Towards Understanding the Smoothed Approximation Ratio of the 2-Opt Heuristic

Marvin Künnemann[1] and Bodo Manthey[2]

[1] Max Planck Institute for Informatics, Saarbrücken, Germany and
Saarbrücken Graduate School of Computer Science, `marvin@mpi-inf.mpg.de`
[2] University of Twente, Enschede, The Netherlands, `b.manthey@utwente.nl`

Abstract. The 2-Opt heuristic is a very simple, easy-to-implement local search heuristic for the traveling salesman problem. While it usually provides good approximations to the optimal tour in experiments, its worst-case performance is poor.

In an attempt to explain the approximation performance of 2-Opt, we analyze the smoothed approximation ratio of 2-Opt. We obtain a bound of $O(\log(1/\sigma))$ for the smoothed approximation ratio of 2-Opt. As a lower bound, we prove that the worst-case lower bound of $\Omega(\frac{\log n}{\log \log n})$ for the approximation ratio holds for $\sigma = O(1/\sqrt{n})$.

Our main technical novelty is that, different from existing smoothed analyses, we do not separately analyze objective values of the global and the local optimum on all inputs, but simultaneously bound them on the same input.

1 2-Opt and Smoothed Analysis

The traveling salesman problem (TSP) is one of the best-studied combinatorial optimization problems. Euclidean TSP is the following variant: given points $X \subseteq [0,1]^d$, find the shortest Hamiltonian cycle that visits all points in X (also called a *tour*). Even this restricted variant is NP-hard for $d \geq 2$ [17].

While Euclidean TSP admits a polynomial-time approximation scheme [1, 16], heuristics that are simpler and easier to implement are often used in practice. A very simple and popular heuristic for finding near-optimal tours quickly is the 2-Opt heuristic: starting from an initial tour, we iteratively replace two edges by two other edges to obtain a shorter tour until we have found a local optimum. Experiments indicate that 2-Opt converges to near-optimal solutions quickly and produces solutions that are within a few percent of the optimal solution [10,11]. In contrast to its success on practical instances, 2-Opt performs poorly in the worst case: the worst-case running-time is exponential even for $d = 2$ [8] and its worst-case approximation ratio of $O(\log n)$ has an almost matching lower bound of $\Omega(\log n/\log \log n)$ for Euclidean instances [6].

In order to explain the performance of algorithms whose worst-case performance guarantee does not reflect the observed performance, smoothed analysis has been introduced [19], which is a hybrid of worst-case analysis (which is often too pessimistic) and average-case analysis (which is often dominated by

completely random instances that have special properties not shared by typical instances). In smoothed analysis, an adversary specifies an instance, and then this instance is slightly randomly perturbed. The smoothed performance is the expected performance, where the expected value is taken over the random perturbation. The motivating assumption of smoothed analysis is that practical instances are often subjected to a small amount of random noise that can, e.g., come from measurement errors or numerical imprecision. Smoothed analysis often allows more realistic conclusions about the performance of an algorithm than mere worst-case or average-case analysis.

Smoothed analysis has been applied successfully to explain the running time of the 2-Opt heuristic [8, 15] as well as other local search algorithms [2, 3, 14]. We refer to two surveys for an overview of smoothed analysis [13, 20].

Much less is known about the smoothed approximation performance of algorithms. Karger and Onak have shown that multi-dimensional bin packing can be approximated arbitrarily well for smoothed instances [12] and there are frameworks to approximate Euclidean optimization problems such as TSP for smoothed instances [4, 7]. However, these approaches mostly consider algorithms tailored to solving smoothed instances.

With respect to concrete algorithms, we are only aware of analyses of the jump and lex-jump heuristics for scheduling [5, 9] and an upper bound of $O(\phi^{1/d})$ for the smoothed approximation ratio of 2-Opt in the so-called one-step model [8]. Here, ϕ is an upper bound on the density functions according to which the points are drawn. Translated to Gaussian perturbation, we would obtain an upper bound of $O(1/\sigma)$ if we truncate the Gaussian distribution such that all points lie in a hypercube of constant sidelength.

In order to explain the practical approximation performance of 2-Opt, we provide an improved smoothed analysis of its approximation ratio. More precisely, we provide bounds on the quality of the worst local optimum, when the n data points from $[0, 1]^d$ are perturbed by Gaussian distributions of standard deviation σ. Our bound of $O(\log(1/\sigma))$ improves significantly upon the direct translation of the bound of Englert et al. [8] to Gaussian perturbations (see Section 3 for how to translate the bound to Gaussian perturbations). It smoothly interpolates between the average-case constant approximation ratio and the worst-case bound of $O(\log n)$.

In order to obtain our improved bound for the smoothed approximation ratio, we take into account the origins of the points, i.e., their unperturbed positions. Although this information is not available to the algorithm, it can be exploited in the analysis. The smoothed analyses of approximation ratios so far [4, 5, 7–9, 12] essentially ignored this information. While this simplifies the analysis, being oblivious to the unperturbed positions seems to be too pessimistic. In fact, we see that the bound of Englert et al. [8] cannot be improved beyond $O(1/\sigma)$ by ignoring the positions of the points (Section 3). The reason for this limitation is that the lower bound for the global optimum is obtained if all points have the same origin, which corresponds to an average-case rather than a smoothed analysis. On the other hand, the upper bound for the local optimum has to

hold for all choices of the unperturbed points, most of which yield higher costs for the global optimum than the average-case analysis. Taking this into account carefully yields our bound of $O(\log(1/\sigma))$ (Section 4).

