# Smoothed Analysis of the 2-Opt Heuristic for the TSP: Polynomial Bounds for Gaussian Noise*

Bodo Manthey    Rianne Veenstra

University of Twente, Enschede, Netherlands, Department of Applied Mathematics
b.manthey@utwente.nl, h.j.veenstra@student.utwente.nl

The 2-opt heuristic is a very simple local search heuristic for the traveling salesman problem. While it usually converges quickly in practice, its running-time can be exponential in the worst case.

Englert, Röglin, and Vöcking (*Algorithmica*, 2014) provided a smoothed analysis in the so-called one-step model in order to explain the performance of 2-opt on $d$-dimensional Euclidean instances. However, translating their results to the classical model of smoothed analysis, where points are perturbed by Gaussian distributions with standard deviation $\sigma$, yields only bounds that are polynomial in $n$ and $1/\sigma^d$.

We prove bounds that are polynomial in $n$ and $1/\sigma$ for the smoothed running-time with Gaussian perturbations. In addition, our analysis for Euclidean distances is much simpler than the existing smoothed analysis.

## 1 2-Opt and Smoothed Analysis

The traveling salesman problem (TSP) is one of the classical 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 $d \geq 2$ [21]. We consider Euclidean TSP with Manhattan and Euclidean distances as well as squared Euclidean distances to measure the distances between points. For the former two, there exist polynomial-time approximation schemes (PTAS) [2, 20]. The latter, which has applications in power assignment problems for wireless networks [13], admits a PTAS for $d = 2$ and is APX-hard for $d \geq 3$ [24].

As it is unlikely that there are efficient algorithms for solving Euclidean TSP optimally, heuristics have been developed in order to find near-optimal solutions quickly. One very simple and popular heuristic is 2-opt: 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 quite quickly [14, 15], but its worst-case performance is bad: the worst-case running-time is exponential even for $d = 2$ [11] and the approximation ratio can be $\Omega(\log n/ \log\log n)$ for Euclidean instances [7].

*An extended abstract of a preliminary version of this work appeared in the proceedings of ISAAC 2013 [19].

An alternative to worst-case analysis is average-case analysis, where the expected performance with respect to some probability distribution is measured. The average-case running-time for Euclidean and random instances and the average-case approximation ratio for non-metric instances of 2-opt were analyzed [5, 7, 10, 16]. However, while worst-case analysis is often too pessimistic because it is dominated by artificial instances that are rarely encountered in practice, average-case analysis is dominated by random instances, which have often very special properties with high probability that they do not share with typical instances.

In order to overcome the drawbacks of both worst-case and average-case analysis and to explain the performance of the simplex method, Spielman and Teng invented smoothed analysis [22]: 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 underlying assumption is that real-world instances are often subjected to a small amount of random noise. This noise can come from measurement or rounding errors, or it might be a realistic assumption that the instances are influenced by unknown circumstances, but we do not have any reason to believe that these are adversarial. Smoothed analysis often allows more realistic conclusions about the performance than worst-case or average-case analysis. Since its invention, it has been applied successfully to explain the performance of a variety of algorithms [3, 4, 6, 8, 18, 25]. We refer to two surveys for an overview [17, 23].

Englert, Röglin, and Vöcking [11] provided a smoothed analysis of 2-opt in order to explain its performance. They used the *one-step model*: an adversary specifies $n$ density functions $f_1, \ldots, f_n : [0,1]^d \to [0, \phi]$. Then the $n$ points $x_1, \ldots, x_n$ are drawn independently according to the densities $f_1, \ldots, f_n$, respectively. Here, $\phi$ is the perturbation parameter. If $\phi = 1$, then the only possibility is the uniform distribution on $[0,1]^d$, and we obtain an average-case analysis. The larger $\phi$, the more powerful the adversary. Englert et al. [11] proved that the expected number of iterations of 2-opt is $O(n^4\phi)$ and $O(n^{4+\frac{1}{3}}\phi^{\frac{8}{3}}\log(n\phi))$ for Manhattan and Euclidean distances, respectively. These bounds can be improved slightly by choosing the initial tour with an insertion heuristic. However, if we transfer these bounds to the classical model of points perturbed by Gaussian distributions of standard deviation $\sigma$, we obtain bounds that are polynomial in $n$ and $1/\sigma^d$ [11, Section 6]. This is because the maximum density of a $d$-dimensional Gaussian with standard deviation $\sigma$ is $\Theta(\sigma^{-d})$. While this is polynomial for any fixed $d$, it is unsatisfactory that the degree of the polynomial depends on $d$.

## 1.1 Our Contribution

We provide a smoothed analysis of the running-time of 2-opt in the classical model, where points in $[0,1]^d$ are perturbed by independent Gaussian distributions of standard deviation $\sigma$. The bounds that we prove for Gaussian perturbations are polynomial in $n$ and $1/\sigma$, and the degree of the polynomial is independent of $d$. As distance measures, we consider Manhattan (Section 3), Euclidean (Section 5), and squared Euclidean distances (Section 4).

The analysis for Manhattan distances is essentially an adaptation of the existing analysis by Englert et al. [11, Section 4.1]. Note that our bound does not have any factor that is exponential in $d$.

Our analysis for Euclidean distances is considerably simpler than the one by Englert et al., which is rather technical and takes more than 20 pages [11, Section 4.2 and Appendix C].

The analysis for squared Euclidean distances is, to our knowledge, not preceded by a smoothed analysis in the one-step model. Because of the nice properties of squared Euclidean

| | Manhattan | Euclidean | squared Euclidean |
|---|---|---|---|
| Englert et al. [11] | $2^{O(d)} n^4 \phi$ | $O_d\big(n^{4+\frac{1}{3}} \phi^{\frac{8}{3}} \log(n\phi)\big)$ | – |
| general | $O\big(\frac{d^2 n^4 D_{\max}}{\sigma}\big)$ | $O\big(\frac{\sqrt{d} n^4 D_{\max}^4}{\sigma^4}\big)$ | $O\big(\frac{\sqrt{d} n^4 D_{\max}^2}{\sigma^2}\big)$ |
| $\sigma = O(1/\sqrt{n \log n})$ | $O\big(\frac{d^2 n^4}{\sigma}\big)$ | $O\big(\frac{\sqrt{d} n^4}{\sigma^4}\big)$ | $O\big(\frac{\sqrt{d} n^4}{\sigma^2}\big)$ |
| $\sigma = \Omega(1/\sqrt{n \log n})$ | $O\big(d^2 n^5 \sqrt{\log n}\big)$ | $O\big(\sqrt{d} n^6 \log^2 n\big)$ | $O\big(\sqrt{d} n^5 \log n\big)$ |

Table 1: Our bounds compared to the bounds obtained by Englert et al. [11] for the one-step model. The bounds can roughly be transferred to Gaussian noise by replacing $\phi$ with $\sigma^{-d}$. For convenience, we added our bounds for small and large values of $\sigma$: for $\sigma = O(1/\sqrt{n \log n})$, we have $D_{\max} = \Theta(1)$, for larger $\sigma$, we have $D_{\max} = \Theta(\sigma\sqrt{n \log n})$. The notation $O_d$ means that terms depending on $d$ are hidden in the $O$.

distances and Gaussian perturbations, this smoothed analysis is relatively compact and elegant (see in particular Section 4.2).

Table 1 summarizes our bounds.

## 2 Preparation and Technical Lemmas

Throughout the rest of this paper, $X$ denotes a set of $n$ points in $\mathbb{R}^d$, where each point is drawn according to an independent $d$-dimensional Gaussian distribution with mean in $[0,1]^d$ and standard deviation $\sigma$.

We assume that $\sigma \leq 1$. This is justified by two reasons. First, small $\sigma$ are actually the interesting case, i.e., when the order of magnitude of the perturbation is relatively small. Second, the smoothed number of iterations that 2-opt needs is a monotonically decreasing function of $\sigma$: if we have $\sigma > 1$, then this is equivalent to adversarial instances in $[0, 1/\sigma]^d$ that are perturbed with standard deviation 1. This in turn is dominated by adversarial instances in $[0,1]^d$ that are perturbed with standard deviation 1, as $[0,1/\sigma]^d \subseteq [0,1]^d$. Thus, any bound for $\sigma = 1$ holds also for larger $\sigma$.

