Smoothed Analysis of the 2-Opt Algorithm for the General TSP

MATTHIAS ENGLERT, University of Warwick
HEIKO RÖGLIN, University of Bonn
BERTHOLD VÖCKING, RWTH Aachen University

2-Opt is a simple local search heuristic for the traveling salesperson problem, which performs very well in experiments both with respect to running time and solution quality. In contrast to this, there are instances on which 2-Opt may need an exponential number of steps to reach a local optimum. To understand why 2-Opt usually finds local optima quickly in experiments, we study its expected running time in the model of smoothed analysis, which can be considered as a less pessimistic variant of worst-case analysis in which the adversarial input is subject to a small amount of random noise.

In our probabilistic input model an adversary chooses an arbitrary graph G and additionally a probability density function for each edge according to which its length is chosen. We prove that in this model the expected number of local improvements is $O(mn\phi \cdot 16^{\sqrt{\ln m}}) = m^{1+o(1)}n\phi$, where n and m denote the number of vertices and edges of G, respectively, and ϕ denotes an upper bound on the density functions.

CCS Concepts: •**Theory of computation** → **Design and analysis of algorithms; Graph algorithms analysis;**

Additional Key Words and Phrases: Traveling salesperson problem, local search, 2-Opt, probabilistic analysis, smoothed analysis

ACM Reference Format:
Matthias Englert, Heiko Röglin, and Berthold Vöcking, 2016. Smoothed Analysis of the 2-Opt Algorithm for the General TSP. *ACM Trans. Algor.* V, N, Article A (January YYYY), 15 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

An instance of the traveling salesperson problem (TSP) consists of a set of cities and the pairwise distances between these cities. The goal is to find the shortest tour that visits every city exactly once and returns to the starting city in the end. The TSP is one of the most studied optimization problems and numerous theoretical and experimental results have been obtained. In experiments the most successful heuristics for the TSP are based on the principle of local search. These heuristics start with some solution and improve it by local operations until a local optimum is reached. Even though the TSP is NP-hard to approximate, in many cases these heuristics quickly compute very good solutions.

The 2-Opt algorithm is a particularly simple local search heuristic for the TSP. It starts with an arbitrary initial tour and incrementally improves this tour by ex-

This work was supported in part by the EU within the 6th Framework Programme under contract 001907 (DELIS) and by DFG grants VO 889/2 and WE 2842/1.
An extended abstract appeared in Proc. of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA 2007). This extended abstract also contains additional results that have already appeared in another journal article [Englert et al. 2014].
Author's addresses: M. Englert, DIMAP and Dept. of Computer Science, University of Warwick, englert@dcs.warwick.ac.uk; H. Röglin, Dept. of Computer Science, University of Bonn, roeglin@cs.uni-bonn.de; B. Vöcking, Dept. of Computer Science, RWTH Aachen University.

DOI: http://dx.doi.org/10.1145/0000000.0000000

changing two edges from the current tour with two edges that are not in the current tour (ensuring that after the exchange another tour is obtained and that this tour is shorter than the current tour). We will call such a local improvement an *improving 2-change*. 2-Opt terminates if the current tour admits no improving 2-change anymore. Lueker [Lueker 1975] has constructed instances for the general TSP on which 2-Opt can make an exponential number of local improvements. In contrast to this, in experiments the 2-Opt heuristic needs only a subquadratic number of local improvements until it reaches a local optimum [Johnson and McGeoch 1997].

The reason for the big discrepancy between Lueker's result and the experimental observations is that worst-case instances for 2-Opt have a very artificial structure and do not occur naturally in applications. In order to provide a theoretical underpinning of this statement, we study the running time of the 2-Opt algorithm in the framework of smoothed analysis, which has originally been invented by Spielman and Teng [Spielman and Teng 2004] to explain the practical success of the simplex method. This model can be considered as a less pessimistic variant of worst-case analysis in which the adversarial input is subject to a small amount of random noise and it is by now a well-established alternative to worst-case analysis.

In the model we consider, an adversary specifies an arbitrary graph $G = (V, E)$ with n nodes and m edges. The nodes represent the cities and the edges represent the roads between the cities along which the salesperson can travel. Every edge $e \in E$ has a certain length $d(e) \geq 0$. Instead of fixing each edge length deterministically, the adversary can only specify, for each edge $e \in E$, a probability density function $f_e : [0,1] \to [0,\phi]$ according to which the length $d(e)$ is chosen independently of the other edge lengths. The parameter $\phi \geq 1$ determines how powerful the adversary is. The adversary can, for example, choose for each edge length an interval of length $1/\phi$ from which it is chosen uniformly at random. This shows that in the limit for $\phi \to \infty$ the adversary is as powerful as in a classical worst-case analysis, whereas the case $\phi = 1$ constitutes an average-case analysis with uniformly chosen edge lengths. We call an instance of this form a *ϕ-perturbed graph*. Note that the restriction to the interval $[0,1]$ is merely a scaling issue and entails no loss of generality. In particular, the restriction $d(e) \geq 0$ is without loss of generality as well because negative distances can be avoided by adding the same sufficiently large number to each distance. This does neither affect the behavior of 2-Opt nor does it change the relative order of different tours because every tour contains exactly n edges.

The TSP is often defined only for complete graphs, in which the distance between every pair of cities is finite. In contrast to this, we do not need to assume that the graph G is complete. This model is slightly more general because by leaving out edges, one can explicitly forbid the salesperson to travel directly between certain cities. However, it makes only sense to apply the 2-Opt algorithm to graphs for which at least some tour is known because for general graphs it is already NP-hard to find an initial tour.

When talking about the number of local improvements, it is convenient to consider the *state graph*. For a given graph G, the nodes of this directed graph correspond to the possible tours in G and an arc from a node v to a node u is contained if and only if u can be obtained from v by performing one improving 2-change. Observe that the state graph is acyclic because the tour length is strictly decreasing on any path. We study the length of the longest path in the state graph because this is the maximal number of local improvements the 2-Opt algorithm can make, regardless of the initial tour and regardless of which local improvement is chosen if multiple are possible in the current tour.

THEOREM 1.1. *For every ϕ-perturbed graph with n vertices and m edges the expected length of the longest path in the 2-Opt state graph is $O(mn\phi \cdot 16^{\sqrt{\ln m}}) = m^{1+o(1)}n\phi$.*

This theorem provides an explanation why worst-case examples do not occur in experiments. It shows that already a small amount of randomness in the edge lengths makes it very unlikely to obtain an instance on which 2-Opt can take more than a polynomial number of steps. In practice, random noise can originate, for example, from measurement errors. We can also use random noise to model influences that we cannot quantify exactly but for which we do not have any reason to believe that they are adversarial.