To complement our upper bound, we show that the lower bound by Chandra et al. [6] remains true for $\sigma = O(1/\sqrt{n})$ (Section 5). We conclude our paper by discussing our results and pointing out open questions (Section 6). Due to lack of space some proofs had to be omitted, which we defer to a full version of this article.

2 Preliminaries

Throughout the paper, we consider input in the Euclidean space $[0,1]^d$ and assume the dimension d to be a fixed constant. Given a sequence of points $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^d$, we call a collection $T \subseteq [n] \times [n]$ of edges a *tour*, if T is connected and every $i \in [n] = \{1, \ldots, n\}$ has in- and outdegree exactly one in T. (Note that we consider directed tours, which is useful in the analysis, but our distances are always symmetric.) Given any collection of edges S, its length is denoted by $L(S) = \sum_{(u,v) \in S} d(u,v)$, where $d(u,v)$ denotes the Euclidean distance between points X_u and X_v. We call a tour T *2-optimal*, if $d(u,v) + d(w,z) \leq d(u,w) + d(v,z)$ for all edge pairs $(u,v), (w,z) \in T$. Equivalently, it is not possible to obtain a shorter tour by replacing (u,v) and (w,z) in a 2-optimal tour T by two new edges. The 2-Opt heuristic replaces a pair of edges (u,v) and (w,z) by (u,w) and (v,z) if this decreases the tour length while this is possible. Thus, it terminates with a 2-optimal tour.

We call a collection $T \subseteq [n]^2$ a *partial 2-optimal tour* if T is a subset of a tour and $d(u,v) + d(w,z) \leq d(u,w) + d(v,z)$ for all edges $(u,v), (w,z) \in T$. Our main interests are the traveling salesman functional $\mathsf{TSP}(X) := \min_{\text{tour } T} L(T)$ and the following functional mapping the point set X to the length of the longest 2-optimal tour through X: $\mathsf{2OPT}(X) := \max_{\text{2-optimal tour } T} L(T)$.

We note that the results in Section 3 hold for metrics induced by arbitrary norms in $\mathbb{R}^d$ (Lemma 2 and 3) or typical ℓ_p norms (Lemma 4 and 5), not only for the Euclidean metric. We conjecture that also the upper bound in Section 4 holds for more general metrics, while the lower bound in Section 5 is probably specific for the Euclidean metric. Still, we think that the construction can be adapted to work for most natural metrics.

Perturbation models. In the Gaussian perturbation model (also called *two-step model*) for smoothed analysis, an adversary specifies points $x_1, \ldots, x_n$ in $[0,1]^d$ that serve as unperturbed *origins*. Each such point x_i is perturbed independently by adding a normally distributed random variable of mean 0 and standard deviation σ independently to each coordinate. Equivalently, we draw n random noise vectors $Z_i \sim \mathcal{N}(0, \sigma^2)$, where by abuse of notation $\mathcal{N}(0, \sigma^2)$ refers to the multivariate normal distribution with covariance matrix $\text{diag}(\sigma^2)$, to obtain the *perturbed input* $X_1 = x_1 + Z_1, \ldots, X_n = x_n + Z_n$. For compactness, we denote the set of unperturbed points by $\overline{X} = \{x_1, \ldots, x_n\}$ and the set of perturbed

points by $X = \{X_1, \ldots, X_n\}$. We write $X \leftarrow \mathrm{pert}_\sigma(\overline{X})$ to make explicit from which point set $\overline{X}$ the points in X are obtained.

Note that we may assume $\sigma \leq 1$ without loss of generality. If $\sigma > 1$, we can rescale the instance to be contained in $[0, 1/\sigma]^d$ and perturb the points by Gaussians with standard deviation 1 instead, which gives an equivalent instance. Thus, every upper bound for $\sigma = 1$ carries over to larger values of σ.

The ϕ-bounded perturbation model (also called *one-step model*) lets the adversary directly specify (not necessarily identical) distributions by choosing probability density functions $f_1, \ldots, f_n : [0, 1]^d \to [0, \phi]$. The perturbed input is then generated by independently sampling $X_1 \sim f_1, \ldots, X_n \sim f_n$. Note that the resulting input is always contained in $[0, 1]^d$ and with higher ϕ, the adversary can concentrate points to smaller regions of the input space. Roughly speaking, when translating Gaussian perturbations to the one-step model, ϕ is proportional to σ^{-d} for fixed d.

The following technical lemma provides a convenient way to bound the deviation of a perturbed point from its mean in the two-step model.