Note that we make the dependence on all parameters (the number $n$ of points, the dimension $d$, and the perturbation parameter $\sigma$) explicit, i.e., the $O$ or $\Omega$ do not hide factors depending on, e.g., the dimension $d$.

### 2.1 2-Opt State Graph and Linked 2-Changes

Given a tour $H$ that visits all points in $X$, a *2-change* replaces two edges $\{x_1, x_2\}$ and $\{x_3, x_4\}$ of $H$ by two new edges $\{x_1, x_3\}$ and $\{x_2, x_4\}$, provided that this yields again a tour (this is the case if $x_1, x_2, x_3, x_4$ appear in this order in the tour) and that this decreases the length of the tour, i.e., $d(x_1, x_2) + d(x_3, x_4) - d(x_1, x_3) - d(x_2, x_4) > 0$, where $d(a,b) = \|a-b\|_2$ (Euclidean distances), $d(a,b) = \|a-b\|_1$ (Manhattan distances), or $d(a,b) = \|a-b\|_2^2$ (squared Euclidean distances). The 2-opt heuristic iteratively improves an initial tour by applying 2-changes until it reaches a local optimum.

The number of iterations that 2-opt needs depends of course heavily on the initial tour and on which 2-change is chosen in each iteration. We do not make any assumptions about the initial tour and about which 2-change is chosen. Following Englert et al. [11], we consider the

*2-opt state graph*: we have a node for every tour and a directed edge from tour $H$ to tour $H'$ if $H'$ can be obtained by one 2-change. The 2-opt state graph is a directed acyclic graph, and the length of the longest path in the 2-opt state graph is an upper bound for the number of iterations that 2-opt needs.

In order to improve the bounds, we also consider *pairs of linked 2-changes* [11]. Two 2-changes form a pair of linked 2-changes if there is one edge added in one 2-change and removed in the other 2-change. Formally, one 2-change replaces $\{x_1, x_2\}$ and $\{x_3, x_4\}$ by $\{x_1, x_3\}$ and $\{x_2, x_4\}$ and the other 2-change replaces $\{x_1, x_3\}$ and $\{x_5, x_6\}$ by $\{x_1, x_5\}$ and $\{x_2, x_6\}$. The edge $\{x_1, x_3\}$ is the one that appears and disappears again (or the other way round). It can happen that $\{x_2, x_4\}$ and $\{x_5, x_6\}$ intersect. Englert et al. [11] called a pair of linked 2-changes a *type $i$ pair* if $|\{x_2, x_4\} \cap \{x_5, x_6\}| = i$. As type 2 pairs, which involve in fact only four nodes, are difficult to analyze because of dependencies, we ignore them. Fortunately, the following lemma states that we will find enough disjoint pairs of linked 2-changes of type 0 and 1 in any sufficiently long sequence of 2-changes.

**Lemma 2.1** (Englert et al. [11, Lemma 9])**.** *Every sequence of $t$ consecutive 2-changes contains at least $t/6 - 7n(n-1)/24$ disjoint pairs of linked 2-changes of type 0 or type 1.*

Following Englert et al. [11, Figure 8], we subdivide type 1 pairs into type 1a and type 1b depending on how $\{x_2, x_4\}$ and $\{x_5, x_6\}$ intersect. With one of the 2-changes being $\{x_1, x_2\}$ and $\{x_3, x_4\}$ being replaced by $\{x_1, x_3\}$ and $\{x_2, x_4\}$, we obtain the following types according to the other 2-change:

*Type 0:* $\{x_1, x_3\}$ and $\{x_5, x_6\}$ are replaced by $\{x_1, x_5\}$ and $\{x_3, x_6\}$.

*Type 1a:* $\{x_1, x_3\}$ and $\{x_2, x_5\}$ are replaced by $\{x_1, x_5\}$ and $\{x_2, x_3\}$.

*Type 1b:* $\{x_1, x_3\}$ and $\{x_2, x_5\}$ are replaced by $\{x_1, x_2\}$ and $\{x_3, x_5\}$.

The main idea in the proofs by Englert et al. [11] and also in our proofs is to bound the minimal improvement of any 2-change or the minimal improvement of any pair of linked 2-changes. We denote the smallest improvement of any 2-change by $\Delta_{\min}$ and the smallest improvement of any pair of linked 2-changes by $\Delta_{\min}^{\text{link}}$. It will be clear from the context which distance measure is used for $\Delta_{\min}$ and $\Delta_{\min}^{\text{link}}$.

Suppose that the initial tour has a length of at most $L$, then 2-opt cannot run for more than $L/\Delta_{\min}$ iterations and not for more than $\Theta(L/\Delta_{\min}^{\text{link}})$ iterations, provided that $L/\Delta_{\min}^{\text{link}} = \Omega(n^2)$ because of Lemma 2.1.

The following lemma formalizes this and shows how to bound the expected number of iterations using a tail bound for $\Delta_{\min}$ or $\Delta_{\min}^{\text{link}}$.

**Lemma 2.2.** *Suppose that, with a probability of at least $1 - 1/n!$, any tour has a length of at most $L$. Let $\gamma > 1$. Then*

*(1) If $\mathbb{P}(\Delta_{\min} \leq \varepsilon) = O(P\varepsilon)$, then the expected length of the longest path in the 2-opt state graph is bounded from above by $O(PLn \log n)$.*

*(2) If $\mathbb{P}(\Delta_{\min} \leq \varepsilon) = O(P\varepsilon^\gamma)$, then the expected length of the longest path in the 2-opt state graph is bounded from above by $O(P^{1/\gamma}L)$.*

*(3) The same bounds as (1) and (2) hold if we replace $\Delta_{\min}$ by $\Delta_{\min}^{\text{link}}$, provided that $PL = \Omega(n^2)$ for Case 1 and $P^{1/\gamma}L = \Omega(n^2)$ for Case 2.*

*Proof.* If the length of the longest tour is longer than $L$, then we use the trivial upper bound of $n!$. This contributes only $O(1)$ to the expected value.

Consider the first statement. Let $T$ be the longest path in the 2-opt state graph. If $T \geq t$, then $\Delta_{\min} \leq L/t$. Plugging this in and observing the $n!$ is an upper bound for $T$ yields

$$\mathbb{E}(T) = \sum_{t=1}^{n!} \mathbb{P}(T \geq t) \leq \sum_{t=1}^{n!} O(PL/t) = O(\log(n!) \cdot PL) = O(PLn \log n).$$

Now consider the second statement, and let $T$ be as above. Let $K = O(L \cdot P^{1/\gamma})$. Then

$$\begin{aligned}\mathbb{E}(T) &= \sum_{t=1}^{n!} \mathbb{P}(T \geq t) \leq \sum_{t=1}^{n!} \min\big(1, O\big(P(L/t)^\gamma\big)\big) \\ &= K + PL^\gamma \cdot \sum_{t \geq K} O(t^{-\gamma}) = K + PL^\gamma \cdot O(K^{1-\gamma}) = O(K).\end{aligned}$$

Finally, we consider the third statement. The statement follows from the observation that the maximal number of disjoint pairs of linked 2-changes and the length of the longest path in the 2-opt state graph are asymptotically equal if they are of length at least $\Omega(n^2)$ (Lemma 2.1) and the probability statements become nontrivial only for $t = \Omega(PL)$ in the first and $t = \Omega(P^{1/\gamma}L)$ in the second case. □

## 2.2 Basic Probability Theory

In order to get an upper bound for the length of the initial tour, we need an upper bound for the diameter of the point set $X$. Such an upper bound is also necessary for the analysis of 2-changes with Euclidean distances (Section 5). We choose $D_{\max}$ such that $X \subseteq [-D_{\max}, D_{\max}]^d$ with a probability of at least $1 - 1/n!$. For fixed $d$ and $\sigma \leq 1$, we can choose $D_{\max} = \Theta(1 + \sigma\sqrt{n \log n})$ according to the following lemma. For $\sigma = O(1/\sqrt{n \log n})$, we have $D_{\max} = \Theta(1)$.

