

$$\mathbf{Pr}\left[\Delta_1,\Delta_2 \in [0,\varepsilon] \mid d(v_2,v_3)=\tau\right] \le \frac{\kappa \cdot \kappa'}{\tau^{3/2}} \cdot \varepsilon^{3/2} \cdot \left(\ln\left(\frac{1}{\varepsilon}\right)+1\right) \quad .$$

In order to get rid of the condition $d(v_2,v_3)=\tau$, we integrate over all possible values the random variable $d(v_2,v_3)$ can take. For $0 \le \tau \le \sqrt{2}$, the density of $f_{d(v_2,v_3)}(\tau)$ equals τ. Hence, we get

$$\mathbf{Pr}\left[\Delta_1,\Delta_2 \in [0,\varepsilon]\right] \le \int_{\tau=0}^{\sqrt{2}} \frac{\kappa \cdot \kappa'}{\sqrt{\tau}} \cdot \varepsilon^{3/2} \cdot \left(\ln\left(\frac{1}{\varepsilon}\right)+1\right)\, dr = O\left(\varepsilon^{3/2}\left(\ln\left(\frac{1}{\varepsilon}\right)+1\right)\right) \quad .$$

Applying a union bound over all $O(n^5)$ possible pairs of type 2 concludes the proof when one takes into account the factor $(2\pi\phi)^4$ due to considering the simplified random experiment. $\square$

4.2.4 The Expected Number of 2-Changes on ϕ-perturbed L_2 instances

Based on Lemmas 8, 11, and 12, we are now able to prove part b) of Theorem 2.

Proof of Theorem 2 b). We use the same notations as in the proof of part a) of the theorem. For $t > n^2$, we have $t/6 - 5n(n-1)/48 > t/16$ and hence due to Lemmas 11 and 12,

$$\begin{aligned}\mathbf{Pr}\left[T \geq t\right] &\leq \mathbf{Pr}\left[\Delta_{\min}^{(1)} \leq \frac{16\sqrt{2}n}{t}\right] + \mathbf{Pr}\left[\Delta_{\min}^{(2)} \leq \frac{16\sqrt{2}n}{t}\right] \\ &= O\left(\min\left\{\frac{n^8(\log^2(t/n)+1)}{t^2}\phi^5, 1\right\}\right) + O\left(\min\left\{\frac{n^{13/2}(\log(t/n)+1)}{t^{3/2}}\phi^4, 1\right\}\right) .\end{aligned}$$

This implies the following bound on the expected length of the longest path in the state graph

$$\mathbf{E}\left[T\right] = n^2 + \sum_{t=1}^{n!} O\left(\min\left\{\frac{n^8 \log^2 t}{t^2}\phi^5, 1\right\}\right) + O\left(\min\left\{\frac{n^{13/2}\log t}{t^{3/2}}\phi^4, 1\right\}\right) .$$

Splitting the sums at $t = n^4 \cdot \log(n\phi) \cdot \phi^{5/2}$ and $t = n^{13/3} \cdot \log^{2/3}(n\phi) \cdot \phi^{8/3}$, respectively, yields

$$\mathbf{E}\left[T\right] = O\left(n^4 \cdot \log(n\phi) \cdot \phi^{5/2}\right) + O\left(n^{13/3} \cdot \log^{2/3}(n\phi) \cdot \phi^{8/3}\right) .$$

This concludes the proof of part b) of the theorem. $\square$

Using the same observations as in the proof of part a) of Theorem 3 yields also part b).

Proof of Theorem 3 b). Since the initial tour has length $O(\sqrt{n} \cdot \log n)$, we obtain for an appropriate constant c and $t > n^2$,

$$\begin{aligned}\mathbf{Pr}\left[T \geq t\right] &\leq \mathbf{Pr}\left[\Delta_{\min}^{(1)} \leq \frac{c \cdot \sqrt{n} \cdot \log n}{t}\right] + \mathbf{Pr}\left[\Delta_{\min}^{(2)} \leq \frac{c \cdot \sqrt{n} \cdot \log n}{t}\right] \\ &= O\left(\min\left\{\frac{n^7 \cdot \log^2 n \cdot \log^2 t \cdot \phi^5}{t^2}, 1\right\}\right) + O\left(\min\left\{\frac{n^{23/4} \cdot \log^{3/2} n \cdot \log t \cdot \phi^4}{t^{3/2}}, 1\right\}\right) .\end{aligned}$$

This implies

$$\mathbf{E}\left[T\right] = n^2 + \sum_{t=1}^{n!} O\left(\min\left\{\frac{n^7 \cdot \log^2 n \cdot \log^2 t \cdot \phi^5}{t^2}, 1\right\}\right) + O\left(\min\left\{\frac{n^{23/4} \cdot \log^{3/2} n \cdot \log t \cdot \phi^4}{t^{3/2}}, 1\right\}\right) .$$

Splitting the sums at $t = n^{7/2} \cdot \log^2(n\phi) \cdot \phi^{5/2}$ and $t = n^{23/6} \cdot \log^{5/3}(n\phi) \cdot \phi^{8/3}$, respectively, yields

$$\mathbf{E}\left[T\right] = O\left(n^{7/2} \cdot \log^2(n\phi) \cdot \phi^{5/2}\right) + O\left(n^{23/6} \cdot \log^{5/3}(n\phi) \cdot \phi^{8/3}\right) .$$

This concludes the proof of part b) of the theorem. $\square$

4.3 General Graphs

In this section, we analyze the expected number of 2-changes on ϕ-perturbed graphs. Observe that ϕ-perturbed graphs contain more randomness than ϕ-perturbed L_1 or L_2 instances since each edge length is a random variable that is independent of the other edge lengths. Similar to L_1 instances, it is easy to obtain a polynomial bound on the expected number of steps by just estimating the improvement made by any of the 2-changes. For L_1 and L_2 instances we improved this simple bound by considering pairs of linked 2-changes. For ϕ-perturbed graphs we pursue the same approach but due to the larger amount of randomness, we are now able to consider not only pairs of linked steps but longer sequences of linked steps.

We know that every sequence of steps which contains k distinct 2-changes shortens the tour by at least $\Delta^{(k)} = \Delta_{\min}^{(1)} + \cdots + \Delta_{\min}^{(k)}$, where $\Delta_{\min}^{(i)}$ denotes the i-th smallest improvement. This observation alone, however, does not suffice to improve the simple bound substantially. Instead we show that one can identify in every long enough sequence of consecutive 2-changes, subsequences that are *linked*, where a sequence $S_1, \dots, S_k$ of 2-changes is called *linked* if for every $i < k$, there exists an edge which is added to the tour in step S_i and removed from the tour in step S_{i+1}. We analyze the smallest improvement of a linked sequence that consists of k distinct 2-Opt steps. Obviously, this improvement must be at least $\Delta^{(k)}$ as in the worst-case, the linked sequence consists of the k smallest improvements. Intuitively, one can hope that it is much larger than $\Delta^{(k)}$ since it is unlikely that the k smallest improvements form a sequence of linked steps. We show that this is indeed the case and use this result to prove the desired upper bound on the expected number of 2-changes.

To make this more precise, we introduce the notion of *witness sequences*, i. e., linked sequences of 2-changes that satisfy some additional technical properties. We show that the smallest total improvement made by a witness sequence yields an upper bound on the running time, that is, whenever the 2-Opt heuristic needs many local improvement steps to find a locally optimal solution, there must be a witness sequence whose total improvement is small. Furthermore, our probabilistic analysis reveals that it is unlikely that there exists a witness sequence whose total improvement is small. Together, these results yield the desired bound on the expected number of 2-changes.

4.3.1 Definition of Witness Sequences

In this section, we give a formal definition of the notion of a *k-witness sequence.* As mentioned above, a witness sequence $S_1, \dots, S_k$ has to be linked, i. e., for $i < k$, there must exist an edge which is added to the tour in step S_i and removed from the tour in step S_{i+1}. Let m denote the number of edges in the graph. Then there are at most $4^{k-1} \cdot m^{k+1}$ such linked sequences as there are at most m^2 different choices for S_1 and once S_i is fixed, there are at most $4m$ different choices for S_{i+1}. For a fixed 2-change the probability that it is an improvement by at most ε is bounded by $\varepsilon\phi$. We would like to show an upper bound of $(\varepsilon\phi)^k$ on the probability that each step in the witness sequence $S_1, \dots, S_k$ is an improvement by at most ε. For general linked sequences, this is not true as the steps can be dependent in various ways. Hence, we need to introduce further restrictions on witness sequences.

In the following definitions, we assume that a linked sequence $S_1, \dots, S_k$ of 2-changes is given. In step S_i the edges e_{i-1} and f_{i-1} are removed from the tour and the edges e_i and g_i are added to the tour, i. e., for $i < k$, e_i denotes an edge added to the tour in step S_i and removed from the tour in step S_{i+1}. These definitions are illustrated in Figure 6.

Definition 21 (witness sequences of type 1)**.** *If for every $i \le k$, the edge e_i does not occur in any step S_j with $j < i$, then $S_1, \dots, S_k$ is called a k-*witness sequence of type 1.

Intuitively, witness sequences of type 1 possess enough randomness as every step introduces an edge which has not been seen before. Based on this observation, we prove in Lemma 24 that one obtains the desired bound of $(\varepsilon\phi)^k$ on the probability that every step is an improvement by at most ε for these sequences.