Lemma 1 (Chi-square bound [19, Cor. 2.19]). *Let x be a Gaussian random vector in $\mathbb{R}^d$ of standard deviation σ centered at the origin. Then, for $t \geq 3$, we have $\Pr\big[\|x\| \geq \sigma 3\sqrt{d \ln t}\big] \leq t^{-2.9d}$.*

3 Length of 2-optimal Tours under Perturbations

In this section, we provide an upper bound for the length of any 2-optimal tour and a lower bound for the length of any global optimum. These two results yield an upper bound of $O(1/\sigma)$ for the approximation ratio.

Chandra et al. [6] proved a bound on the worst-case length of 2-optimal tours that, in fact, already holds for the more general notion of *partial* 2-optimal tours. For an intuition why this is true, let us point out that their proof strategy is to argue that not too many long arcs in a tour may have similar directions due to the 2-optimality of the edges, while short edges do not contribute much to the length. The claim then follows from a packing argument. It can be verified that it is never required that the collection of edges is closed or connected.

Lemma 2. *Let $d \geq 2$. There exists a constant c_d such that for every sequence X of n points in $[0, 1]^d$, any partial 2-optimal tour has length less than $c_d \cdot n^{1-1/d}$.*

While this bound directly applies to any perturbed instance under the one-step model, Gaussian perturbations fail to satisfy the premise of bounded support in $[0, 1]^d$. However, Gaussian tails are sufficiently light to enable us to translate the result to the two-step model by carefully taking care of outliers.

Lemma 3. *Let $d \geq 2$. There exists a constant b_d such that for any $\sigma \leq 1$ the following statement holds. The probability that any partial 2-optimal tour on X has length greater than $b_d \cdot n^{1-1/d}$, i.e., $\mathsf{2OPT}(X) \geq b_d \cdot n^{1-1/d}$, is bounded by $\exp(-\Omega(\sqrt{n}))$. Furthermore,*

$$\mathrm{E}_{X \leftarrow \mathrm{pert}_\sigma(\overline{X})}\big[\mathsf{2OPT}(X)\big] \leq b_d \cdot n^{1-1/d}.$$

We complement the bound above by a lower bound on tour lengths of perturbed inputs, making use of the following result by Englert et al. [8] for the one-step model.

Lemma 4. *Let $X_1, \ldots, X_n$ be a ϕ-perturbed instance. Then with probability $1 - \exp(-\Omega(n))$, any tour on $X_1, \ldots, X_n$ has length at least $\Omega(n^{1-1/d}/\sqrt[d]{\phi})$.*

It also follows from their results that this bound translates to the two-step model consistently with the intuitive correspondence of $\phi \sim \sigma^{-d}$ between the one-step and the two-step model.

Lemma 5. *Let $X_1, \ldots, X_n$ be an instance of points in the unit cube perturbed by Gaussians of standard deviation $\sigma \leq 1$. Then with probability $1 - \exp(-\Omega(n))$ any tour on $X_1, \ldots, X_n$ has length at least $\Omega(\sigma n^{1-1/d})$.*

Note that Lemmas 3 and 5 almost immediately yield the following bound on the approximation performance for the two-step model.[3]

Observation 1 *Let $X_1, \ldots, X_n$ be an instance of points in the unit cube perturbed by Gaussians of standard deviation $\sigma \leq 1$. Then the approximation performance of 2-Opt is bounded by $O(1/\sigma)$ in expectation and with probability $1 - \exp(-\Omega(\sqrt{n}))$.*

We remark that this bound is best possible for an analysis of perturbed instances that separately bounds the lengths of any 2-optimal tour from above and gives a lower bound on any optimal tour.

4 Upper Bound on the Approximation Performance

In this section, we establish an upper bound on the approximation performance of 2-Opt under Gaussian perturbations. We achieve a bound of $O(\log 1/\sigma)$. Due to the lower bound presented in Section 5, we cannot expect an approximation ratio of $o(\log(1/\sigma)/\log\log(1/\sigma))$. Thus, our bound is almost tight.

As noted in the previous section, to beat $O(1/\sigma)$ it is essential to exploit the structure of the unperturbed input. This will be achieved by classifying edges of a tour into *long* and *short* edges and bounding the length of long edges by a (worst-case) global argument and short edges locally against the partial optimal tour on subinstances (by a reduction to an (almost-)average case). The local arguments for short edges will exploit how many unperturbed origins lie in the vicinity of a given region.

The global argument bounding long edges follows from the worst-case $O(\log n)$ bound on the worst-case approximation performance [6] that we rephrase here for our purposes.

Lemma 6. *Let T be a 2-optimal tour and* OPT *denote the length of the optimal traveling salesman tour T_{OPT}. Let T_i contain the set of all edges in T whose length is in $[\text{OPT}/2^i, \text{OPT}/2^{i-1}]$. Then $L(T_i) = O(\text{OPT})$. In particular, it follows that $L(T) = O(\log n) \cdot \text{OPT}$.*

---

[3] To show the expected approximation ratio, we additionally make use of Lemma 6.

In the proof of our bound of $O(\log 1/\sigma)$, the above lemma accounts for all edges of length $[\Omega(\sigma), O(1)]$. A central idea to bound all shorter edges is to apply the one-step model result to small parts of the input space. In particular, we will condition sets of points to be perturbed into cubes of side length σ. The following technical lemma helps to capture what values of ϕ suffice to express the conditional density function of these points depending on the distance of their unperturbed origins to the cube. This allows for appealing to the one-step model result of Lemma 4.