**Lemma 2.3.** *Let $c \geq 2$ be a sufficiently large constant, and let $D_{\max} = c \cdot (\sigma\sqrt{n \log n} + 1)$. Then $\mathbb{P}(X \not\subseteq [-D_{\max}, D_{\max}]^d) \leq 1/n!$.*

*Proof.* We have $X \not\subseteq [-D_{\max}, D_{\max}]^d$ only if there is a point $x_i$ and a coordinate of $x_i$ that is perturbed by more than $D_{\max} - 1 \geq c\sigma \cdot \sqrt{n \log n}$. According to Durrett [9, Theorem 1.2.3], the probability that a 1-dimensional Gaussian of standard deviation $\sigma$ is more than $c\sigma\sqrt{n \log n}$ away from its mean is bounded from above by $2 \cdot \frac{\exp(-c^2 n \log n/2)}{c\sqrt{2\pi n \log n}}$. Thus, the probability that $X \not\subseteq [-D_{\max}, D_{\max}]^d$ is bounded from above by $2dn \cdot \frac{\exp(-c^2 n \log n/2)}{c\sqrt{2\pi n \log n}}$. For sufficiently large $c$, this is at most $1/n!$. □

The following lemma is well known and follows from the fact that the density of a $d$-dimensional Gaussian with standard deviation $\sigma$ is bounded from above by $(2\sigma)^{-d}$.

**Lemma 2.4.** *Let $a \in \mathbb{R}^d$ be drawn according to a $d$-dimensional Gaussian distribution of standard deviation $\sigma$, and let $B = \{b \in \mathbb{R}^d \mid \|b - c\|_2 \leq \varepsilon\}$ be a $d$-dimensional hyperball of radius $\varepsilon$ centered at $c \in \mathbb{R}^d$. Then $\mathbb{P}(a \in B) \leq (\varepsilon/\sigma)^d$.*

For $x, y \in \mathbb{R}^d$ with $x \neq y$, let $L(x, y) = \{\xi \cdot (y - x) + x \mid \xi \in \mathbb{R}\}$ denote the straight line through $x$ and $y$.

**Lemma 2.5.** *Let $a, b \in \mathbb{R}^d$ be arbitrary with $a \neq b$. Let $c \in \mathbb{R}^d$ be drawn according to a $d$-dimensional Gaussian distribution with standard deviation $\sigma$. Then the probability that $c$ is $\varepsilon$-close to $L(a,b)$, i.e., $\min_{c^\star \in L(a,b)} \|c - c^\star\|_2 \leq \varepsilon$, is bounded from above by $(\varepsilon/\sigma)^{d-1}$.*

*Proof.* We divide drawing $c$ into drawing a 1-dimensional Gaussian $c^\star$ in the direction of $a-b$ and drawing a $(d-1)$-dimensional Gaussian $c'$ in the hyperplane orthogonal to $a-b$ and containing $c^\star$. Then the distance of $c$ to $L(a,b)$ is $\|c - c^\star\|_2$. For every $c^\star$, the point $c$ is $\varepsilon$-close to $L(a,b)$ only if $c'$ falls into a $(d-1)$-dimensional hyperball of radius $\varepsilon$ around $c^\star$. The lemma follows by applying Lemma 2.4. □

We need the following lemma in Section 5.

**Lemma 2.6.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function whose derivative is bounded from above by $B$, let $c$ be distributed according to Gaussian distribution with standard deviation $\sigma$. Let $I$ be an interval of size $\varepsilon$, and let $f(I) = \{f(x) \mid x \in I\}$ be the image of $I$. Then $\mathbb{P}(c \in f(I)) = O(B\varepsilon/\sigma)$.*

*Proof.* Since the derivative of $f$ is bounded by $B$, the set $f(I)$ is contained in some interval of length $B\varepsilon$. The lemma follows since the density of $c$ is bounded from above by $O(1/\sigma)$. □

### 2.3 Chi Distribution

The chi distribution [12, Section 8] is the distribution of the Euclidean length of a $d$-dimensional Gaussian random vector of standard deviation $\sigma$ and mean 0. In the following, we denote its density function by $\chi_d$. It is given by

$$\chi_d(x) = \frac{2^{1-\frac{d}{2}} \cdot \left(\frac{x}{\sigma}\right)^{d-1} \cdot \exp\bigl(-(x/\sigma)^2/2\bigr)}{\sigma \cdot \Gamma(d/2)}, \tag{1}$$

where $\Gamma$ denotes the gamma function. We need the following lemma a couple of times.

**Lemma 2.7.** *Assume that $c \in \mathbb{N}$ is a fixed constant and $d \in \mathbb{N}$ is arbitrary with $d > c$. Then we have*

$$\int_0^\infty \chi_d(x) x^{-c} \, \mathrm{d}x = \frac{2^{-c/2} \Gamma\left(\frac{d-c}{2}\right)}{\sigma^c \cdot \Gamma\left(\frac{d}{2}\right)} = \Theta\left(\frac{1}{d^{c/2} \cdot \sigma^c}\right).$$

*Proof.* The first equality follows by integration. For the second inequality, we observe $2^{-c/2}$ is a fixed constant and that

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x+\mu(x)}$$

for some function $\mu$ with $\mu(x) \in \left[0, \frac{1}{12x}\right]$ according to Stirling's formula [1, 6.1.37]. We have

$\frac{d-c}{2} \geq \frac{1}{2}$ as $d > c$ and both are integers. Then

$$\begin{aligned}
\frac{\Gamma(\frac{d-c}{2})}{\Gamma(\frac{d}{2})} &= \frac{\sqrt{2\pi} \cdot \left(\frac{d-c}{2}\right)^{\frac{d-c-1}{2}} \cdot \exp\left(-\frac{d-c}{2} + \mu\left(\frac{d-c}{2}\right)\right)}{\sqrt{2\pi} \cdot \left(\frac{d}{2}\right)^{\frac{d-1}{2}} \cdot \exp\left(-\frac{d}{2} + \mu\left(\frac{d}{2}\right)\right)} \\
&= \frac{\left(\frac{d-c}{2}\right)^{\frac{d-c-1}{2}}}{\left(\frac{d}{2}\right)^{\frac{d-1}{2}}} \cdot \underbrace{\exp\left(\frac{c}{2} + \mu\left(\frac{d-c}{2}\right) - \mu\left(\frac{d}{2}\right)\right)}_{=\Theta(1)} \\
&= \underbrace{\left(\frac{d-c}{d}\right)^{\frac{d-1}{2}}}_{=A} \cdot \underbrace{\left(\frac{d-c}{2}\right)^{-\frac{c}{2}}}_{=B} \cdot \Theta(1).
\end{aligned}$$

Here, the third equality follows from two facts: $c$ is a fixed constant, thus $\exp(c/2) = \Theta(1)$ and $\frac{d-c}{2}, \frac{d}{2} \geq \frac{1}{2}$. Thus, $\mu(\frac{d-c}{2})$ and $\mu(\frac{d}{2})$ lie between 0 and a constant. Hence, the exponential term is $\Theta(1)$.

Analyzing $A$ and $B$ remains to be done: We have $B \leq (d/2)^{-c/2}$, thus $B = O(d^{-c/2})$. If $d \leq 2c$, then $B$ is bounded from below by a constant and so is $d^{-c/2}$. If $d \geq 2c$, then $B \geq (d/4)^{-c/2} = \Omega(d^{-c/2})$.

We have $A = \left(1 - \frac{c}{d}\right)^{\frac{d-1}{2}} \leq \exp\left(-\frac{(d-1)\cdot c}{2d}\right) = O(1)$. Distinguishing the cases $d \leq 2c$ and $d > 2c$ as for $B$ yields $A = \Omega(1)$. □

The analysis with Euclidean and squared Euclidean distances depends on the distribution of the distance between two points perturbed by Gaussians, where larger distance is better for the analysis. The following two lemmas show that, given that larger distance is better, we can replace the distribution of the distance by the corresponding chi distribution.

**Lemma 2.8.** *Assume that $a$ is drawn according to a $d$-dimensional Gaussian distribution with standard deviation $\sigma$ and mean 0. Assume that $b$ is drawn according to a $d$-dimensional Gaussian distribution with standard deviation $\sigma$ and mean $\mu$. Then $\|b\|_2$ stochastically dominates $\|a\|_2$, i.e., $\mathbb{P}(\|b\|_2 \leq t) \leq \mathbb{P}(\|a\|_2 \leq t)$ for all $t \in \mathbb{R}$.*

*Proof.* For $d = 1$, we have the following:

$$\begin{aligned}
\mathbb{P}\big(\|b\|_2 \leq t\big) &= \mathbb{P}\big(b \in [-t, t]\big) = \mathbb{P}\big(a \in [-t-\mu, t-\mu]\big) \\
&= \mathbb{P}\big(a \in [-t, t]\big) + \underbrace{\mathbb{P}\big(a \in [-t-\mu, -t]\big) - \mathbb{P}\big(a \in [t-\mu, t]\big)}_{\leq 0} \leq \mathbb{P}\big(\|a\|_2 \leq t\big)
\end{aligned}$$

Now we prove the lemma for larger $d$. Since Gaussian distributions are rotation symmetric, we can assume that $\mu = (\delta, 0, \ldots, 0)$ for some $\delta \geq 0$.

