Smoothed Analysis of the TSP Algorithms

Jens Ziegler

Geboren am 11. Februar 1989 in Siegburg

September 22, 2012

Bachelorarbeit Mathematik

Betreuer: Prof. Dr. Nitin Saxena

Mathematisches Institut

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Contents

1 Introduction

Smoothed Analysis was first introduced by Daniel A. Spielman and Shang-Hua Teng in 2001. Up until that time, there were only two standard ways of measuring an algorithms complexity: worst-case and average-case-analysis. However, these two techniques are not always adequate in explaining phenomena that occur in practice. The simplex algorithm, for example, has worst-case and average-case exponential running time. Yet, in practice the simplex algorithm often operates in polynomial time. These phenomena can be explained by smoothed complexity: "the smoothed complexity of an algorithm is the maximum over its inputs of the expected running time of the algorithm under slight perturbations of that input" [3]. And if the smoothed complexity of an algorithm is low, it is unlikely that the algorithm has a high running time on realistic instances.

A *travelling salesman problem* (TSP) consists of a set of vertices $\{v_1, v_2, \ldots, v_n\}$ and of a distance $d(v_i, v_j)$ for each pair of vertices $\{v_i, v_j\}$. Generally, a TSP is visualised as graph, in which the edges between two vertices correspond to the distance between those vertices. In this thesis we only consider the symmetric TSP, for which $d(v_i, v_j) = d(v_j, v_i)$. To solve the problem, one has to find a tour of minimum length, that visits each vertex exactly once and returns to the first vertex again at it's end.

The TSP is known to be a NP-hard problem [13], but there exist several algorithms which can approximate the optimal tour to any arbitrary accuracy in polynomial time for Euclidean TSP instances, as shown by Arora [14]. The first approximation algorithm of the TSP was the Christofides algorithm, which first creates a minimum spanning tree on the graph $G = (V, E)$ and finds a minimum weight perfect matching in the complete graph over all vertices with odd degree in the minimum spanning tree. Combining the edges of the matching and the tree to one multigraph allows to find an Eulerian circuit in this multigraph. Then one only has to make the circuit Hamiltonian by skipping already visited nodes. This algorithm achieves an approximation ratio of $1,5$.

Nearest neighbour algorithms chose for every vertex which is being processed the nearest unvisited neighbour. In average, this algorithms have an approximation ratio of $1,25$, but there exist classes of instances on which this algorithms always find the worst possible solution.

Solving a TSP problem exact, however, will result in exponentially large running times. If one would try every possible tour and determine the tour with the shortest length, the running time would be $O(n!)$. Naturally, this method is highly impracticable, even for rather small instances.

For slightly larger instances one can use branch-and-bound algorithms, which have been introduced by Land and Doig [15]. If one combines the branch-and-bound method with techniques of linear programming, one can solve instances up to 200 vertices fairly well. The best results, however, are

achieved by branch-and-cut algorithms, which introduce problem specific cut planes to the branch-and-bound tree, allowing to prune the tree far earlier and more effective than by using mere branch-and-bound procedures. This method also holds the current record in solving TSP instances exactly, with 85,900 vertices [16].
Experts all over the world are currently trying to solve the TSP world tour problem, meaning they are trying to solve a 1,904,711 vertices instance, where the vertices correspond to locations throughout the world. The best solution so far was just recently discovered in October 2011 with a length of 7,515,778,188. Whether this solution is optimal is still unknown.

The *2-Opt-algorithm* (2-Opt) is a local search heuristic for the TSP. It starts with an arbitrary initial tour and modifies the tour in the following way: Let u_1, u_2, v_1, v_2 be vertices of the tour which are distinct and appear in this order. If there are edges $\{u_1, u_2\}$ and $\{v_1, v_2\}$ so that removing these edges and inserting the edges $\{u_1, v_1\}$ and $\{u_2, v_2\}$ would reduce the total length of the tour, a new tour is constructed in that way. Englert, Röglin and Vöcking refer to this as an *improving step.* If there are no improving steps possible, the algorithm terminates with a local optimum, which is in most cases very close to the global optimum itself. A local improvement made by the algorithm will be denoted as a *2-change*[1]. It has been shown by Lueker that it can take 2-Opt an exponential number of steps before finding a locally optimal solution[4], when the edge lengths do not satisfy the triangle inequality. Whether that is also the case when metric instances are considered, has been answered by Englert, Röglin and Vöcking for L_1 and L_2 instances. In [1] they construct such metric instances, for which 2-Opt takes an exponential number of steps. Since this is not relevant for the smoothed analysis of 2-Opt, I won't go into details about this construction.
The distribution of the n vertices, or points, plays an important role for the analysis of 2-Opts complexity. For example, Chandra, Karloff and Tovey show that the expected running time of 2-Opt for Euclidean instances, where the points are placed uniformly at random in the unit square, is bounded by $O(n^{10} \log n)$[5]. For instances, where the distances are measured with the Manhattan metric and the points are distributed in the same way as above, they show that the expected running time is bounded by $O(n^6 \log n)$.
However, Englert, Röglin and Vöcking use a different, more general probabilistic distribution model of the n points. For each vertex v_i the distribution is given by a density function $f_i : [0,1]^2 \to [0, \phi]$ for some given $\phi \geq 1$. The resulting upper bounds for 2-Opt will only depend on the number of vertices and the upper bound ϕ of the density functions. The instances created with such input are denoted as *ϕ-perturbed Euclidean* or *Manhattan instances*, based on which metric has been used. For $\phi = 1$ we get a uniform distribution in the unit square as before, and the larger we chose ϕ, the better we can approximate worst case instances by the distributions of the vertices. This will play an important role in the smoothed analysis of 2-Opt, as we

set $\phi \sim \frac{1}{\sigma^2}$, where σ is a small standard deviation of a Gaussian random variable. The σ is used to perturb the position of the n points after the points have been placed.

Another input model that we will consider is one where the distances of the edges are perturbed instead of the points, so that the structure of the input graph doesn't change. In this model, the perturbation of the length of an edge does not influence the other lengths in the graph. These perturbations are restricted by the density functions $f_e : [0,1] \to [0,\phi]$ for each edge $e \in E$, when the graph G is given by G = (V,E). In this case the maximum density is ϕ for a given $\phi \geq 1$. Such inputs will be denoted as *ϕ-perturbed graphs*.

A *state graph* is a directed graph that contains a vertex for every possible tour and an arc between two vertices exists only when the one tour can be obtained from the other tour by performing one improving 2-Opt step.
Now to Englert, Röglin and Vöcking's first theorem with relevance to the smoothed analysis of 2-Opt:

Theorem 1. *The expected length of the longest path in the 2-Opt state graph*

a) is $O(n^4\phi)$ for ϕ-perturbed Manhattan instances with n points.
b) is $O(n^{4+\frac{1}{3}}\log(n\phi)\phi^{\frac{8}{3}})$ for ϕ-perturbed Euclidean instances with n points.
c) is $O(mn^{1+o(1)}\phi)$ for ϕ-perturbed graphs with n vertices and m edges.

The second theorem confirms an experimental study[6], which indicated that the approximation ratio and the running time of 2-Opt can be improved, if the initial tour is chosen with an insertion heuristic instead of arbitrarily.

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

a) is $O(n^{3.5} \log n\phi)$ on ϕ-perturbed Manhattan instances with n points when one starts with a tour obtained by an arbitrary insertion heuristic.
b) is $O(n^{3+\frac{5}{6}} \log^2(n\phi)\phi^{\frac{8}{3}})$ on ϕ-perturbed Euclidean instances with n points when one starts with a tour obtained by an arbitrary insertion heuristic.
c) is $O(mn^{1+o(1)}\phi)$ on ϕ-perturbed graphs with n vertices and m edges.

The third and last theorem that is required for the smoothed analysis pertains to the approximation ratio of 2-Opt. In general this ratio is close to 1, which is contrary to the theoretical result on the algorithm's worst-case analysis by Chandra, Karloff and Tovey[5]. Yet, they have shown that the approximation ratio is bounded from above by a constant when one considers n points uniformly at random distributed on the unit square. This result applies to Englert, Röglin and Vöcking's model as follows:

Theorem 3. *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 $O(\sqrt{\phi})$.*

In the next section, we will give all the necessary definitions and notations that are required for the understanding of this thesis. Then we will prove theorems 1,2 and 3 and finally come to the smoothed analysis of 2-Opt, and then go over to Karp's partitioning scheme.

2 Preliminaries

We start with some definitions:

Definition A pair (V, d) of a non-empty set V and a function $d : V \times V \to R_{\geq 0}$ 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*).

d is called a metric on V when (V, d) is a metric space and a TSP instance, in which the distances are measured according to a metric, is called a metric TSP instance.

Given two points $X = (x_1, x_2)$ and $Y = (y_1, y_2)$ in $\mathbb{R}^2$, the L_p space provides us with metrics $d_{L_p}(X, Y) = \sqrt[p]{|x_1 - y_1|^p + |x_2 - y_2|^p}$. The L_1 metric is called the *Manhattan metric* and the L_2 metric is often referred to as *Euclidean metric.* An instance of the TSP, where the metric is a L_p metric is called an L_p *instance.* For $p = 2$ we also say *Euclidean instance.* In section 1 we also mentioned *insertion heuristics.* Those create an initial tour for 2-Opt by adding vertex after vertex, until a tour is found, in the following way: Let T_i denote a subtour on a subset S_i of i vertices and let $v \notin S_i$ be the next vertex that is to be inserted into the tour we create. We now chose the edge (x, y) in T_i that minimizes $d(x, v) + d(v, y) - d(x, y)$ and create the tour T_{i+1} by deleting the edge (x, y) from T_i and adding (x, v) and (v, y) to it.

Definition Let A be a set. A σ*-algebra* $\mathcal{A}$ is a set for which

a) $A \in \mathcal{A}$.
b) If a set B is in $\mathcal{A}$ so is its complement A^c.
c) If $A_1, \ldots, A_n \in \mathcal{A}$ then $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

A *measurable space* is a Set A, together with a σ-algebra $\mathcal{A}$ on A. It is denoted by $(A, \mathcal{A})$.

Definition A *probability measure* is a function $P : \mathcal{A} \to \mathbb{R}_0^+$ on a measurable space $(A, \mathcal{A})$ with:

1. $P(A) = 1$
2. For all countable collections $\{A_i\}$ of pairwise disjoint sets the following holds

$$P(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} P(A_i).$$

A *probability space* is a triple $(A, \mathcal{A}, P)$ where $(A, \mathcal{A})$ is a measurable space and P a probability measure.

Definition A *measurable function* is a function $f : A_1 \to A_2$, where $(A_1, \mathcal{A}_1)$ and $(A_2, \mathcal{A}_2)$ are measurable spaces and for every $B \in \mathcal{A}_2 \Rightarrow f^{-1}(B) \in \mathcal{A}_1$.

Let $(A, \mathcal{A}, P)$ be a probability space and $(B, \mathcal{B})$ a measurable space. A *(discrete) random variable* is a function $X : A \to B$ (,B countable,) so that for all $s \in B$ $X^{-1}(s) := \{a \in A : X(a) = s\} \in \mathcal{A}$. In other words: X is a measurable function.

Let $(A, \mathcal{A}, P)$ be a probability space and $X : A \to S$, S countable, a discrete random variable. The *probability mass function* for X is the function $f_X : A \to [0, 1]$ with $f_X(x) = P(X = x) = P(\{s \in S : X(s) = x\})$.

Definition A function $f : \mathbb{R} \to \mathbb{R}$ is called *probability density function*, *density function* or *density* of X if $P(a \leq X \leq b) = \int_a^b f(x)dx$, where the integral is a Lebesgue-integral.
And finally: the *expected value* of a random variable X is defined as $E[X] = \int_A XdP = \int_A X(a)P(da)$.

3 2-Opt

In this section we will go in detail through the proofs of theorems 1 and 2 made by the authors of [1]. We will start with the expected number of 2-Changes regarding random L_1 metric instances, then go over to the L_2 metric and conclude with general TSP instances.