1.1. Related Work

Lueker [Lueker 1975] has constructed TSP instances whose state graphs contain exponentially long paths. This result was generalized to k-Opt, for arbitrary $k \geq 2$, by Chandra, Karloff, and Tovey [Chandra et al. 1999]. These negative results, however, use arbitrary graphs that cannot be embedded into low-dimensional Euclidean space. In [Englert et al. 2014] we have extended these results and constructed two-dimensional Euclidean instances whose 2-Opt state graphs contain exponentially long paths. Also for every other L_p metric, we have constructed two-dimensional instances with exponentially long paths in the 2-Opt state graph.

For Euclidean instances in which n points are placed independently uniformly at random in the unit square, Kern [Kern 1989] has shown that the length of the longest path in the state graph is bounded by $O(n^{16})$ with probability at least $1 - c/n$ for some constant c. Chandra, Karloff, and Tovey [Chandra et al. 1999] have improved this result by bounding the expected length of the longest path in the state graph by $O(n^{10} \log n)$. For instances in which n points are placed uniformly at random in the unit square and the distances are measured according to the Manhattan metric, Chandra, Karloff, and Tovey have shown that the expected length of the longest path in the state graph is $O(n^6 \log n)$.

In [Englert et al. 2014] we have considered a more general probabilistic input model and improved the previously known bounds. The probabilistic model underlying our analysis allows different points to be placed independently according to different continuous probability distributions in the unit hypercube $[0, 1]^d$, for some constant *dimension* $d \geq 2$. The distribution of a point p is determined by a density function $f_p\colon [0, 1]^d \to [0, \phi]$ for some given $\phi \geq 1$. We have proved that in this model the expected length of the longest path in the 2-Opt state graph is $O(n^4\phi)$ for the Manhattan metric and $O(n^{4+1/3} \log(n\phi)\phi^{8/3})$ for the Euclidean metric.

For the case that every point is perturbed by Gaussian noise with standard deviation σ, the results in [Englert et al. 2014] give rise to a bound on the expected length of the longest path in the 2-Opt state graph that is polynomial in n and $1/\sigma^d$ for the Euclidean metric. This has been improved by Manthey and Veenstra [Manthey and Veenstra 2013] who proved for this case an upper bound that is polynomial in n and $1/\sigma$.

2. OUTLINE OF THE ANALYSIS

Before we prove Theorem 1.1, we prove a weaker (yet polynomial) bound on the expected number of 2-changes. The proof of this weaker bound illustrates our proof technique and it sheds light on the problems one has to solve in order to derive a better bound. We discuss these problems and outline our approach in Section 2.2.

In the following we use the notation $[n]$ to denote the set $\{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$.

2.1. A Simple Polynomial Bound

THEOREM 2.1. *For every ϕ-perturbed graph with n vertices and m edges the expected length of the longest path in the 2-Opt state graph is at most $m^2 n^2 \ln(n)\phi$.*

PROOF. First we observe that every tour has length at most n because it contains n edges and every edge has length at most 1 in our probabilistic input model. Let Δ denote the smallest improvement made by any improving 2-change. Then every sequence of ℓ consecutive improving 2-changes decreases the length of the tour by at least $\ell\Delta$. Hence, regardless of the initial tour, after $n/\Delta + 1$ improving 2-changes the length of the tour must have decreased below zero, which is not possible. Thus a lower bound for the smallest possible improvement Δ immediately implies an upper bound of n/Δ on the length of the longest path in the 2-Opt state graph.

In the following we first prove that for any $\varepsilon > 0$,

$$\mathbf{Pr}\left[\Delta \leq \varepsilon\right] \leq m^2 \varepsilon \phi. \tag{1}$$

We denote the improvement made by a 2-change in which the edges e_1 and e_2 are exchanged with the edges e_3 and e_4 by

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

With this notation we can write the smallest possible improvement made by any improving 2-change as

$$\Delta = \min_{\substack{e_1, e_2, e_3, e_4 \\ \Delta(e_1, e_2, e_3, e_4) > 0}} \Delta(e_1, e_2, e_3, e_4),$$

where the minimum is taken over all tuples $(e_1, e_2, e_3, e_4) \in E^4$ for which e_1, e_3, e_2, e_4 is a 4-cycle in G because only these tuples could possibly form a 2-change.

First we bound the probability that a fixed 2-change in which the edges e_1 and e_2 are exchanged with the edges e_3 and e_4 is improving but yields an improvement of at most ε. This corresponds to the event $\Delta(e_1, e_2, e_3, e_4) \in (0, \varepsilon]$. We use the principle of deferred decisions and assume that the lengths $d(e_2)$, $d(e_3)$, and $d(e_4)$ have already been fixed arbitrarily. Then the event $\Delta(e_1, e_2, e_3, e_4) \in (0, \varepsilon]$ is equivalent to the event $d(e_1) \in (\kappa, \kappa+\varepsilon]$, where $\kappa = d(e_4)+d(e_3)-d(e_2)$ is some fixed value. As $d(e_1)$ is a random variable whose density is bounded from above by ϕ, the probability that $d(e_1)$ assumes a value in a fixed interval of length ε is at most $\varepsilon\phi$. Hence,

$$\mathbf{Pr}\left[\Delta(e_1, e_2, e_3, e_4) \in (0, \varepsilon]\right] \leq \varepsilon\phi.$$

We apply a union bound over all possible 2-changes. There are at most $\binom{m}{2} < \frac{m^2}{2}$ choices for the set $\{e_1, e_2\}$ and, once this set is fixed, there are two choices for the set $\{e_3, e_4\}$ because e_1, e_3, e_2, e_4 has to be a 4-cycle. Hence, the total number of different 2-changes is bounded from above by m^2, which yields

$$\mathbf{Pr}\left[\Delta \in (0, \varepsilon]\right] \leq \mathbf{Pr}\left[\exists e_1, e_2, e_3, e_4 : \Delta(e_1, e_2, e_3, e_4) \in (0, \varepsilon]\right] \leq m^2 \varepsilon \phi.$$

This concludes the proof of (1).

