New Results on the Old k-Opt Algorithm for the TSP

Barun Chandra*, Howard Karloff†, Craig Tovey‡

Abstract

Local search with k-change neighborhoods is perhaps the oldest and most widely used heuristic method for the traveling salesman problem, yet almost no theoretical performance guarantees for it were previously known. This paper develops several results, some worst-case and some probabilistic, on the performance of 2- and k-opt local search for the TSP, with respect to both the quality of the solution and the speed with which it is obtained.

1 Introduction

Local search with k-change neighborhoods is perhaps the oldest and most widely used heuristic method for the traveling salesman problem [13, 17]. Given a graph $G = (V, E)$ and a tour T of G ("tour" is synonymous with "Hamiltonian cycle"), a tour T' is said to be obtained from T by an *improving k-change* if T' is shorter than T, and T' is obtained by removing k edges from T and adding k new edges. The k-opt algorithm starts with an arbitrary initial tour and incrementally improves on this tour by making successive improving k'-changes for any $k' \leq k$, terminating when no such improving changes can be made. This paper develops several results, some worst-case and some probabilistic, on the performance of 2- and k-opt algorithms for the TSP, with respect to the two principal criteria, quality and speed.

Quality: how good is a locally optimal solution? The only results on this question that we are aware of are due to Grover [7], Lueker [16] and Plesnik [18]. Grover proves that for any (symmetric) TSP instance, any 2-optimal tour has length at most the average of all tour lengths. (This result also was credited to Edelberg [16, page 7] but without reference). Lueker gives a construction for which this bound is tight when tour lengths differ. Plesnik shows that there are graphs with n vertices satisfying the triangle inequality (i.e., the distances are those in an n-point metric space), whose worst-case performance ratio can be as bad as $\frac{1}{2\sqrt{2}}\sqrt{n}$. Our results regarding solution quality are:

- For TSPs satisfying the triangle inequality the worst-case performance ratio of 2-opt is at most $4\sqrt{n}$ for all n. The k-opt algorithm can have a performance ratio that is at least $\frac{1}{4}n^{\frac{1}{2k}}$ for infinitely many n.

*Department of Computer Science, University of Chicago, Chicago, IL 60637.

†College of Computing, Georgia Tech, Atlanta, Georgia 30332-0280. The author was supported in part by NSF grant CCR 9107349.

‡School of ISyE and College of Computing, Georgia Tech, Atlanta, Georgia 30332-0280. The author was supported in part by NSF grant DDM-9215467.

- For TSPs embedded in the normed space $\mathbb{R}^m$, the worst-case performance ratio of k-opt is $O(\log n)$. If the points are embedded in $\mathbb{R}^2$ and the distances are Euclidean, then there is a $c > 0$ such that the worst-case performance ratio of 2-opt is at least $c \cdot \frac{\log n}{\log \log n}$ for infinitely many n.

- For all norms on $\mathbb{R}^m$, there exists a constant c such that any 2-optimal tour on any TSP instance in the unit hypercube (with norm-induced distances) has length less than $cn^{1-1/m}$. A corollary is that if points are sampled i.i.d. uniformly from the hypercube, 2-opt has $O(1)$ worst-case ratio with high probability and the expected value of this ratio is also $O(1)$.

Speed: how many iterations does local search require? There seem to be three previous results on this question. The first two of these are worst-case. Lueker [16] constructs a TSP instance for which there exists an exponentially long sequence of improving 2-changes. For all sufficiently large k, Johnson, Papadimitriou, and Yannakakis [8] and Krentel [11] prove the existence of instances and initial tours with the stronger property that all improving sequences, starting from the given initial tour, are exponentially long. Krentel [12] claims to have extended this result to all $k \geq 8$.

Here we extend Lueker's construction for all $k > 2$, giving explicit instances for which there exist exponentially long improving sequences.

The third previously known result is probabilistic. Kern [10] shows that for random Euclidean instances on the unit square, the probability is at least $1 - c/n$ that the number of iterations required by 2-opt is $O(n^{16})$, where c is a constant. It was not known if the expected number of iterations is polynomial. Our main probabilistic results regarding speed are:

- For random Euclidean instances in the unit square, the expected number of iterations required by 2-opt is $O(n^{10} \log n)$.