We observe that $\|b\|_2$ dominates $\|a\|_2$ if and only if $\|b\|_2^2$ dominates $\|a\|_2^2$. Let $b' = a + \mu$. It suffices to prove the lemma for this choice of $b'$, as $b'$ follows the same distribution as $b$. Fixing $a_2, \ldots, a_d$ fixes also $a'_2, \ldots, a'_d$. Then $\|b'\|_2^2$ dominates $\|a\|_2^2$ if $|a_1 + \delta|$ dominates $|a_1|$. This is true because the lemma holds for $d = 1$. □

**Lemma 2.9.** *Let $b$ be as in Lemma 2.8, and let $h : [0, \infty] \to [0, \infty)$ be a monotonically decreasing function. Let $g$ be the density function of $\|b\|$. Then*

$$\int_0^\infty g(x)h(x)\,\mathrm{d}x \leq \int_0^\infty f_{d,\sigma}(x)h(x)\,\mathrm{d}x,$$

*provided that both integrals exist.*

*Proof.* Let $a$ denote the $d$-dimensional Gaussian random variable of standard deviation $\sigma$ and mean 0. Then $\|a\|_2$ has density $f_{d,\sigma}$. By Lemma 2.8, $\|a\|_2$ is dominated by $\|b\|_2$. This implies that $h(\|a\|_2)$ dominates $h(\|b\|_2)$ since $h$ is monotonically decreasing. The lemma follows by observing that the two integrals are the two expected values of $h(\|a\|_2)$ and $h(\|b\|_2)$. □

For Euclidean and squared Euclidean distances, it turns out to be useful to study $\Delta_{a,b}(c) = d(c,a) - d(c,b)$ for points $a, b, c \in X$. By abusing notation, we sometimes write $\Delta_{i,j}(k)$ instead of $\Delta_{x_i,x_j}(x_k)$ for short. A 2-change that replaces $\{x_1, x_2\}$ and $\{x_3, x_4\}$ by $\{x_1, x_3\}$ and $\{x_2, x_4\}$ improves the tour length by $\Delta_{1,4}(2) - \Delta_{1,4}(3) = \Delta_{2,3}(1) - \Delta_{2,3}(4)$.

## 3 Manhattan Distances

The essence of our analysis for Manhattan distances is a straightforward adaptation of the analysis in the one-step model. The extra factor of $D_{\max}$ comes from the bound of the initial tour, and the extra factor of $d^2$ stems from stating the dependence on $d$ explicitly and getting rid of the exponential dependence on $d$ [11, Proofs of Theorem 7 and Lemma 10].

**Lemma 3.1.** *$\mathbb{P}(\Delta_{\min}^{\text{link}} \leq \varepsilon) = O(d^2 n^6 \varepsilon^2/\sigma^2)$.*

*Proof.* We consider a pair of linked 2-changes as described in Section 2.1. The improvement of the first 2-change is

$$\Gamma_1 = \sum_{i=1}^{d} |x_{1i} - x_{2i}| + |x_{3i} - x_{4i}| - |x_{1i} - x_{3i}| - |x_{2i} - x_{4i}|,$$

where $x_{ji}$ is the $i$-th coordinate of $x_j \in X$. The improvement of the second 2-change is

$$\Gamma_2 = \sum_{i=1}^{d} |x_{1i} - x_{3i}| + |x_{5i} - x_{6i}| - |x_{1i} - x_{5i}| - |x_{3i} - x_{6i}|.$$

Note that we can have a type 1 pair, i.e., two of the points $x_2, x_4, x_5, x_6$ can be identical.

Each ordering of the $x_{ji}$ gives rise to a linear combination for $\Gamma_1$ and $\Gamma_2$. We have $(6!)^d$ such orderings. If we examine the case distinctions by Englert et al. [11, Lemmas 11, 12, 13] closely, we see that any pair of linear combinations is either impossible (it uses a different ordering of the variables for $\Gamma_1$ and $\Gamma_2$ or one of $\Gamma_1$ and $\Gamma_2$ is non-positive, thus the corresponding 2-change is in fact not a 2-change) or we have one variable $x_{ji}$ that has a non-zero coefficient in $\Gamma_1$ and a coefficient of 0 in $\Gamma_2$ and another variable $x_{j'i'}$ that has a non-zero coefficient in $\Gamma_2$ and a coefficient of 0 in $\Gamma_1$. The absolute values of the non-zero coefficients of $x_{ji}$ and $x_{j'i'}$ is 2. Now $\Gamma_1$ falls into $(0, \varepsilon]$ only if $x_{ji}$ falls into an interval of length $\varepsilon/2$. This happens with a probability of at most $O(\varepsilon/\sigma)$. By independence, the same holds for $\Gamma_2$ and $x_{j'i'}$.

However, we would incur an extra factor of $(6!)^d$ in this way, and we would like to remove all exponential dependence of $d$. In order to do this, we assume that we know $i$ and $i'$ already. This comes at the expense of a factor of $O(d^2)$ for taking a union bound over the choices of $i$ and $i'$. We let an adversary fix values for all $x_{j\tilde{i}}$ with $\tilde{i} \neq i, i'$. Since we know $i$ and $i'$, we are left with at most $(6!)^2 = O(1)$ possible linear combinations.

Finally, the lemma follows by taking a union bound over all $O(n^6)$ possible pairs of linked 2-changes. □

**Theorem 3.2.** *The expected length of the longest path in the 2-opt state graph corresponding to d-dimensional instances with Manhattan distances is at most* $O(d^2 n^4 D_{\max}/\sigma)$*.*

*Proof.* The initial tour has a length of at most $O(ndD_{\max})$ with a probability of at least $1-1/n!$ by Lemma 2.3. We apply Lemma 2.2 for linked 2-changes using Lemma 3.1 and $\gamma = 2$. □

# 4 Squared Euclidean Distances

## 4.1 Preparation

In this section, we have $\Delta_{a,b}(c) = \|c-a\|_2^2 - \|c-b\|_2^2$ for $a, b, c \in \mathbb{R}^d$.

Assume that we have a 2-change that replaces $\{x_1, x_2\}$ and $\{x_3, x_4\}$ by $\{x_1, x_3\}$ and $\{x_2, x_4\}$. The improvement caused by this 2-change is $\Delta_{2,3}(1) - \Delta_{2,3}(4) = \Delta_{1,4}(2) - \Delta_{1,4}(3)$. Given the positions of the four nodes except for a single $i \in \{1,2,3,4\}$, such a 2-change yields a small improvement only if the corresponding $\Delta_{\cdot,\cdot}(i)$ falls into some interval of size $\varepsilon$. The following lemma gives an upper bound for the probability that this happens.

**Lemma 4.1.** *Let* $a, b \in \mathbb{R}^d$*,* $a \neq b$*, and let* $c$ *be drawn according to a Gaussian distribution with standard deviation* $\sigma$*. Let* $I \subseteq \mathbb{R}$ *be an interval of length* $\varepsilon$*. Then*

$$\mathbb{P}\big(\Delta_{a,b}(c) \in I\big) \leq \frac{\varepsilon}{4\sigma \cdot \|a-b\|_2}.$$

*Proof.* Since Gaussian distributions are rotationally symmetric, we can assume without loss of generality that $a = (0,\ldots,0)$ and $b = (\delta, 0, \ldots, 0)$ with $\delta = \|a-b\|_2$. Let $c = (c_1, \ldots, c_d)$. Then $\Delta_{a,b}(c) = c_1^2 - (c_1-\delta)^2 = 2c_1\delta + \delta^2$. Thus, $\Delta_{a,b}(c) \in I$ if and only if $c_1$ falls into an interval of length $\frac{\varepsilon}{2\delta}$. Since $c_1$ is a 1-dimensional Gaussian random variable with a standard deviation of $\sigma$, the probability for this is bounded from above by $\frac{\varepsilon}{4\delta\sigma}$ since the maximum density of a 1-dimensional Gaussian of standard deviation $\sigma$ is bounded from above by $\frac{1}{2\sigma}$. □

## 4.2 Single 2-Changes

In this section, we prove a simple bound for the expected number of iterations of 2-opt with squared Euclidean distances. This bounds holds for all $d \geq 2$. In the next section, we improve this bound for the case $d \geq 3$ using pairs of linked 2-changes.