With the help of (1) we can prove the theorem. We have argued above that the number of steps that 2-Opt can make is bounded from above by n/Δ. Let T denote the length of the longest path in the state graph. This number can only be greater than or equal to $t \in \mathbb{N}$ if $n/\Delta \geq t$, which is equivalent to $\Delta \leq n/t$. Hence, due to (1),

$$\mathbf{Pr}\left[T \geq t\right] \leq \mathbf{Pr}\left[\Delta \leq \frac{n}{t}\right] \leq \frac{m^2 n \phi}{t}.$$

One important observation is that T is always bounded from above by $n!$ because this is an upper bound on the number of different tours, which equals the number of nodes in the state graph. Hence, we obtain the following bound for the expected value of T:

$$\mathbf{E}[T] = \sum_{t=1}^{n!} \mathbf{Pr}[T \geq t] \leq \sum_{t=1}^{n!} \frac{m^2 n\phi}{t} = m^2 n\phi \cdot \sum_{t=1}^{n!} \frac{1}{t} \leq m^2 n^2 \ln(n)\phi.$$

Here we used the inequality $\sum_{t=1}^{n!} \frac{1}{t} \leq 1 + \ln(n!) \leq n\ln(n)$, which holds for $n > 3$. □

2.2. How to Improve the Simple Bound

The bound in Theorem 2.1 is only based on analyzing the smallest improvement Δ made by any of the 2-changes. Intuitively this is too pessimistic because most of the 2-changes might yield a much larger improvement than Δ. For example, two consecutive 2-changes yield an improvement of at least Δ plus the improvement Δ' of the second smallest 2-change. This observation alone, however, does not suffice to improve the bound substantially. In our analysis of the Manhattan and the Euclidean TSP [Englert et al. 2014] we have shown that one can regroup the 2-changes in any sufficiently long path in the state graph to pairs such that each pair of 2-changes is *linked by an edge*, meaning that one edge added to the tour in the first 2-change of the pair is removed from the tour in the second 2-change of the pair. Then we have analyzed the smallest improvement made by any pair of linked 2-changes. This improvement is at least $\Delta+\Delta'$ but one can hope that it is much larger because it is unlikely that the 2-change that yields the smallest improvement and the 2-change that yields the second smallest improvement form a pair of linked steps. We have shown that this is indeed the case and use this result to prove stronger bounds on the expected length of the longest path in the 2-Opt state graph.

The analysis of the Manhattan TSP in [Englert et al. 2014] can easily be adapted to the model of ϕ-perturbed graphs studied in this article. This results in a bound of $O(m^{3/2}n\phi)$ for the expected length of the longest path in the state graph (observe that for complete graphs this coincides with the bound of $O(n^4\phi)$ for the Manhattan TSP proved in [Englert et al. 2014]). In order to prove Theorem 1.1, we will not only consider linked pairs of 2-changes but longer sequences of linked steps. We call a sequence $S_1, \dots, S_k$ of 2-changes linked if for each $i \in [k-1]$ the steps S_i and S_{i+1} are linked by an edge. For the Manhattan and the Euclidean TSP this is not easily possible due to dependencies between the steps in a linked sequence. In ϕ-perturbed graphs these dependencies are less severe because the edge lengths are independent random variables, which makes it possible to study also larger values of k.

In order to control the dependencies, we introduce the notion of *witness sequences* in Section 3.1. These are linked sequences that satisfy some additional technical properties. In Section 3.2 we show that any witness sequence yields a significant improvement with high probability and in Section 3.3 we prove that the steps in any path in the state graph of length $t > n4^{k+1}$ can be grouped into at least $t/4^{k+1}$ disjoint witness sequences of length k. We will see in Section 3.4 that these results together yield the desired bound on the expected length of the longest path in the state graph if one sets $k = \sqrt{\ln m}$.

3. PROOF OF THEOREM 1.1

3.1. Definition of Witness Sequences

In this section, we define three different types of witness sequences. As mentioned above, a witness sequence $S_1, \dots, S_k$ has to be linked, i.e., for $i \in [k-1]$, there must exist an edge that is added to the tour in step S_i and removed from the tour in step S_{i+1}.

LEMMA 3.1. *There are at most $4^{k-1}m^{k+1}$ different linked sequences of length k.*

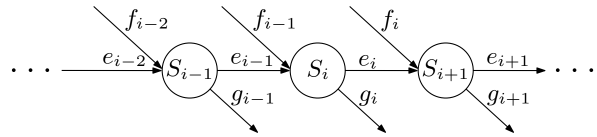

Fig. 1. Illustration of the notation used in Section 3.1. Every node in the shown graph corresponds to a 2-change. The arcs going into a node u represent the edges removed from the tour in step u and the arcs going out of a node u represent the edges added to the tour in step u.

PROOF. There are at most m^2 different choices for the first step S_1 because there are at most $\binom{m}{2} \leq m^2/2$ choices for the two edges that are removed from the tour in step S_1 and, once these are fixed, at most two choices for the edges added to the tour in step S_1 (remember that the edges must form a 4-cycle alternating between edges added and removed from the tour).

Once S_i is fixed, there are at most $4m$ choices for S_{i+1} because there are two choices for the edge that links S_i and S_{i+1}, at most m choices for the other edge removed from the tour in step S_{i+1}, and, once these are fixed, at most two choices for the edges added to the tour in step S_{i+1}. $\square$

We call a sequence of steps *ε-bad* if every step in the sequence is improving but yields an improvement of at most ε. The probability that a fixed 2-change is an improvement by at most ε is bounded from above by $\varepsilon\phi$. Ideally we would like to show an upper bound of $(\varepsilon\phi)^k$ on the probability that each step in a given linked sequence $S_1, \ldots, S_k$ is an improvement by at most ε. However, for general linked sequences this is not true because the steps can be dependent in various ways (some steps might even repeat). We need to introduce further restrictions on linked sequences to obtain a good upper bound on the probability that every step is a small improvement.

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

Definition 3.2 (*witness sequences of type 1*). If for every $i \in [k]$, 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-long witness sequence of type 1*.

A k-long witness sequence of type 1 possesses enough randomness to obtain an upper bound of $(\varepsilon\phi)^k$ for the probability that it is ε-bad because every step introduces an edge that has not occurred in the steps before (see Lemma 3.5).

Definition 3.3 (*witness sequences of type 2*). If for every $i \in [k]$, the edge e_i does not occur in any step S_j with $j < i$ and if each endpoint of f_{k-1} occurs in some step S_j with $j < k$ (not necessarily the same for both endpoints), then $S_1, \ldots, S_k$ is called a *k-long witness sequence of type 2*.

Observe that every k-long witness sequence of type 2 is also a k-long witness sequences of type 1. Hence, also for every witness sequences of type 2, we obtain the desired bound of $(\varepsilon\phi)^k$ for the probability that it is ε-bad. Due to the additional restriction on f_{k-1}, the number of k-long witness sequences of type 2 is at most $k^2 4^k m^k$ (see Lemma 3.5). Even though it seems like a minor detail, it is very important that the exponent of m in this bound is only k and not $k+1$ as for k-long witness sequences of type 1. The reason why this is important is that, as we will see later, the quotient of the exponents of m and ε in the upper bound for the probability that there exists an ε-bad witness sequence determines the exponent of m in the bound for the expected length of the longest path in the state graph. For witness sequences of type 1 this quotient is $(k+1)/k = 1 + 1/k$ while it is only $k/k = 1$ for witness sequences of type 2. Since we aim for the exponent $1 + o(1)$, witness sequences of type 1 are only helpful in our analysis for $k = \omega(1)$ while witness sequences of type 2 of any length yield a good enough bound on the expected length of the longest path in the state graph.