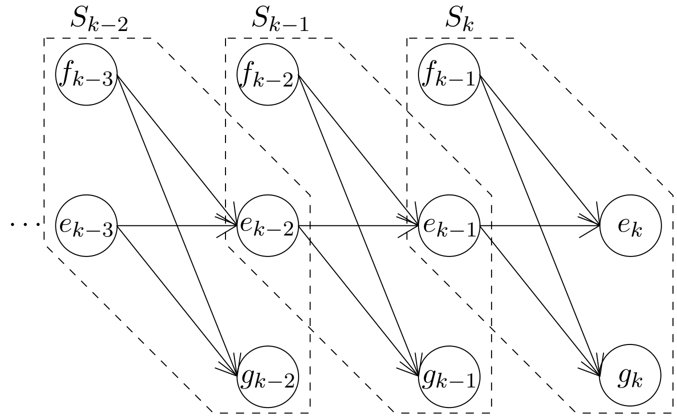

Figure 6: Illustration of the notations used in Definitions 21, 22, and 23. Every node in the DAG corresponds to an edge involved in one of the 2-changes. An arc from a node u to a node v indicates that in one of the 2-changes, the edge corresponding to node u is removed from the tour and the edge corresponding to node v is added to the tour. Hence, every arc is associated with one 2-change.

Definition 22 (witness sequences of type 2)**.** *Assume that for every $i \leq k-1$, the edge e_i does not occur in any step S_j with $j < i$. If the edges e_k, g_k and f_{k-1} occur in steps S_j with $j < k$ and if both endpoints of f_{k-2} occur in steps S_j with $j < k-1$, then $S_1, \ldots, S_k$ is called a* k-witness sequence of type 2*.*

In general, witness sequences of type 2 possess not enough randomness to obtain the desired bound of $(\varepsilon\phi)^k$ as the last step S_k is dependent of the steps before. If we neglect the last step, we obtain a bound of $(\varepsilon\phi)^{k-1}$ on the probability that each of the first $k-1$ steps is an improvement by at most ε. However, due to the additional restrictions, there are much less than $4^{k-1}m^{k+1}$ witness sequences of type 2. As the edges e_k and g_k must be chosen among those edges that occur in the steps S_j with $j < k$, there are only $O(k^2)$ choices for step S_k. Furthermore, as the two endpoints of f_{k-2} must be chosen among those vertices that occur in steps S_j with $j < k-1$, there are only $O(k^2)$ choices for step S_{k-1}, too. Altogether this implies that the number of k-witness sequences of type 2 can be upper bounded by $O(k^4 \cdot m^{k-1})$ which makes up for the missing $\varepsilon\phi$ term in the bound on the probability that each step is an improvement by at most ε.

Definition 23 (witness sequences of type 3)**.** *Assume that for every $i \leq k-1$, the edge e_i does not occur in any step S_j with $j < i$. If the edges e_k and g_k occur in steps S_j with $j < k$ and if f_{k-1} does not occur in any step S_j with $j < k$, then $S_1, \ldots, S_k$ is called a* k-witness sequence of type 3*.*

Analogously to witness sequences of type 1, witness sequences of type 3 possess enough randomness to bound the probability that every step is an improvement by at most ε by $(\varepsilon\phi)^k$ as also the last step introduces a new edge, namely f_{k-1}. Analogously to witness sequences of type 2, one can argue that the number of witness sequences of type 3 is only $O(k^2 m^k)$ as there are only $O(k^2)$ choices for step S_k.

4.3.2 Improvement Made by Witness Sequences

In this section, we analyze the probability that there exists a k-witness sequence in which every step is an improvement by at most ε.

Lemma 24. *The probability that there exists a k-witness sequence in which every step is an improvement by at most ε*

a) of type 1 is bounded from above by $4^{k-1}m^{k+1}(\varepsilon\phi)^k$.

b) of type 2 is bounded from above by $k^4 4^k m^{k-1}(\varepsilon\phi)^{k-1}$.

c) of type 3 is bounded from above by $k^2 4^k m^k(\varepsilon\phi)^k$.

Proof. We use a union bound to estimate the probability that there exists a witness sequence in which every step is a small improvement.

a) We consider k-witness sequences of type 1 first. As already mentioned in the previous section, the number of such sequences is bounded by $4^{k-1}m^{k+1}$ as there are at most m^2 choices for the first step S_1, and once S_i is fixed, there are at most $4m$ choices for step S_{i+1}. The number $4m$ follows since if S_i is fixed, there are two choices for the edge added to the tour in step S_i and removed from the tour in step S_{i+1}, there are at most m choices for the other edge removed in step S_{i+1}, and once these edges are determined, there are two possible 2-Opt steps in which these edges are removed from the tour.

Now fix an arbitrary k-witness sequence $S_1, \ldots, S_k$ of type 1. We use the same notations as in Figure 6 to denote the edges involved in this sequence. In the first step, the edges e_0 and f_0 are replaced by the edges e_1 and g_1. We assume that the lengths of the edges e_0, f_0, and g_1 are determined by an adversary. Analogously to (1), the improvement of step S_1 can be expressed as simple linear combination of the lengths of the involved edges. Hence, for fixed lengths of e_0, f_0, and g_1, the event that S_1 is an improvement by at most ε corresponds to the event that the length $d(e_1)$ of e_1 lies in some fixed interval of length ε. Since the density of $d(e_1)$ is bounded by ϕ, the probability that $d(e_1)$ takes a value in the given interval is bounded by $\varepsilon\phi$. Now consider a step S_i and assume that arbitrary lengths for the edges e_j and f_j with $j < i$ and for g_j with $j \leq i$ are chosen. Since the edge e_i is not involved in any step S_j with $j < i$, its length is not determined. Hence, analogously to the first step, the probability that step S_i is an improvement by at most ε is bounded by $\varepsilon\phi$ independent of the improvements of the steps S_j with $j < i$. Applying this argument to every step S_i yields the desired bound of $(\varepsilon\phi)^k$.

b) In witness sequences of type 2, there are at most m^2 choices for step S_1. Analogously to witness sequences of type 1, the number of possible choices for S_i with $1 < i < k-1$ is at most $4m$. The number of different vertices involved in steps S_j with $j < k-1$ is at most $4 + 2(k-3) < 2k$ as the first step introduces four new vertices and every other step at most 2. Since the endpoints of the edge f_{k-2} must be chosen among those vertices which have been involved in the steps S_j with $j < k-1$, there are less than $4k^2$ possible choices for step S_{k-1}. Furthermore, the number of edges involved in steps S_i with $i < k$ is at most $4 + 3(k-2) < 3k$ as the first step introduces four new edges and every other step at most three. Hence, there are at most $9k^2$ choices for step S_k. Altogether, this implies that the number of different k-witness sequences of type 2 is bounded by $36 \cdot k^4 4^{k-3} m^{k-1} < k^4 4^k m^{k-1}$.

For a fixed witness sequence of type 2, applying the same arguments as for witness sequences of type 1, yields a probability of at most $(\varepsilon\phi)^{k-1}$ that every step is an improvement by at most ε. This bound is worse by a factor of $\varepsilon\phi$ compared to the bound for witness sequences of type 1 since e_k has been involved in a step S_i with $i < k$, and hence, the improvement of step S_k can be completely determined by the improvements of the steps S_i with $i < k$.

c) With the arguments used above, one can easily see that the number of k-witness sequences of type 3 is bounded by $9k^2 4^{k-2} m^k < k^2 4^k m^k$. Furthermore, similar to witness sequences of type 1, we can bound the probability that a fixed k-witness sequence of type 3 consists only of improvements by at most ε by $(\varepsilon\phi)^k$ since the last step introduces an edge which does not occur in the steps S_i with $i < k$, namely f_{k-1}.

□

Definition 25. *In the following, we use the term k-*witness sequence *to denote a k-witness sequence of type 1 or an i-witness sequence of type 2 or 3 with $i \le k$. We call a k-witness sequence* improving *if every 2-change in the sequence is an improvement. Moreover, by $\Delta_{\mathrm{ws}}^{(k)}$ we denote the smallest total improvement made by any improving k-witness sequence.*

Due to Lemma 24, it is unlikely that there exists an improving witness sequence whose total improvement is small.

Corollary 26. *For arbitrarily chosen k and $0 < \varepsilon \le 1/(4^{k-2} m^{k-1} \phi^{k-2})^{1/(k-2)}$,*

$$\mathbf{Pr}\left[\Delta_{\mathrm{ws}}^{(k)} \le \varepsilon\right] \le 10k^5 (4m\varepsilon\phi)^2 \ .$$

Proof. Due to Lemma 24 and the fact that witness sequences of type 2 or 3 must consist of more than two steps, applying a union bound yields the following bound on the probability that there exists an improving k-witness sequence whose total improvement is at most ε

$$\begin{aligned}\mathbf{Pr}\left[\Delta_{\mathrm{ws}}^{(k)} \le \varepsilon\right] &\le 4^{k-1} m^{k+1} (\varepsilon\phi)^k + \sum_{i=3}^{k} i^4 4^i m^{i-1} (\varepsilon\phi)^{i-1} + \sum_{i=3}^{k} i^2 4^i m^i (\varepsilon\phi)^i \\ &\le 4^{k-1} m^{k+1} (\varepsilon\phi)^k + 4k^4 \sum_{i=3}^{k} (4m\varepsilon\phi)^{i-1} + k^2 \sum_{i=3}^{k} (4m\varepsilon\phi)^i \ .\end{aligned}$$

The condition $\varepsilon < 1/(4^{k-2} m^{k-1} \phi^{k-2})^{1/(k-2)}$ implies $4m\varepsilon\phi < 1$. Hence we can bound the sum by

$$\begin{aligned}\mathbf{Pr}\left[\Delta_{\mathrm{ws}}^{(k)} \le \varepsilon\right] &\le 4^{k-1} m^{k+1} (\varepsilon\phi)^k + 4k^5 (4m\varepsilon\phi)^2 + k^3 (4m\varepsilon\phi)^3 \\ &\le 4^{k-1} m^{k+1} (\varepsilon\phi)^k + 5k^5 (4m\varepsilon\phi)^2 \ ,\end{aligned}$$

which implies the corollary since for $\varepsilon \le 1/(4^{k-2} m^{k-1} \phi^{k-2})^{1/(k-2)}$, the second term in the sum is at least as large as first one. □

4.3.3 Finding Witness Sequences

In the previous section, we have shown an upper bound on the probability that there exists an improving k-witness sequence whose total improvement is small. In this section, we show that in every long enough sequence of consecutive 2-changes, one can identify disjoint k-witness sequences, where both k and the number of witness sequences grow with the length of the sequence. This way, we obtain a lower bound on the improvement made by any long enough sequence of consecutive 2-changes in terms of $\Delta_{\mathrm{ws}}^{(k)}$.