**Lemma 4.2.** *For* $d \geq 2$*, we have* $\mathbb{P}(\Delta_{\min} \in (0, \varepsilon]) = O\big(\frac{n^4\varepsilon}{\sigma^2\sqrt{d}}\big)$.

*Proof.* Consider a 2-change where $\{x_1, x_2\}$ and $\{x_3, x_4\}$ are replaced by $\{x_1, x_3\}$ and $\{x_2, x_4\}$. Its improvement is given by $\Delta_{2,3}(1) - \Delta_{2,3}(4)$. We let an adversary fix $x_3$. Then we draw $x_2$. This fixes the distance $\delta = \|x_2 - x_3\|_2$. Now we draw $x_4$. This fixes $\Delta_{2,3}(4)$. The 2-change yields an improvement of at most $\varepsilon$ only if $\Delta_{2,3}(1)$ falls into an interval of size at most $\varepsilon$. According to Lemma 4.1, the probability that this happens is at most $\frac{\varepsilon}{4\delta\sigma}$.

Now let $g$ be the probability density of $\delta = \|x_2 - x_3\|$. Then the probability that the 2-change yields an improvement of at most $\delta$ is bounded from above by

$$\int_0^\infty g(\delta) \cdot \frac{\varepsilon}{4\sigma\delta} \, \mathrm{d}\delta \leq \int_0^\infty \chi_d(\delta) \cdot \frac{\varepsilon}{4\sigma\delta} \, \mathrm{d}\delta = O\left(\frac{\varepsilon}{\sigma^2\sqrt{d}}\right).$$

The first step is due to Lemma 2.9. The second step is due to Lemma 2.7 using $c = 1$ and $d \geq 2$. The lemma follows by a union bound over the $O(n^4)$ possible 2-changes. □

**Theorem 4.3.** *For all $d \geq 2$, the expected length of the longest path in the 2-opt state graph corresponding to d-dimensional instances with squared Euclidean distances is at most $O(n^6 \log n \sqrt{d} D_{\max}^2/\sigma^2)$.*

*Proof.* With a probability of at least $1 - 1/n!$, the instance is contained in a hypercube of sidelength $D_{\max}$. Thus, the longest edge has a length of at most $\sqrt{d} D_{\max}$. Therefore, the initial tour has a length of at most $ndD_{\max}^2$. We combine this with Lemmas 2.2 and 4.2 to complete the proof. □

### 4.3 Pairs of Linked 2-Changes

We can obtain an better bound than in the previous section by analyzing pairs of linked 2-changes. With the following three lemmas, we analyze the probability that pairs of linked 2-changes of type 0, 1a, or 1b yield an improvement of at most $\varepsilon$.

**Lemma 4.4.** *For $d \geq 2$, the probability that there exists a pair of type 0 of linked 2-changes that yields an improvement of at most $\varepsilon$ is bounded from above by $O\left(\frac{n^6 \varepsilon^2}{\sigma^4 d}\right)$.*

*Proof.* Consider a fixed pair of type 0 of linked 2-changes involving the six points $x_1, \ldots, x_6$ as described in Section 2.1. We show that the probability that it yields an improvement of at most $\varepsilon$ is at most $O(\varepsilon\sigma^{-2}/\sqrt{d})$. A union bound over the $O(n^6)$ possibilities of pairs of type 0 yields the lemma.

The basic idea is that we restrict ourselves to analyzing $\Delta_{1,4}(3)$ and $\Delta_{1,6}(5)$ only in order to bound the probability that we have a small improvement. In this way, we use the principle of deferred decision to show that we can analyze the improvements of the two 2-changes as if they were independent:

1. We let an adversary fix $x_1$ arbitrarily.
2. We draw $x_4$, which determines the distance $\|x_1 - x_4\|$.
3. We draw $x_2$. This fixes the position of the interval of size bad interval for $\Delta_{1,4}(3)$, and $x_3$ is still random.
4. We draw $x_3$. The probability that $x_3$ assumes a position such that the first 2-change yields an improvement of at most $\varepsilon$ is thus at most $\frac{\varepsilon}{4\sigma \cdot \|x_1 - x_4\|}$.
5. We draw $x_6$. This determines the distance $\|x_1 - x_6\|$.
6. We draw $x_5$. The probability that $x_5$ assumes a position such that the second 2-change yields an improvement of at most $\varepsilon$ is thus at most $\frac{\varepsilon}{4\sigma \cdot \|x_1 - x_6\|}$.

Let $g$ be the probability density function of the distance between $x_1$ and $x_4$, and let $g'$ be the probability density function of the distance between $x_1$ and $x_6$. By independence of the points, the probability that both 2-changes of the pair yield an improvement of at most $\varepsilon$ is bounded from above by

$$\int_{\delta=0}^{\infty} g(\delta) \cdot \frac{\varepsilon}{4\sigma\delta} \, \mathrm{d}\delta \cdot \int_{\delta=0}^{\infty} g'(\delta) \cdot \frac{\varepsilon}{4\sigma\delta} \, \mathrm{d}\delta.$$

We observe that $\frac{\varepsilon}{4\sigma\delta}$ is monotonically decreasing in $\delta$. Thus, by Lemma 2.9, we can replace $g$ and $g'$ by the density $\chi_d$ of the chi distribution to get the following upper bound for the probability that a pair of type 0 yields an improvement of at most $\varepsilon$:

$$\left(\int_0^\infty \chi_d(\delta) \cdot \frac{\varepsilon}{4\delta\sigma}\,\mathrm{d}\delta\right)^2 = O\left(\frac{\varepsilon^2}{\sigma^4 d}\right).$$

Here, we use Lemma 2.7 with $c = 1$, which is allowed since $d \geq 2$. □

**Lemma 4.5.** *For $d \geq 2$, the probability that there exists a pair of type 1a of linked 2-changes that yields an improvement of at most $\varepsilon$ is bounded from above by $O\big(\frac{n^5\varepsilon^2}{\sigma^4 d}\big)$.*

*Proof.* We can analyze pairs of type 1a in the same way as type 0 pairs in Lemma 4.4. To do this, we analyze $\Delta_{2,3}(4)$ and $\Delta_{1,2}(5)$:

1. We fix $x_2$.
2. We draw $x_3$. This fixes $\|x_2 - x_3\|$.
3. We draw $x_1$. This fixes $\|x_1 - x_2\|$. In addition, this fixes the positions of the intervals into which $\Delta_{2,3}(4)$ and $\Delta_{1,2}(5)$ must fall if the first or second 2-change yield an improvement of at most $\varepsilon$.
4. We draw $x_4$.
5. We draw $x_5$.

The remainder of the proof is identical to the proof of Lemma 4.4. □

**Lemma 4.6.** *For $d \geq 3$, the probability that there exists a pair of type 1b of linked 2-changes that yields an improvement of at most $\varepsilon$ is bounded from above by $O\big(\frac{n^5\varepsilon^2}{\sigma^4 d}\big)$.*

*Proof.* Again, we proceed similarly to Lemma 4.4. We analyze a fixed pair of type 1b, where $\{x_1, x_2\}$ and $\{x_3, x_4\}$ are replaced by $\{x_1, x_3\}$ and $\{x_2, x_4\}$ in one step and $\{x_1, x_3\}$ and $\{x_2, x_5\}$ are replaced by $\{x_1, x_2\}$ and $\{x_3, x_5\}$, and apply a union bound over the $O(n^5)$ possible type 1a pairs. We analyze the probability that $\Delta_{2,3}(4)$ or $\Delta_{2,3}(5)$ assume a bad value.

We draw the points in the following order:

1. We fix $x_2$.
2. We draw $x_3$. This fixes the distance $\|x_2 - x_3\|_2$, which is crucial for both 2-changes.
3. We draw $x_1$.
4. We draw $x_4$. The probability that the first 2-change yields an improvement of at most $\varepsilon$ is at most $\frac{\varepsilon}{4\sigma \cdot \|x_2 - x_3\|}$.
5. We draw $x_5$. The probability that the second 2-change yields an improvement of at most $\varepsilon$ is at most $\frac{\varepsilon}{4\sigma \cdot \|x_2 - x_3\|}$.