Lemma 7. *Let $c \in [0, \sigma]^d$ and $k = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$. Let Y be the random variable $X \sim \mathcal{N}(c, \sigma^2)$ conditioned on $X \in Q := [k_1\sigma, (k_1+1)\sigma] \times \cdots \times [k_d\sigma, (k_d+1)\sigma]$ and f_Y be the corresponding probability density function. Then f_Y is bounded from above by $\exp(\|k\|_1 + (3/2)d)\sigma^{-d}$.*

The main result of this section is the following theorem.

Theorem 2. *Let $X = (X_1, \ldots, X_n)$ be an instance of points in $[0,1]^d$ perturbed by Gaussians of standard deviation $\sigma \leq 1$. With probability $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$ for any constant $\varepsilon > 0$, we have $\mathsf{2OPT}(X) \leq O(\log(1/\sigma)) \cdot \mathsf{TSP}(X)$. Furthermore, $\mathrm{E}\left[\frac{\mathsf{2OPT}(X)}{\mathsf{TSP}(X)}\right] = O(\log(1/\sigma))$.*

Since the approximation performance of 2-Opt is bounded by $O(\log n)$ in the worst-case, we may assume that $1/\sigma = O(n^\varepsilon)$ for all $\varepsilon > 0$, since otherwise our smoothed result is superseded by Lemma 6. In what follows, let T_{OPT} and T be any optimal and 2-optimal, respectively, traveling salesman tour on $X_1, \ldots, X_n$.

4.1 Outliers and Long Edges

We will first show that the contribution of almost all points outside $[0,1]^d$ is bounded by $O(\sigma n^{1-1/d})$ with high probability and in expectation, similar to Lemma 3. For this, we subdivide C into growing cubes $A_i := [-a_i, 1 + a_i]^d$. Here, we set $a_i := 3\sigma\sqrt{di\ln(3/\sigma)}$ for $i \geq 1$ and $A_0 = [0,1]^d$. Let n_i be the number of points not contained in A_{i-1}. For every point X_j, Lemma 1 with $t := (3/\sigma)^i$ bounds $\Pr[X_j \notin A_i] \leq (\sigma/3)^{2.9d(i-1)}$ (note that we have chosen the a_i such that $t \geq 3$). Thus, $\mathrm{E}[n_i] \leq n(\sigma/3)^{2.9d(i-1)}$. For any tour T, we define E_i as the set of edges of T contained in A_i with at least one endpoint in $A_i \setminus A_{i-1}$. We first bound the contribution of the E_i with $i \geq 2$.

Lemma 8. *With probability $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$ for any constant $\varepsilon > 0$, we have $\sum_{i=2}^\infty L(E_i) = O(\sigma n^{1-1/d})$. Additionally, $\mathrm{E}[\sum_{i=2}^\infty L(E_i)] = O(\sigma n^{1-1/d})$.*

In the remainder of the proof, we bound the total length of edges inside A_1. Define $C := A_1$ and note that all edges in C have bounded length $\sqrt{d}(1 + a_1) = O(1)$. Recall that for any 2-optimal tour T, T_i contains the set of all edges in T whose length is in $[\mathrm{OPT}/2^i, \mathrm{OPT}/2^{i-1}]$. Let k_1 be such that $\sqrt{d}(1 + a_1) \in [\mathrm{OPT}/2^{k_1}, \mathrm{OPT}/2^{k_1-1}]$. Then $L(T_k) = 0$ for all $k < k_1$, since no longer edges exist. Let k_2 be such that $\sigma \in [\mathrm{OPT}/2^{k_2}, \mathrm{OPT}/2^{k_2-1}]$. Then $\sum_{k=k_1}^{k_2} L(T_k) =$

$O((k_2 - k_1) \cdot \text{OPT}) = O(\log(1/\sigma)\text{OPT})$ by Lemma 6. This argument bounds the contribution of *long edges*, i.e., edges longer than σ, in the worst case, after observing the perturbation of the input points.

4.2 Short Edges

To account for the length of the remaining edges, we take a different route: Call an edge that is shorter than σ a *short edge* and partition the bounding box C into a grid of $(\sigma \times \cdots \times \sigma)$-cubes $C_1, \ldots, C_M$ with $M = \Theta((\sigma/(1+a_1))^{-d}) = \Theta(\sigma^{-d})$, which we call *cells*. All edges in T_k for $k \geq k_2$, i.e., short edges, are completely contained in a single cell or run from some cell C_i to one of its $3^d - 1$ neighboring cells. For a given tour T, let $E_{C_i}(T)$ denote the short edges of T for which at least one of the endpoints lies in C_i.

We aim to relate the length of the edges $E_{C_i}(T)$ for any 2-optimal tour T to the length of the edges $E_{C_i}(T_{\text{OPT}})$ of the optimal tour T_{OPT}. This local approach is justified by the following property.