3.1 Expected Number of 2-Changes on the L_1 Metric

We start with a theorem that makes use of weaker bounds than the ones in Theorem 1:

Theorem 4. *Starting with an arbitrary tour, the expected number of steps performed by 2-Opt on ϕ-perturbed L_1 instances is $O(n^6 \log(n)\phi)$.*

Proof. We start with an initial tour. This tour can have a maximum length of $2n$, as it consists of n edges and the length of every edge is bounded from above by 2, due to the fact that we use the L_1 metric and the n points are all in $[0,1]^2$. Now one has to show, that every 2-Opt step increases the tour by a polynomially large amount. In order to do so, we take a fixed 2-Opt step S. Let e_1 and e_2 be the edges that are removed from the tour in this step. e_3 and e_4 denote the newly added edges. Now the improvement of the tour $\Delta(S)$ in step S can be written as

$$\Delta(S) = d(e_1) + d(e_2) - d(e_3) - d(e_4).$$

Now we say that $e_1 = (v_1, v_2)$, $e_2 = (v_3, v_4)$, $e_3 = (v_1, v_3)$ and $e_4 = (v_2, v_4)$. We denote the Cartesian coordinates of v_i with (x_i, y_i), allowing us to rewrite the above equation as

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

And we can even write $\Delta(S)$ as a linear combination of the coordinates. For example, if $x_1 \geq x_2 \geq x_3 \geq x_4$ and $y_1 \geq y_2 \geq y_3 \geq y_4$ we can write it as

$$\Delta(S) = -2x_2 + 2x_3 - 2y_2 + 2x_3.$$

We can put the coordinates in $(4!)^2$ different orders as there are 4! possibilities for the x-coordinates, and 4! possible orders for the y-coordinates. For each of the $(4!)^2$ orders there is a corresponding linear combination like above. Now suppose that all of the non-zero variables except one in one of this linear combinations have already been determined by an adversary. Regardless how those variables are chosen, the linear combination can only take a value in the interval $(0, \epsilon]$ when the last variable x_i lies in an interval of length at most ϵ, for some $\epsilon > 0$. This interval is given by the already determined variables. But we know that

$$P[0 < x_i \leq \epsilon] \leq \epsilon\phi,$$

since we consider a ϕ-perturbed instance and the density of x_i is bound from above by ϕ. That means we can bound the probability that $\Delta(S)$ takes a value in $(0, \epsilon]$ as follows:

$$P[0 < \Delta(S) \leq \epsilon] \leq (4!)^2,$$

because for the improvement to be in that interval, one of the linear combinations has to be. Now we define $\Delta_{min} := min\{\Delta(S)|\Delta(S) > 0\}$, the smallest improving 2-Opt step, and we get

$$P[\Delta_{min} \leq \epsilon] \leq (4!)^2 \epsilon n^4 \phi,$$

as there are $n \cdot (n-1) \cdot (n-2) \cdot (n-3) < n^4$ possibilities to chose the four vertices for a single 2-Opt step, bounding the number of different 2-opt steps from above by n^4. Now let T be a random variable that describes how many 2-Opt steps are done before the algorithm reaches a local optimum. Observe that $T \geq t$ can only hold for a given number t if $\Delta_{min} \leq \frac{2n}{t}$, otherwise we would have a tour that is longer than $2n$ after t steps, and since that is not possible, T couldn't exceed t. This yields

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

Since the number of possible different TSP tours is always bounded from above by $n!$, and non of these tours can appear twice in the local search, T is discrete random variable, which allows us to bound the expected value of T via

$$\begin{aligned} E[T] &= \sum_{t=1}^{n!} P[T \geq t] \leq \sum_{t=1}^{n!} \frac{2(4!)^2 n^5 \phi}{t} \\ &= 2(4!)^2 n^5 \phi \sum_{t=1}^{n!} \frac{1}{t} \leq 2(4!)^2 n^5 \phi \ln(n!) + 1 \\ &= 2(4!)^2 n^5 \phi \cdot O(n \log n) + 1 = O(n^6 \log n \cdot \phi), \end{aligned}$$

where we have bounded the $n!$-th harmonic number by $\ln(n!) + 1$ and with $\ln(n!) = O(n \log n)$. $\square$

This bound only depends on the smallest improvement made by any possible 2-Opt step. But in general one would think that the improvement of a step is larger than Δ_{min}. We achieve this by considering pairs of 2-changes linked by an edge, meaning that one of the added edges of the first 2-change is removed from the tour again by the second 2-change. Then we analyse the smallest improvement of such a pair of 2-changes. It will be shown that the probability of this improvement being much larger than the sum of the smallest and second smallest improving 2-Opt step is very high, and this result will help proving theorem 1.

3.1.1 Pairs of Linked 2-Changes

Now we consider an arbitrary sequence of consecutive 2-changes $S_1, \dots, S_t$ and start with

Lemma 5. *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 $\frac{t}{3} - \frac{n(n-1)}{12}$.*

Proof. Let $S_1, \dots, S_t$ be an arbitrary sequence of consecutive 2-changes and assume that a list $\mathcal{L}$ of linked pairs of 2-changes is created. We can not assume that there are no non-disjoint pairs in $\mathcal{L}$, so we need to modify the list. Let $S_1, \dots, S_{i-1}$ denote the 2-changes that have already been done by the algorithm for $1 \le i \le t$, meaning that the next 2-change in the process would be S_i. Now we denote the edges that are removed from the tour in this step with e_1 and e_2, and the added ones with e_3 and e_4. If there exists a $j > i$ so that e_3 is removed from the tour in S_j, we chose the smallest of such j's and add the pair (S_i, S_j) to $\mathcal{L}$. We do the same for e_3 and our list of linked pairs of 2-changes is created. Now we need to assure that the elements of our list are pairwise disjoint. We do that by creating another list $\mathcal{L}'$, which is empty. Now we go through the first list and check for every element if it is disjoint with all elements of $\mathcal{L}'$. If an element is disjoint, we add it to the new list, which is obviously disjoint in the end. As we have seen above, every 2-change allows 2 possible different pairs for the list $\mathcal{L}$. Minus the cases in which an edge is added to the tour but never removed, are $\frac{n(n-1)}{2}$, since there are n possibilities for the first vertex and $n-1$ for the second and divide by 2 because the TSP is symmetric. That means $\mathcal{L}$ contains at least $2t - \frac{n(n-1)}{2}$ elements. Every of the 2-changes can only be in 4 different pairs of $\mathcal{L}$, since there are only 4 edges to consider in one 2-change. This means each pair in $\mathcal{L}$ is non-disjoint from at most 6 other pairs in the list. We can now conclude the number of elements in $\mathcal{L}'$ as at least a sixth of the elements in $\mathcal{L}$. $\square$

Now we will evaluate the possible cases of added edges in a fixed linked pair of 2-changes. Let $\{v_1, v_2\}$ and $\{v_3, v_4\}$ be the edges removed from the tour in the first 2-change of such a fixed linked pair, and let $\{v_1, v_3\}$ and $\{v_2, v_4\}$ be the added edges. We 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 $\{v_3, v_6\}$. While the vertices v_1, v_2, v_3, v_4 have to be distinct, the vertices v_5 and v_6 are not necessarily distinct from the vertices v_2 and v_4. There are 3 cases we have to consider:

1. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 0$. In this case the edges $\{v_1, v_5\}$ and $\{v_3, v_6\}$ are added to the tour in the second step.
2. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 1$. We assume $v_2 \in \{v_5, v_6\}$ and obtain two subcases: In the second step of the tour the edges a) $\{v_1, v_5\}$ and $\{v_2, v_3\}$ are added to the tour, b) the edges $\{v_1, v_2\}$ and $\{v_3, v_5\}$ are added to the tour.
3. $|\{v_2, v_4\} \cap \{v_5, v_6\}| = 2$. If $v_2 = v_5$ and $v_4 = v_6$ the tour would be the same after the pair of linked 2-changes as before, resulting in the only possibility: $v_2 = v_6$ and $v_4 = v_5$.

We observe that pairs of type 3 can result in too little randomness when the distances are measured via the L_2 metric, meaning the steps are too dependent on each other, so we can't bound the probability that both steps improve the tour at most by ϵ appropriately. It is necessary, to consider that the list in the lemma above is created without pairs of type 3, in order to analyse ϕ-perturbed L_2 instances. We will now show, that there are always enough pairs of type 1 or 2 to complete the analysis.

Let S_i and S_j for $i < j$ be a pair of type 3. Further let $\{v_1, v_2\}$ and $\{v_3, v_4\}$ be the edges exchanged with $\{v_1, v_3\}$ and $\{v_2, v_4\}$ in S_i, and the latter shall be the ones exchanged with $\{v_1, v_4\}$ and $\{v_2, v_3\}$ in S_j. Now consider steps S_l and $S_{l'}$ with $l > j$ and $l' > j$ where the edge $\{v_1, v_4\}$ in S_l and $\{v_2, v_3\}$ in $S_{l'}$ is removed, if such l, l' exist. Neither the pair (S_j, S_l) nor the pair $(S_j, S_{l'})$ can be of type 3, because then we would simply add the removed edges to the tour again. This means that for every pair (S_i, S_j) of type 3 we have two pairs (S_j, S_l) and $(S_j, S_{l'})$ of type 1 or 2, unless such l, l' don't exist. When we consider a pair of type 2 we see that it has at most two pairs of type 3 associated with it. In case a) that would be the two pairs we get when we remove either one of the beforehand added edges and remove $\{v_7, v_8\}$ with $v_2 = v_8$ and $v_3 = v_7$ or $v_1 = v_8$ and $v_5 = v_7$ and add the two resulting edges to the tour. Analogously one can show the same for case b) and for pairs of type 1. Let x be the total number of pairs of type 3, and y be the total number of pairs of type 1 and 2. We have $2y \geq 2x - \frac{n(n-1)}{2}$, because we have to subtract the possible cases of an edge added to the tour, but never removed, and respectively $y \geq x - \frac{n(n-1)}{4}$. The total number of pairs is still at most $2t$ and $2x - \frac{n(n-1)}{4} \leq 2t$, because there cannot be more pairs of type 1, 2 and 3 than there are pairs in total. Which leads to

$$(\star) \qquad x \leq t + \frac{n(n-1)}{8}.$$

As we established earlier, the total number of pairs is at least $2t - \frac{n(n-1)}{2}$ and this yields $y \geq t - \frac{5n(n-1)}{8}$ when we subtract $(\star)$ from it. When we construct the list $\mathcal{L}'$ now, the following lemma must hold:

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

3.1.2 Analysing the Pairs of Linked 2-Changes

To analyse the pairs we first prove two lemmas about pairs of type 1 and 2.

Lemma 7. *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 $O(n^6\epsilon^2\phi^2)$.*

Proof. Let $\{v_1, v_2\}$ and $\{v_3, v_4\}$ be the edges that are replaced by $\{v_1, v_3\}$ and $\{v_2, v_4\}$ in the first step of the pair and let $\{v_1, v_3\}$ and $\{v_5, v_6\}$ are exchanged with $\{v_1, v_5\}$ and $\{v_3, v_6\}$. We can again write the improvements Δ_1 of step one and Δ_2 of step two 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

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

where x_i is the x-coordinate of vertex i and y_i its y-coordinate. Observe, that the improvements might be negative. And again, we can write these improvements as linear combinations of the coordinates, which depend on the order of those coordinates. Let σ_x and σ_y denote a fixed order, and $\Delta_1^{\sigma_x,\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y}$ the corresponding linear combinations. Further, let $\mathcal{A}$ be the event that Δ_1 and Δ_2 takes a value in some interval $(0, \epsilon]$, for $\epsilon > 0$, and analogously $\mathcal{A}^{\sigma_x,\sigma_y}$ for σ_x and σ_y. As we already know, $\mathcal{A}$ only occurs when $\mathcal{A}^{\sigma_x,\sigma_y}$ occurs for at least one pair of σ_x and σ_y, which leads to

$$P[\mathcal{A}] \leq \sum_{\sigma_x,\sigma_y} P[\mathcal{A}^{\sigma_x,\sigma_y}].$$

The number of total orders is $(4!)^2$ and if we can show that $P[\mathcal{A}^{\sigma_x,\sigma_y}]$ for such an order is bounded by $O(\epsilon^2\phi^2)$, the fact that there are $n(n-1)(n-2)(n-3)(n-4)(n-5) < n^6$ possible vertices to chose in order to form a pair of type 1 yields the lemma. To show this bound we need the following lemma:

Lemma 8. *Let $X_1,\dots,X_{2n}$ be random variables and assume that for $i \in \{1,\dots,n\}$, 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 $(X_1,X_2),\dots,(X_{2n-1},X_{2n})$ are independent, and let $X = (X_1,\dots,X_{2n})$ be a vector. Furthermore, let $k \leq n$ and, for $i \in \{1,\dots,2_k\}$, let $\lambda^{(i)} \in \mathbb{Z}^{2n}$ be a row vector such that the vectors $\lambda^{(1)},\dots,\lambda^{(2k)}$ are linearly independent. For a fixed ϵ, we denote by $\mathcal{A}_i$ the event that $\lambda^{(i)} \cdot X \in [0,\epsilon]$ occurs, i.e. the linear combination of the variables $X_1,\dots,X_{2n}$ with the coefficients $\lambda^{(i)}$ takes a value in the interval $[0,\epsilon]$. Then we have*

$$P[\bigcap_{i=1}^{2k} \mathcal{A}_i] \leq (\epsilon\phi)^{2k}.$$

□

The proof of this lemma can be found in [1] and it is denoted by Lemma 31.