The main difference to Lemma 4.4 is that the sizes of the bad intervals are not independent. However, once the size of the bad intervals is fixed, we can analyze the probabilities that $\Delta_{2,3}(4)$ or $\Delta_{2,3}(5)$ fall into their bad intervals as independent. Given that $\|x_2 - x_3\| = \delta$ is fixed, the probability that the first and the second 2-change yield an improvement of at most $\varepsilon$ is bounded from above by $\frac{\varepsilon^2}{16\delta^2\sigma^2}$. Since this is decreasing in $\delta$, we can replace the distribution of $\delta$ by the chi distribution to obtain an upper bound according to Lemma 2.9. Thus, using Lemma 2.7 with $c = 2$ and $d \geq 3$, we obtain the following upper bound for the probability that a pair of type 1b yields an improvement of at most $\varepsilon$:

$$\int_{\delta=0}^{\infty} \chi_d(\delta) \cdot \frac{\varepsilon^2}{16\delta^2\sigma^2} \,\mathrm{d}\delta = O\left(\frac{\varepsilon^2}{d\sigma^4}\right).$$

□

With the three lemmas above, we can obtain a bound on the expected number of iterations of 2-opt for TSP with squared Euclidean distances.

**Theorem 4.7.** *For $d \geq 3$, the expected length of the longest path in the 2-opt state graph corresponding to $d$-dimensional instances with squared Euclidean distances is at most $O\big(\frac{n^4\sqrt{d}D_{\max}^2}{\sigma^2}\big)$.*

*Proof.* The probability that any pair of linked 2-changes of type 0, 1a, or 1b yields an improvement of at most $\varepsilon$ is bounded from above by $O\big(\frac{\varepsilon^2 n^6}{\sigma^4 d}\big)$. We apply Lemma 2.2 with $\gamma = 2$ and observe that the initial tour has a length of at most $O(ndD_{\max}^2)$ with a probability of at least $1 - 1/n!$. □

# 5 Euclidean Distances

## 5.1 Difference of Euclidean Distances

In this section, we have $\Delta_{a,b}(z) = \|z-a\|_2 - \|z-b\|_2$ for $a, b, z \in \mathbb{R}^d$. Analyzing $\|z-a\|_2 - \|z-b\|_2$ turns out to be more difficult than analyzing $\|z-a\|_2^2 - \|z-b\|_2^2$ in the previous section. In particular the case when $\|z-a\|_2 - \|z-b\|_2$ is close to the maximal possible value of $\|a-b\|_2$ requires special attention.

We observe that $\eta = \Delta_{a,b}(z)$ behaves essentially 2-dimensionally: it depends only on the distance of $z$ from $L(a,b)$ (this is $x$ in the following lemma) and on the position of the projection $z$ onto $L(a,b)$ (this is $y$ in the following lemma). Furthermore, it depends on the distance $\|a-b\|_2$ between $a$ and $b$ (this is $\delta$ in the following lemma). The following lemma makes the connection between $x$ and $y$ explicit for a given $\eta$. Figure 1 depicts the situation described in the lemma.

**Lemma 5.1.** *Let $z = (x,y) \in \mathbb{R}^2$, $x \geq 0$, $y \geq 0$. Let $a = (0, -\delta/2)$ and $b = (0, \delta/2)$ be two points at a distance of $\delta$. Let $\eta = \|z-a\|_2 - \|z-b\|_2$. Then we have*

$$y^2 = \frac{\eta^2\delta^2 + 4\eta^2x^2 - \eta^4}{4\delta^2 - 4\eta^2} = \frac{\eta^2}{4} + \frac{\eta^2x^2}{\delta^2 - \eta^2} \tag{2}$$

*for $0 \leq \eta < \delta$ and*

$$x^2 = \frac{y^2 \cdot \left(4\delta^2 - 4\eta^2\right) + \eta^4 - \eta^2\delta^2}{4\eta^2} = \frac{y^2 \cdot \left(\delta^2 - \eta^2\right)}{\eta^2} - \frac{\delta^2 - \eta^2}{4}. \tag{3}$$

*for $\delta \geq \eta > 0$. Furthermore, $\eta > \delta$ is impossible.*

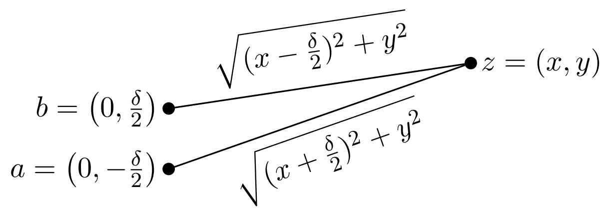


Figure 1: The situation for Lemma 5.1.

*Proof.* The last statement follows from the triangle inequality.

We have $\eta = \sqrt{(y+\delta/2)^2+x^2} - \sqrt{(y-\delta/2)^2+x^2}$. Rearranging terms and squaring implies

$$\begin{aligned} \eta^2 + (y-\delta/2)^2 + x^2 + 2\eta\sqrt{(y-\delta/2)^2+x^2} &= (y+\delta/2)^2 + x^2 \\ \Leftrightarrow 2\eta\sqrt{(y-\delta/2)^2+x^2} &= 2y\delta - \eta^2. \end{aligned}$$

Squaring again implies

$$\begin{aligned} 4\eta^2 \cdot \left((y-\delta/2)^2 + x^2\right) &= 4y^2\delta^2 - 4y\delta\eta^2 + \eta^4 \\ \Leftrightarrow 4\eta^2 \cdot \left(y^2 - \delta y + \delta^2/4 + x^2\right) &= 4y^2\delta^2 - 4y\delta\eta^2 + \eta^4 \\ \Leftrightarrow 4\eta^2 y^2 - 4\eta^2\delta y + \eta^2\delta^2 + 4\eta^2 x^2 &= 4y^2\delta^2 - 4y\delta\eta^2 + \eta^4 \\ \Leftrightarrow 4\eta^2 y^2 + \eta^2\delta^2 + 4\eta^2 x^2 &= 4y^2\delta^2 + \eta^4. \end{aligned}$$

Again rearranging terms implies

$$y^2 \cdot \left(4\delta^2 - 4\eta^2\right) = \eta^2\delta^2 + 4\eta^2 x^2 - \eta^4.$$

Using the assumption $\eta < \delta$ or $\eta > 0$ implies the two claims. □

As said before, the difficult case in analyzing $\Delta_{a,b}(c) = \eta$ is when $\eta \approx \|a-b\|_2$. In terms of the previous lemma, this can only happen if $x$ is small, i.e., if $c$ . The following lemmas makes this precise.

**Lemma 5.2.** *Let $a, b, z \in [-D_{\max}, D_{\max}]^d$. Assume that $\|a-b\| = \delta$ and that $z$ has a distance of $x$ from $L(a,b)$. Then*

$$|\Delta_{a,b}(z)| \le \delta - \frac{x^2\delta}{8dD_{\max}^2} \tag{4}$$

*Proof.* Let $y$ be the distance of $z$ from $m = \frac{a+b}{2}$, and let $\eta = \Delta_{a,b}(z)$. Then, according to (3), we have

$$x^2 = \frac{y^2 \cdot \left(\delta^2 - \eta^2\right)}{\eta^2} - \frac{\delta^2 - \eta^2}{4}.$$

The second term on the right-hand side is non-positive as $\eta \le \delta$. We omit it and use the upper bound $y \le \sqrt{d}D_{\max}$ to obtain the following weaker bound:

$$x^2 \le \frac{dD_{\max}^2 \cdot \left(\delta^2 - \eta^2\right)}{\eta^2}. \tag{5}$$

We distinguish two cases. The first case is that $\eta < \delta/2$. In this case, it suffices to show that $\delta/2 \leq \delta - \frac{x^2\delta}{8dD_{\max}^2}$ in order to prove (4). Since $|x| \leq 2\sqrt{d}D_{\max}$, this holds because $\delta/2 \leq \delta - \delta/2$.