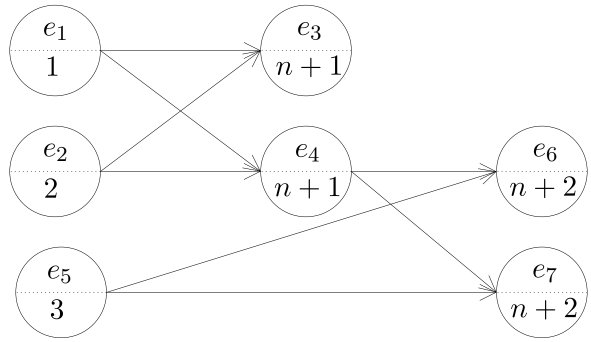

Figure 7: A witness DAG: Initially the edges e_1, e_2, e_5 and $n-3$ other edges not shown in this figure are contained in the tour. First, e_1 and e_2 are exchanged with e_3 and e_4, then e_4 and e_5 are exchanged with e_6 and e_7.

Lemma 27. *Let k be chosen arbitrarily, and let $S_1, \dots, S_t$ denote a sequence of consecutive 2-changes performed by the 2-Opt heuristic with $t > n2^k$. The sequence $S_1, \dots, S_t$ shortens the tour by at least $t/4^{k+3} \cdot \Delta_{\mathrm{ws}}^{(k)}$.*

Basically we have to show that one can find $t/4^{k+3}$ disjoint k-witness sequences in the given sequence of consecutive 2-changes. Therefore, we first introduce a so-called *witness DAG* (directed acyclic graph) which represents the sequence $S_1, \dots, S_t$ of 2-changes. In order to not confuse the constructed witness DAG W with the input graph G, we use the terms *nodes* and *arcs* when referring to the DAG W and the terms *vertices* and *edges* when referring to G. Nodes of W correspond to edges of G combined with a time stamp. The construction is started by adding the edges of the initial tour as nodes into W. These nodes get the time stamps $1, \dots, n$ in an arbitrary order. Then the sequence $S_1, \dots, S_t$ is processed step by step. Assume that the steps $S_1, \dots S_{i-1}$ have already been processed and that step S_i is to be processed next. Furthermore, assume that in step S_i the edges e_{i-1} and f_{i-1} are exchanged with the edges e_i and g_i. Since the edges e_{i-1} and f_{i-1} are contained in the tour after the steps $S_1, \dots, S_{i-1}$, there are nodes in W corresponding to these edges. Let u_1 and u_2 denote the nodes with the most recent time stamps corresponding to e_{i-1} and f_{i-1}, respectively. We create two new nodes u_3 and u_4 corresponding to edges e_i and f_i, each with time stamp $n+i$. Finally, four new arcs are added to W, namely the arcs (u_1, u_3), (u_1, u_4), (u_2, u_3), and (u_2, u_4). We refer to these four arcs as *twin arcs*. Observe that each node in W has indegree and outdegree at most 2. We call the resulting DAG W a *t-witness DAG*. An example of such a DAG is depicted in Figure 7.

By the *height* of a node u, we denote the length of a shortest path from u to a leaf of W. After the witness DAG has been constructed completely, we associate with each node u with height at least k a sub-DAG of W. The sub-DAG W_u associated with such a node u is the induced sub-DAG of those nodes of W that can be reached from u by traversing at most k arcs. The following two lemmas imply Lemma 27.

Lemma 28. *For every sub-DAG W_u, the 2-changes represented by the arcs in W_u yield a total improvement of at least $\Delta_{\mathrm{ws}}^{(k)}$.*

Lemma 29. *For $t > n4^{k+2}$, every t-witness DAG contains at least $t/4^{k+3}$ nodes u whose corresponding sub-DAGs W_u are pairwise disjoint.*

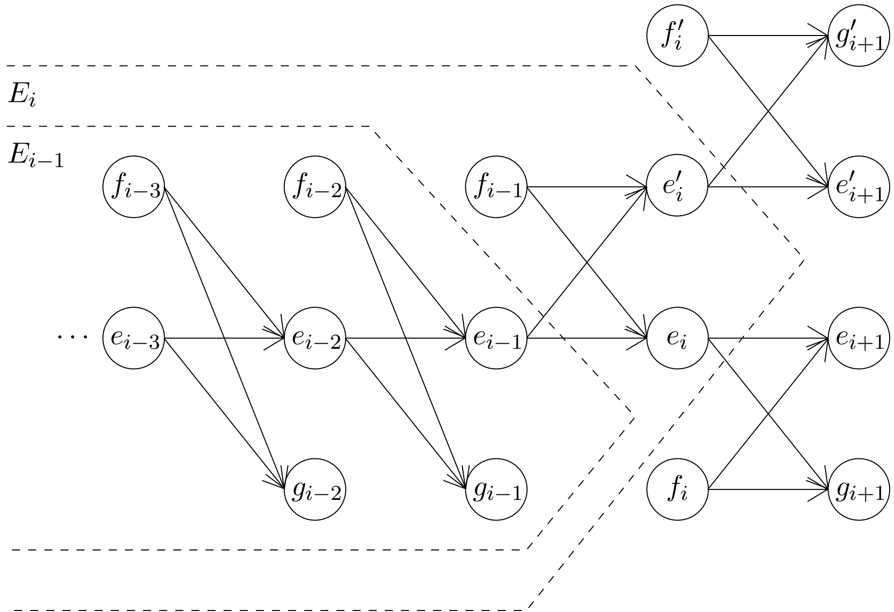

Figure 8: Construction of a path in the witness DAG: The path has been constructed up to step S_i and now it has to be decided whether to continue it along e_i or e'_i.

Proof of Lemma 28. It remains to prove Lemma 28. Assume that a sub-DAG W_u with root u is given. Since node u has height k, one can identify 2^{k-2} distinct sequences of linked 2-changes of length k in the sub-DAG W_u. In the following, we show that at least one of these sequences is a k-witness sequence or a sequence whose total improvement is as large as the total improvement of one of the k-witness sequences. We give a recursive algorithm that constructs such a sequence step by step. It is initialized with the sequence which consists only of the first step S_1 that is represented by the two outgoing arcs of the root u.

Assume that the algorithm is called with a sequence of steps $S_1, \ldots, S_i$ that has been constructed so far. Given this sequence, it has to decide if the sequence is continued with a step S_{i+1} such that S_i and S_{i+1} are linked or if the construction is stopped since a k-witness sequence is found. In Figure 8, we summarize the notations which we use in the following. In step S_j for $j \leq i+1$, the edges e_{j-1} and f_{j-1} are exchanged with the edges e_j and g_j. In step S'_{i+1}, the edges e'_i and f'_i are exchanged with the edges e'_{i+1} and g'_{i+1}. We denote by E_i all edges that are involved in steps S_j with $j \leq i$. Similar by E_{i-1} we denote all edges that are involved in steps S_j with $j \leq i-1$.

Our construction ensures that whenever the algorithm is called with a sequence $S_1, \ldots, S_i$ as input, then at least one of the edges that is added to the tour in step S_i is not contained in E_{i-1}. In the following, assume without loss of generality that $e_i \notin E_{i-1}$. When we call the algorithm recursively with sequence $S_1, \ldots, S_{i+1}$ or with sequence $S_1, \ldots, S_i, S'_{i+1}$, then either the recursive call never gives back a return value since a witness sequence is found in the recursive call which immediately stops the construction or a 2-change S is returned. Whenever a 2-change S is returned, the meaning is as follows: There exists a sequence of linked 2-changes in the sub-DAG W_u starting with S_{i+1} or S'_{i+1}, respectively, whose net effect equals the 2-change S, that is, after all steps in the sequence have been performed, the same two edges as in S have been removed from the tour, the same two edges have been added to the tour, and all other edges either stayed in or out of the tour.

In this case, we can virtually replace step S_{i+1} or S'_{i+1}, respectively, by the new step S.

When the algorithm is called with the sequence $S_1, \dots, S_i$, it performs the following steps:

1. Based on the last step S_i, identify the steps S_{i+1} and S'_{i+1}.
2. If $i = k$, then $S_1, \dots, S_i$ is a k-witness sequence of type 1.
3. If $e_{i+1} \notin E_i$ or $g_{i+1} \notin E_i$, then make a recursive call with the sequence $S_1, \dots, S_{i+1}$ as input. If a step S is returned, replace S_{i+1} virtually by the returned step, that is, in the following steps of the algorithm, assume that S_{i+1} equals step S. In this case the edges e_{i+1} and g_{i+1} that are added to the tour in the new step S are always chosen from the set E_i.
4. If $e'_i \notin E_{i-1}$ and ($e'_{i+1} \notin E_i$ or $g'_{i+1} \notin E_i$), then make a recursive call with the sequence $S_1, \dots, S_i, S'_{i+1}$ as input. If a step S is returned, replace S'_{i+1} virtually by the returned step, that is, in the following steps of the algorithm, assume that S'_{i+1} equals step S. In this case the edges e'_{i+1} and g'_{i+1} that are added to the tour in the new step S are always chosen from the set E_i.
5. If $e'_i \in E_{i-1}$ and $e_{i+1}, g_{i+1} \in E_i$:
   (a) If $f_{i-1} \in E_{i-1}$, then $S_1, \dots, S_{i+1}$ is a witness sequence of type 2.
   (b) If $f_i \notin E_i$, then $S_1, \dots, S_{i+1}$ is a witness sequence of type 3.
   (c) If $e_{i+1}, g_{i+1} \in E_{i-1}$, then $S_1, \dots, S_{i+1}$ is witness sequence of type 2 since one endpoint of f_{i-1} equals one endpoint of e'_i and the other one equals one endpoint of either e_{i+1} or g_{i+1}.
   (d) If $f_i \in E_i$ and ($e_{i+1} \in E_i \setminus E_{i-1}$ or $g_{i+1} \in E_i \setminus E_{i-1}$), then one can assume w. l. o. g. that $g_{i+1} = f_{i-1}$ and $e_{i+1} \in E_{i-1}$ since $e_{i+1} \neq e'_i$ and $g_{i+1} \neq e'_i$ (e_{i+1} and g_{i+1} share one endpoint with e_i, e'_i does not share an endpoint with e_i.) In this case, return the step $S = (e_{i-1}, f_i) \to (e_{i+1}, e'_i)$.
6. If $e'_i \notin E_{i-1}$ and $e_{i+1}, g_{i+1}, e'_{i+1}, g'_{i+1} \in E_i$:
   (a) If $e_{i+1}, g_{i+1} \in E_{i-1}$ and $e'_{i+1}, g'_{i+1} \in E_{i-1}$, then $S_1, \dots, S_{i+1}$ is a witness sequence of type 2.
   (b) If $f'_i \notin E_i$, then $S_1, \dots, S_i, S'_{i+1}$ is a witness sequence of type 3.
   (c) If $f_i, f'_i \in E_i$ and ($e_{i+1} \in E_i \setminus E_{i-1}$ or $g_{i+1} \in E_i \setminus E_{i-1}$) and ($e'_{i+1} \in E_i \setminus E_{i-1}$ or $g'_{i+1} \in E_i \setminus E_{i-1}$), then as in case 5 d), assume w. l. o. g. $g_{i+1} = g'_{i+1} = f_{i-1}$ and $e_{i+1}, e'_{i+1} \in E_{i-1}$. In this case, it must be $f_i \neq e'_i$ and $f'_i \neq e_i$ as otherwise step S_i would be reversed in step S_{i+1} or S'_{i+1}, respectively. Hence, $f_i, f'_i \in E_{i-1}$ and $S_1, \dots, S_{i+1}$ is a witness sequence of type 2 since one endpoint of f_{i-1} equals one endpoint of f_i and the other endpoint equals one endpoint of f'_i.
   (d) If $|\{e_{i+1}, e'_{i+1}, g_{i+1}, g'_{i+1}\} \cap (E_i \setminus E_{i-1})| = 1$, assume w. l. o. g. $e_{i+1}, g_{i+1}, e'_{i+1} \in E_{i-1}$ and $g'_{i+1} = f_{i-1}$. As in the previous case, it must $f'_i \in E_{i-1}$. We replace step S_i by the step $S = (e_{i-1}, f'_i) \to (e_i, e'_{i+1})$. Then the sequence $S_1, \dots, S_{i+1}$ is a witness sequence of type 2 as $f'_i \in E_{i-1}$. Observe that the original sequence $S_1, \dots, S_{i+1}$ together with the step S'_{i+1} yields the same net effect and hence the same improvement as the sequence with the modified step $S_i = S$.

Observe that basically this algorithm just constructs a path through the DAG starting at node u. When a path corresponding to the sequence $S_1, \dots, S_i$ of 2-changes has been constructed, the algorithm decides to either stop the construction since a witness sequence has been found, or, if possible, to continue the path with an arc corresponding to a step S_{i+1} or S'_{i+1}. In some situations, it can happen that the algorithm has not found a witness sequence yet but cannot continue the construction since S_i adds two edges to the tour that have been seen before. In such cases, step S_i is pruned and the algorithm reconsiders the path $S_1, \dots, S_{i-1}$ again. Based on the pruned step S_i it can then either decide that a witness sequence has been found, that also S_{i-1} has to be pruned, or it can decide to continue the path with S'_i instead of S_i.

This concludes the proof as the presented algorithm always identifies a k-witness sequence whose total improvement is at most as large as the improvement made by the steps in the sub-DAG W_u. □

Proof of Lemma 29. A t-witness DAG W consists of $n + 2t$ nodes and n of these nodes are leaves. Since the indegree and the outdegree of every node is bounded by 2, there are at most $n2^k$ nodes in W whose height is less than k. Hence, there are at least $n + 2t - n2^k \geq t$ nodes in W with an associated sub-DAG. We construct a set of disjoint sub-DAGs in a greedy fashion: We take an arbitrary sub-DAG W_u and add it to the set of disjoint sub-DAGs that we construct. After that, we remove all nodes, arcs, and twin arcs of W_u from the DAG W. We repeat these steps until no sub-DAG W_u is left in W.

In order to see that the constructed set consists of at least $t/4^{k+3}$ disjoint sub-DAGs, observe that each sub-DAG consists of at most 2^{k+1} nodes as its height is k. Furthermore, observe that each of these nodes can be contained in at most 2^{k+2} sub-DAGs W_v. Hence, every sub-DAG W_u can only be non-disjoint from at most $2^{2k+3} < 4^{k+2}$ other sub-DAGs W_v. Hence, the number of disjoint sub-DAGs must be at least $\lfloor t/4^{k+2} \rfloor > t/4^{k+3}$. □

4.3.4 The Expected Number of 2-Changes on ϕ-perturbed Graphs

Now we can prove Theorem 2 c) which directly implies Theorem 3 c).

Proof of Theorem 2 c). We combine Corollary 26 and Lemma 27 to obtain an upper bound on the probability that the length T of the longest path in the state graph exceeds t. For $t \geq n4^{k+2}$, the tour is shortened by the sequence of 2-changes by at least $t/4^{k+3} \cdot \Delta_{\text{ws}}^{(k)}$. Hence, for $t \geq n4^{k+2}$,

$$\mathbf{Pr}\left[T \geq t\right] \leq \mathbf{Pr}\left[\frac{t}{4^{k+3}} \cdot \Delta_{\text{ws}}^{(k)} \leq n\right] = \mathbf{Pr}\left[\Delta_{\text{ws}}^{(k)} \leq \frac{n \cdot 4^{k+3}}{t}\right] .$$

Combining this inequality with Corollary 26 yields for $t \geq 4^{k+4} \cdot n \cdot \phi \cdot m^{(k-1)/(k-2)}$,

$$\mathbf{Pr}\left[T \geq t\right] \leq 10k^5 \left(\frac{4^{k+4} \cdot n \cdot m \cdot \phi}{t}\right)^2 .$$

Hence, we can bound the expected number of 2-changes by

$$\mathbf{E}\left[T\right] \leq 4^{k+4} \cdot n \cdot \phi \cdot m^{(k-1)/(k-2)} + \sum_{t=1}^{n!} \min\left\{10k^5 \left(\frac{4^{k+4} \cdot n \cdot m \cdot \phi}{t}\right)^2, 1\right\} .$$

Splitting the sum at $t = n \cdot m \cdot \phi \cdot k^{5/2} \cdot 4^{k+4}$ yields

$$\mathbf{E}\left[T\right] = O(k^{5/2} \cdot 4^k \cdot n \cdot m^{(k-1)/(k-2)} \cdot \phi) .$$

Setting $k = \sqrt{\log m}$ yields the theorem. □

5 The Expected Approximation Ratio on L_1 and L_2 Instances

In this section, we consider the expected approximation ratio of the solution found by 2-Opt on ϕ-perturbed L_1 and L_2 instances. Chandra, Karloff, and Tovey show that if one has a set of n points in the unit square $[0,1]^2$ and the distances are measured according to a metric that is induced by a norm, then every locally optimal solution has length at most $c \cdot \sqrt{n}$ for an appropriate constant c depending on the metric [CKT99]. L_p metrics are induced by norms and hence, it follows that 2-Opt yields a tour of length at most $O(\sqrt{n})$ on ϕ-perturbed L_1 and L_2 instances. Hence, in order to bound the expected approximation ratio of 2-Opt on these instances, we just need to lower bound the expected value of $1/\text{OPT}$, where OPT denotes the length of the shortest tour.

Lemma 30. *For ϕ-perturbed L_1 and L_2 instances, it holds*

$$\mathbf{E}\left[\frac{1}{\text{OPT}}\right] = O\left(\sqrt{\frac{\phi}{n}}\right) .$$

Proof. Let $p_1, \ldots, p_n$ denote the points of the ϕ-perturbed instance. We partition the unit square into $k = \lceil n\phi \rceil$ smaller squares with area $1/k$ each and analyze how many of these smaller squares contain at least one of the points. Assume that X of these squares contain a point, then the optimal tour must have length at least $X/(9\sqrt{k})$. In order to see this, we construct a set $P \subseteq \{p_1, \ldots, p_n\}$ of points as follows: Consider the points $p_1, \ldots, p_n$ one after another, and insert a point p_i into P if P does not contain a point in the same square as p_i or in one of its 8 neighboring squares yet. Due to the triangle inequality, the optimal tour on P is at most as long as the optimal tour on $p_1, \ldots, p_n$. Furthermore, P contains at least $X/9$ points and every edge between two points from P has length at least $1/\sqrt{k}$ since P does not contain two points in the same or in two neighboring squares.