We divide all possible pairs of linear combinations $(\Delta_1^{\sigma_x,\sigma_y},\Delta_2^{\sigma_x,\sigma_y})$ into three classes. Class A contains every such pair where one linear combination of it equals 0. A pair of linear combinations is in class B if $\Delta_1^{\sigma_x,\sigma_y} = -\Delta_2^{\sigma_x,\sigma_y}$, and if both linear combinations are linearly independent the pair belongs to class C. So, for a pair $(\Delta_1^{\sigma_x,\sigma_y},\Delta_2^{\sigma_x,\sigma_y})$ of class A or B, the event $\mathcal{A}^{\sigma_x,\sigma_y}$ can not occur, because at least one linear combination of such a pair is ≤ 0. For pairs of class C we can use Lemma 8 and we get $P[\mathcal{A}^{\sigma_x,\sigma_y}] \leq \epsilon^2\phi^2$. What we now have to show is, that the three classes cover every pair of the linear combinations. Let σ_x and σ_y be a fixed order. We can express the improvements as sums of their x and y parts, meaning $\Delta_1^{\sigma_x,\sigma_y} = X_1^{\sigma_x} + Y_1^{\sigma_y}$ and $\Delta_2^{\sigma_x,\sigma_y} = X_2^{\sigma_x} + Y_2^{\sigma_y}$. If we show for the pair of linear combinations $(X_1^{\sigma_x},X_2^{\sigma_x})$ that it belongs to one of the three classes, the pair $(Y_1^{\sigma_y},Y_2^{\sigma_y})$ does too, because they are symmetric. And since the properties we defined for the classes are preserved when we add x and y parts, the pair of linear combination $(\Delta_1^{\sigma_x,\sigma_y},\Delta_2^{\sigma_x,\sigma_y})$ would be also in class A,B, or C. We assume that $(X_1^{\sigma_x},X_2^{\sigma_x})$ is linear dependent and not in class A or B and lead this to contradiction. Luckily, there are only a few cases to consider. Our assumption can only be true, if $X_1^{\sigma_x}$ does not contain x_2 and x_4 $(\star)$, and

$X_2^{\sigma_x}$ does not contain x_5 and x_6 $(\star\star)$, otherwise they would be automatically linear independent. We see that $(\star)$ can only be, if

$$x_3 \geq x_4 \wedge x_2 \geq x_4 \wedge x_2 \geq x_1 \tag{1}$$

$$\vee$$

$$x_3 \leq x_4 \wedge x_2 \leq x_4 \wedge x_2 \leq x_1, \tag{2}$$

and $(\star\star)$ only if

$$x_5 \geq x_6 \wedge x_3 \geq x_6 \wedge x_5 \geq x_1 \tag{3}$$

$$\vee$$

$$x_5 \leq x_6 \wedge x_3 \leq x_6 \wedge x_5 \leq x_1. \tag{4}$$

In case (1) we get

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

and for case (2)

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

and for case (3)

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

and finally for case (4)

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

So we see that for $x_1 \geq x_3$ we get $X_1^{\sigma_x} \in \{0, -2x_1 + 2x_3\}$ and $X_2^{\sigma_x} \in \{0, 2x_1 - 2x_3\}$ in which case the pair of linear combinations $(X_1^{\sigma_x}, X_2^{\sigma_x})$ would be of class A or B. For $x_3 \geq x_1$ we get $X_1^{\sigma_x} \in \{0, 2x_1 - 2x_3\}$ and $X_2^{\sigma_x} \in \{0, -2x_1 + 2x_3\}$ and again, the pair $(X_1^{\sigma_x}, X_2^{\sigma_x})$ would be in either class A or B. This is a contradiction to our assumption, meaning the assumption was wrong which proves the lemma. □

Lemma 9. *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 $O(n^5 \epsilon^2 \phi^2)$.*

Proof. The proof is very similar to the one above, so by using the same notations we now consider pairs of type 2 a) first. We get

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

and

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

Again we assume, that the pair $(X_1^{\sigma_x}, X_2^{\sigma_x})$ is linear dependent and does not belong to either class A or B. With the same argument as above we know that x_2 cannot be in $X_1^{\sigma_x}$ $(\star)$ and x_5 cannot be in $X_2^{\sigma_x}$ $(\star\star)$, which leads to the following conditions for $(\star)$:

$$x_3 \geq x_4 \wedge x_2 \geq x_4 \tag{1}$$

$$\vee$$

$$x_3 \leq x_4 \wedge x_2 \leq x_4, \tag{2}$$

and for $(\star\star)$:

$$x_2 \geq x_5 \wedge x_1 \geq x_5 \tag{3}$$

$$\vee$$

$$x_2 \leq x_5 \wedge x_1 \leq x_5. \tag{4}$$

We can write for case (1):

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

and for case (2)

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

and for case (3)

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

and finally for case (4)

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

If $x_1 \geq x_3$ we have $X_1^{\sigma_x} \in \{0, -2x_1 + 2x_2, -2x_1 + 2x_3, -2x_2 + 2x_3\}$ and $X_2^{\sigma_x} \in \{0, 2x_1 - 2x_2, 2x_1 - 2x_3, 2x_2 - 2x_3\}$, so the pair of those two linear combinations is in class A, B, or C. For $x_3 \geq x_1$ we have $X_1^{\sigma_x} \in \{0, 2x_1 - 2x_2, 2x_1 - 2x_3, 2x_2 - 2x_3\}$ and $X_2^{\sigma_x} \in \{0, -2x_1 + 2x_2, -2x_1 + 2x_3, -2x_2 + 2x_3\}$ and again the pair is in one of the three classes. This is a contradiction to our assumption, so the lemma is proven for pairs of type 2 a). The proof for pairs of type 2 b) is done analogously. □

3.1.3 Proofs of Theorem 1 a) and Theorem 2 a)

With the Lemmas 6,7 and 9 we can now prove Theorem 1 a).

Proof of Theorem 1 a). Let T be the discrete random variable that describes the length of the longest path in the state graph. T is discrete, because the number of possible tours is bounded from above by $n!$ and no tour can appear twice during the local search. For $T \geq t$ there has to be a sequence of t consecutive 2-changes in the state graph. From Lemma 6 we know, that there exist at least $\frac{t}{6} - \frac{5n(n-1)}{48}$ linked pairs of type 1 and 2 in the state graph. Let Δ_{min}^i for $i \in \{1, 2\}$ be the smallest improvement of a pair of improving 2-Opt steps of type i. And for $t > n^2$ we get $\frac{t}{6} - \frac{5n(n-1)}{48} > \frac{t}{6} - \frac{5t}{48} = \frac{t}{16}$, because $n(n-1) < n^2 < t$. With the use of Lemmas 7 and 9 and the same argument we used in the proof of Theorem 4, that $T \geq t$ can only be if the smallest improvement is at most $\frac{2n}{t}$, we have

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

We simply used $\epsilon = \frac{2n}{(t/16)}$ in the second row above. For the expected value of T we get

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

since $T > n^2$ and for $t \geq n^4\phi$ the minimum would always be 1. $\square$

Proof of Theorem 2 a). For an arbitrary set of n points in the unit square, we know that for every metric on $\mathbb{R}^2$ the optimal tour has length $O(\sqrt{n})$ [5] and every insertion heuristic finds an $O(\log n)$-approximation [7]. So if we start 2-Opt with an insertion heuristic the initial tour has length $O(\sqrt{n}\log n)$. This means that for an appropriate constant c and $t > n^2$ we get

$$\begin{aligned} P[T \geq t] &\leq P[\Delta_{min}^{(1)} \leq \frac{c\sqrt{n}\log n}{t}] + P[\Delta_{min}^{(2)} \leq \frac{c\sqrt{n}\log n}{t}] \\ &= O(\min\{\frac{n^7\log^2 n \cdot \phi^2}{t^2}, 1\}) + O(\min\{\frac{n^6\log^2 n \cdot \phi^2}{t^2}, 1\}) \\ &= O(\min\{\frac{n^7\log^2 n \cdot \phi^2}{t^2}, 1\}), \end{aligned}$$

with the same arguments as in the proof above. Again with the same arguments we have

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

what concludes this part of the proof. $\square$

3.2 Expected Number of 2-Changes on the L_2 Metric

Before we can start to prove Theorems 1 b) and 2 b) there has to be done a lot of work. First, we are going to analyse the improvement of a single 2-change.

3.2.1 Analysis of a Single 2-Change

The improvement of a single 2-change can be described by a random variable. Let $\{O, Q_1\}$ and $\{P, Q_2\}$ be the edges that are replaced by $\{O, Q_2\}$ and $\{P, Q_1\}$ in a 2-change. We will first assume slightly weaker conditions for this input model and expand the analysis later to our general input model. Assume that $O = (0,0)$ and P ,Q_1 and Q_2 are chosen independently and uniformly from the interior of a circle with radius $\sqrt{2}$ around O. Let $P = (0,T)$, where $T = d(O,P)$ and let $Z_1 = d(O,Q_1) - d(P,Q_1)$ and $Z_2 = d(O,Q_2) - d(P,Q_2)$. So we have $\Delta = Z_1 - Z_2$, where Δ denotes the improvement of the 2-change. Z_1 and Z_2 are random variables, of which we need to show some properties first.

Lemma 10. *Let $i \in \{1,2\}$, $Q = Q_i$, and $R = d(O,Q)$. 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*

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

For $z \notin [-\tau, \min\{\tau, 2r-\tau\}]$ the density is 0.

Proof. By using polar coordinates for Q we obtain

$$\begin{aligned} Z = d(O,Q) - d(P,Q) &= r - \sqrt{(0 - x_Q)^2 + (\tau - y_Q)^2} \\ &= r - \sqrt{x_Q^2 + y_Q^2 + \tau^2 - 2\tau y_Q} \\ &= r - \sqrt{r^2 + \tau^2 - 2r\tau\cos\alpha}, \end{aligned}$$

where α is the angle between the x-axis and the line between O and Q, which yields $y_Q = r\cos\alpha$. When we chose α uniformly from the interval $[0,\pi]$ instead of $[0, 2\pi)$ the density of Z doesn't change, because it is on the interval $[0,\pi]$ symmetric to itself on $[\pi, 2\pi]$. With this observation we see that for $\alpha = 0$ we get

$$Z = r - \sqrt{r^2 - 2r\tau + \tau^2} = r - \sqrt{(r-\tau)^2} \vee r - \sqrt{(\tau - r)^2} = -\tau \vee 2r - \tau.$$

For $\alpha = \pi$ we have

$$Z = r - \sqrt{r^2 + 2r\tau + \tau^2} = r - \sqrt{(r+\tau)^2} = -\tau.$$

Hence, Z can only take values in $[-\tau, min\{\tau, 2r-\tau\}]$. Since α is restricted to $[0,\pi]$ we can provide an unique inverse function that maps Z to α:

$$\begin{aligned} z = r - \sqrt{r^2 + \tau^2 - 2r\tau\cos\alpha} &\Leftrightarrow (z-r)^2 = r^2 + \tau^2 - 2r\tau\cos\alpha \\ &\Leftrightarrow \cos\alpha = \frac{\tau^2 + 2zr - z^2}{2r\tau}, \end{aligned}$$

which leads to

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

Due to a simple density transformation we can write

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

where f_α is the density of α, namely the uniform density over $[0,\pi]$. Now we only have to bound α'. We know that for $|x| < 1$ we have $(\arccos(x)^{'}) = -\frac{1}{\sqrt{1-x^2}}$, so we get

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

For $r \geq \tau$ we show

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

what proves this case of the lemma as

$$\begin{aligned}
f_{Z|T=\tau,R=r}(z) &= \frac{1}{\pi} \cdot \frac{2(z-r)}{\sqrt{4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4}} \\
&\leq \frac{2(z-r)}{\sqrt{2(z-r)^2(\tau^2 - z^2)}} \\
&= \sqrt{\frac{2}{\tau^2 - z^2}}.
\end{aligned}$$

It is

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

and for $r \leq \tau$ we show (1)

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

to prove this case of the lemma as

$$\begin{aligned}
f_{Z|T=\tau,R=r}(z) &= \frac{1}{\pi} \cdot \frac{2(z-r)}{\sqrt{4r^2\tau^2 - 4r^2z^2 - 4r\tau^2 z + 4rz^3 - \tau^4 + 2\tau^2 z^2 - z^4}} \\
&\leq \frac{2(z-r)}{\sqrt{2(z-r)^2(\tau + z)(2r - \tau - z)}} \\
&= \sqrt{\frac{2}{(\tau + z)(2r - \tau - z)}}.
\end{aligned}$$

It is

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

and

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

□

The next lemma provides us with a bound for the conditional density $f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}$ where $R_1 = d(O, Q_1)$ and $R_2 = d(O, Q_2)$. Notice, that changing T will influence Z_1 and Z_2 while changing R_i, $i \in \{1, 2\}$, will only affect Z_i.