Definition 3.4 (*witness sequences of type 3*). If for every $i \in [k-1]$, the edge e_i does not occur in any step S_j with $j < i$, if e_k and g_k both occur in steps S_j with $j < k$ (not necessarily the same), 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-long witness sequence of type 3*.

Also every witness sequences of type 3 possesses enough randomness to bound the probability that it is ε-bad by $(\varepsilon\phi)^k$ because every step introduces a new edge. The number of witness sequences of type 3 is bounded from above by $k^2 4^{k+1} m^k$ (see Lemma 3.5). Hence, the same reasoning as for witness sequences of type 2 applies and witness sequences of type 3 of any length yield a good enough bound on the expected length of the longest path in the state graph.

3.2. Probability of the Existence of a Bad Witness Sequence

In this section, we analyze the probability that there exists an ε-bad k-long witness sequence.

LEMMA 3.5. *The probability that there exists*

a) an ε-bad k-long witness sequence of type 1 is bounded from above by $4^{k-1} m^{k+1} (\varepsilon\phi)^k$,
b) an ε-bad k-long witness sequence of type 2 is bounded from above by $k^2 4^k m^k (\varepsilon\phi)^k$,
c) an ε-bad k-long witness sequence of type 3 is bounded from above by $k^2 4^{k+1} m^k (\varepsilon\phi)^k$.

PROOF. a) We consider k-long witness sequences of type 1 first. In accordance with Lemma 3.1 the number of such sequences is at most $4^{k-1} m^{k+1}$. Now fix an arbitrary k-long witness sequence $S_1, \ldots, S_k$ of type 1. We use the same notation as in Figure 1 to denote the edges involved in this sequence. In the first step, the edges e_0 and f_0 are replaced by the edges e_1 and g_1. As in the proof of Theorem 2.1, we use the principle of deferred decisions and assume that the lengths of the edges e_0, f_0, and g_1 are determined by an adversary. The improvement of step S_1 can be expressed as a simple linear combination of the lengths of the involved edges. Hence, for fixed lengths of e_0, f_0, and g_1, the event that S_1 is an improvement by at most ε corresponds to the event that the length $d(e_1)$ of e_1 lies in some fixed interval of length ε. Since the density of $d(e_1)$ is bounded by ϕ, the probability that $d(e_1)$ takes a value in this interval is bounded by $\varepsilon\phi$.

Now we consider a step S_i with $i \geq 2$ and apply again the principle of deferred decisions. We assume that arbitrary lengths for the edges e_j and f_j with $j < i$ and for g_j with $j \leq i$ are chosen. Since the edge e_i is not involved in any step S_j with $j < i$, its length is not determined. Hence, analogously to the first step, the probability that step S_i is an improvement by at most ε is bounded from above by $\varepsilon\phi$ for every realization of the steps S_j with $j < i$. Applying this argument to every step S_i yields the desired bound of $(\varepsilon\phi)^k$. A union bound over all witness sequences of type 1 concludes the proof of a).

b) Since $S_1, \ldots, S_{k-1}$ is a $(k-1)$-long witness sequence, there are at most $4^{k-2} m^k$ choices for these steps. The number of different vertices involved in steps S_i with $i < k$ is at most $4 + 2(k-2) = 2k$ because the first step introduces four new vertices and every other step at most two. Since the endpoints of the edge f_{k-1} must be chosen among those vertices that have been involved in the steps S_i with $i < k$, there are at most $\binom{2k}{2} < 2k^2$ choices for f_{k-1}. Furthermore, for fixed S_{k-1} there are two choices for the edge e_{k-1} that links S_{k-1} and S_k. If the edges e_{k-1} and f_{k-1} are determined, there are two choices for e_k and g_k. Hence, in total there are at most $8k^2$ possible choices for step S_k. This implies that the number of different k-long witness sequences of type 2 is bounded by $8k^2 4^{k-2} m^k < k^2 4^k m^k$.

Applying the same arguments as for witness sequences of type 1, yields for every witness sequence of type 2 that it is ε-bad only with a probability of at most $(\varepsilon\phi)^k$. A union bound over all witness sequences of type 2 concludes the proof of b).

c) Since $S_1, \ldots, S_{k-1}$ is a $(k-1)$-long witness sequence, there are at most $4^{k-2}m^k$ choices for these steps. The number of different edges involved in steps S_i with $i < k$ is at most $4 + 3(k-2) < 3k$ because the first step introduces four new edges and every other step at most three. Hence, when the steps $S_1, \ldots, S_{k-1}$ are fixed, there are at most two choices for the edge e_{k-1} that links S_{k-1} and S_k and there are at most $\binom{3k}{2} \le 9k^2/2$ choices for the set $\{e_k, g_k\}$. Once e_{k-1} and $\{e_k, g_k\}$ are fixed, there are two choices for f_{k-1}. The total number of k-long witness sequences of type 3 can thus be bounded from above by $18k^2 4^{k-2} m^k < k^2 4^{k+1} m^k$.

Similar to witness sequences of type 1, we can bound the probability that a fixed k-long witness sequence of type 3 is ε-bad from above by $(\varepsilon\phi)^k$ because also the last step introduces an edge that does not occur in the steps before, namely f_{k-1}. □

Definition 3.6. In the following, we use the term *k-witness sequence* to denote a k-long witness sequence of type 1 or an i-long witness sequence of type 2 or 3 with $i \le k$.

Observe that in general a k-witness sequence can contain non-improving 2-changes, which increase the length of the tour. As 2-Opt does not make such 2-changes, we are only interested in k-witness sequences in which every 2-change is improving.

Definition 3.7. We call a k-witness sequence *improving* if every 2-change in the sequence is an improvement. Moreover, by $\Delta_{\text{ws}}^{(k)}$ we denote the smallest total improvement made by any improving k-witness sequence.

The reason why the previous definition treats witness sequences of type 1 differently than those of type 2 or 3 is that, as discussed above, witness sequences of type 1 are only helpful in our analysis if they are long enough while witness sequences of type 2 or 3 of any length are helpful. Lemma 3.5 shows that it is unlikely that there exists an improving k-witness sequence whose total improvement is small.