Lemma 9. *For any tour T, the contribution $L(E_{C_i}(T))$ of cell C_i is lower bounded by $\mathsf{TSP}(X \cap C_i) - O(\sigma|X \cap C_i|^{\frac{d-2}{d-1}})$.*

Intuitively, a cell C_i is of one of two kinds: either few points are expected to be perturbed into it and hence it cannot contribute much to the length of any 2-optimal tour (a *sparse cell*), or many unperturbed origins are close to the cell (a *heavy cell*). In the latter case, either the conditional densities of points perturbed into C_i are small, hence any optimal tour inside C_i has a large value by Lemma 4, or we find another cell close to C_i that has a very large contribution to the length of any tour.

To formalize this intuition, fix a cell C_i and let n_i be the expected number of points X_j with $X_j \in C_i$. Assume for convenience that a_1/σ and $(1 + a_1)/\sigma$ are integer. We describe the position of a cube C_i canonically by indices $\text{pos}(C_i) \in \{-\frac{a_i}{\sigma}, \ldots, \frac{1+a_i}{\sigma}\}^d$. For two cubes C_i and C_j, we define their distance as $\text{dist}(C_i, C_j) = \|\text{pos}(C_i) - \text{pos}(C_j)\|_1$. For $k \geq 0$, let D_k denote all cells of distance k to C_i and let $n(D_k)$ denote the cardinality of unperturbed origins located in a cell in D_k. We call a perturbed point $X_\ell \in C_i$ with unperturbed origin $x_\ell \in C_j$, for some $C_j \in D_k$, a *k-successful point*. Let S_k denote the set of all k-successful points. Then $n_i = \sum_{k=0}^{\infty} \text{E}[|S_k|]$.

Lemma 10. *Let $K \geq 0$ and define $S_{\leq K} := S_0 \cup \cdots \cup S_K$ as the set of k-successful points for $k \leq K$. Let $\mu := \text{E}[|S_{\leq K}|]$. If $K = o(\log \mu)$, then with probability $1 - \exp(\mu)$, we have*

$$L(E_{C_i}(T_{\text{OPT}})) \geq \frac{\sigma\mu^{1-1/d}}{\exp(O(K+1))}.$$

Proof (Sketch). The claim follows from Lemma 9 and by regarding $S_{\leq K}$ as a ϕ-perturbed instance. For this, Lemma 7 bounds the maximum density of the distributions and Lemma 4 bounds the optimal tour length from below. □

Lemma 11. *Let $\alpha := M^{\frac{d}{d-1}}$, $k_1 := \gamma \log\log(1/\sigma)$ and $k_2 := (1/\gamma')\sqrt{\log 1/\sigma}$ for sufficiently small constants γ, γ'. Then we can classify each cell C_i with $n_i \geq \frac{n}{\alpha}$ into one of the following two types.*

(T1) With probability $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$ for any constant $\varepsilon > 0$, we have

$$L(E_{C_i}(T)) = O(\log 1/\sigma)L(E_{C_i}(T_{\mathrm{OPT}})).$$

(T2) There is some $C_j \in D_{k_1} \cup \cdots \cup D_{k_2}$ such that for any $f(1/\sigma) = \mathrm{polylog}(1/\sigma)$, we have

$$L(E_{C_i}(T)) = \frac{L(E_{C_j}(T_{\mathrm{OPT}}))}{f(1/\sigma)},$$

with probability $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$ for any constant $\varepsilon > 0$.

Proof (Sketch). By Lemma 2, we can bound $L(E_{C_i}(T)) = O(\sigma n_i^{1-1/d})$. If we have $\mathrm{E}[|S_{\leq k_1}|] = \Omega(n_i)$, then Lemma 10 already proves C_i to have type T1. Otherwise, by tail bounds for the Gaussian distribution, we argue that some cell C_j in a cell of distance at most k_2 contains at least $n_i \exp(\Omega((\log\log n)^2))$ unperturbed origins. These are sufficiently many to let C_j contribute $f(1/\sigma)\sigma n_i^{1-1/d}$, for any $f(1/\sigma) = \mathrm{polylog}(1/\sigma)$, to the optimal tour length. $\qquad\square$

4.3 The Total Length of 2-optimal Tours

To bound the total length of short edges, consider first sparse cells C_i, i.e., $n_i \leq n/\alpha$. For each such cell, Chernoff bounds yield that with probability $1 - \exp(-\Omega(n/\alpha))$, at most $2n/\alpha$ points are contained in C_i, since each point is perturbed independently. By union bound, no sparse cell contains more than $2n/\alpha$ points with probability at least $1 - M\exp(-\Omega(n/\alpha))$. In this event, Lemma 2 allows for bounding the contribution of sparse cells by

$$\sum_{i:n_i \leq n/\alpha} L(E_{C_i}(T)) \leq M(3\sigma)c_d\left(\frac{6n}{\alpha}\right)^{1-\frac{1}{d}} = O\left(\frac{M\sigma n^{1-\frac{1}{d}}}{\alpha^{1-\frac{1}{d}}}\right) = O(\sigma n^{1-\frac{1}{d}}). \quad (1)$$