Lemma 11. *Let τ, r_1 and r_2 be distances with $r_1 \leq r_2$ and $0 \leq \tau, r_1, r_2 \leq \sqrt{2}$. 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 $\delta \geq 0$, given τ, r_1, and r_2, is bounded from above by*

$$
\begin{cases}
\frac{\kappa}{\tau}(\ln(\frac{1}{\delta}) + 1) & \text{if } \tau \leq r_1 \wedge \tau \leq r_2 \\
\frac{\kappa}{\sqrt{r_1 r_2}}(\ln(\frac{1}{\delta}) + \ln(\frac{1}{|2(r_1 - r_2) - \delta|}) + 1) & \text{if } r_1 \leq \tau \wedge r_2 \leq \tau \\
\frac{\kappa}{\sqrt{r_1 \tau}}(\ln(\frac{1}{\delta}) + 1) & \text{if } r_1 \leq \tau \leq r_2 \\
\frac{\kappa}{\sqrt{r_2 \tau}}(\ln(\frac{1}{\delta}) + \ln(\frac{1}{|2(\tau - r_2) - \delta|}) + 1) & \text{if } r_2 \leq \tau \leq r_1
\end{cases}.
$$

The following two identities are used in the proof:
For every $c > 0$ and $a > 0$,

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

and

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

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

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

Proof. We can express $f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}$ as convolution of $f_{Z|T=\tau,R=r_1}$ and $f_{Z|T=\tau,R=r_2}$:

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

Let κ be a sufficiently large constant.
In the case $\tau \leq r_1 \wedge \tau \leq r_2$ we know that Z_i can only take values in the interval $[-\tau, \tau]$, which allows the assumption $0 \leq \delta \leq 2\tau$. This leads to

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

The previous lemma gives us the following bound of the densities of Z_i, $i \in \{1, 2\}$:

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

This leads to

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

as we used the above mentioned identities in the last step.
In the case $r_1 \leq \tau \wedge r_2 \leq \tau$ Z can only take values in $[-\tau, 2r_i-\tau]$, so the assumption $0 \leq \delta \leq 2r_1$ can be made, and

$$
f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)=\int\limits_{z=-\tau+\delta}^{min\{-\tau,2r_i-\tau\}} f_{Z|T=\tau,R=r_1}(z)\cdot f_{Z|T=\tau,R=r_2}(z-\delta)dz,
$$

and again we can use our already known bounds, yielding

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

Now we have to consider two subcases, in the first we assume $\delta \in [\max\{0, 2(r_1-r_2)\}, 2r_1]$ and we get

$$
\begin{aligned}
&f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)\\
&\leq \mathcal{R} \int\limits_{-\tau+\delta}^{2r_1-\tau} \Big(\frac{1}{\sqrt{\tau+z}} + \frac{1}{\sqrt{2r_1-\tau-z}}\Big)\Big(\frac{1}{\sqrt{\tau+z-\delta}} + \frac{1}{\sqrt{2r_2-\tau-z+\delta}}\Big) dz\\
&= \mathcal{R}\Big(\int\limits_{-\tau+\delta}^{2r_1-\tau} \frac{dz}{\sqrt{(\tau+z)(\tau+z-\delta)}} + \int\limits_{-\tau+\delta}^{2r_1-\tau} \frac{dz}{\sqrt{(2r_1-\tau-z)(\tau+z-\delta)}}\\
&+ \int\limits_{-\tau+\delta}^{2r_1-\tau} \frac{dz}{\sqrt{(\tau+z)(2r_2-\tau-z+\delta)}}\\
&+ \int\limits_{-\tau+\delta}^{2r_1-\tau} \frac{dz}{\sqrt{(2r_1-\tau-z)(2r_2-\tau-z+\delta)}}\Big)\\
&= \mathcal{R}\Big(\int\limits_{0}^{2r_1-\delta} \frac{1}{\sqrt{z(z+\delta)}} dz + \int\limits_{0}^{2r_1-\delta} \frac{1}{\sqrt{z(2r_1-\delta-z)}} dz\\
&+ \int\limits_{\delta}^{2r_1} \frac{1}{\sqrt{z(2r_2+\delta-z)}} dz + \int\limits_{0}^{2r_1-\delta} \frac{1}{\sqrt{z(2(r_2-r_1)+\delta+z)}} dz\Big)\\
&\leq \begin{cases} \frac{\kappa}{\sqrt{r_1 r_2}}(\ln(\frac{1}{\delta})+1) & \text{if } r_1 \leq r_2 \\ \frac{\kappa}{\sqrt{r_1 r_2}}(\ln(\frac{1}{\delta})+\ln(\frac{1}{2(r_2-r_1)+\delta})+1) & \text{if } r_2 \leq r_1 \end{cases},
\end{aligned}
$$

with $\mathcal{R} = \frac{2}{\sqrt{r_1 r_2}}$.

The second subcase we have to consider is $\delta \in [0, \max\{0, 2(r_1 - r_2)\}]$. Since this case is only relevant if $r_2 \leq r_1$, we obtain

$$
\begin{aligned}
&f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)\\
&\le \mathcal{R}\int\limits_{-\tau+\delta}^{2r_2-\tau+\delta}\left(\frac{1}{\sqrt{\tau+z}}+\frac{1}{\sqrt{2r_1-\tau-z}}\right)\left(\frac{1}{\sqrt{\tau+z-\delta}}+\frac{1}{\sqrt{2r_2-\tau-z+\delta}}\right)dz\\
&=\mathcal{R}\Big(\int\limits_{-\tau+\delta}^{2r_2-\tau+\delta}\frac{dz}{\sqrt{(\tau+z)(\tau+z-\delta)}}+\int\limits_{-\tau+\delta}^{2r_2-\tau+\delta}\frac{dz}{\sqrt{(2r_1-\tau-z)(\tau+z-\delta)}}\\
&+\int\limits_{-\tau+\delta}^{2r_2-\tau+\delta}\frac{dz}{\sqrt{(\tau+z)(2r_2-\tau-z+\delta)}}\\
&+\int\limits_{-\tau+\delta}^{2r_2-\tau+\delta}\frac{dz}{\sqrt{(2r_1-\tau-z)(2r_2-\tau-z+\delta)}}\Big)\\
&=\mathcal{R}\Big(\int\limits_{0}^{2r_2}\frac{dz}{\sqrt{z(z+\delta)}}+\int\limits_{0}^{2r_2}\frac{dz}{\sqrt{z(2r_1-\delta-z)}}\\
&+\int\limits_{0}^{2r_2}\frac{dz}{\sqrt{z(2r_2+\delta-z)}}+\int\limits_{0}^{2r_2}\frac{dz}{\sqrt{z(2(r_2-r_1)-\delta+z)}}\Big)\\
&\le \frac{\kappa}{\sqrt{r_1r_2}}\left(\ln\left(\frac{1}{\delta}\right)+\ln\left(\frac{1}{2(r_1-r_2)-\delta}\right)+1\right),
\end{aligned}
$$

with $\mathcal{R}=\frac{2}{\sqrt{r_1r_2}}$.

The third case is $r_1\le\tau\le r_2$ and we observe that Z_1 only takes values in $[-\tau,2r_1-\tau]$ and Z_2 only takes values in $[-\tau,\tau]$. Again we can assume $0\le\delta\le 2r_1$ and

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

This time we obtain

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

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

In the fourth and last case $r_2 \leq \tau \leq r_1$ Z_1 takes only values in $[-\tau, \tau]$ and Z_2 takes only values in $[-\tau, 2r_2 - \tau]$, what leads to the assumption $0 \leq \delta \leq 2(\tau - r_2)$ and

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

One last time we estimate the bound

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

,

which finally concludes the longest proof of this thesis. □

Now we prove some lemmas about some weaker conditions:

Lemma 12. *Let τ and r_1 be distances with $0 \leq \tau, r_1 \leq \sqrt{2}$. 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*

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

Proof. Using the law of total probability for densities allows us to write the conditional density as

$$
\begin{aligned}
f_{\Delta|T=\tau,R_1=r_1}(\delta) &= \int\limits_{r_2=0}^{\sqrt{2}} f_{R_2}(r_2)\cdot f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)dr_2 \\
&= \int\limits_{r_2=0}^{\sqrt{2}} r_2\cdot f_{\Delta|T=\tau,R_1=r_1,R_2=r_2}(\delta)dr_2 \\
&\le \frac{\kappa}{\sqrt{r_1}\cdot\tau}(\ln(\frac{1}{|\delta|})+1)\int\limits_{r_2=0}^{\sqrt{2}} r_2 \\
&= \frac{\kappa}{\sqrt{r_1}\cdot\tau}(\ln(\frac{1}{|\delta|})+1),
\end{aligned}
$$

where the last inequality follows from Lemma 11. □

Lemma 13. *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*

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

The proof is analogously to the proof of Lemma 12. □

Lemma 14. *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*

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

To proof this lemma one just hast to integrate over all possible values of T and make use of the Lemmas above. □

Analogously one proves the following lemma by integrating over all possible values of R_1 and R_2.

Lemma 15. *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 $f_{\Delta|T=\tau}(\delta)$ of Δ for $\delta \ge 0$ under the condition $T = \tau$ can be bounded by*

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

□

The following lemma is proven by integrating the conditional density in Lemma 10 over all possible values for R.

Lemma 16. *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 $f_{Z_i|T=\tau}(z)$ of Z_i for $z \in [-\tau, \tau]$ under the condition $T = \tau$ can be bounded by*

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

For $z \notin [-\tau, \tau]$ the density is 0.

□

And finally we only need one last lemma to conclude this part. Only the case where we consider a single 2-change is missing, in which we don't need the conditional density of Δ, but the density:

Lemma 17. *For a sufficiently large constant κ, the density $f_\Delta(\delta)$ of Δ for $\delta \ge 0$ can be bounded by*

$$f_\Delta(\delta) \le \kappa(\ln(\frac{1}{\delta}) + 1).$$

We use the bound of Lemma 15 and integrate over all values T can take to prove this lemma. □

3.2.2 Simplified Random Experiments

Now we are going to compare the simplified version from before with the actual situation. We will denote the event $\Delta \in [0, \epsilon]$ with $\mathcal{E}_O$ when we consider the original experiment, and $\mathcal{E}_S$ when we consider the simplified experiment.We will show that $P[\mathcal{E}_O]$ does not differ much from $P[\mathcal{E}_S]$. In the original experiment O isn't necessary the origin, instead let $O = (x, y) \in [0, 1]^2$ and let $\mathcal{R}_{(x,y)}$ denote the region with

$$(P, Q_1, Q_2) \in \mathcal{R}_{(x,y)} \Leftrightarrow \mathcal{E} occurs.$$

The shape of this region is independent from the choice of O, merely the position of this regions changes. Let $\mathcal{V} = \sup_{(x,y)\in[0,1]^2} vol(\mathcal{R}_{(x,y)} \cap [0,1]^6)$, the supremum of the biggest possible volume of the region lying in $[0,1]^6$. Since the density functions of P, Q_1 and Q_2 are bounded from above by ϕ, we can bound $P[\mathcal{E}_O] \leq \phi^3\mathcal{V}$ in the original experiment. We make the simple observation

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

let $\mathcal{V}' = vol(\mathcal{R}_{(0,0)} \cap [-1,1]^6)$, and it is $\mathcal{V} \leq \mathcal{V}'$. Since the circle around the origin with radius $\sqrt{2}$ contains the square $[-1,1]^2$ completely and the density functions of P, Q_1 and Q_2 in the simplified experiment are $\frac{1}{2\pi}$, we get $P[\mathcal{E}_S] \geq \frac{1}{2\pi}$ in this case. Meaning, that $P[\mathcal{E}_S]$ is smaller by a factor of at most $(2\pi\phi)^3$ than $P[\mathcal{E}_O]$.

Lemma 18. *The probability that there exists an improving 2-change whose improvement is at most ϵ is bounded from above by*

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

Proof. We use Lemma 17, a union bound over all possible 2-changes, namely $n^4 > n(n-1)(n-2)(n-3)$ and the factor $(2\pi\phi)^3$ and we have proven the lemma. □

Theorem 19. *Starting with an arbitrary tour, the expected number of steps performed by 2-Opt on ϕ-perturbed L_2 instances is*

$$O(n^7 \cdot \log^2(n) \cdot \phi^3).$$

Proof. The proof is almost entirely the same as the proof of Theorem 4. This time we have to bound $P[\Delta_{min} \leq \epsilon]$ with Lemma 18. A simple substitution in the formulas given in the proof of Theorem 4 and a short calculation yields the theorem. □

In order to improve this bound, we will now consider again pairs of linked 2-changes. But since the results we have for those are restricted for the original experiment, we have to stretch the results to the simplified version first.

Pairs of Type 1. Let for a fixed pair of type one v_3 be the origin and let v_1, v_2, v_4, v_5 and v_6 be chosen uniformly from a circle of radius $\sqrt{2}$ centred at the origin. Let $\mathcal{E}_O$ denote the event, that Δ_1 and Δ_2 lie in $[0,\epsilon]$ for some given ϵ in the original experiment, and let $\mathcal{E}_S$ be the same for the simplified experiment. Using the same methods as above we get that $P[\mathcal{E}_S]$ is smaller than $P[\mathcal{E}_O]$ by at most a factor of $(2\pi\phi)^5$.