COROLLARY 3.8. *For any natural number $k \ge 3$ and $0 < \varepsilon \le \left(64 m^{(k-1)/(k-2)}\phi\right)^{-1}$,*

$$\mathbf{Pr}\left[\Delta_{\text{ws}}^{(k)} \le \varepsilon\right] \le 800 \cdot (m\varepsilon\phi)^2.$$

PROOF. Due to Lemma 3.5 and the fact that witness sequences of type 2 or 3 must consist of at least two steps, a union bound over all k-witness sequences yields

$$\begin{aligned}\mathbf{Pr}\left[\Delta_{\mathrm{ws}}^{(k)} \le \varepsilon\right] &\le 4^{k-1}m^{k+1}(\varepsilon\phi)^k + \sum_{i=2}^{k} i^2 4^i m^i (\varepsilon\phi)^i + \sum_{i=2}^{k} i^2 4^{i+1} m^i (\varepsilon\phi)^i \\ &\le 4^{k-1}m^{k+1}(\varepsilon\phi)^k + 5\sum_{i=2}^{\infty} i^2 (4m\varepsilon\phi)^i \\ &= 4^{k-1}m^{k+1}(\varepsilon\phi)^k + 5\cdot\frac{(4m\varepsilon\phi)^2(4-12m\varepsilon\phi+(4m\varepsilon\phi)^2)}{(1-4m\varepsilon\phi)^3} \\ &\le 4^{k-1}m^{k+1}(\varepsilon\phi)^k + 5\cdot\left(\frac{16}{15}\right)^3\cdot(4m\varepsilon\phi)^2\cdot\left(4+\frac{1}{16^2}\right) \\ &\le 4^{k-1}m^{k+1}(\varepsilon\phi)^k + 25\cdot(4m\varepsilon\phi)^2.\end{aligned}$$

Here we used, in the third step, that $\sum_{i=1}^{\infty} i^2 a^i = \frac{a(a+1)}{(1-a)^3}$ for any $a \in [0,1)$. In the fourth and fifth step, we used that the upper bound on ε in the corollary implies $4m\varepsilon\phi \le 1/16$.

The above inequality implies the corollary because for $\varepsilon \le \left(64m^{(k-1)/(k-2)}\phi\right)^{-1}$, the second term in the sum is at least as large as the first one. □

3.3. Finding Witness Sequences

In the previous section, we have shown an upper bound on the probability that there exists an ε-bad k-witness sequence. In this section, we show that in every long enough sequence of consecutive 2-changes, one can identify a certain number of disjoint k-witness sequences. In this way, we obtain a lower bound on the improvement made by any long enough sequence of consecutive 2-changes in terms of $\Delta_{\mathrm{ws}}^{(k)}$.

LEMMA 3.9. *Let $n \ge 8$, $k \in \mathbb{N}$, and let $S_1, \dots, S_t$ denote a sequence of consecutive 2-changes performed by the 2-Opt heuristic with $t \ge n4^{k-1}$. The sequence $S_1, \dots, S_t$ shortens the tour by at least $t/4^{k+1} \cdot \Delta_{\mathrm{ws}}^{(k)}$.*

Basically, we have to show that one can find $t/4^{k+1}$ disjoint k-witness sequences in the given sequence $S_1, \dots, S_t$ of consecutive 2-changes. To do this, we first introduce a so-called *witness DAG* (directed acyclic graph) which represents the sequence $S_1, \dots, S_t$ of 2-changes. In order to not confuse the constructed witness DAG W with the input graph G, we use the terms *nodes* and *arcs* when referring to the DAG W and the terms *vertices* and *edges* when referring to G. For every step S_i in the given sequence there is one node in W. Every node has at most two incoming and either zero or two outgoing arcs and every arc is labeled with an edge of the graph G. Consider a node that corresponds to a step S_i in which the edges e and e' are exchanged with the edges f and f'. If there exists a step S_j with $j > i$ in which the edge f is removed from the tour then let $j_1 > i$ denote the smallest such index, i.e., the edge f is removed from the tour in step S_{j_1} and does not occur in the steps $S_{i+1}, \dots, S_{j_1-1}$. Similarly if there exists a step S_j with $j > i$ in which the edge f' is removed from tour then let $j_2 > i$ denote the smallest such index. Only if both j_1 and j_2 are defined, the node that corresponds to S_i has outgoing arcs. It has one outgoing arc to the node that corresponds to S_{j_1} and that is labeled with f and it has a second outgoing arc to the node that corresponds to S_{j_2} and that is labeled with f'.

We call nodes of W without outgoing arcs *leaves*. By the *height* of a node u, we denote the length of a shortest path from u to a leaf of W. We associate with each node u of height at least $k-1$ a sub-DAG W_u of W. The sub-DAG W_u associated with such a node u is the induced sub-DAG of those nodes of W that can be reached from u by traversing at most $k-1$ arcs. The following two lemmas directly imply Lemma 3.9.

LEMMA 3.10. *Let u be a node of height at least $k-1$ in W. The 2-changes represented by the nodes in the sub-DAG W_u yield a total improvement of at least $\Delta_{\text{ws}}^{(k)}$.*

LEMMA 3.11. *Let $n \geq 8$. Every witness DAG that represents a sequence of $t \geq n4^{k-1}$ 2-changes contains at least $t/4^{k+1}$ nodes of height at least $k-1$ whose corresponding sub-DAGs are pairwise disjoint.*

PROOF PROOF OF LEMMA 3.10. Assume that a sub-DAG W_u with root u of height at least $k-1$ in W is given. Any path from u to some other node in W_u corresponds to a sequence of 2-changes. Let P be such a path. From the definition of W it follows that every node on P corresponds to a step S_i where the indices are strictly increasing along P (in particular, every node on P corresponds to a step with a different index). In the following, we show that at least one path in W_u corresponds to a k-witness sequence or a sequence whose total improvement is at least as large as the total improvement of one of the k-witness sequences.

In order to identify such a path, we unroll the sub-DAG W_u to a complete binary tree T of height $k-1$. The root of T is the node u and every node in T whose distance to the root is smaller than $k-1$ has two children, namely (copies of) its two direct successors in W_u. In general, the binary tree T contains multiple nodes that represent the same step S_i. However, if P is a downward path in T from the root u to some other node, then it is still the case that each node on P corresponds to a step S_i where the indices are strictly increasing along P.

Let v be an inner node of T, let a be one of its outgoing arcs, let e be the label of a, and let P be the downward path from the root u to the node v in T, not including v itself. We say that the arc a is *non-continuable* if the edge e occurs in one of the steps that are represented by the nodes of P and *continuable* otherwise. (Observe that this does not necessarily mean that one of the arcs on the path from u to v has label e.) The intuition underlying this definition is as follows: We would like to find a downward path in W_u starting at the root u whose nodes correspond to a witness sequence. Only paths in which all arcs are continuable can correspond to witness sequences of type 1 or 2. For witness sequences of type 3 all arcs except for the last one must be continuable.

Now let v be a leaf of T. Then v does not have any outgoing arcs. Nevertheless, as every node of T, it corresponds to a step in which two edges are added to the tour. We call the leaf v *non-continuable* if both these edges occur in steps that are represented by the nodes of the downward path from the root u to the node v in T, not including v itself, and *continuable* otherwise. The intuition underlying this definition is as follows: Any downward path in W_u starting at the root u to a leaf v can only correspond to a witness sequence of type 1 if all its arcs are continuable and if the leaf v is also continuable.