For bounding the length in the remaining cells, the heavy cells, let $\mathcal{T}_1 := \{i \mid C_i \text{ has type T1}\}$ and $\mathcal{T}_2 := \{i \mid C_i \text{ has type T2}\}$. We observe that with probability at least $1 - M\exp(-\Omega(n^{1-\varepsilon})) = 1 - \exp(-\Omega(n^{1-\varepsilon}))$, all type-T1 cells C_i satisfy $L(E_{C_i}(T)) = O(\log 1/\sigma)L(E_{C_i}(T_{\mathrm{OPT}}))$. Thus,

$$\sum_{i \in \mathcal{T}_1} L(E_{C_i}(T)) \leq \sum_{i \in \mathcal{T}_1} O(\log 1/\sigma)L(E_{C_i}(T_{\mathrm{OPT}})) \leq O(\log 1/\sigma)\mathrm{OPT}, \quad (2)$$

where the last inequality follows from $\sum_{i=1}^M L_{C_i}(T_{\mathrm{OPT}}) \leq 2 \cdot \mathrm{OPT}$, which holds since every edge in OPT (inside C) is counted twice on the left-hand side.

Let $A : \mathcal{T}_2 \to \{1, \ldots, M\}$ be any function that assigns to each type-T2 cell C_i a corresponding cell $C_{A(i)} \in D_{k_1} \cup \cdots \cup D_{k_2}$ satisfying the condition in (T2). We say that C_i *charges* $C_{A(i)}$. We can choose any $f(1/\sigma) = \mathrm{polylog}(1/\sigma)$ and

have with probability at least $1 - M \exp(-\Omega(n^{1-\varepsilon})) = 1 - \exp(-\Omega(n^{1-\varepsilon}))$ that $L(E_{C_i}(T)) = \frac{L(E_{C_{A(i)}}(T_{\mathrm{OPT}}))}{f(1/\sigma)}$ for all $i \in \mathcal{T}_2$. Assume that this event occurs. Since every cell C_i can only be charged by cells in distance $k_1 \le k \le k_2$, each cell can only be charged $\sum_{k=k_1}^{k_2} |D_k| = O(k_2^d)$ times. Hence,

$$\sum_{i \in \mathcal{T}_2} L(E_{C_{A(i)}}(T_{\mathrm{OPT}})) \le O(k_2^d) \sum_{i=1}^{M} L(E_{C_i}(T_{\mathrm{OPT}})) = O(k_2^d)\mathrm{OPT}.$$

Since $k_2^d = \mathrm{polylog}(1/\sigma)$, choosing $f(1/\sigma) = \mathrm{polylog}(1/\sigma)$ sufficiently large yields

$$\sum_{i \in \mathcal{T}_2} L(E_{C_i}(T)) \le \sum_{i \in \mathcal{T}_2} \frac{L(E_{C_{A(i)}}(T_{\mathrm{OPT}}))}{f(1/\sigma)} \le \frac{O(k_2^d)\mathrm{OPT}}{f(1/\sigma)} = O(\mathrm{OPT}). \quad (3)$$

Proof (of Theorem 2). By a union bound, we can bound by $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$, for any constant $\varepsilon > 0$, the probability that (i) $\mathrm{OPT} = \Omega(\sigma n^{1-1/d})$ (by Lemma 5), (ii) all edges outside C contribute $O(\sigma n^{1-1/d}) = O(\mathrm{OPT})$ (by Lemma 8), (iii) all sparse cells contribute $O(\sigma n^{1-1/d}) = O(\mathrm{OPT})$ (by (1)), (iv) the type-T1 cells C_i induce a cost of $O(\log 1/\sigma)\mathrm{OPT}$ (by (2)), and (v) the type-2 cells induce a cost of $O(\mathrm{OPT})$ (by (3)). Since the remaining edges are long edges and contribute only $O(\log(1/\sigma) \cdot \mathrm{OPT})$, we obtain that every 2-optimal tour has a length of at most $O(\log 1/\sigma)\mathrm{OPT}$ with probability $1 - \exp(-\Omega(n^{1/2-\varepsilon}))$.

Since a 2-optimal tour always constitutes a $O(\log n)$-approximation to the optimal tour length by Lemma 6, we also obtain that the expected cost of the worst 2-optimal tour is bounded by

$$O(\log 1/\sigma) \cdot \mathrm{OPT} + \exp(-\Omega(n^{1/2-\varepsilon})) \cdot O(\log n) \cdot \mathrm{OPT} = O(\log 1/\sigma) \cdot \mathrm{OPT}.$$

$\square$

5   Lower Bound on the Approximation Ratio

We complement our upper bound on the approximation performance by the following lower bound: for $\sigma = O(1/\sqrt{n})$, the worst-case lower bound is robust against perturbations. For this, we face the technical difficulty that in general, a single outlier might destroy the 2-optimality of a desired long tour, potentially cascading into a series of 2-Opt iterations that result in a substantially different or even optimal tour.