Hence, it remains to analyze the random variable X. For each square i with $1 \le i \le k$, we define a random variables X_i which takes value 0 if square i is empty and value 1 if square i contains at least one point. The density functions that specify the locations of the points induce for each pair of square i and point j a probability p_i^j such that point j falls into square i with probability p_i^j. Hence, one can think of throwing n balls into k bins in a setting where each ball has its own probability distribution over the bins. Due to the bounded density, we have $p_i^j \le \phi/k$. For each square i, let M_i denote the probability mass associated with the square, that is

$$M_i = \sum_{j=1}^{n} p_i^j \le \frac{n\phi}{k} .$$

We can write the expected value of the random variable X_i as

$$\mathbf{E}[X_i] = \mathbf{Pr}[X_i = 1] = 1 - \prod_{j=1}^{n}(1 - p_i^j) \ge 1 - \left(1 - \frac{M_i}{n}\right)^n$$

as, under the constraint $\sum_j (1 - p_i^j) = n - M_i$, the term $\prod_j (1 - p_i^j)$ is maximized if all the p_i^j are equal. Due to linearity of expectation, the expected value of X is

$$\mathbf{E}[X] \ge \sum_{i=1}^{k} 1 - \left(1 - \frac{M_i}{n}\right)^n = k - \sum_{i=1}^{k}\left(1 - \frac{M_i}{n}\right)^n .$$

Observe that $\sum_i M_i = n$. Thus, the sum $\sum_i (1 - M_i/n)$ becomes maximal if the M_i's are chosen as unbalanced as possible. Hence, we assume that $\lceil k/\phi \rceil$ of the M_i's take their maximal value of $n\phi/k$ and the other M_i' are zero. This yields, for sufficiently large n,

$$\begin{aligned} \mathbf{E}\left[X\right] &\geq k - \left(\left\lceil \frac{k}{\phi} \right\rceil \left(1 - \frac{\phi}{k}\right)^n + \left(k - \left\lceil \frac{k}{\phi} \right\rceil \right) \right) \\ &\geq \frac{k}{\phi} - \frac{2k}{\phi}\left(1 - \frac{\phi}{k}\right)^n \\ &\geq \frac{k}{\phi}\left(1 - 2\left(1 - \frac{1}{n+1}\right)^n\right) \geq \frac{n}{5} \ . \end{aligned}$$

Hence, we obtain the following bound on the expected length of the optimal tour

$$\mathbf{E}\left[\mathrm{OPT}\right] \geq \frac{\mathbf{E}\left[X\right]}{9\sqrt{k}} \geq \frac{n}{45\sqrt{k}} \geq \frac{\sqrt{n}}{45\sqrt{\phi+1}} \ .$$

We still need to determine the expected value of the random variable $1/\mathrm{OPT}$. Therefore, we first show that X and hence also OPT are sharply concentrated around their mean values. The random variable X is the sum of n 0-1-random variables. If these random variables were independent, we could simply use a Chernoff bound to bound the probability that X takes a value that is smaller than its mean value. The X_i's are negatively associated, in the sense that whenever we already know that some of the X_i's are zero, then the probability of the event that another X_i's also take the value zero becomes smaller. Hence, intuitively, the dependencies can only help when one wants to bound the probability that X takes a value smaller than its mean value. Dubhashi and Ranjan formalize this intuition by introduction the notion of negative dependence and by showing that in the case of negative dependent random variables one can still apply the Chernoff bound [DR98]. This yields

$$\mathbf{Pr}\left[X \leq \frac{n}{10}\right] \leq e^{-n/40} \ .$$

Thus, as $1/X \leq 1$ with certainty,

$$\mathbf{E}\left[\frac{1}{X}\right] \leq (1 - e^{-n/40}) \cdot \frac{10}{n} + e^{-n/40} \leq \frac{11}{n}$$

for sufficiently large n. Altogether, this implies

$$\mathbf{E}\left[\frac{1}{\mathrm{OPT}}\right] \leq \mathbf{E}\left[\frac{9 \cdot \sqrt{\lceil n\phi \rceil}}{X}\right] = O\left(\sqrt{\frac{\phi}{n}}\right) \ .$$

□

If one combines Lemma 30 with the result of Chandra, Karloff, and Tovey that every locally optimal solution has length $O(\sqrt{n})$, one obtains Theorem 4.

6 Extensions

6.1 Smoothed Analysis

Smoothed Analysis was introduced by Spielman and Teng as a hybrid of worst case and average case analysis [ST04]. The semi-random input model in a smoothed analysis is designed to capture the behavior of algorithms on typical inputs better than a worst case or an average case analysis alone as it allows an adversary to specify a worst case input which is randomly perturbed afterwards. In Spielman and Teng's analysis of the Simplex algorithm the adversary specifies an arbitrary linear program which is perturbed by adding independent Gaussian random variables to each number in the linear program.

We suggest the following perturbation model for L_1 and L_2 instances for the TSP. First an adversary chooses a set of n points in the unit square. Then the coordinates of these points are perturbed by adding independent random variables to them. The random variables we add are basically Gaussian random variables with standard deviation $\sigma \leq 1$. If, however, one of the added random variables has an absolute value larger than some given $\alpha \geq 1$, then we draw another Gaussian random variable with standard deviation σ until the absolute value is bounded by α. Let X denote one such random variable and let Y denote a Gaussian random variable with standard deviation σ and density function f_Y. Then the density f_X of X can be bounded by

$$f_X(x) \leq \frac{\sup_{y\in\mathbb{R}} f_Y(y)}{\mathbf{Pr}\left[|Y| \leq \alpha\right]} \leq \frac{1/(\sigma\sqrt{2\pi})}{1-\sigma/\sqrt{2\pi}\cdot\exp(-\alpha^2/(2\sigma^2))} \ .$$

Observe that after the perturbation all points lie in the square $[-\alpha, 1+\alpha]^2$. Hence, in order to apply Theorems 2, 3, and 4, we first have to scale and shift the instance such that every point lies in the unit square. This can increase the density f_X of X by at most a factor of $(2\alpha+1)^2$. Thus with

$$\phi = \frac{(2\alpha+1)^2}{(\sigma\sqrt{2\pi}-\sigma^2\exp(-\alpha^2/(2\sigma^2)))^2} = O\left(\frac{\alpha^2}{\sigma^2}\right). \tag{5}$$

we can apply the aforementioned theorems.

Finally, let us remark that if the standard deviation is small enough, then it is not necessary to redraw the Gaussian random variables until they lie in the interval $[-\alpha,\alpha]$. For $\sigma \leq \min\{\alpha/\sqrt{4n\ln n}, 1\}$, the probability that one of the Gaussian random variables has an absolute value larger than $\alpha \geq 1$ can be bounded by

$$\frac{2n}{\sigma\sqrt{2\pi}}\cdot\int_{x=\alpha}^{\infty}\exp\left(-x^2/(2\sigma^2)\right)\,dx \leq \frac{2n\sigma}{\sqrt{2\pi}}\cdot\exp\left(-\alpha^2/(2\sigma^2)\right) \leq n^{-2n} \ .$$

In this case, even if one does not redraw the random variables outside $[-\alpha,\alpha]$, the Theorems 2, 3, and 4 can be applied with the corresponding ϕ given in (5). To see this, one must only observe that the worst case bound for the number of 2-changes is $(n!)$ and the worst case approximation ratio is $O(\log n)$ [CKT99]. Multiplying these values with the failure probability of n^{-2n} constitutes less than 1 to the expected values. In particular, this implies that the expected length of the longest path in the state graph is bounded by $O(\mathrm{poly}(n, 1/\sigma))$.

6.2 TSP in Higher Dimensions

So far we considered only two-dimensional Euclidean and Manhattan instances as this is the most relevant case from a practical point of view. In fact, our analysis for the Manhattan metric goes through for every dimension d without major modifications. To be more precise, we consider the model in which for every point a probability distribution $f : [0,1]^d \to [0,\phi]$ is given. The bound given in Theorem 2 is only worse by a factor of d since the length of the initial tour can grow linearly with d. The bound in Theorem 3 is worse by a factor of $\sqrt{d} \cdot n^{(d-2)/2d}$ since the length of the optimal tour in $[0,1]^d$ can only be bounded by $\sqrt{d} \cdot n^{(d-1)/d}$. The same is also true for Euclidean instances. In order to see that our analysis for Euclidean instances goes through for higher dimensions, observe that we consider arrangements of only three points in the core of the analysis in Section 4.2. These three points span a two dimensional subspace of the Euclidean space, and one can apply the analysis in the Lemmas 13 and 14 to this subspace. With the same arguments as in the proof of Theorem 4, we obtain that the expected approximation ratio for d-dimensional ϕ-perturbed Manhattan and Euclidean instances is $O(\sqrt[d]{\phi})$.

References

[Aro98] Sanjeev Arora. Polynomial Time Approximation Schemes for Euclidean Traveling Salesman and other Geometric Problems. *Journal of the ACM*, 45(5):753–782, 1998.

[CKT99] Barun Chandra, Howard J. Karloff, and Craig A. Tovey. New Results on the Old k-Opt Algorithm for the Traveling Salesman Problem. *SIAM Journal on Computing*, 28(6):1998–2029, 1999.

[DR98] Devdatt P. Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998.

[JM97] David S. Johnson and Lyle A. McGeoch. The Traveling Salesman Problem: A Case Study in Local Optimization. In E.H.L. Aarts and J.K. Lenstra, editors, *Local Search in Combinatorial Optimization*. John Wiley and Sons, 1997.

[Ker89] Walter Kern. A Probabilistic Analysis of the Switching Algorithm for the Euclidean TSP. *Mathematical Programming*, 44(2):213–219, 1989.

[Lue75] George S. Lueker. Unpublished Manuscript, 1975. Princeton University.

[Pap77] Christos H. Papadimitriou. The Euclidean Traveling Salesman Problem is NP-Complete. *Theoretical Computer Science*, 4(3):237–244, 1977.

[Rei91] Gerhard Reinelt. TSPLIB – A Traveling Salesman Problem Library. *ORSA Journal on Computing*, 3:376–384, 1991.

[RSI77] Daniel J. Rosenkrantz, Richard Edwin Stearns, and Philip M. Lewis II. An Analysis of Several Heuristics for the Traveling Salesman Problem. *SIAM Journal on Computing*, 6(3):563–581, 1977.

[ST04] Daniel A. Spielman and Shang-Hua Teng. Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. *Journal of the ACM*, 51(3):385–463, 2004.

[vLS80] Jan van Leeuwen and Anneke A. Schoon. Untangling a traveling salesman tour in the plane. Technical report, University of Utrecht, 1980. Report No. RUU-CS-80-11.