If T contains a downward path that corresponds to a k-witness sequence of type 1 then we are done. Assume that T does not contain such a path. Then the following property must be true for any path P from the root u to a leaf v of T: at least one of the arcs of P is non-continuable or v is non-continuable. This is the case because any path from u to a continuable leaf v that contains only continuable arcs corresponds to a k-witness sequence of type 1 (the continuable arcs correspond to the edges $e_1, e_2, \ldots, e_k$ in Definition 3.2). Now we remove all nodes from T below non-continuable arcs to get a subtree T' of T. To be more precise, a node v of T is contained in T' if and only if the downward path from u to v in T does not contain a non-continuable arc. We will show that we can find a witness sequence of type 2 or 3 in T'.

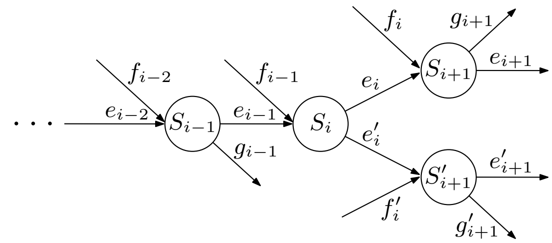

Fig. 2. Summary of our notation. We assume that the nodes corresponding to S_{i+1} and S'_{i+1} are leaf nodes of T' and non-continuable.

Let v_{i+1} be one node with maximum distance from the root in T' and let v_i be its parent. Let the 2-changes represented by the nodes on the downward path P from the root u to v_{i+1} be $S_1, \ldots, S_{i+1}$. If v_i has two children in T' then let v'_{i+1} denote the child different from v_{i+1} and let S'_{i+1} denote the step that is represented by v'_{i+1}. In Figure 2, we summarize the notation that we use in the following. In step S_j for $j \leq i-1$ and $j = i+1$, the edges e_{j-1} and f_{j-1} are exchanged with the edges e_j and g_j. In step S_i, the edges e_{i-1} and f_{i-1} are exchanged with the edges e_i and e'_i, and in step S'_{i+1}, the edges e'_i and f'_i are exchanged with the edges e'_{i+1} and g'_{i+1}. We denote by E_i all edges that are involved in steps S_j with $j \leq i$. Similarly, by E_{i-1} we denote all edges that are involved in steps S_j with $j \leq i-1$.

Observe that all leaves in T' must be non-continuable. For leaves of height smaller than $k-1$ this follows from the definition of T'. If any such leaf v had a continuable arc in T then this arc and the corresponding child of v would also be contained in T'. Leaves of height $k-1$ in T' cannot be continuable because otherwise the path to such a leaf would represent a k-witness sequence of type 1, as discussed above. Our construction ensures that $S_1, \ldots, S_i$ is an i-witness sequence of type 1 because the path from the root u to the leaf v_{i+1} consists of continuable arcs only. The sequence $S_1, \ldots, S_{i+1}$, however, is not a witness sequence of type 1 because all leaves of T' are non-continuable, which implies $e_{i+1}, g_{i+1} \in E_i$.

In the following we will shrink the tree T' until a witness sequence of type 2 or 3 is found. For this, we define the operation **contract**(S_i, S_{i+1}). This operation will only be applied if the node that corresponds to S_i has only a single child in T' (namely the one that corresponds to S_{i+1}) and if the net effect of S_i and S_{i+1} together corresponds to a single 2-change S. In this case the operation **contract**(S_i, S_{i+1}) replaces the nodes v_i and v_{i+1} that represent the steps S_i and S_{i+1} by a node the represents the 2-change S. We call the tree that results from this operation again T'. The following invariant will remain true throughout the construction: The only nodes that were produced by a contract operation are leaves in the current tree T'. Furthermore each leaf that was created by a contract operation has the same net effect as the contracted steps and it is non-continuable. For every leaf that was produced by contract operations, the steps contracted form a descending path in the original tree T' in which every node has at most one child.

The following case analysis shows that it is always possible to either identify a witness sequence of type 2 or 3 or to apply the operation **contract**(S_i, S_{i+1}). We use the notation **return**$_j(R_1, \ldots, R_\ell)$ to denote that $R_1, \ldots, R_\ell$ is an ℓ-long witness sequence of type j.

Since v_{i+1} is non-continuable, we can assume $e_{i+1}, g_{i+1} \in E_i$.
If $e'_i \notin E_{i-1}$, then v'_{i+1} exists in T'. Since v_{i+1} is a node of maximum distance to the

root, v'_{i+1} must also be a leaf in T' and hence it is also non-continuable due to the invariant. This is equivalent to $e'_{i+1}, g'_{i+1} \in E_i$.

(1) If $f_{i-1} \in E_{i-1}$, then $\mathbf{return}_2(S_1, \ldots, S_i)$.
From now on we assume $f_{i-1} \notin E_{i-1}$.

(2) If $e'_i \in E_{i-1}$, then consider the following cases.
   (a) If $f_i \notin E_i$, then $\mathbf{return}_3(S_1, \ldots, S_{i+1})$.
   (b) If $e_{i+1}, g_{i+1} \in E_{i-1}$, then $\mathbf{return}_2(S_1, \ldots, S_i)$.
   $S_1, \ldots, S_i$ is a witness sequence of type 2 because 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}.
   (c) If $f_i \in E_i$ and ($e_{i+1} \in E_i \setminus E_{i-1}$ or $g_{i+1} \in E_i \setminus E_{i-1}$), then $\mathbf{contract}(S_i, S_{i+1})$.
   *In this case one can assume w.l.o.g. that $g_{i+1} = f_{i-1}$ and $e_{i+1} \in E_{i-1}$ since $E_i \setminus E_{i-1} = \{e_i, f_{i-1}\}$ and the edges e_i, e_{i+1}, and g_{i+1} are pairwise distinct because they occur in the same 2-change S_{i+1}. The **contract**-operation replaces v_i and v_{i+1} by a node that represents the 2-change $S := (e_{i-1}, f_i) \to (e'_i, e_{i+1})$.*