Theorem 3. *Let $\sigma = O(1/\sqrt{n})$. For infinitely many n, there is an instance X of points in $\mathbb{R}^2$ perturbed by normally distributed noise of standard deviation σ such that with probability $1 - O(n^{-s})$ for any constant $s > 0$, we have $\mathsf{2OPT}(X) = \Omega(\log n / \log \log n) \cdot \mathsf{TSP}(X)$. This also yields*

$$\mathrm{E}\left[\frac{\mathsf{2OPT}(X)}{\mathsf{TSP}(X)}\right] = \Omega\left(\frac{\log n}{\log \log n}\right).$$
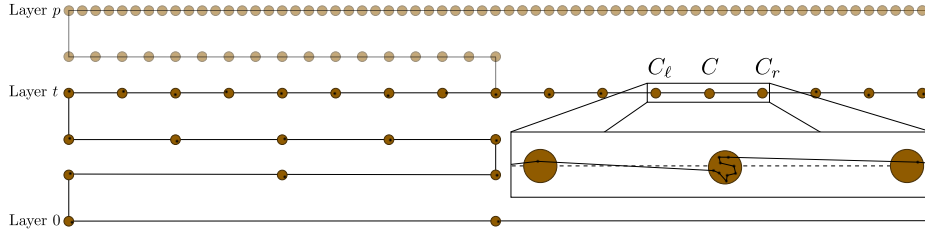
9

Fig. 1. Parts V_1 and V_3 of the lower bound instance. Each point is contained in a corresponding small container (depicted as brown circle) with high probability. The black lines indicate the constructed 2-optimal tour, which on V_2 runs analogously.

We remark that our result transfers naturally to the one-step model with $\phi = \Omega(n)$ and interestingly, holds *with probability 1* over the random perturbations.

Furthermore, even when we initialize the tour using the *nearest neighbor heuristic*, 2-Opt might, with probability $O(1)$, return a 2-optimal tour of length $\Omega(\log n / \log \log n) \cdot \mathsf{TSP}(X)$ on perturbed inputs. For space reasons, the necessary changes to the construction below are deferred to a full version of this article.

Proof of Theorem 3. We alter the construction of Chandra et al. [6] to strengthen it against Gaussian perturbations with standard deviation $\sigma = O(1/\sqrt{n})$ (see Figure 1). Let $p \geq 3$ be an odd integer and $P := 3p^{2p}$. The original instance of [6] is a subset of the $(P \times P)$-grid, which we embed into $[0,1]^2$ by scaling by $1/P$, and consists of three parts V_1, V_2 and V_3. The vertices in V_1 are partitioned into the layers $L_0, \ldots, L_p$. Layer i consists of $p^{2i} + 1$ equidistant vertices, each of which has a vertical distance of $c_i = p^{2p-2i-1}/P$ to the point above it in Layer $i+1$ and a horizontal distance of $a_i = p^{2p-2i}/P$ to the nearest neighbor(s) in the same layer. The set V_2 is a copy of V_1 shifted to the right by a distance of $2/3$. The remaining part V_3 consists of a copy of Layer p of V_1 shifted to the right by $1/3$ to connect V_1 and V_2 by a path of points. We regard L_i as the set of Layer-i points in $V_1 \cup V_2 \cup V_3$.

As in the original construction, we will construct an instance of $n = \Theta(p^{2p})$ points, which implies $p = \Theta(\log n / \log \log n)$. Let $0 \leq t \leq p$ be the largest odd integer such that $p^{2t+1} \leq (3\sigma)^{-1}$. In our construction, we drop all Layers $t+1, \ldots, p$ in both V_1 and V_2, as well as Layer p in V_3. Instead, we connect V_1 and V_2 already in Layer t by an altered copy of Layer t of V_1 shifted to the right by $1/3$. Let C be an arbitrary point of our construction, for convenience we will use the central point of Layer t in V_3. We introduce $p^{2p} - 1$ additional copies of this point C. These surplus points serve as a "padding" of the instance to ensure $n = \Theta(p^{2p})$. Note that the resulting instance has $t + 1$ layers $L_0, \ldots, L_t$. We chose t such that the magnitude of perturbation is negligible compared to the pairwise distances of all non-padding points. Furthermore, the restriction on σ ensures that incorporating the padding points increases the optimal tour length only by a constant.

10

Lemma 12. *With probability $1 - O(n^{-s})$ for any constant $s > 0$, the optimal tour has length $O(1)$.*

We find a long 2-optimal tour on all non-padding points analogously to the original construction by taking a shortcut of the original 2-optimal tour, which connects V_1 and V_2 already in Layer t (see Figure 1).

Consider the padding points, which are yet to be connected. Let C_ℓ denote the nearest point in Layer t of V_3 that is to the left of C. Symmetrically, C_r is the nearest point to the right of C. Let T^p be any 2-optimal path from C_ℓ to C_r that passes through all the padding points (including C). We replace the edges (C_ℓ, C) and (C, C_r) by the path T^p, completing the construction of our tour T.

Lemma 13. *Let $s > 0$ be arbitrary. With probability $1 - O(n^{-s})$, T is 2-optimal and has a length of $\Omega(\log n / \log \log n)$.*

By Lemmas 12 and 13, Theorem 3 follows.

6 Discussions and Open Problems

We have proved an upper bound of $O(\log 1/\sigma)$ for the smoothed approximation ratio of 2-Opt. Furthermore, we have proved that the lower bound of Chandra et al. [6] remains robust even for $\sigma = O(1/\sqrt{n})$ and even if it is initialized with the nearest-neighbor heuristic. We leave as an open problem to generalize our upper bounds to the one-step model to improve the current bound of $O(\sqrt[d]{\phi})$ [8], but conjecture that this might be difficult.