A Some Probability Theory

Lemma 31. *Let $X_1, \ldots, X_{2n}$ be random variables and assume that for $i \in \{1, \ldots, n\}$, the random variables X_{2i-1} and X_{2i} are described by a joint density $f_i : [0,1]^2 \to [0, \phi]$ for some given $\phi \geq 1$. Assume that the random vectors $(X_1, X_2), \ldots, (X_{2n-1}, X_{2n})$ are independent, and let $X = (X_1, \ldots, X_{2n})^T$. Furthermore, let $k \leq n$ and, for $i \in \{1, \ldots, 2k\}$, let $\lambda^{(i)} \in \mathbb{Z}^{2n}$ be a row vector such that the vectors $\lambda^{(1)}, \ldots, \lambda^{(2k)}$ are linearly independent. For a fixed ε, we denote by $\mathcal{A}_i$ the event that $\lambda^{(i)} \cdot X \in [0, \varepsilon]$ occurs, i. e., the linear combination of the variables $X_1, \ldots, X_{2n}$ with the coefficients $\lambda^{(i)}$ takes a value in the interval $[0, \varepsilon]$. Then we have*

$$\mathbf{Pr}\left[\bigcap_{i=1}^{2k} \mathcal{A}_i\right] \leq (\varepsilon\phi)^{2k}.$$

Proof. The main tool for proving the lemma is a change of variables. Instead of using the canonical basis of the $2n$-dimensional vector space $\mathbb{R}^{2n}$, we use the given linear combinations as basis vectors. To be more precise, the basis $\mathcal{B}$ which we are using consists of two parts: it contains the vectors $\lambda^{(1)}, \ldots, \lambda^{(2k)}$ (that is why they are assumed to be linearly independent) and it is completed by some vectors from the canonical basis $\{e^{(1)}, \ldots, e^{(2n)}\}$, where $e^{(i)}$ denotes the i-th canonical row vector, i. e.,

$$e_j^{(i)} = \begin{cases} 0 & \text{if} \quad i \neq j \\ 1 & \text{if} \quad i = j \end{cases}.$$

That is, we have $\mathcal{B} = \{\lambda^{(1)}, \ldots, \lambda^{(2k)}, e^{(\pi(1))}, \ldots, e^{(\pi(2n-2k))}\}$, for some injective function $\pi : [2n - 2k] \to [2n]$ where $[i]$ denotes the set $\{1, \ldots, i\}$.

Let $\Phi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be defined by $\Phi(X) = AX$ where A denotes the $(2n) \times (2n)$-matrix

$$A = \begin{pmatrix} \lambda^{(1)} \\ \ldots \\ \lambda^{(2k)} \\ e^{(\pi(1))} \\ \ldots \\ e^{(\pi(2n-2k))} \end{pmatrix}.$$

Observe, that Φ is a bijection since $\mathcal{B}$ is a basis of $\mathbb{R}^{2n}$.

Let $Y = (Y_1, \ldots, Y_{2n})^T$ be defined as $Y = \Phi(X)$, let $f : \mathbb{R}^{2n} \to \mathbb{R}$ denote the joint density of the random variables $X_1, \ldots, X_{2n}$, and let $g : \mathbb{R}^{2n} \to \mathbb{R}$ denote the joint density of the random variables $Y_1, \ldots, Y_{2n}$. Due to the independence of the random vectors $(X_1, X_2), \ldots, (X_{2n-1}, X_{2n})$, we have $f(x_1, \ldots, x_{2n}) = f_1(x_1, x_2) \cdot \cdots \cdot f_n(x_{2n-1}, x_{2n})$. We can express the joint density g as follows

$$g(y_1, \ldots, y_{2n}) = |\det_{\partial} \Phi^{-1}(y_1, \ldots, y_{2n})| \cdot f(\Phi^{-1}(y_1, \ldots, y_{2n})),$$

where $\det_\partial$ denotes the determinant of the Jacobian matrix of Φ^{-1}.

The matrix A is invertible as $\mathcal{B}$ is a basis of $\mathbb{R}^{2n}$. Hence, it holds $\Phi^{-1}(Y) = A^{-1}Y$ and the Jacobian matrix of Φ^{-1} equals A^{-1}. Thus, $\det_\partial \Phi^{-1} = \det A^{-1} = (\det A)^{-1}$. Since all entries of A are integers, also its determinant must be an integer, and since it has rank $2n$, we know that

$\det A \neq 0$. Hence $|\det A| \geq 1$ and $|\det A^{-1}| \leq 1$. Thus,

$$\begin{aligned} & g(y_1, \ldots, y_{2n}) \\ \leq \; & f(\Phi^{-1}(y_1, \ldots, y_{2n})) \\ = \; & f_1(\Phi_1^{-1}(y_1, \ldots, y_{2n}), \Phi_2^{-1}(y_1, \ldots, y_{2n})) \cdot \cdots \cdot f_n(\Phi_{2n-1}^{-1}(y_1, \ldots, y_{2n-1}), \Phi_{2n}^{-1}(y_1, \ldots, y_{2n})) \ . \end{aligned}$$

The probability we want to estimate can be written as

$$\mathbf{Pr}\left[\bigcap_{i=1}^{2k} \mathcal{A}_i\right] = \int_{y_1=0}^{\varepsilon} \cdots \int_{y_{2k}=0}^{\varepsilon} \int_{y_{2k+1}=-\infty}^{\infty} \cdots \int_{y_n=-\infty}^{\infty} g(y_1, \ldots, y_{2n}) \, dy_{2n} \ldots \, dy_1 \ .$$

Since all variables X_i take values in the interval $[0,1]$ and since for $i \in \{2k+1, \ldots, 2n\}$, we have $Y_i = X_{\pi(i-2k)}$, this simplifies to

$$\mathbf{Pr}\left[\bigcap_{i=1}^{2k} \mathcal{A}_i\right] = \int_{y_1=0}^{\varepsilon} \cdots \int_{y_{2k}=0}^{\varepsilon} \int_{y_{2k+1}=0}^{1} \cdots \int_{y_n=0}^{1} g(y_1, \ldots, y_{2n}) \, dy_{2n} \ldots \, dy_1 \ , \tag{6}$$

Without loss of generality assume that $\{i \,|\, \nexists j \in [2n-2k] : \pi(j) = i\} \subseteq [4k]$, that is, only vectors $e^{(i)}$ from the canonical basis with $i \leq 4k$ are replaced by the vectors $\lambda^{(j)}$. Furthermore, assume w. l. o. g. that for $i > 4k$, we have $\pi(i) = i$. Then we can bound the joint density g as follows

$$g(y_1, \ldots, y_{2n}) \leq \phi^{2k} \cdot f_{2k+1}(y_{4k+1}, y_{4k+2}) \cdot \cdots \cdot f_n(y_{2n-1}, y_{2n}) \ . \tag{7}$$

Putting together (6) and (7) yields

$$\begin{aligned} \mathbf{Pr}\left[\bigcap_{i=1}^{k} \mathcal{A}_i\right] \; \leq \; & (\varepsilon\phi)^{2k} \cdot \int_{y_{4k+1}=0}^{1} \int_{y_{4k+2}=0}^{1} f_{2k+1}(y_{4k+1}, y_{4k+2}) \cdots \\ & \cdots \int_{y_{2n-1}=0}^{1} \int_{y_{2n}=0}^{1} f_{2n}(y_{2n-1}, y_{2n}) \, dy_{2n} \ldots \, dy_{4k+1} \\ = \; & (\varepsilon\phi)^{2k} \ , \end{aligned}$$

where the last equation follows because the f_i's denote density functions. □

B Proofs of some Lemmas from Section 4.2

B.1 Proof of Lemma 13

Proof of Lemma 13. To make the calculations simpler, we use polar coordinates to describe the location of the point Q. In fact, since the radius r is given, the point Q is completely determined by the angle α between the x-axis and the line between O and Q. Hence, the random variable Z can be written as

$$Z = r - \sqrt{r^2 + \tau^2 - 2r\tau \cdot \cos\alpha} \ .$$

Based on the observation that the angle α is chosen uniformly at random from the interval $[0, 2\pi)$, we prove the desired bound on density of the random variable Z.

First of all, observe that choosing the angle α uniformly from the interval $[0, \pi]$ does not change the density of Z since Z behaves symmetrically in the interval $[\pi, 2\pi]$. Furthermore, observe that certain constraints are imposed upon Z implicitly. Namely, Z can only take values in the interval $[-\tau, \min\{\tau, 2r - \tau\}]$.

When α is restricted to the interval $[0, \pi]$, then there exists a unique inverse function mapping Z to α, namely

$$\alpha(z) = \arccos\left(\frac{\tau^2 + 2zr - z^2}{2r\tau}\right) \ .$$

The conditional density $f_{Z|T=\tau,R=r}$ can be expressed as

$$\begin{aligned} f_{Z|T=\tau,R=r}(z) &= f_\alpha(\alpha(z)) \cdot \frac{d}{dz}\alpha(z) \\ &= \frac{1}{\pi} \cdot \frac{d}{dz}\alpha(z) \ , \end{aligned}$$

where f_α denotes the density of α, i. e., the uniform density over $[0, \pi]$.