(3) If $e'_i \notin E_{i-1}$, then $e_{i+1}, g_{i+1}, e'_{i+1}, g'_{i+1} \in E_i$. Consider the following cases.
   (a) If $f_i \notin E_i$ or $f'_i \notin E_i$, then $\mathbf{return}_3(S_1, \ldots, S_{i+1})$ or $\mathbf{return}_3(S_1, \ldots, S_i, S'_{i+1})$, respectively.
   From now on we assume $f_i, f'_i \in E_i$.
   (b) If $e_{i+1}, g_{i+1}, e'_{i+1}, g'_{i+1} \in E_{i-1}$, then $\mathbf{return}_2(S_1, \ldots, S_i)$.
   $S_1, \ldots, S_i$ is a witness sequence of type 2 due to the invariant and the fact that the endpoints of f_{i-1} coincide with some endpoints of $e_{i+1}, g_{i+1}, e'_{i+1}, g'_{i+1}$.
   (c) If $|\{e_{i+1}, e'_{i+1}, g_{i+1}, g'_{i+1}\} \cap (E_i \setminus E_{i-1})| \geq 1$, then assume w.l.o.g. $g_{i+1} \in E_i \setminus E_{i-1}$ and $\mathbf{return}_2(S_1, \ldots, S_{i-1}, S)$ for the 2-change S defined below.
   In this case $E_i \setminus E_{i-1} = \{e_i, e'_i, f_{i-1}\}$. Furthermore, $e_{i+1} \neq e'_i$ and $g_{i+1} \neq e'_i$ because e_{i+1} and g_{i+1} both share one endpoint with e_i whereas e'_i and e_i do not share any endpoints. Furthermore, the edges e_i, e_{i+1}, and g_{i+1} are pairwise distinct because they occur in the same 2-change S_{i+1}. As in case 2 (c), we assume w.l.o.g. that $g_{i+1} = f_{i-1}$ and $e_{i+1} \in E_{i-1}$.
   It must be $f_i \neq e'_i$ as otherwise step S_i would be reversed in step S_{i+1}. Furthermore, the edges f_i, e_i, and, $f_{i-1} = g_{i+1}$ are pairwise distinct because they occur in the same 2-change S_{i+1}. Hence, $f_i \in E_{i-1}$. For the step $S := (e_{i-1}, f_i) \to (e'_i, e_{i+1})$, the sequence $S_1, \ldots, S_{i-1}, S$ is a witness sequence of type 2 because $f_i \in E_{i-1}$ and $e'_i \notin E_{i-1}$. Observe that the original sequence $S_1, \ldots, S_{i+1}$ yields the same net effect and hence the same improvement as the sequence $S_1, \ldots, S_{i-1}, S$.

If the operation $\mathbf{contract}(S_i, S_{i+1})$ is performed in Case 2 (c) then the invariant stays true. If the node v_{i+1} was not created by a previous contraction this follows easily because Case 2 is only reached if v_i has only one child and $\mathbf{contract}(S_i, S_{i+1})$ replaces v_i and v_{i+1} by a node that represents the 2-change $(e_{i-1}, f_i) \to (e_{i+1}, e'_i)$, which is the net effect of S_i and S_{i+1} together. Furthermore $e_{i+1}, e'_i \in E_{i-1}$ and hence the new node is non-continuable. With the same arguments it also follows that the invariant stays true if the node v_{i+1} was created by previous contractions.

We repeatedly apply the case analysis above to a node in T' with maximum distance to the root until a witness sequence is found. Observe that the operation $\mathbf{contract}(S_i, S_{i+1})$ is only performed in Case 2 (c) and that each time it is performed the number of nodes in T' decreases by one. Furthermore it is not possible that T' shrinks to a single node because this node would be a leaf that must be non-continuable due to the invariant. However, the root of T' is always continuable because there are no previous steps in which the edges added to the tour can occur. Hence after finitely many occurrences of the operation $\mathbf{contract}(S_i, S_{i+1})$ one of the other cases must be reached. In all other cases, immediately a witness sequence of type 2 or 3 is returned.

The witness sequence returned is in general not a sequence of steps that are contained in the DAG W_u because the last step S (and only the last step) in the returned sequence is potentially the result of some contract operations. This is, in particular, true for Case 3 (c) in which the steps S_i and S_{i+1} are contracted without explicitly calling the operation **contract**. Due to the invariant, we know that the steps that are contracted into the last step S have the same net effect as S. Furthermore these steps have pairwise distinct indices because they lie on a downward path in T. So the improvement of every step is counted at most once. Hence, the improvement of the witness sequence returned always equals the total improvement of some steps that are contained in W_u. This concludes the proof. □

PROOF PROOF OF LEMMA 3.11. Let W be a witness DAG that consists of t nodes that represent the steps $S_1, \dots, S_t$. By definition a node in W has either two direct successors or none at all. The case that a node has no successors can only occur if at least one of the edges that is added to the tour in the corresponding step is not removed anymore in later steps. Since the final tour that is obtained after performing the steps $S_1, \dots, S_t$ contains exactly n edges, at most n of the nodes of W can be leaves. Hence W contains at least $t - n$ nodes with two outgoing arcs.

We defined the height of a node v in W to be the minimum distance from v to one of the leaves of W. Since every node in W has an indegree of at most two, there are at most $n2^{k-1}$ nodes in W whose height is smaller than $k - 1$. Hence, there are at least $t - n2^{k-1}$ nodes in W with an associated sub-DAG of depth $k - 1$. We construct a set of disjoint sub-DAGs in a greedy fashion: We take an arbitrary sub-DAG W_u and add it to the set of disjoint sub-DAGs that we construct. After that, we remove all nodes of W_u from the DAG W. We repeat these steps until no complete sub-DAG W_u is left in W.

In order to see that the constructed set consists of at least $t/4^{k+2}$ disjoint sub-DAGs, observe that each sub-DAG of depth $k - 1$ contains at most $2^k - 1$ nodes because the outdegree of every node is at most two. Each node can be contained in at most $2^k - 1$ sub-DAGs of depth $k - 1$ because the indegree of every node is at most two. Hence, every sub-DAG W_u can only intersect with at most $(2^k - 1)^2 \leq 4^k$ other sub-DAGs. Thus, the number of pairwise disjoint sub-DAGs must be at least

$$\left\lfloor \frac{t - n2^{k-1}}{4^k} \right\rfloor \geq \left\lfloor \frac{t/2}{4^k} \right\rfloor \geq \frac{t}{4^{k+1}},$$

where both inequalities follow from the assumption $t \geq n4^{k-1}$. For the second inequality we additionally used the assumption $n \geq 8$. □

3.4. The Expected Number of 2-Changes

Now we can prove Theorem 1.1.