The second case is that $\eta \geq \delta/2$. We have

$$\delta^2 - \eta^2 = (\delta - \eta) \cdot (\delta + \eta) \leq 2\delta \cdot (\delta - \eta).$$

Replacing $\delta^2 - \eta^2$ by $2\delta \cdot (\delta - \eta)$ in the numerator and $\eta^2$ by $\delta^2/4$ in the denominator of (5), we obtain

$$x^2 \leq \frac{dD_{\max}^2 \cdot \left(\delta^2 - \eta^2\right)}{\eta^2} \leq \frac{8dD_{\max}^2 \cdot (\eta - \delta)}{\delta}.$$

Rearranging terms completes the proof. □

In order to be able to apply Lemma 2.6, we need the following upper bound on the derivative of $y$ with respect to $\eta$, given that $x$ is fixed.

**Lemma 5.3.** *For $x, y \geq 0$, let $y = \sqrt{\frac{\eta^2}{4} + \frac{\eta^2 x^2}{\delta^2 - \eta^2}}$ with $0 < \eta < \delta$. Assume further that $\eta \leq \delta - \frac{x^2\delta}{8dD_{\max}^2}$ and that $x \leq 2\sqrt{d}D_{\max}$. Then the derivative of $y$ with respect to $\eta$ is bounded by*

$$\frac{1}{2} + \frac{8D_{\max}^3 d^{3/2}}{\delta x^2}.$$

*Proof.* The derivative of $y$ with respect to $\eta$ is given by

$$\begin{aligned}
y' = \frac{\mathrm{d}y}{\mathrm{d}\eta} &= \frac{\delta^4 - 2\delta^2\eta^2 + \eta^4 + 4\delta^2 x^2}{2 \cdot (\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}} \\
&= \frac{(\delta^2 - \eta^2)^2 + 4\delta^2 x^2}{2 \cdot (\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}} \\
&= \frac{(\delta^2 - \eta^2)^2}{2 \cdot (\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}} + \frac{4\delta^2 x^2}{2 \cdot (\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}} \\
&\leq \frac{1}{2} + \frac{2\delta^2 x^2}{(\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}}.
\end{aligned}$$

We observe that $y' \geq 0$ for all $x$ and allowed choices of $\eta$ and $\delta$. For the second term, we have

$$\frac{2\delta^2 x^2}{(\delta^2 - \eta^2)^{3/2} \cdot \sqrt{\delta^2 - \eta^2 + 4x^2}} \leq \frac{2\delta^2 x^2}{(\delta^2 - \eta^2)^{3/2} \cdot \sqrt{4x^2}} = \frac{\delta^2 x}{(\delta^2 - \eta^2)^{3/2}}.$$

By assumption, we have $\delta - \eta \geq \frac{x^2\delta}{8dD_{\max}^2}$ and $\eta \leq \delta$. Thus, we have

$$\frac{\delta^2 x}{(\delta^2 - \eta^2)^{3/2}} = \frac{\delta^2 x}{\big((\delta - \eta) \cdot (\delta + \eta)\big)^{3/2}} \leq \frac{\delta^2 x}{(2\delta)^{3/2} \cdot \left(\frac{x^2\delta}{8dD_{\max}^2}\right)^{3/2}} = \frac{8D_{\max}^3 d^{3/2}}{\delta x^2}$$

□

Using Lemmas 5.3 and 2.6, we can bound the probability that $\Delta_{a,b}(z)$ assumes a value in an interval of size $\varepsilon$.

**Lemma 5.4.** *Let $d \geq 4$. Let $a, b \in [-D_{\max}, D_{\max}]^d$ be arbitrary, $a \neq b$, and let $z$ be drawn according to a Gaussian distribution with standard deviation $\sigma$. Let $\delta = \|a - b\|_2$. Let $I$ be an interval of length $\varepsilon$. Then*

$$\mathbb{P}\big(\Delta_{a,b}(z) \in I \text{ or } z \notin [-D_{\max}, D_{\max}]^d\big) = O\left(\frac{\varepsilon D_{\max}^3 \sqrt{d}}{\sigma^3 \delta}\right).$$

*Proof.* Let $x$ denote the distance of $z$ to $L(a, b)$, and let $y$ denote the position of the projection of $z$ onto $L(a, b)$. First, let us assume that $x$ is fixed. Then, by Lemmas 5.3 and 2.6, the probability that $\Delta_{a,b}(z) \in I$ is bounded from above by

$$O\left(\left(1 + \frac{D_{\max}^3 d^{3/2}}{\delta x^2}\right) \cdot \frac{\varepsilon}{\sigma}\right).$$

Here, the requirements of Lemma 5.3 are satisfied because of Lemma 5.2, or we have $z \notin [-D_{\max}, D_{\max}]^d$.

We observe that this probability is decreasing in $x$. Thus, in order to get an upper bound for the probability with random $x$, we can use the $(d-1)$-dimensional chi distribution for $x$ according to Lemma 2.9. We obtain

$$\begin{aligned}\int_{x=0}^{\infty} \chi_{d-1}(x) \cdot O\left(\left(1 + \frac{D_{\max}^3 d^{3/2}}{\delta x^2}\right) \cdot \frac{\varepsilon}{\sigma}\right) \mathrm{d}x &= O\left(\frac{\varepsilon}{\sigma}\right) + \int_{x=0}^{\infty} \chi_{d-1}(x) \cdot O\left(\frac{D_{\max}^3 d^{3/2} \varepsilon}{\sigma \delta x^2}\right) \mathrm{d}x \\ &= O\left(\frac{\varepsilon}{\sigma} + \frac{D_{\max}^3 \sqrt{d} \varepsilon}{\sigma^3 \delta}\right).\end{aligned}$$

by Lemma 2.7 using $c = 2$ and $d - 1 \geq 3$. Since $\delta \leq 2\sqrt{d} D_{\max}$, the lemma follows. □

## 5.2 Analysis of Pairs of 2-Changes

We immediately go to pairs of linked 2-changes, as these yield the better bounds.

**Lemma 5.5.** *For $d \geq 4$, the probability that a pair of linked 2-changes of type 0 yields an improvement of at most $\varepsilon$ or some point lies outside $[-D_{\max}, D_{\max}]^d$ is bounded from above by*

$$O\left(\frac{n^6 \varepsilon^2 D_{\max}^6}{\sigma^8}\right).$$

*Proof.* We proceed similarly as in the proof of Lemma 4.4 for type 0 pairs for squared Euclidean distances. We draw the points of a fixed pair of linked 2-changes as in the proof of Lemma 4.4.

In the same way as in the proof of Lemma 4.4, using Lemma 5.4 instead of Lemma 4.1, we obtain that the probability that one fixed of the two 2-changes yields an improved of at most $\varepsilon$ is bounded from above by

$$\int_{\delta=0}^{\infty} \chi_d(\delta) \cdot O\left(\frac{\varepsilon D_{\max}^3 \sqrt{d}}{\sigma^3 \delta}\right) \mathrm{d}\delta = O\left(\frac{\varepsilon D_{\max}^3}{\sigma^4}\right).$$

Here, we applied Lemma 2.7 with $c = 1$.

Again in the same way as in the proof of Lemma 4.4, we can analyze both 2-changes of the type 0 pair is if they are independent. Finally, the lemma follows by a union bound over the $O(n^6)$ possibilities for a type 0 pair. □

**Lemma 5.6.** *For $d \geq 4$, the probability that a pair of linked 2-changes of type 1a yields an improvement of at most $\varepsilon$ or some point lies outside $[-D_{\max}, D_{\max}]^d$ is bounded from above by*

$$O\left(\frac{n^5 \varepsilon^2 D_{\max}^6}{\sigma^8}\right).$$

*Proof.* The lemma can be proved in the same way as Lemma 4.5 with differences analogous to the proof of Lemma 5.5. □

**Lemma 5.7.** *For $d \geq 4$, the probability that a pair of linked 2-changes of type 1b yields an improvement of at most $\varepsilon$ or some point lies outside $[-D_{\max}, D_{\max}]^d$ is bounded from above by*

$$O\left(\frac{n^5 \varepsilon^2 D_{\max}^6}{\sigma^8}\right).$$

*Proof.* Similar to the proof of Lemma 4.6 and using Lemma 5.4, the probability that the two 2-changes of the pair both yield an improvement of at most $\varepsilon$ is bounded from above by

$$\int_{\delta=0}^{\infty} \chi_d(\delta) \cdot O\left(\frac{\varepsilon D_{\max}^3 \sqrt{d}}{\sigma^3 \delta}\right)^2 \mathrm{d}\delta = \int_{\delta=0}^{\infty} \chi_d(\delta) \cdot O\left(\frac{\varepsilon^2 D_{\max}^6 d}{\sigma^6 \delta^2}\right) \mathrm{d}\delta.$$