Hence, in order to estimate $f_{Z|T=\tau,R=r}(z)$, we only need to bound the derivative of the function $\alpha(z)$. For $|x| < 1$, the derivative of arc cosine is

$$(\arccos(x))' = -\frac{1}{\sqrt{1 - x^2}} \ .$$

Hence, the derivative of $\alpha(z)$ can be calculated as follows:

$$\begin{aligned} \alpha'(z) &= \frac{r - z}{r\tau} \cdot \frac{-1}{\sqrt{1 - \frac{(\tau^2 + 2zr - z^2)^2}{4r^2\tau^2}}} \\ &= \frac{2(z - r)}{\sqrt{4\tau^2 r^2 - (\tau^2 + 2zr - z^2)^2}} \\ &= \frac{2(z - r)}{\sqrt{4\tau^2 r^2 - (\tau^4 + 4z^2 r^2 + z^4 + 4\tau^2 zr - 2\tau^2 z^2 - 4z^3 r)}} \\ &= \frac{2(z - r)}{\sqrt{4r^2\tau^2 - 4r^2 z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4}} \ . \end{aligned}$$

We distinguish between the cases $r \geq \tau$ and $r \leq \tau$:

- **First case: $r \geq \tau$.**
  In this case, we prove

  $$4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4 \geq 2(z - r)^2(\tau^2 - z^2)$$

  yielding the lemma for $r \geq \tau$. A simple calculation shows

  $$\begin{aligned} & 4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4 - 2(z - r)^2(\tau^2 - z^2) \\ =\ & 2r^2\tau^2 - 2r^2z^2 - \tau^4 + z^4 \\ =\ & 2r^2(\tau^2 - z^2) - \tau^4 + z^4 \\ \geq\ & 2\tau^2(\tau^2 - z^2) - \tau^4 + z^4 \\ =\ & (\tau^2 - z^2)^2 \geq 0 \ . \end{aligned}$$

- **Second case:** $r \leq \tau$.
  In this case, we prove

$$4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4 \geq 2(z-r)^2(\tau+z)(2r-\tau-z)$$

yielding the lemma for $r \leq \tau$. A simple calculation shows

$$\begin{aligned} & 4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4 - 2(z-r)^2(\tau+z)(2r-\tau-z) \geq 0 \\ \iff \quad & (-2r+z+\tau)(\tau+z)(z^2+2\tau z - 2rz + 2r^2 - \tau^2 - 2\tau r) \geq 0 \\ \iff \quad & z^2 + 2\tau z - 2rz + 2r^2 - \tau^2 - 2\tau r \leq 0 \ , \end{aligned}$$

which is implied by

$$\begin{aligned} & z^2 + 2\tau z - 2rz + 2r^2 - \tau^2 - 2\tau r \\ = \ & z^2 + 2z(\tau - r) + 2r^2 - \tau^2 - 2\tau r \\ \leq \ & (2r-\tau)^2 + 2(2r-\tau)(\tau - r) + 2r^2 - \tau^2 - 2\tau r \\ = \ & 2(r^2 - \tau^2) \leq 0 \ . \end{aligned}$$

This concludes the proof. □

B.2 Proof of Lemma 14

In the following proof, we implicitly use the following two identities. For every $c > 0$ and $a > 0$,

$$\int_{z=0}^{c} \frac{1}{\sqrt{z(c-z)}}\, dz = \pi$$

and

$$\int_{z=0}^{a} \frac{1}{\sqrt{z(z+c)}}\, dz = \ln\left(\frac{c}{2} + a + \sqrt{a(a+c)}\right) - \ln\left(\frac{c}{2}\right) \ .$$

If a is bounded from above by some constant, then we can find a constant κ such that

$$\int_{z=0}^{a} \frac{1}{\sqrt{z(z+c)}}\, dz \leq \ln\left(\frac{1}{c}\right) + \kappa \ .$$

Proof of Lemma 14. The conditional density of Δ can be calculated as convolution of the conditional densities of Z_1 and Z_2 as follows:

$$\begin{aligned} & f_{\Delta \mid T=\tau, R_1=r_1, R_2=r_2}(\delta) \\ = \ & \int_{z=-\infty}^{\infty} f_{Z \mid T=\tau, R=r_1}(z) \cdot f_{Z \mid T=\tau, R=r_2}(z-\delta)\, dz \ . \end{aligned}$$

In order to estimate this integral, we distinguish between several cases. In the following, let κ denote a sufficiently large constant.

- **First case:** $\tau \le r_1$ **and** $\tau \le r_2$**.**
In this case, Z_i takes only values in the interval $[-\tau, \tau]$. Hence we can assume $0 \le \delta \le 2\tau]$ and

$$\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
= & \int_{z=-\tau+\delta}^{\tau} f_{Z|T=\tau,R=r_1}(z) \cdot f_{Z|T=\tau,R=r_2}(z-\delta)\,dz \ .
\end{aligned}$$

Due to Lemma 13, we can estimate the densities of Z_1 and Z_2 by

$$\begin{aligned}
f_{Z|T=\tau,R=r_i}(z) & \le \sqrt{\frac{2}{\tau^2 - z^2}} \le \sqrt{\frac{2}{\tau(\tau - |z|)}} \\
& \le \sqrt{\frac{2}{\tau}} \left(\frac{1}{\sqrt{\tau - z}} + \frac{1}{\sqrt{\tau + z}} \right) \ .
\end{aligned}$$

For $\delta \in [0, 2\tau]$, we obtain the following upper bound on the density of Δ:

$$\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
\le & \frac{2}{\tau} \int_{-\tau+\delta}^{\tau} \left(\frac{1}{\sqrt{\tau - z}} + \frac{1}{\sqrt{\tau + z}} \right) \left(\frac{1}{\sqrt{\tau - z + \delta}} + \frac{1}{\sqrt{\tau + z - \delta}} \right) dz \\
= & \frac{2}{\tau} \Bigg(\int_{-\tau+\delta}^{\tau} \frac{1}{\sqrt{(\tau - z)(\tau - z + \delta)}}\,dz + \int_{-\tau+\delta}^{\tau} \frac{1}{\sqrt{(\tau + z)(\tau - z + \delta)}}\,dz \\
& \int_{-\tau+\delta}^{\tau} \frac{1}{\sqrt{(\tau - z)(\tau + z - \delta)}}\,dz + \int_{-\tau+\delta}^{\tau} \frac{1}{\sqrt{(\tau + z)(\tau + z - \delta)}}\,dz \Bigg) \\
= & \frac{2}{\tau} \Bigg(\int_{0}^{2\tau-\delta} \frac{1}{\sqrt{z(z + \delta)}}\,dz + \int_{\delta}^{2\tau} \frac{1}{\sqrt{z(2\tau + \delta - z)}}\,dz \\
& \int_{0}^{2\tau-\delta} \frac{1}{\sqrt{z(2\tau - \delta - z)}}\,dz + \int_{0}^{2\tau-\delta} \frac{1}{\sqrt{z(z + \delta)}}\,dz \Bigg) \\
\le & \frac{\kappa}{\tau} \left(\ln\left(\frac{1}{\delta}\right) + 1 \right) \ .
\end{aligned}$$

- **Second case:** $r_1 \le \tau$ **and** $r_2 \le \tau$**.**
In this case, Z_i takes only values in the interval $[-\tau, 2r_i - \tau]$. Hence we can assume $0 \le \delta \le 2r_1$ and

$$\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
= & \int_{z=-\tau+\delta}^{\min\{2r_1-\tau, 2r_2-\tau+\delta\}} f_{Z|T=\tau,R=r_1}(z) \cdot f_{Z|T=\tau,R=r_2}(z-\delta)\,dz \ .
\end{aligned}$$

Due to Lemma 13, we can estimate the densities of Z_1 and Z_2 by

$$
\begin{aligned}
f_{Z|T=\tau,R_i=r_i}(z) &\leq \sqrt{\frac{2}{(\tau+z)(2r_i-\tau-z)}} \\
&\leq \begin{cases} \sqrt{\frac{2}{r_i(\tau+z)}} & \text{if} \quad z \leq r_i-\tau \\ \sqrt{\frac{2}{r_i(2r_i-\tau-z)}} & \text{if} \quad z \geq r_i-\tau \end{cases} \\
&\leq \sqrt{\frac{2}{r_i}}\left(\frac{1}{\sqrt{\tau+z}}+\frac{1}{\sqrt{2r_i-\tau-z}}\right) \quad .
\end{aligned}
$$

- **Case 2.1:** $\delta \in [\max\{0, 2(r_1-r_2)\}, 2r_1]$.
  We obtain the following upper bound on the density of Δ:

$$
\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
\leq\ & \frac{2}{\sqrt{r_1r_2}}\int_{-\tau+\delta}^{2r_1-\tau}\left(\frac{1}{\sqrt{\tau+z}}+\frac{1}{\sqrt{2r_1-\tau-z}}\right)\left(\frac{1}{\sqrt{\tau+z-\delta}}+\frac{1}{\sqrt{2r_2-\tau-z+\delta}}\right)dz \\
=\ & \frac{2}{\sqrt{r_1r_2}}\left(\int_{-\tau+\delta}^{2r_1-\tau}\frac{1}{\sqrt{(\tau+z)(\tau+z-\delta)}}\,dz+\int_{-\tau+\delta}^{2r_1-\tau}\frac{1}{\sqrt{(2r_1-\tau-z)(\tau+z-\delta)}}\,dz\right. \\
& \left.\int_{-\tau+\delta}^{2r_1-\tau}\frac{1}{\sqrt{(\tau+z)(2r_2-\tau-z+\delta)}}\,dz+\int_{-\tau+\delta}^{2r_1-\tau}\frac{1}{\sqrt{(2r_1-\tau-z)(2r_2-\tau-z+\delta)}}\,dz\right) \\
=\ & \frac{2}{\sqrt{r_1r_2}}\left(\int_{0}^{2r_1-\delta}\frac{1}{\sqrt{z(z+\delta)}}\,dz+\int_{0}^{2r_1-\delta}\frac{1}{\sqrt{z(2r_1-\delta-z)}}\,dz\right. \\
& \left.\int_{\delta}^{2r_1}\frac{1}{\sqrt{z(2r_2+\delta-z)}}\,dz+\int_{0}^{2r_1-\delta}\frac{1}{\sqrt{z(2(r_2-r_1)+\delta+z)}}\,dz\right) \quad .
\end{aligned}
$$

  For $r_1 \leq r_2$ this can be upper bounded by

$$
\frac{\kappa}{\sqrt{r_1r_2}}\left(\ln\left(\frac{1}{\delta}\right)+1\right) \quad .
$$

  For $r_1 \geq r_2$ this can be upper bounded by

$$
\frac{\kappa}{\sqrt{r_1r_2}}\left(\ln\left(\frac{1}{\delta}\right)+\ln\left(\frac{1}{2(r_2-r_1)+\delta}\right)+1\right) \quad .
$$