Pairs of Type 2. Let for a fixed pair of type two v_2 be the origin and let v_1, v_3, v_4, and v_5 be chosen uniformly from a circle of radius $\sqrt{2}$ centred at the origin. Using the same methods as above we get that $P[\mathcal{E}_S]$ is smaller than $P[\mathcal{E}_O]$ by at most a factor of $(2\pi\phi)^4$.

3.2.3 Analysis of Pairs of Linked 2-Changes

Lemma 20. *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 $O(n^6 \cdot \epsilon^2 \cdot (\log^2(\frac{1}{\epsilon}) + 1) \cdot \phi^5)$.*

Proof. We consider the simplified case first, use the notations from above and observe that the events $\Delta_1 \in [0, \epsilon]$ and $\Delta_2 \in [0, \epsilon]$ are independent, since the coordinates of v_1 are fixed and only v_1 and v_3 play a role in both steps of the linked pair of type 1. The densities of v_2, v_4, v_5 and v_6 are rotationally symmetric, so we don't need to care about the position of v_1, but the distance $d(v_1, v_3)$ gives us all we need. We have

$$\begin{aligned}
P[\Delta_i \in [0, \epsilon] | d(v_1, v_3) = r] &= \int_0^\epsilon f_{\Delta_i | d(v_1,v_3)=r}(\delta) d\delta \\
&\leq \frac{\kappa}{\sqrt{r}} \cdot \epsilon \cdot (\ln(\frac{1}{\epsilon}) + 1),
\end{aligned}$$

where we used the bound from Lemma 15 and κ is a sufficiently large constant. Since Δ_1 and Δ_2 are independent when the distance between v_1 and v_3 is fixed, we obtain

$$\begin{aligned}
P[\Delta_1, \Delta_2 \in [0, \epsilon] | d(v_1, v_3) = r] &\leq \frac{\kappa^2}{r} \cdot \epsilon^2 \cdot (\ln(\frac{1}{\epsilon}) + 1)^2 \\
&\leq \frac{\kappa'}{r} \cdot \epsilon^2 \cdot (\ln^2(\frac{1}{\epsilon}) + 1),
\end{aligned}$$

for a sufficiently large constant κ'. Now we can use the law of total probability again and we get

$$\begin{aligned} P[\Delta_1, \Delta_2 \in [0, \epsilon]|] &= \int_0^{\sqrt{2}} f_{d(v_1,v_3)}(r) \cdot P[\Delta_1, \Delta_2 \in [0, \epsilon]| d(v_1, v_3) = r] dr \\ &= \int_0^{\sqrt{2}} r \cdot P[\Delta_1, \Delta_2 \in [0, \epsilon]| d(v_1, v_3) = r] dr \\ &\leq \sqrt{2}\kappa' \cdot \epsilon^2 \cdot (\ln^2(\frac{1}{\epsilon}) + 1). \end{aligned}$$

Since there are $O(n^6)$ different pairs of type one, because there are six nodes to chose, we can conclude the proof using a union bound over all this pairs and the factor $(2\pi\phi)^5$ we obtained from simplifying the experiment. $\square$

Lemma 21. *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 $O(n^5 \cdot \epsilon^{\frac{3}{2}} \cdot (\log(\frac{1}{\epsilon}) + 1) \cdot \phi^4)$.*

Proof. Let for a fixed pair of type two v_2 be the origin and let v_1, v_3, v_4, and v_5 be chosen uniformly from a circle of radius $\sqrt{2}$ centred at the origin. Since there are only 5 vertices instead of 6 involved now, the distances of the vertices affect the analysis stronger than it was the case with pairs of type 1. The nodes v_1, v_2, and v_3 take part in both 2-changes of the first step of the pair. Because of that, and the fact that there is only one new node, v_5, introduced in the second pair, we haven't the circumstance for pairs of type 2 that fixing one distance makes Δ_1 and Δ_2 independent. We start with pairs of type 2 a). When we analyse $P[\Delta_1 \in [0, \epsilon]]$ we can consider both distances $d(v_1, v_3)$ and $d(v_2, v_3)$ as random variables to calculate the mentioned probability. However, for the second step, given that $v_1, \ldots, v_4$ are already chosen, we can't do this, because we used the randomness of the distances in the first step. This holds not true for $d(v_1, v_5)$ and $d(v_2, v_5)$. So we assume, that $d(v_1, v_3)$ and $d(v_2, v_3)$ are already chosen, meaning $Z = d(v_2, v_5) - d(v_1, v_5)$ has to take a value in an already predetermined interval of length ϵ for Δ_2 to take a value in $[0, \epsilon]$. We can use Lemma 14 to obtain for a sufficiently large constant κ

$$P[\Delta_1 \in [0, \epsilon]| d(v_1, v_2) = r] \leq \frac{\kappa}{\sqrt{r}} \cdot \epsilon \cdot (\ln(\frac{1}{\epsilon}) + 1).$$

Lemma 16 gives us for $|z| \leq r$

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

Since the integral

$$\int_a^b \frac{\kappa}{\sqrt{r^2 - z^2}} dz$$

is biggest for $a = -r \wedge b = -r + \epsilon$ or $a = r - \epsilon \wedge b = r$, remember that the interval has to have length ϵ, Z has the highest probability to take a value in $[-r, -r + \epsilon]$ and $[r - \epsilon, r]$. Thus we have for a sufficiently large constant κ'

$$\begin{aligned} P[\Delta_2 \in [0, \epsilon] | d(v_1, v_2) = r] &\leq \int_{\max\{r-\epsilon, -r\}}^{r} \frac{\kappa}{\sqrt{r^2 - z^2}} dz \\ &\leq \frac{\kappa}{\sqrt{r}} \cdot \int_{\max\{r-\epsilon, -r\}}^{r} \frac{1}{\sqrt{r - |z|}} dz \leq \frac{\kappa' \sqrt{\epsilon}}{\sqrt{r}} \end{aligned}$$

for fixed vertices v_3 and v_4. Together with the bound for $P[\Delta_1 \in [0, \epsilon]]$ we get

$$P[\Delta_1, \Delta_2 \in [0, \epsilon] | d(v_1, v_2) = r] \leq \frac{\kappa\kappa'}{\sqrt{r}} \cdot \epsilon^{\frac{3}{2}} \cdot (\ln(\frac{1}{\epsilon}) + 1).$$

Now we eliminate the condition by integrating over all possible values of r, using the law of total probability:

$$P[\Delta_1, \Delta_2 \in [0, \epsilon]] \leq \int_0^{\sqrt{2}} r \cdot \frac{\kappa\kappa'}{\sqrt{r}} \cdot \epsilon^{\frac{3}{2}} \cdot (\ln(\frac{1}{\epsilon}) + 1) dr = O(\epsilon^{\frac{3}{2}} (\ln(\frac{1}{\epsilon}) + 1)).$$

As there are 5 vertices to chose, we have $O(n^5)$ possible pairs of type 2. With a union bound over them, and taking the $(2\pi\phi)^4$ we get from the simplified experiment into consideration, we have proven the lemma for pairs of type 2 a).
For pairs of type 2 b) the distances $d(v_2, v_5)$ and $d(v_3, v_5)$ are random variables in the second step, and analogously we get the following results:

$$P[\Delta_1 \in [0, \epsilon] | d(v_2, v_3) = \tau] \leq \frac{\kappa}{\tau} \cdot \epsilon \cdot (\ln(\frac{1}{\epsilon}) + 1),$$

for a sufficiently large constant κ using Lemma 15, and

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

for $|z| \leq \tau$ with $Z = d(v_2, v_5) - d(v_3, v_5)$, using Lemma 16. Using the same arguments as above we obtain for fixed vertices v_1 and v_3:

$$\begin{aligned} P[\Delta_2 \in [0,\epsilon] | d(v_2, v_3) = \tau] &\leq \int_{\max\{\tau-\epsilon,-\tau\}}^{r} \frac{\kappa}{\sqrt{\tau^2 - z^2}} dz \\ &\leq \frac{\kappa}{\sqrt{\tau}} \cdot \int_{\max\{\tau-\epsilon,-\tau\}}^{r} \frac{1}{\sqrt{\tau - |z|}} dz \leq \frac{\kappa' \sqrt{\epsilon}}{\sqrt{\tau}}, \end{aligned}$$

for a sufficiently large constant κ'. This leads again to

$$P[\Delta_1, \Delta_2 \in [0,\epsilon]] \leq \int_0^{\sqrt{2}} \tau \cdot \frac{\kappa\kappa'}{\sqrt{\tau}} \cdot \epsilon^{\frac{3}{2}} \cdot (\ln(\frac{1}{\epsilon}) + 1) d\tau = O(\epsilon^{\frac{3}{2}} (\ln(\frac{1}{\epsilon}) + 1)).$$

With the same argument as for pairs of type 2 a) the proof is concluded. □

3.2.4 Proofs of Theorem 1 b) and 2 b)

Proof of Theorem 1 b). Let T be the discrete random variable that describes the length of the longest path in the state graph. T is discrete, because the number of possible tours is bounded from above by $n!$ and no tour can appear twice during the local search. For $T \geq t$ there has to be a sequence of t consecutive 2-changes in the state graph. From Lemma 6 we know, that there exist at least $\frac{t}{6} - \frac{5n(n-1)}{48}$ linked pairs of type 1 and 2 in the state graph. Let Δ^i_{min} for $i \in \{1, 2\}$ be the smallest improvement of a pair of improving 2-Opt steps of type i. And for $t > n^2$ we get $\frac{t}{6} - \frac{5n(n-1)}{48} > \frac{t}{6} - \frac{5t}{48} = \frac{t}{16}$, because $n(n-1) < n^2 < t$. With the use of Lemmas 20 and 21 and the same argument we used in the proof of Theorem 4, that $T \geq t$ can only be if the smallest improvement is at most $\frac{2n}{t}$, we have

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

This leads to

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

When we split the two sums at $t = n^4 \cdot \log(n\phi) \cdot \phi^{\frac{5}{2}}$ and $t = n^{\frac{13}{3}} \cdot \log^{\frac{2}{3}}(n\phi) \cdot \phi^{\frac{8}{3}}$, respectively, we see that the minimum would be 1 for greater values of t, so we conclude

$$
\begin{aligned}
E[T] &= O(n^4 \cdot \log(n\phi) \cdot \phi^{\frac{5}{2}}) + O(n^{\frac{13}{3}} \cdot \log^{\frac{2}{3}}(n\phi) \cdot \phi^{\frac{8}{3}}) \\
&= O(n^{\frac{13}{3}} \cdot \log(n\phi) \cdot \phi^{\frac{8}{3}})
\end{aligned}
$$

and are done. □

Proof of Theorem 2 b). For an arbitrary set of n points in the unit square, we know that for every metric on $\mathbb{R}^2$ the optimal tour has length $O(\sqrt{n})$ [5] and every insertion heuristic finds an $O(\log n)$-approximation [7]. So if we start 2-Opt with an insertion heuristic the initial tour has length $O(\sqrt{n}\log n)$. So for an appropriate constant c and $t > n^2$ we get

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

We get

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

When we split the two sums at $t = n^{\frac{7}{2}} \cdot \log^2(n\phi) \cdot \phi^{\frac{5}{2}}$ and $t = n^{\frac{23}{6}} \cdot \log^{\frac{5}{3}}(n\phi) \cdot \phi^{\frac{8}{3}}$, respectively, we see that the minimum would be 1 for greater values of t, so we conclude

$$\begin{aligned} E[T] &= O(n^{\frac{7}{2}} \cdot \log^2(n\phi) \cdot \phi^{\frac{5}{2}}) + O(n^{\frac{23}{6}} \cdot \log^{\frac{5}{3}}(n\phi) \cdot \phi^{\frac{8}{3}}) \\ &= O(n^{\frac{23}{6}} \cdot \log^2(n\phi) \cdot \phi^{\frac{8}{3}}) \end{aligned}$$

and are done. □

3.3 Expected Number of 2-Changes on General Graphs

In ϕ-perturbed graphs every edge can be considered a random variable, thus they are much more random than ϕ-perturbed L_1 or L_2 instances. This is the main reason to move on from pairs of linked 2-changes to sequences of linked 2-changes in order to analyse the expected number of 2-changes of ϕ-perturbed graphs. We call a sequence $S_1, \ldots, S_k$ *linked* if there exists for every $i < k$ an edge that has been added to the tour in S_i and which is removed again in S_{i+1}. The improvement to the tour of such a linked sequence must be at least $\Delta^{(k)}$, the sum of the smallest improvement of all steps. But in general one can expect the improvement, like in the case of ϕ-perturbed L_1 and L_2 instances, to be much higher. To prove this we need to define linked sequences with additional properties.