Now the lemma follows by applying Lemma 2.7 with $c = 2$. □

**Theorem 5.8.** *For $d \geq 4$, the expected length of the longest path in the 2-opt state graph corresponding to $d$-dimensional instances with Euclidean distances is at most $O(\frac{\sqrt{d} n^4 D_{\max}^4}{\sigma^4})$.*

*Proof.* We have $\mathbb{P}(\Delta_{\min}^{\text{link}} \leq \varepsilon) = O(\frac{n^6 \varepsilon^2 D_{\max}^6}{\sigma^8})$ by Lemmas 5.5, 5.6, and 5.7. If all points are in $[-D_{\max}, D_{\max}]^d$, then the longest edge has a length of $O(\sqrt{d} D_{\max})$. Thus, the initial tour has a length of at most $O(n\sqrt{d} D_{\max})$. Plugging this into Lemma 2.2 yields the result. □

# 6 Concluding Remarks

**Polynomial bound for Euclidean distances for all $d$.** Our approach for Euclidean distances does not work for $d = 2$ and $d = 3$. However, we can use the bound of Englert et al. [11] for Euclidean distances, which yields a bound polynomial in $n$ and $1/\sigma$ for both cases.

**Initial tour.** One reason that we obtain worse bounds is that our upper bound for the length of the initial tour is worse because we do not truncate the Gaussian distributions. This effect is even stronger for Euclidean distances, where the maximum distance between points plays a role also in the analysis of the 2-changes. Only for $\sigma = O(1/\sqrt{n \log n})$, this effect is negligible, as then $D_{\max} = O(1)$.

In the same way as Englert et al. [11], we can slightly improve the smoothed number of iterations by using an insertion heuristic to choose the initial tour. We save a factor of $n^{1/d}$ for Manhattan and Euclidean distances and a factor of $n^{2/d}$ for squared Euclidean distances. The reason is that there always exist tours of length $O(D_{\max} n^{1-\frac{1}{d}})$ for $n$ points in $[-D_{\max}, D_{\max}]^d$ for Euclidean and Manhattan distances and of length $O(D_{\max}^2 n^{1-\frac{2}{d}})$ for squared Euclidean distances for $d \geq 2$ [26].

**Approximation ratio.** Using the fact that any local optimum of 2-opt yields a tour of length at most $O(D_{\max} n^{1-\frac{1}{d}})$ [7] and that the optimal tour has a length of $\Omega(n^{1-\frac{1}{d}}\sigma)$ [11], we obtain a smoothed approximation ratio of $O(D_{\max}/\sigma)$. This, however, is worse than the worst-case ratio of $O(\log n)$ [7] as $D_{\max}/\sigma = \Omega(\sqrt{n/\log n})$. The reason for this bound is that the upper bound for the local optimum involves $D_{\max}$.

We conjecture an approximation ratio of $O(1/\sigma)$, which is what we would obtain if plugging $\sigma = \Theta(\phi^{-d})$ into the bound of Englert et al. [11] were allowed.

# References


[1] Milton Abramowitz and Irene A. Stegun, editors. *Pocketbook of Mathematical Functions.* Harri Deutsch, 1984.

[2] Sanjeev Arora. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *Journal of the ACM*, 45(5):753–782, 1998.

[3] David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the $k$-means method. *Journal of the ACM*, 58(5), 2011.

[4] David Arthur and Sergei Vassilvitskii. Worst-case and smoothed analysis of the ICP algorithm, with an application to the $k$-means method. *SIAM Journal on Computing*, 39(2):766–782, 2009.

[5] Karl Bringmann, Christian Engels, Bodo Manthey, and B. V. Raghavendra Rao. Random shortest paths: Non-euclidean instances for metric optimization problems. In Krishnendu Chatterjee and Jiří Sgall, editors, *Proc. of the 38th Int. Symp. on Mathematical Foundations of Computer Science (MFCS)*, volume 8087 of *Lecture Notes in Computer Science*, pages 219–230. Springer, 2013.

[6] Tobias Brunsch, Kamiel Cornelissen, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the successive shortest path algorithm. In *Proc. of the 24th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 1180–1189. SIAM, 2013.

[7] Barun Chandra, Howard Karloff, and Craig Tovey. New results on the old $k$-opt algorithm for the traveling salesman problem. *SIAM Journal on Computing*, 28(6):1998–2029, 1999.

[8] Valentina Damerow, Bodo Manthey, Friedhelm Meyer auf der Heide, Harald Räcke, Christian Scheideler, Christian Sohler, and Till Tantau. Smoothed analysis of left-to-right maxima with applications. *ACM Transactions on Algorithms*, 8(3), 2012.

[9] Richard Durrett. *Probability: Theory and Examples.* Cambridge University Press, 2013.

[10] Christian Engels and Bodo Manthey. Average-case approximation ratio of the 2-opt algorithm for the TSP. *Operations Research Letters*, 37(2):83–84, 2009.

[11] Matthias Englert, Heiko Röglin, and Berthold Vöcking. Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP. *Algorithmica*, 68(1):190–264, 2014.

[12] Merran Evans, Nicholas Hastings, and Brian Peacock. *Statistical Distributions.* John Wiley & Sons, 3rd edition, 2000.

[13] Stefan Funke, Sören Laue, Zvi Lotker, and Rouven Naujoks. Power assignment problems in wireless communication: Covering points by disks, reaching few receivers quickly, and energy-efficient travelling salesman tours. *Ad Hoc Networks*, 9(6):1028–1035, 2011.

[14] David S. Johnson and Lyle A. McGeoch. The traveling salesman problem: A case study. In Emile Aarts and Jan Karel Lenstra, editors, *Local Search in Combinatorial Optimization*, chapter 8. John Wiley & Sons, 1997.

[15] David S. Johnson and Lyle A. McGeoch. Experimental analysis of heuristics for the STSP. In Gregory Gutin and Abraham P. Punnen, editors, *The Traveling Salesman Problem and its Variations*, chapter 9. Kluwer Academic Publishers, 2002.

[16] Walter Kern. A probabilistic analysis of the switching algorithm for the TSP. *Mathematical Programming*, 44(2):213–219, 1989.

[17] Bodo Manthey and Heiko Röglin. Smoothed analysis: Analysis of algorithms beyond worst case. *it – Information Technology*, 53(6):280–286, 2011.

[18] Bodo Manthey and Heiko Röglin. Worst-case and smoothed analysis of $k$-means clustering with Bregman divergences. *Journal of Computational Geometry*, 4(1):94–132, 2013.

[19] Bodo Manthey and Rianne Veenstra. Smoothed analysis of the 2-Opt heuristic for the TSP: Polynomial bounds for Gaussian noise. In Leizhen Cai, Siu-Wing Cheng, and Tak-Wah Lam, editors, *Proc. of the 24th Ann. Int. Symp. on Algorithms and Computation (ISAAC)*, volume 8283 of *Lecture Notes in Computer Science*, pages 579–589. Springer, 2013.

[20] Joseph S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for Geometric TSP, $k$-MST, and related problems. *SIAM Journal on Computing*, 28(4):1298–1309, 1999.

[21] Christos H. Papadimitriou. The Euclidean traveling salesman problem is NP-complete. *Theoretical Computer Science*, 4(3):237–244, 1977.

[22] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

[23] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: An attempt to explain the behavior of algorithms in practice. *Communications of the ACM*, 52(10):76–84, 2009.

[24] Fred van Nijnatten, René Sitters, Gerhard J. Woeginger, Alexander Wolff, and Mark de Berg. The traveling salesman problem under squared Euclidean distances. In Jean-Yves Marion and Thomas Schwentick, editors, *Proc. of the 27th Int. Symp. on Theoretical Aspects of Computer Science (STACS)*, volume 5 of *LIPIcs*, pages 239–250. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2010.

[25] Roman Vershynin. Beyond Hirsch conjecture: Walks on random polytopes and smoothed complexity of the simplex method. *SIAM Journal on Computing*, 39(2):646–678, 2009.

[26] Joseph E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*, volume 1675 of *Lecture Notes in Mathematics*. Springer, 1998.