PROOF PROOF OF THEOREM 1.1. We combine Corollary 3.8 and Lemma 3.9 to obtain an upper bound on the probability that the length T of the longest path in the state graph exceeds t. Let $n \geq 8$. For $t \geq n4^{k-1}$, the tour is shortened by the sequence of 2-changes by at least $t/4^{k+1} \cdot \Delta_{\text{ws}}^{(k)}$. Hence, for $t \geq n4^{k-1}$,

$$\mathbf{Pr}\left[T \geq t\right] \leq \mathbf{Pr}\left[\frac{t}{4^{k+1}} \cdot \Delta_{\text{ws}}^{(k)} \leq n\right] = \mathbf{Pr}\left[\Delta_{\text{ws}}^{(k)} \leq \frac{n4^{k+1}}{t}\right].$$

Combining this inequality with Corollary 3.8 yields for $t \geq t' := \lceil 4^{k+4} n m^{(k-1)/(k-2)} \phi \rceil$,

$$\mathbf{Pr}\left[T \geq t\right] \leq 800 \left(\frac{4^{k+1} n m \phi}{t}\right)^2.$$

Note that the restriction $t \geq t'$ is necessary to apply Corollary 3.8. We can bound the expected number of 2-changes as follows:

$$\begin{aligned}\mathbf{E}\left[T\right] &= \sum_{t=1}^{\infty}\mathbf{Pr}\left[T \geq t\right] \leq t' + \sum_{t=t'+1}^{\infty} 800\left(\frac{4^{k+1}nm\phi}{t}\right)^2 \\ &\leq t' + \int_{t'}^{\infty} 800\left(\frac{4^{k+1}nm\phi}{t}\right)^2 dt \\ &\leq t' + \frac{800(4^{k+1}nm\phi)^2}{t'} \\ &= O\left(4^k nm^{(k-1)/(k-2)}\phi\right).\end{aligned}$$

Setting $k = \sqrt{\ln m}$ yields

$$\begin{aligned}\mathbf{E}\left[T\right] &= O\left(4^{\sqrt{\ln m}} m^{\frac{1}{\sqrt{\ln m}-2}} nm\phi\right) \\ &= O\left(4^{2\sqrt{\ln m}} nm\phi\right),\end{aligned}$$

where the last equation holds for sufficiently large m. □

4. UPPER BOUND FOR THE SECOND MOMENT

Our method does not yield strong concentration bounds for the expected length of the longest path in the state graph. The reason is that the exponent of ε in Corollary 3.8 is only 2. It is, however, possible to bound the second moment of T.

THEOREM 4.1. *For every ϕ-perturbed graph with n vertices and m edges*

$$\mathbf{E}\left[T^2\right] = O\left(\left(16^{\sqrt{\ln m}} m\phi\right)^2 \cdot n^3\right).$$

PROOF. The proof follows along the same lines as the proof of Theorem 1.1. Let $n \geq 8$. For $t \geq n4^{k-1}$, the tour is shortened by the sequence of 2-changes by at least $t/4^{k+1} \cdot \Delta_{\text{ws}}^{(k)}$. Hence, for $t \geq n4^{k-1}$,

$$\mathbf{Pr}\left[T^2 \geq t\right] = \mathbf{Pr}\left[T \geq \sqrt{t}\right] \leq \mathbf{Pr}\left[\frac{\sqrt{t}}{4^{k+1}} \cdot \Delta_{\text{ws}}^{(k)} \leq n\right] = \mathbf{Pr}\left[\Delta_{\text{ws}}^{(k)} \leq \frac{n4^{k+1}}{\sqrt{t}}\right].$$

Combining this inequality with Corollary 3.8 yields for $t \geq t' := (\lceil 4^{k+4} nm^{(k-1)/(k-2)}\phi\rceil)^2$,

$$\mathbf{Pr}\left[T \geq \sqrt{t}\right] \leq 800\left(\frac{4^{k+1}nm\phi}{\sqrt{t}}\right)^2.$$

Note that the restriction $t \geq t'$ is necessary to apply Corollary 3.8. Using that T^2 cannot be larger than $(n!)^2$, we can bound the expected value of T^2 as follows:

$$\begin{aligned}\mathbf{E}\left[T^2\right] &= \sum_{t=1}^{(n!)^2} \mathbf{Pr}\left[T^2 \geq t\right] \leq t' + \sum_{t=t'+1}^{(n!)^2} 800\left(\frac{4^{k+1}nm\phi}{\sqrt{t}}\right)^2 \\ &\leq t' + \int_{t'}^{(n!)^2} 800\left(\frac{4^{k+1}nm\phi}{\sqrt{t}}\right)^2 dt \\ &\leq t' + 800(4^{k+1}nm\phi)^2 \int_1^{(n!)^2} \frac{1}{t}\, dt \\ &\leq t' + 800(4^{k+1}nm\phi)^2 \cdot \ln((n!)^2) \\ &= O\left(\left(4^k nm^{(k-1)/(k-2)}\phi\right)^2 n\ln(n)\right).\end{aligned}$$

Setting $k = \sqrt{\ln m}$ yields

$$\begin{aligned}\mathbf{E}\left[T^2\right] &= O\left(\left(4^{\sqrt{\ln m}} m^{\frac{1}{\sqrt{\ln m}-2}} nm\phi\right)^2 n\ln(n)\right) \\ &= O\left(\left(16^{\sqrt{\ln m}} nm\phi\right)^2 n\right),\end{aligned}$$

where the last equation holds for sufficiently large m. □

Let $B = c \cdot 16^{\sqrt{\ln m}} nm\phi$, where c is the constant from the Big O notation in Theorem 1.1. Then Markov's inequality implies for every $a \geq 1$ that $\mathbf{Pr}[T \geq aB] \leq 1/a$. Theorem 4.1 implies the following concentration bound, which is stronger for large a.

COROLLARY 4.2. *If m is sufficiently large, there exists a constant κ such that $\mathbf{Pr}[T \geq aB] \leq n\kappa/a^2$ for any $a \geq 1$.*

PROOF. Let κ be chosen such that $\mathbf{E}\left[T^2\right] \leq \kappa B^2 n$. Such a constant κ must exist due to Theorem 4.1. Then

$$\mathbf{Pr}[T \geq aB] = \mathbf{Pr}\left[T^2 \geq a^2B^2\right] = \mathbf{Pr}\left[T^2 \geq \frac{a^2}{n\kappa}\kappa B^2 n\right] \leq \frac{n\kappa}{a^2},$$

which proves the corollary. □

REFERENCES

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

Matthias Englert, Heiko Röglin, and Berthold Vöcking. 2014. Worst Case and Probabilistic Analysis of the 2-Opt Algorithm for the TSP. *Algorithmica* 68, 1 (2014), 190–264. DOI:http://dx.doi.org/10.1007/s00453-013-9801-4

David S. Johnson and Lyle A. McGeoch. 1997. The Traveling Salesman Problem: A Case Study in Local Optimization. In *Local Search in Combinatorial Optimization*, E. H. L. Aarts and J. K. Lenstra (Eds.). John Wiley and Sons, New York, NY, USA.

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

George S. Lueker. 1975. Unpublished Manuscript. (1975). Princeton University.

Bodo Manthey and Rianne Veenstra. 2013. Smoothed Analysis of the 2-Opt Heuristic for the TSP. In *Proc. of the 24th International Symposium on Algorithms and Computation (ISAAC)*. Springer, Berlin, 579–589.

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

Received ; revised ; accepted