3.3.1 Definition of Witness Sequences

Before we define the three different types of witness sequences, there are some general things to say. When m is the number of edges in the graph, then there are $O(m^2)$ different possible choices for S_1. For S_2 there is already one edge predetermined to be removed in the first step, so there are just m possible choices for the second edge of this step, and there are 4 ways to add a new edge in the second step. Together this makes $4m$ possible choices for S_2. Since we can do the same for any S_i, $3 \leq i \leq k$, we can bound the number of linked sequences from above by $m^2 \cdot (4m)^{k-1} = 4^{k-1}m^{k+1}$.

In the definitions of the different types of witness sequences we will use the following notations: e_{i-1} and f_{i-1} denote the edges that are removed from the tour in S_i. e_i and f_i denote the edges that are added to the tour in S_i.

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

Definition 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.*

Definition 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.*

3.3.2 Improvement to the Tour Made by Witness Sequences

As in the sections before, we want to know the probability that there exists a k-witness sequence in which every step improves the tour by at most ϵ. The following lemma tells us exactly that.

Lemma 22. *The probability that there exists a k-witness sequence in which every step is an improvement by at most ϵ is*

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

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

c) bounded from above by $k^2 4^k m^k(\epsilon\phi)^k$ for sequences of type 3.

Proof. a) As we already know, the number of k-witness sequences of type 1 is bounded from above by $4^{k-1}m^{k+1}$. Let $S_1, \ldots, S_k$ be a fixed k-witness sequence of type 1. We assume that in the first step the edges e_0 and f_0 are removed from the tour, and e_1 and g_1 are added. Further, we assume that the lengths e_0, f_0 and g_1 are predetermined by an adversary. We can write the improvement Δ_1 of step one as

$$\Delta_1 = d(e_0) + d(f_0) - d(e_1) - d(g_1),$$

meaning that if this step shall be an improvement to the tour of at most ϵ, $d(e_1)$ has to take a value in an interval with length at most ϵ. This leads to

$$P[d(e_1) \in [a,b]] = \int_a^b f_{d(e_1)}(x)dx \leq \epsilon\phi,$$

since the density of $d(e_1)$ is bounded from above by ϕ and $b-a=\epsilon$. Now we consider a step S_i with $i>1$ and assume that the lengths of e_j, f_j for $j<i$, and g_l for $l \leq i$ are predetermined arbitrarily. Per definition, e_i is not involved in any step S_h with $h<i$, meaning the length of this edge has still to be determined. So we have the same situation as we had for S_1, meaning the probability that S_i improves the tour at most by ϵ is bounded from above by $\epsilon\phi$. Since we got k steps, and for each one we get the same bound, the lemma follows when we combine this result with the bound of the number of possible k-witness sequences of type 1.

b) For k-witness sequences of type 2 there are also $O(m^2)$ possible choices for S_1, and $4m$ possible choices for steps S_i with $1<i<k-1$ The first step introduces 4 new vertices to the sequence and every step S_i with $1<i<k$ at most 2, so the number of different vertices involved in the steps S_j with $j<k-1$ is at most $4+2(k-3)<2k$. Which leaves us with less than $4k^2$ choices for S_{k-1}, because both points contained in f_{k-2} have to be chosen from the vertices that had already been chosen in the steps S_j with $j<k-1$. Now we can do the same for S_k: With the first step there are 4 new edges introduced to the sequence, in every other step S_j with $1<j<k$ there are at most 3 new edges introduced. So there are at most $4+3(k-2)<3k$ edges involved in every step before S_k, which leaves us at most $9k^2$ different choices for S_k. When we put all the results together, we get a bound on the number of different possible k-witness sequences of type 2, namely $36k^4 4^{k-3} m^{k-1} < k^4 4^k m^{k-1}$.
Now we can use the same arguments as for k-witness sequences of type 1, except for the last step S_k, because the edge e_k has already been used before. Thus we get the bound $(\epsilon\phi)^{k-1}$ on the improvement of at most ϵ of every step S_i with $i<k$. Combined with the bound on the number of possible sequences of type 2 yields this part of the lemma.

c) This part of the proof is done analogously, as one shows, that the number of different possible k-witness sequences of type 3 is bounded by $9k^2 4^{k-2} m^k < k^2 4^k m^{k-1}$. Contrary to sequences of type 2, the last step introduces with f_{k-1} a new edge to the tour. So we get again a factor of $(\epsilon\phi)^k$, and combined with the number of possible sequences, the proof of the lemma is complete.

□

Definition 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 $i \leq k$. We call a k-witness sequence improving if every 2-change in the sequence is an improvement. Moreover, by $\Delta_{WS}^{(k)}$ we denote the smallest total improvement made by any improving k-witness sequence.

Using the lemma from above, we can now show that it is highly unlikely for an improving witness sequence to have a small total improvement.

Corollary 23. *For* $0 < \epsilon \leq 1/(4^{k-2}m^{k-1}\phi^{k-2})^{1/(k-2)}$ *and arbitrarily chosen* k *it is*

$$P[\Delta_{WS}^{(k)} \leq \epsilon] \leq 10k^5(4m\epsilon\phi)^2.$$

Proof. According to the lemma above, the fact that every witness sequence of type 2 or 3 must consist of three or more steps, and using the notation from the definition above, we get

$$\begin{aligned} P[\Delta_{WS}^{(k)} \leq \epsilon] &\leq 4^{k-1}m^{k+1}(\epsilon\phi)^k + \sum_{i=3}^{k} i^4 4^i m^{i-1}(\epsilon\phi)^{i-1} + \sum_{i=3}^{k} i^2 4^i m^i (\epsilon\phi)^i \\ &\leq 4^{k-1}m^{k+1}(\epsilon\phi)^k + 4k^4 \sum_{i=3}^{k}(4m\epsilon\phi)^{i-1} + k^2 \sum_{i=3}^{k}(4m\epsilon\phi)^i. \end{aligned}$$

And it is

$$\begin{aligned} \epsilon < \frac{1}{(4^{k-2}m^{k-1}\phi^{k-2})^{\frac{1}{k-2}}} \Leftrightarrow 4m\epsilon\phi < \frac{4m\phi}{4\phi m^{\frac{k-1}{k-2}}} \Leftrightarrow 4m\epsilon\phi < m^{\frac{-1}{k-2}} \\ \Rightarrow 4m\epsilon\phi < 1. \end{aligned}$$

This allows us to bound the two sums in the following way

$$\begin{aligned} P[\Delta_{WS}^{(k)} \leq \epsilon] &\leq 4^{k-1}m^{k+1}(\epsilon\phi)^k + 4k^5(4m\epsilon\phi)^2 + k^3(4m\epsilon\phi)^3 \\ &\leq 4^{k-1}m^{k+1}(\epsilon\phi)^k + 5k^5(4m\epsilon\phi)^2. \end{aligned}$$

A simple calculation shows, that for our choice of ϵ the first term of the sum is smaller or equal to the second term. This concludes the proof. □

3.3.3 Identifying Witness Sequences

Now we have to show, that we can find disjoint k-witness sequences in every sequence of consecutive 2-changes of enough length. In order to do so, we first have to define a witness directed acyclic graph (DAG), that will represent the sequence of 2-changes $S_1, \ldots, S_t$ for $t > n2^k$. To clarify notations, we will speak of nodes and arcs when we consider a DAG W, and of vertices and edges when we consider the input graph G.

W has a node for every edge of the initial tour. Every one of this nodes gets an arbitrary time stamp $i \in \{1, \ldots, n\}$, with the only condition that no two nodes have the same time stamp. Now assume $S_1, \ldots, S_{i-1}$ have already been processed, meaning S_i would be the next step. Let e_{i-1} and f_{i-1} be the edges that are exchanged in S_i with e_i and f_i. The edges e_{i-1} and f_{i-1} were in the tour, so there are nodes u_1 and u_2 in W corresponding to those edges. Now two new nodes u_3 and u_4, which correspond to e_i and f_i, are added to W with the time stamp $n + i$ and we add four new arcs: (u_1, u_3), (u_1, u_4), (u_2, u_3), and (u_2, u_4), which we shall call *twin arcs*. This leads to the fact that every node in W has indegree and outdegree of at most 2 and we call W a *t-witness* DAG. The *height* of a node u of W shall be defined as the length of the shortest path from the node to one of W's leafs. If a node u has a height $\geq k$, we can identify a so called sub-DAG W_u of W, which contains all nodes of W, that can be reached from u while using no more than k arcs, and the arcs between those nodes are induced by W.

Lemma 24. *For every sub-DAG W_u, the 2-changes represented by the arcs in W_u yield a total improvement of at least $\Delta_{WS}^{(k)}$.*

Proof. Let W_u be a fixed sub-DAG with root u. We know that the height of u is at least k, which means that every path from u to a leaf in W_u has length $k-1$. Since there are always two possible arcs to chose at nodes which are no leafs, one can identify 2^{k-1} different sequences of linked 2-changes of length k in W_u. We will now show that at least one of those sequences is either a k-witness sequence, or a sequence which improves the tour as much as the total improvement of one the k-witness sequences would do. In order to do so, we give an recursive algorithm that is initialized with the sequence S_1, which is represented by the two outgoing arcs of u.

Let the algorithm be called by a sequence $S_1, \ldots, S_i$ that has already been constructed. Now the algorithm has to determine whether there will be added a step S_{i+1}, which is linked to S_i, to the sequence, or the sequence is already a k-witness sequence. For that matter, let e_{j-1} and f_{j-1} be the edges that are replaced by e_j and f_j in S_j with $j \leq i+1$. Further, let e_i' and f_i' be the edges that are replaced by e_{i+1}' and g_{i+1}' in S_{i+1}', where S_{i+1}' is the alternative to S_{i+1} in being the next step of the sequence. Let E_i denote all edges involved in steps S_l with $l \leq i$ and E_{i-1} all edges involved in steps S_l with $l \leq i - 1$.

When the algorithm is called with a sequence $S_1, \ldots, S_i$ there is at least one edge, that is new to the tour, added in S_i. Let w.l.o.g. e_i be that edge, meaning $e_i \notin E_{i-1}$. We will see, that when the algorithm is called recursively with a sequence $S_1, \ldots, S_{i+1}$ or $S_1, \ldots, S_{i+1}'$, it will either not have a return value, since it found a witness sequence, or it will return a 2-change S. Whenever that happens, there is a sequence of linked 2-changes in W_u that starts with S_{i+1} if the algorithm was called with $S_1, \ldots, S_i$, or with S_{i+1}' otherwise. After all steps of this sequences are performed, they

introduce exactly the same edges as S to the tour. Analogously, they remove the same edges from the tour, and every other edge either never left the tour, or was never added to the tour. So if this happens, we can replace S_{i+1}, or S'_{i+1} respectively, with S.
Now, this is what the algorithm does when called with a sequence $S_1, \ldots, S_i$:

1. Based on the last step S_i, identify the steps S_{i+1} and S'_{i+1}.

2. If $i = k$, then $S_1, \ldots, S_i$ is a k-witness sequence of type 1.

3. If $e_i \notin E_i$ or $g_{i+1} \notin E_i$, then make a recursive call with the sequence $S_1, \ldots, S_{i+1}$ 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 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 $S_1, \ldots, S'_{i+1}$ 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 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 $S_1, \ldots, S_{i+1}$ is a witness sequence of type 2.
   (b) If $f_i \notin E_i$, then $S_1, \ldots, S_{i+1}$ is a witness sequence of type 3.
   (c) If $e_{i+1}, g_{i+1} \in E_{i-1}$, then $S_1, \ldots, S_{i+1}$ is a 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 ($e_{i+1} \in E_i \quad E_{i-1}$ or $g_{i+1} \in E_i \quad E_{i-1}$) 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 $e_i \in E_{i-1}$ and $e_{i+1}, g_{i+1}, e'_{i+1}, g'_{i+1} \in E_i$:
   (a) If $e_{i+1}, g_{i+1} \notin E_{i-1}$ and $e'_{i+1}, g'_{i+1} \in E_{i-1}$, then $S_1, \ldots, S_i$ is a witness sequence of type 2.
   (b) If $f'_i \notin E_i$, then $S_1, \ldots, S_i, S'_{i+1}$ is a witness sequence of type 3.
   (c) If $f_i, f'_i \in E_i$ and ($e_{i+1} \in E_i \quad E_{i-1}$ or $g_{i+1} \in E_i \quad E_{i-1}$) and ($e'_{i+1} \in E_i \quad E_{i-1}$ or $g'_{i+1} \in E_i \quad E_{i-1}$), 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 $S_1, \ldots, S_{i+1}$ 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 $|\{e_{i+1}, e'_{i+1}, g_{i+1}, g'_{i+1}\} \cap (E_i \; E_{i-1})| = 1$, 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 $S_1, \ldots, S_{i+1}$ is a witness sequence of type 2 as $f'_i \in E_{i-1}$. Observe: The original sequence $S_1, \ldots, S_{i+1}$ together with the step S'_{i+1} yields the same net effect and hence the same improvement as the sequence with the modified step $S_i = S$.