- **Case 2.2:** $\delta \in [0, \max\{0, 2(r_1-r_2)\}]$.
  This case is only relevant if $r_2 \leq r_1$. We obtain the following upper bound on the density of

Δ:

$$
\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
\leq\ & \frac{2}{\sqrt{r_1 r_2}} \int_{-\tau+\delta}^{2r_2-\tau+\delta} \left(\frac{1}{\sqrt{\tau+z}} + \frac{1}{\sqrt{2r_1-\tau-z}} \right) \left(\frac{1}{\sqrt{\tau+z-\delta}} + \frac{1}{\sqrt{2r_2-\tau-z+\delta}} \right) dz \\
=\ & \frac{2}{\sqrt{r_1 r_2}} \Bigg(\int_{-\tau+\delta}^{2r_2-\tau+\delta} \frac{1}{\sqrt{(\tau+z)(\tau+z-\delta)}}\, dz + \int_{-\tau+\delta}^{2r_2-\tau+\delta} \frac{1}{\sqrt{(2r_1-\tau-z)(\tau+z-\delta)}}\, dz \\
& \int_{-\tau+\delta}^{2r_2-\tau+\delta} \frac{1}{\sqrt{(\tau+z)(2r_2-\tau-z+\delta)}}\, dz + \int_{-\tau+\delta}^{2r_2-\tau+\delta} \frac{1}{\sqrt{(2r_1-\tau-z)(2r_2-\tau-z+\delta)}}\, dz \Bigg) \\
=\ & \frac{2}{\sqrt{r_1 r_2}} \Bigg(\int_0^{2r_2} \frac{1}{\sqrt{z(z+\delta)}}\, dz + \int_0^{2r_2} \frac{1}{\sqrt{z(2r_1-\delta-z)}}\, dz \\
& \int_0^{2r_2} \frac{1}{\sqrt{z(2r_2+\delta-z)}}\, dz + \int_0^{2r_2} \frac{1}{\sqrt{z(2(r_1-r_2)-\delta+z)}}\, dz \Bigg) \\
\leq\ & \frac{\kappa}{\sqrt{r_1 r_2}} \left(\ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{1}{2(r_1-r_2)-\delta}\right) + 1 \right) \quad .
\end{aligned}
$$

- **Third case: $r_1 \leq \tau \leq r_2$.**
  In this case, Z_1 takes only values in the interval $[-\tau, 2r_1-\tau]$ and Z_2 takes only values in the interval $[-\tau, \tau]$. Hence we can assume $0 \leq \delta \leq 2r_1$ and

$$
\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
=\ & \int_{z=-\tau+\delta}^{2r_1-\tau} f_{Z|T=\tau,R=r_1}(z) \cdot f_{Z|T=\tau,R=r_2}(z-\delta)\, dz \quad .
\end{aligned}
$$

For $\delta \in [0, 2r_1]$, we obtain the following upper bound on the density of Δ:

$$
\begin{aligned}
& f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\
\leq\ & \frac{2}{\sqrt{\tau r_1}} \int_{-\tau+\delta}^{2r_1-\tau} \left(\frac{1}{\sqrt{\tau+z}} + \frac{1}{\sqrt{2r_1-\tau-z}} \right) \left(\frac{1}{\sqrt{\tau-z+\delta}} + \frac{1}{\sqrt{\tau+z-\delta}} \right) dz \\
=\ & \frac{2}{\sqrt{\tau r_1}} \Bigg(\int_{-\tau+\delta}^{2r_1-\tau} \frac{1}{\sqrt{(\tau+z)(\tau-z+\delta)}}\, dz + \int_{-\tau+\delta}^{2r_1-\tau} \frac{1}{\sqrt{(2r_1-\tau-z)(\tau-z+\delta)}}\, dz \\
& \int_{-\tau+\delta}^{2r_1-\tau} \frac{1}{\sqrt{(\tau+z)(\tau+z-\delta)}}\, dz + \int_{-\tau+\delta}^{2r_1-\tau} \frac{1}{\sqrt{(2r_1-\tau-z)(\tau+z-\delta)}}\, dz \Bigg) \\
=\ & \frac{2}{\sqrt{\tau r_1}} \Bigg(\int_{\delta}^{2r_1} \frac{1}{\sqrt{z(2\tau+\delta-z)}}\, dz + \int_0^{2r_1-\delta} \frac{1}{\sqrt{z(2(\tau-r_1)+\delta+z)}}\, dz \\
& \int_0^{2r_1-\delta} \frac{1}{\sqrt{z(z+\delta)}}\, dz + \int_0^{2r_1-\delta} \frac{1}{\sqrt{z(2r_1-\delta-z)}}\, dz \Bigg) \\
\leq\ & \frac{\kappa}{2\sqrt{\tau r_1}} \left(\ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{1}{2(\tau-r_1)+\delta}\right) + 1 \right) \leq \frac{\kappa}{\sqrt{\tau r_1}} \left(\ln\left(\frac{1}{\delta}\right) + 1 \right) \quad ,
\end{aligned}
$$

where the last inequality follows since $\tau \geq r_1$.

- **Fourth case: $r_2 \le \tau \le r_1$.**
  In this case, Z_1 takes only values in the interval $[-\tau, \tau]$ and Z_2 takes only values in the interval $[-\tau, 2r_2 - \tau]$. Hence we can assume $0 \le \delta \le 2(\tau - r_2)$ and

$$\begin{aligned} & f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\ = & \int_{z=-\tau+\delta}^{2r_2-\tau+\delta} f_{Z|T=\tau,R=r_1}(z) \cdot f_{Z|T=\tau,R=r_2}(z-\delta)\,dz \ . \end{aligned}$$

  For $\delta \in [0, 2(\tau - r_2)]$, we obtain the following upper bound on the density of Δ:

$$\begin{aligned} & f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta) \\ \le & \frac{2}{\sqrt{\tau r_2}} \int_{-\tau+\delta}^{2r_2-\tau} \left(\frac{1}{\sqrt{\tau - z}} + \frac{1}{\sqrt{\tau + z}}\right)\left(\frac{1}{\sqrt{\tau + z - \delta}} + \frac{1}{\sqrt{2r_2 - \tau - z + \delta}}\right) dz \\ = & \frac{2}{\sqrt{\tau r_2}} \left(\int_{-\tau+\delta}^{2r_2-\tau} \frac{1}{\sqrt{(\tau+z)(\tau+z-\delta)}}\,dz + \int_{-\tau+\delta}^{2r_2-\tau} \frac{1}{\sqrt{(\tau-z)(\tau+z-\delta)}}\,dz\right. \\ & \left.\int_{-\tau+\delta}^{2r_2-\tau} \frac{1}{\sqrt{(\tau+z)(2r_2-\tau-z+\delta)}}\,dz + \int_{-\tau+\delta}^{2r_2-\tau} \frac{1}{\sqrt{(\tau-z)(2r_2-\tau-z+\delta)}}\,dz\right) \\ = & \frac{2}{\sqrt{\tau r_2}} \left(\int_{0}^{2r_2-\delta} \frac{1}{\sqrt{z(z+\delta)}}\,dz + \int_{0}^{2r_2-\delta} \frac{1}{\sqrt{z(2\tau-\delta-z)}}\,dz\right. \\ & \left.\int_{\delta}^{2r_2} \frac{1}{\sqrt{z(2r_2+\delta-z)}}\,dz + \int_{\delta}^{2r_2} \frac{1}{\sqrt{z(2(\tau-r_2)-\delta+z)}}\,dz\right) \\ \le & \frac{\kappa}{\sqrt{\tau r_2}} \left(\ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{1}{2(\tau-r_2)-\delta}\right) + 1\right) \ . \end{aligned}$$

Altogether, this yields the lemma. □

B.3 Proofs of Lemmas 15 and 16

Lemma 14 yields the following two lemmas which give bounds on the conditional density of the random variable Δ when only one of the radii R_1 and R_2 is given by integrating over all possible values of R_2 or R_1, respectively.

Lemma 32. *Let τ and r_1 be distances with $0 \le \tau, r_1 \le \sqrt{2}$. We are interested in the conditional density $f_{\Delta|T=\tau,R_1=r_1}$ of the random variable Δ, when the distance T and the radius R_1 are given. For a sufficiently large constant κ, this conditional density is bounded by*

$$f_{\Delta|T=\tau,R_1=r_1}(\delta) \le \frac{\kappa}{\sqrt{r_1} \cdot \tau} \left(\ln\left(\frac{1}{|\delta|}\right) + 1\right) \ .$$

Proof. We can write the conditional density of Δ as

$$\begin{aligned} f_{\Delta|T=\tau,R_1=r_1}(\delta) &= \int_{r_2=0}^{\sqrt{2}} f_{R_2}(r_2) \cdot f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)\,dr_2 \\ &= \int_{r_2=0}^{\sqrt{2}} r_2 \cdot f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)\,dr_2 \ . \end{aligned}$$

Bounding this integral with the help of Lemma 14 yields the lemma. □

The next lemma follows similarly.

Lemma 33. *Let τ and r_2 be distances with $0 \le \tau, r_2 \le \sqrt{2}$. We are interested in the conditional density $f_{\Delta|T=\tau,R_2=r_2}$ of the random variable Δ, when the distance T and the radius R_2 are given. For a sufficiently large constant κ, this conditional density is bounded by*

$$f_{\Delta|T=\tau,R_2=r_2}(\delta) \quad \le \quad \begin{cases} \frac{\kappa}{\tau}\left(\ln\left(\frac{1}{|\delta|}\right)+1\right) & \text{if} \quad r_2 \ge \tau \\ \frac{\kappa}{\sqrt{r_2\tau}}\left(\ln\left(\frac{1}{|\delta|}\right)+\ln\left(\frac{1}{|2(\tau-r_2)-\delta|}\right)+1\right) & \text{if} \quad r_2 \le \tau \end{cases} .$$

Now Lemma 15 and Lemma 16 follow easily from the Lemmas 32 and 33 by integrating over all possible values of T or R_i, respectively. The Lemmas 17 and 18 follow analogously.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de/