While our bound significantly improves the previously known bound for the smoothed approximation ratio of 2-Opt, we readily admit that it still does not explain the performance observed in practice. A possible explanation is that when the initial tour is not picked by an adversary or the nearest neighbor heuristic, but using a construction heuristic such as the spanning tree heuristic or an insertion heuristic, an approximation factor of 2 is guaranteed even before 2-OPT has begun to improve the tour [18]. However, a smoothed analysis of the approximation ratio of 2-Opt initialized with a good heuristic might be difficult: even in the average-case, it is only known that the length of an optimal TSP is concentrated around $\gamma_d \cdot n^{\frac{d-1}{d}}$ for some constant $\gamma_d > 0$. But the precise value of γ_d is unknown [21]. Since experiments suggest that 2-Opt even with good initialization does not achieve an approximation ratio of $1 + o(1)$ [10, 11], one has to deal with the precise constants, which seems challenging.

Finally, we conjecture that many examples for showing lower bounds for the approximation ratio of concrete algorithms for Euclidean optimization such as the TSP remain stable under perturbation for $\sigma = O(1/\sqrt{n})$. The question remains whether such small values of σ, although they often suffice to prove polynomial smoothed running-time, are essential to explain practical approximation ratios or if already slower decreasing σ provide a sufficient explanation.

References

1. Arora, S.: Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. Journal of the ACM 45(5), 753–782 (1998)
2. Arthur, D., Manthey, B., Röglin, H.: Smoothed analysis of the k-means method. Journal of the ACM 58(5) (2011)
3. Arthur, D., Vassilvitskii, S.: Worst-case and smoothed analysis of the ICP algorithm, with an application to the k-means method. SIAM J. Comp. 39(2), 766–782 (2009)
4. Bläser, M., Manthey, B., Rao, B.V.R.: Smoothed analysis of partitioning algorithms for Euclidean functionals. Algorithmica 66(2), 397–418 (2013)
5. Brunsch, T., Röglin, H., Rutten, C., Vredeveld, T.: Smoothed performance guarantees for local search. Mathematical Programming 146(1–2), 185–218 (2014)
6. Chandra, B., Karloff, H., Tovey, C.: New results on the old k-opt algorithm for the traveling salesman problem. SIAM J. Comp. 28(6), 1998–2029 (1999)
7. Curticapean, R., Künnemann, M.: A quantization framework for smoothed analysis of Euclidean optimization problems. In: Proc. of the 21st Ann. European Symp. on Algorithms (ESA). pp. 349–360 (2013)
8. Englert, M., Röglin, H., Vöcking, B.: Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP. Algorithmica 68(1), 190–264 (2014)
9. Etscheid, M.: Performance guarantees for scheduling algorithms under perturbed machine speeds. Discrete Applied Mathematics (to appear)
10. Johnson, D.S., McGeoch, L.A.: The traveling salesman problem: A case study. In: Aarts, E., Lenstra, J.K. (eds.) Local Search in Combinatorial Optimization, chap. 8. John Wiley & Sons (1997)
11. Johnson, D.S., McGeoch, L.A.: Experimental analysis of heuristics for the STSP. In: Gutin, G., Punnen, A.P. (eds.) The Traveling Salesman Problem and its Variations, chap. 9. Kluwer Academic Publishers (2002)
12. Karger, D., Onak, K.: Polynomial approximation schemes for smoothed and random instances of multidimensional packing problems. In: Proc. of the 18th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA). pp. 1207–1216. SIAM (2007)
13. Manthey, B., Röglin, H.: Smoothed analysis: Analysis of algorithms beyond worst case. it – Information Technology 53(6), 280–286 (2011)
14. Manthey, B., Röglin, H.: Worst-case and smoothed analysis of k-means clustering with Bregman divergences. J. of Comp. Geom. 4(1), 94–132 (2013)
15. Manthey, B., Veenstra, R.: Smoothed analysis of the 2-Opt heuristic for the TSP: Polynomial bounds for Gaussian noise. In: Proc. of the 24th Ann. Int. Symp. on Algorithms and Computation (ISAAC). pp. 579–589 (2013)
16. Mitchell, J.S.B.: Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for Geometric TSP, k-MST, and related problems. SIAM J. Comp. 28(4), 1298–1309 (1999)
17. Papadimitriou, C.H.: The Euclidean traveling salesman problem is NP-complete. Theoretical Computer Science 4(3), 237–244 (1977)
18. Rosenkrantz, D.J., Stearns, R.E., Lewis II, P.M.: An analysis of several heuristics for the traveling salesman problem. SIAM J. Comp. 6(3), 563–581 (1977)
19. Spielman, D.A., Teng, S.H.: Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. Journal of the ACM 51(3), 385–463 (2004)
20. Spielman, D.A., Teng, S.H.: Smoothed analysis: An attempt to explain the behavior of algorithms in practice. Communications of the ACM 52(10), 76–84 (2009)
21. Yukich, J.E.: Probability Theory of Classical Euclidean Optimization Problems, Lecture Notes in Mathematics, vol. 1675. Springer (1998)