This part of the proof has been taken exactly out of [1]. The algorithm does find k-witness sequences, which are constructed via the sub-DAG, meaning that the improvement made by this 2-changes can't be smaller then the total improvement of any of the k-witness sequences, and that yields the lemma. ☐

Lemma 25. *For $t > n4^{k+2}$, every t-witness DAG contains at least $\frac{t}{4^{k+3}}$ nodes u whose corresponding sub-DAGs W_u are pairwise disjoint.*

Proof. Let W be a t-witness DAG, then W consists of $n+2t$ nodes, for every edge of the initial tour one, and every step S_i introduces two new nodes to W. Further, n of these nodes are leafs, because there still have to be n edges in the tour afterwards. As we already mentioned, the in- and outdegree of every node is bounded from above by 2, which leads to the conclusion, that the number of nodes of height less than k is at most $n2^k$. That implies, there are at least $n + 2t - n2^k \geq t$ nodes in W with height at least k, meaning there is sub-DAG W_u associated to this nodes. Now we construct a list of disjoint sub-DAGs by first adding an arbitrary sub-DAG to it, and then we delete every node, arc and every twin arc of that sub-DAG from W. We add another arbitrary sub-DAG to the list, and repeat the procedure. We do this until there are no more sub-DAGs left in W. Obviously, the sub-DAGs are disjoint.

Each sub-DAG of W consists of at most 2^{k+1} nodes, as the height is k and the in- and outdegree is bound from above 2. Let v be one node of the sub-DAG. As we can only have at most 2^{k+1} nodes to either left or right from v, which induce a sub-DAG that also includes v, each node can be contained in at most $2 \cdot 2^{k+1} = 2^{k+2}$ sub-DAGs. This means, that every sub-DAG can only be non-disjoint from at most $2^{k+1} \cdot 2^{k+2} = 2^{2k+3} < 4^{k+2}$ other sub-DAGs. As there were at least t nodes in W with associated sub-DAGs, the number of disjoint sub-DAGs must be at least $\lfloor \frac{t}{4^{k+2}} \rfloor > \frac{t}{4^{k+3}}$, what proves the lemma. ☐

Lemma 26. *Let k be chosen arbitrarily, and let $S_1, \ldots, S_t$ denote a sequence of consecutive 2-changes performed by the 2-Opt heuristic with $t > n2^k$. The sequence $S_1, \ldots, S_t$ shortens the tour by at least $\frac{t}{4^{k+4}} \cdot \Delta : WS^k$.*

Proof. If we combine Lemmas 24 and 25 they yield this lemma. □

3.3.4 Proof of Theorem 1 c) and 2 c)

Proof of Theorem 1 c). For $t \geq n4^{k+2}$, Lemma 26 tells us, that the tour is shortened by a sequence of 2-changes by at least $\frac{t}{4^{k+3}} \cdot \Delta_{WS}^{(k)}$. Let T once again denote the random variable describing the longest path in the state graph. We have

$$P[T \geq t] \leq P[\frac{t}{4^{k+3}} \cdot \Delta_{WS}^{(k)} \leq n] = P[\Delta_{WS}^{(k)} \leq \frac{n4^{k+3}}{t}].$$

For $t \geq 4^{k+4} n\phi m^{\frac{k-1}{k-2}}$ Corollary 23 gives us

$$P[T \geq t] \leq 10k^5 (\frac{4^{k+4} nm\phi}{t})^2$$

.

Which leaves us with the following expected value of T:

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

with the same arguments as in the proofs of the other parts. We split the sum at $t = nm\phi k^{\frac{5}{2}}$, since for bigger or equal values the minimum is always one. This yields

$$E[T] = O(k^{\frac{5}{2}} 4^k nm^{\frac{k-1}{k-2}})$$

.

With $k = \sqrt{\log m}$ the proof is concluded. □

This directly implies Theorem 2 c), as the expected number of steps equals the expected length of the longest path in the state graph.

3.4 Proof of Theorem 3

At this point we only need to prove Theorem 3 in order to make the smoothed analysis. As it has been shown in [5], that for TSP instances with n points in $[0,1]^2$, where the distances are measured according to a metric that is induced by a norm, every locally optimal solution of 2-Opt is not larger than $c\sqrt{n}$, where c is a constant which depends on the metric. The L_1 and L_2 metrics are both induced by a norm, so the tour 2-Opt finds on corresponding ϕ-perturbed L_1 and L_2 instances is not larger than $O(\sqrt{n})$. This put together with the following lemma will prove the theorem:

Lemma 27. *For ϕ-perturbed L_1 and L_2 instances, it holds*

$$E[\frac{1}{OPT}] = O(\sqrt{\frac{\phi}{n}}),$$

where OPT is the length of the shortest possible tour.

Proof. We denote the vertices of the ϕ-perturbed L_1 or L_2 instance with $p_1, \ldots, p_n$ and partition the unit square into $k = \lceil n\phi \rceil$ smaller squares with length $\frac{1}{k}$. Now we construct a set $P \subseteq \{p_1, \ldots, p_n\}$ in the following way: We start with $P = \{p_1\}$ and check for p_2 if it is in the same square as p_1 or if p_1 is in one of p_2's 8 neighbour squares. If it isn't, we insert p_2 into P. We continue this for every vertex, meaning we check a vertex p_i in the same way with all vertices in P. The shortest tour on P is at most as long as the shortest tour on $\{p_1, \ldots, p_n\}$. To see this, one just has to consider the fact that the triangle inequality holds, because if there is an edge in the optimal tour on P, which is not used in the optimal tour on $\{p_1, \ldots, p_n\}$, then this edge has lower costs then the two edges needed in the tour on $\{p_1, \ldots, p_n\}$. The worst case is obviously $P = \{p_1, \ldots, p_n\}$, where both optimal tours have the same length.

Let X be the number of squares which contain at least one vertex, then P contains at least $\frac{X}{9}$ vertices, due to the construction of P. And because of the construction of the squares, every edge between two vertices in P has length at least $\frac{1}{\sqrt{k}}$. This yields, that the length of the optimal tour on P is at least $\frac{X}{9} \cdot \frac{1}{\sqrt{k}} = \frac{X}{9\sqrt{k}}$.
For $1 \leq i \leq k$ let X_i denote the random variable which equals 1 if square i contains at least one vertex, and 0 otherwise. The probability p_i^j, that vertex j is in square i is induced by the given density function of vertex j for every $1 \leq j \leq n$. Since the densities are bounded from above by ϕ, the probability p_i^j is bounded from above by $\frac{\phi}{k}$, since it is the probability that a point falls in an interval of length $\frac{1}{k}$. Further let for $1 \leq i \leq k$ M_i be the probability mass for square i. It is

$$M_i = \sum_{j=1}^{n} p_i^j \leq \frac{n\phi}{k},$$

and we have

$$E[X_i] = P[X_i = 1] = 1 - P[X_i = 0] = 1 - \prod_{j=1}^{n}(1 - p_i^j) \geq 1 - (1 - \frac{M_i}{n})^n,$$

as $\sum\limits_{j=1}^{n}(1 - p_i^j = \sum\limits_{j=1}^{n} 1 - \sum\limits_{j=1}^{n} p_i^j = n - M_i$, which means that $\prod\limits_{j=1}^{n}(1 - p_i^j)$ is maximized, when all p_i^j are equal. Since the expected value is linear, we get

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

It is $\sum\limits_{i=1}^{k} M_i = n$, because every vertex is per definition in one of the squares, and this means the sum $\sum\limits_{i=1}^{k}(1 - \frac{M_i}{n})^n$ is maximized, when the M_i's are chosen as unbalanced as possible. So we assume that $\lceil \frac{k}{\phi} \rceil$ of them take their biggest possible value of $\frac{n\phi}{k}$ and all the others are zero. This implies for sufficiently large n

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

This provides us with

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

But since it is our goal to prove a bound on the expected value of $\frac{1}{OPT}$, we will show that X, and that implies OPT, too, are sharply concentrated around their mean values. X is the sum of n 0-1-random variables, meaning the X_i are Bernoulli random variables. If they were independent, we could apply the Chernoff bound on them to bound the probability of X taking a value that is smaller than its mean value. However, it was shown in [8], that we can still apply a Chernoff bound, if the X_i are negative dependent, meaning that if some of them are zero, this decreases the probability of the other X_i taking the value zero. This is obviously the case here, and we can use the bound:

$$P[X \leq \frac{n}{10}] \leq e^{\frac{-n}{40}},$$

and since $\frac{1}{X} \leq 1$ we get the following bound on the expected value of $\frac{1}{X}$:

$$\begin{aligned} E[\frac{1}{X}] &< \frac{10}{n} \cdot P[\frac{1}{X} < \frac{10}{n}] + P[\frac{1}{X} \geq \frac{10}{n}] \\ &\leq (1 - e^{\frac{-n}{40}}) \cdot \frac{10}{n} + e^{\frac{-n}{40}} < \frac{11}{n}, \end{aligned}$$

for sufficiently large n. If we combine this with the results from above, this implies

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

□

3.5 Smoothed Analysis of 2-Opt

We will analyse a perturbation model for L_1 and L_2 instances, where the n points are at first chosen by an adversary in the unit square, but then the coordinates of the vertices are perturbed by adding independent random variables to them. This random variables are Gaussian random variables with standard deviation $\sigma \leq 1$. After that, we check if one of the random variable's absolute value is larger than some given $\alpha \geq 1$. If this is the case, we replace the random variable with a new Gaussian random variable with standard deviation σ. We do this as long as it takes to bound all absolute values of the random variables by α. Let X be one of the just drawn random variables, and let Y be an arbitrary Gaussian random variable with standard deviation σ and density function f_Y. As $f_X(x) \leq \sup_{y \in \mathbb{R}} f_Y(y)$, since Y is not bounded by α, we can bound the density of X by

$$f_X(x) \leq \frac{\sup_{y\in\mathbb{R}} f_Y(y)}{P[|Y| \leq \alpha]} \leq \frac{1/(\sigma\sqrt{2\phi})}{1 - \sigma/\sqrt{2\pi} \cdot e^{-\alpha^2/(2\sigma^2)}}.$$

Since the points do not lie in the unit square after the perturbation, but in $[-\alpha, 1+\alpha]^2$, we can't apply Theorems 1-3. In order to be able to apply them, we have to rescale and shift the instance in a way such that it lies again in the unit square. Doing so can increase the density of X by at most a factor of $(2\alpha+1)^2$, as becomes obvious when considering the corresponding integral to the density. When we chose

$$\phi = \frac{(2\alpha+1)^2}{(\sigma\sqrt{2\phi} - \sigma^2 e^{-\alpha^2/(2\sigma^2)})^2} = O(\frac{\alpha^2}{\sigma^2})$$

we can apply the theorems.
We get for L_1 instances:

- Expected length of the longest path in the 2-Opt state graph: $O(\frac{n^4\alpha^2}{\sigma^2})$.
- The expected number of steps performed by 2-Opt: $O(\frac{n^{3.5}\alpha^2}{\sigma^2}\log n)$.
- The approximation ratio: $O(\frac{\alpha}{\sigma})$.

And for L_2 instances:

- Expected length of the longest path in the 2-Opt state graph: $O(\frac{n^{4+1/3}\alpha^{16/3}}{\sigma^{16/3}}\log(\frac{n\alpha^2}{\sigma}))$.
- The expected number of steps performed by 2-Opt: $O(\frac{n^{3+5/6}\alpha^{16/3}}{\sigma^{16/3}}\log^2(\frac{n\alpha^2}{\sigma^2}))$.
- The approximation ratio: $O(\frac{\alpha}{\sigma})$.

If the standard deviation is small enough, meaning $\sigma \leq \min\{\frac{\alpha}{\sqrt{4n\ln n}}, 1\}$, one has not to redraw the Gaussian random variables until their absolute values are bound by α to apply the theorems. The proof can be found in [1].

4 Karp's Partitioning Scheme

Karp's Partitioning Scheme is, like 2-Opt, a heuristic, but only for Euclidean TSP instances. Given a set V consisting of n points in the unit square, the algorithm first partitions the unit square into $k = \sqrt{\frac{n}{log(n)}}$ stripes, with the restraint that each stripe has to contain exactly $\sqrt{n \cdot \log n}$ points. Then the algorithm partitions this stripes again into k cells with the constraint that each cell has to contain exactly $\frac{n}{k^2} = \log n$ points and computes for each cell an optimal TSP tour. In a final step the algorithm joins all computed tours to obtain a tour for V, which shall be denoted by $KP(V)$. This is a very specific version of Karp's Partitioning Scheme, in general it works for every $k = \sqrt{\frac{n}{s}}$ with $s! \leq n$.
When the n points are chosen uniformly and independently in the unit square, it has be shown by Steele that $KP(V)$ converges completely to $TSP(V)$, where $TSP(V)$ denotes the optimal tour on V. We want to extend this to ϕ-perturbed L_1 instances, using methods from [2]. A smoothed analysis of the running time of Karp's Partitioning Scheme has not been done so far, leaving this question still open. We can, however, analyse the expected approximation ratio of the algorithm, which we will do in this section.

4.1 Preliminaries

Definition An *Euclidean functional* is a function $F : ([0,1]^2)^* \to \mathbb{R}$ which maps a finite set $V \subseteq [0,1]^2$ to a real number $F(V)$. In our case $F(V)$ equals $TSP(V)$, the length of the optimal TSP tour over a set V consisting of n points.
We call an Euclidean functional *smooth*, if there is a constant C such that

$$|F(V \cup W) - F(V)| \leq C \cdot \sqrt{|Y|},$$

for all finite $V, W \subseteq [0,1]^2$.
We call an Euclidean functional, according to Frieze and Yukich [11], *near-additive*, if for all partitions $C_1, \ldots, C_s$ of the unit square into cells we have

$$|F(V) - \sum_{i=1}^{l} F(V_i)| \leq O(\sum_{i=1}^{s} diameter(C_i)), \tag{5}$$

for all finite sets $V \subseteq [0,1]^2$, and where V_i is the set of points in cell C_i, and the diameter of a cell is the greatest distance between any pair of vertices in that cell.

Definition The *Nearest neighbour graph* for a set V of n points in a metric space is a directed graph which consists of a vertex for every point in V, and edges (v, w) between all $v, w \in V$, with $v \neq w$, whenever the distance $d(v, w)$ is shorter than $d(v, x)$ for all $x \in V$ with $x \neq v$ and $x \neq w$.

The total edge length $NN(V)$ of the nearest neighbour graph is

$$NN(V) = \sum_{v \in V} \min_{w \in V, w \neq v} ||v - w||,$$

and NN is an Euclidean functional.
In the following, $\mu_F(n, \phi)$ will denote a lower bound on the expected value of an Euclidean functional F, which maps a set of n points.

4.2 Necessary Lemmas and Theorems

We start with a theorem, which has been proven by Rhee in [10] for n identically distributed points, but she also mentions that the proof can be extended to the situation when the points are drawn independently with not necessarily identical distributions.

Theorem 28. *Let V be a set of n points drawn independently according to identical distributions from $[0, 1]^2$. Let F be a smooth Euclidean functional. Then there exist constants C and C' such that for all $t > 0$, we have*

$$P[|F(V) - E[F(V)]| > t] \leq C \cdot e^{\frac{-C' t^4}{n}}.$$

□

The next theorem gives us a first expected approximation ratio of Karp's Partitioning Scheme, but in fact the theorem is much more general. We assume that A is some algorithm, which divides $[0, 1]^2$ into s cells $C_1, \ldots, C_s$, computes optimal solutions for each cell and joins them to a solution for V. Let F be a smooth and near-additive functional, then the value computed by A can be bounded by

$$A(V) \leq \sum_{i=1}^{s} F(V_i) + J',$$

where J' upper bounds the costs of joining the solutions of the cells and V_i is the set of points in C_i. Because F is near-additive we get

$$A(V) \leq F(V) + J,$$

for $J = J' + O(\sum_{i=1}^{s} diameter(C_i))$, which implies

$$\frac{A(V)}{F(V)} \leq 1 + O(\frac{J}{F(V)}),$$

and since $E[F(V)] \geq \mu_F(n, \phi)$ we obtain

$$\frac{E[A(V)]}{E[F(V)]} \leq 1 + O(\frac{J}{\mu_F(n, \phi)}).$$

Theorem 29. *Assume that A has a worst-case approximation ratio of $\alpha(n)$ for instances V consisting of n points. Then the expected approximation ratio of A for ϕ-perturbed L_2 instances on V is*

$$E[\frac{A(V)}{F(V)}] \leq 1 + O(\frac{J}{\mu_F(n, \phi)} + \alpha(n) \cdot e^{\frac{-C\mu_F(n,\phi)^4}{n}},$$

for some constant $C > 0$ and J chosen as above.

Proof. If $F(V) \geq \frac{\mu_F(n,\phi)}{2}$ then we obtain an approximation ratio of $1 + O(\frac{J}{\mu_F(n,\phi)})$ as shown above. If $F(V) < \frac{\mu_F(n,\phi)}{2}$ the ratio is $\alpha(n)$, thus we have

$$\frac{A(V)}{F(V)} \leq \min\{1 + O(\frac{J}{\mu_F(n, \phi)}), \alpha(n)\}.$$

Applying Theorem 28 yields

$$P[F(V) < \frac{\mu_F(n, \phi)}{2}] \leq P[|F(V) - E[F(V)]| > \mu_F(n, \phi)] \leq C' \cdot e^{\frac{-C\mu_F(n,\phi)}{n}},$$

for some constants $C, C' > 0$. Combining this two results gives us

$$E[\frac{A(V)}{F(V)}] \leq 1 + O(\frac{J}{\mu_F(n, \phi)} + \alpha(n) \cdot e^{\frac{-C\mu_F(n,\phi)^4}{n}}),$$

and the proof is complete. ☐

The following lemma will help us to choose $\mu_{TSP}(n, \phi)$ for our smoothed analysis.

Lemma 30. *For ϕ-perturbed L_2 instances we have*

$$E[NN(V)] = \Omega(\sqrt{\frac{n}{\phi}}),$$

where NN is the nearest neighbour functional.

Proof. As we have

$$NN(V) = \sum_{v \in V} \min_{w \in V, w \neq v} ||v - w||,$$

we obtain

$$E[NN(V)] = n \cdot E[min_{i \geq 2} ||v_1 - v_i||],$$

due to the linearity of the expected value. Now we assume, that v_1 has been chosen by an adversary, and $v_2, \ldots, v_n$ have been drawn to their respective density function. We get

$$\begin{aligned}
E[min_{i \geq 2} ||v_1 - v_i||] &= \int_0^\infty P[min_{i \geq 2} ||v_1 - v_i|| \geq r] dr \\
&= \int_0^\infty \prod_{i=2}^n (1 - P[||v_1 - v_i|| \leq r]) dr \\
&\geq \int_0^{1/\sqrt{\phi \pi n}} \prod_{i=2}^n (1 - P[||v_1 - v_i|| \leq r]) dr.
\end{aligned}$$

We can bound the probability that the distance from v_1 to v_i is less or equal to r from above by ϕ times the area of a circle of radius r, since ϕ is an upper bound on the density of every point. This lets us continue with

$$E[min_{i\geq 2}||v_1 - v_i||] \geq \int_0^{1/\sqrt{\phi\pi n}} (1-\phi\pi r^2)^{n-1} dr$$
$$\geq \int_0^{1/\sqrt{\phi\pi n}} (1-\frac{1}{n})^{n-1} dr \geq \frac{1}{e\sqrt{\phi\pi n}},$$

as $1 - \phi\pi r^2 \geq 1 - \frac{1}{n}$ for $r \in [0, 1/\sqrt{\phi\pi n}]$ and $(1-\frac{1}{n})^{n-1} \geq \frac{1}{e}$. This implies $E[min_{i\geq 2}||v_1 - v_i||] = \Omega(1/\sqrt{n\phi})$, which completes the proof. □

4.3 Smoothed Analysis of Karp's Partitioning Scheme

Since we now have the tools to complete the smoothed analysis of the algorithm's approximation ratio, we still need to figure out how to choose J, $\mu_{TSP}(n, \phi)$ and $\alpha(n)$ to be able to apply Theorem 29.

Since the nearest neighbour functional is a lower bound for the TSP, we can apply Lemma 30 to obtain $\mu_{TSP}(n, \phi) = \Omega(\sqrt{n/\phi})$. We also have $\mu_{TSP}(n, \phi) = O(\sqrt{n/\phi})$, due to the results of Chandra, Karloff and Tovey ([5]), which state that the length of an optimal TSP tour is $O(\sqrt{n})$. This implies

$$\mu_{TSP}(n, \phi) = \Theta(\sqrt{n/\phi}).$$

Further we use the following bound provided by Karp [11] and Steele [12]:

$$KP(V) \leq TSP(V) + 6k = TSP(V) + 6\sqrt{n/\log n},$$

for $k^2 = n/\log n$. This bound gives us

$$J = O(\sqrt{\frac{n}{\log n}}).$$

Now only $\alpha(n)$ is left to be determined. Let $v, w \in V$ be two arbitrary points. Every tour has to visit both points, so according to the triangle inequality the tour must have a length of at least $2||v - w||$. Since every tour has exactly n edges, we obtain an upper bound on the length of every possible tour: $\frac{n}{2}TSP(V)$. Thus we get

$$\alpha(n) = \frac{n}{2},$$

which leads us to the final theorem of this thesis:

Theorem 31. *For $\phi \in o(\sqrt{n/\log n})$, the expected approximation ratio of Karp's Partitioning Scheme is*

$$E[\frac{KP(V)}{TSP(V)} \leq 1 + O(\sqrt{\frac{\phi}{\log n}}).$$

Proof. We use $\mu_{TSP}(n, \phi) = \Theta(\sqrt{n/\phi})$, $J = O(\sqrt{\frac{n}{\log n}})$ and $\alpha(n) = \frac{n}{2}$ with Theorem 29 to obtain

$$E[\frac{KP(V)}{TSP(V)}] \leq 1 + O(\sqrt{\frac{\phi}{\log n}}) + O(n \cdot e^{-\Omega(\frac{n}{\phi^2})})..$$

$\phi \in o(\sqrt{n/\log n})$ implies that $O(\sqrt{\frac{\phi}{\log n}})$ is an upper bound for $O(n \cdot e^{-\Omega(\frac{n}{\phi^2})})$, which concludes the proof. □

5 Conclusions and Open Questions

The gap between theoretical and practical results concerning the TSP is still rather big, and therefore it is still much work left to be done. For example, we have analysed 2-Opt by constructing lower bounds on the smallest possible improvement of sequences of linked 2-changes. This pivot rule, however, is in practice much too pessimistic, as the improvement can be expected to be much better per step. Furthermore, not even the complexity of computing locally optimal solutions for 2-Opt is known yet. As until today there has not been any published work on locally optimal solutions which can be applied to our case.

Another interesting fact is, that the approximation ratio of 2-Opt in experiments were sometimes very close to 1. Not any theoretical work comes even close to this results.

The smoothed analysis for ϕ-perturbed graphs is still an open question. If one finds a suitable perturbation model for such instances, he would be provided by this thesis with the necessary theorems to analyse ϕ-perturbed graphs.

For Karp's Partitioning Scheme we have only seen a smoothed analysis of the approximation ratio, but no work has been done on the general running time. The total running time is $2^{O(n/k^2)} poly(n/k^2) + O(k^2)$, which is polynomial in n for $k^2 = n/\log n$. Furthermore, the results we showed are for a very specific choice of k. It would be interesting to see results for general possible values of k.

In general, there have not been many results for the TSP in the field of smoothed analysis, especially on the known exact algorithms, e.g. branch-and-cut algorithms. Though the running time of such algorithms is of course exponential, a smoothed analysis could provide new methods to improve this algorithms further. But in order for a successful analysis, a suitable approach has to be found first.

References

[1] Matthias Englert, Heiko Röglin, and Berthold Vöcking. *Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP.* In Proc. of the 18th ACM-SIAM Symp. on Discrete Algorithms (SODA), pages 12951304, 2007.

[2] Markus Bläser, Bodo Manthey, and B. V. Raghavendra Rao. *Smoothed analysis of partitioning algorithms for Euclidean functionals.* In Frank Dehne, John Iacono, and Jörg-Rüdiger Sack, editors, Proc. of the 12th Algorithms and Data Structures Symposium (WADS), volume 6844 of Lecture Notes in Computer Science, pages 110-121. Springer, 2011.

[3] www.cs.yale.edu/homes/spielman/SmoothedAnalysis/framework.html - the so called homepage of Smoothed Analysis.

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

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

[6] David S. Johnson and Lyle A. McGeoch. *The Travelling 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.

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

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

[9] J. Michael Steele. *Subadditive Euclidean functionals and nonlinear growth in geometric probability*. Ann. Probab., 9(3):365-376, 1981.

[10] Wansoo T. Rhee. *A matching problem and subadditve euclidean functionals*. Ann. Appl. Probab., 3(3):794-801, 1993.

[11] J. Michael Steele. *Complete convergence of short paths in Karp's algorithm for the TSP*. Math. Oper. Res., 6:374-378, 1981.

[12] Richard M. Karp. *Probabilistic analysis of partitioning algorithms for the travelling salesman problem in the plane*. Math. Oper. Res., 2(3):209-224, 1977.

[13] Richard M. Karp. *Reducibility Among Combinatorial Problems*. In R. E. Miller and J. W. Thatcher (editors). Complexity of Computer Computations. New York: Plenum. pp. 85103, 1972.

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

[15] A. H. Land and A. G. Doig. *An automatic method of solving discrete programming problems*. Econometrica 28 (3): pp. 497520, 1960.

[16] D. L. Applegate, R. M. Bixby, V. Chvtal, W. J. Cook. *The Traveling Salesman Problem*. ISBN 0-691-12993-2, 2006.