- For random L_1 instances in the unit hypercube, the expected number of iterations required by 2-opt is $O(n^6 \log n)$.

Taken together, our results provide the first theoretical proof of the quality of 2-opt as a heuristic for random TSP instances in the unit square. In particular, the expected time is polynomial, and the expected worst-case performance ratio is bounded by a constant.

2 Preliminaries

We begin by stating some definitions and notation that will be used throughout the paper.

A *metric space* (V, d) is a nonempty set V of points and a function $d : V \times V \rightarrow \mathbb{R}$, called *distance*, satisfying the following properties for all $x, y, z \in V$:

(i) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

(ii) $d(x, y) = d(y, x)$;

(iii) $d(x, z) \leq d(x, y) + d(y, z)$.

A *norm* N on $\mathbb{R}^m$ is a function $||\cdot|| : \mathbb{R}^m \to \mathbb{R}$ satisfying the following properties.

(i) For all $x \in \mathbb{R}^m$, $||x|| \geq 0$ and $||x|| = 0$ if and only if $x = 0$;

(ii) $||cx|| = |c| \cdot ||x||$ for every $c \in \mathbb{R}$ and $x \in \mathbb{R}^m$;

(iii) $||x + y|| \leq ||x|| + ||y||$ for every $x, y \in \mathbb{R}^m$.

A norm N induces a metric space where the distance function is defined as $d_N(x, y) = ||x - y||$. If the norm of a m-dimensional vector $x = (x_1, x_2, \ldots, x_d)$ is defined as $\sqrt[p]{|x_1^p| + |x_2^p| + \cdots + |x_d^p|}$, where p is a positive integer, then the corresponding metric is called the L_p metric. The L_2 metric is called the *Euclidean metric*.

A *geometric graph* is given by a finite nonempty set V of points in $\mathbb{R}^m$ and a norm N on $\mathbb{R}^m$. The graph is a complete weighted graph on V with the weight of edge $\{x, y\}$ being $d_N(x, y) = ||x - y||$. When the metric considered is the Euclidean metric, the graph is called a *Euclidean graph*.

Given a weighted graph $G = (V, E)$ we refer to the weight or length of an edge $e \in E$ by $wt(e)$. Given a collection of edges $E' \subseteq E$, the weight $wt(E')$ of E' is the sum of the weights of the edges in E'; if the edges in E' form a tour T', we also refer to $wt(T')$ as the *length* of the tour. We denote an optimal tour by $OPT(G)$. Since we work only in complete graphs in a metric space (so given V, $G = (V, E)$ is completely determined), we also abuse notation slightly and refer to $OPT(V)$.

Given a weighted graph $G = (V, E)$ and a tour T of G, a tour T' is said to be obtained from T by an *improving k-change* if T' is shorter than T, and T' is obtained by removing k edges from T and adding k new edges. A tour T is said to be *k-optimal* if for all $k' \leq k$, no improving k'-change can be made to T. The *k-opt algorithm* starts with an arbitrary initial tour T_0 and incrementally improves on this tour by finding $T_1, T_2, \ldots, T_z$ where T_{i+1} is obtained from T_i by an improving k'-change for some $k' \leq k$, and T_z is k-optimal.

The k-opt algorithm can start from many different initial tours, and even starting from the same initial tour, k-opt can end up in many different k-optimal tours. All the upper bounds in this paper are proved for the worst possible outcome of k-opt.

3 Bounds on Performance Ratios in Metric Spaces

We first prove that the performance ratio of k-opt cannot be bounded by a function of n if the triangle inequality is not imposed.

Theorem 3.1. For all $k \geq 2$, for all $n \geq 2k + 10$, for all $M > 0$, there exists a complete weighted graph G on n vertices, with strictly positive weights, containing a k-optimal tour T' such that $wt(T')/OPT(G) > M$.

Proof. We prove the result for all k even and $n \geq 2k + 8$. The result will follow for k odd since k-optimality implies $(k-1)$-optimality. The idea of the construction is to take a pair of Hamiltonian cycles in G which differ by a $(k+1)$-change. We set the weights of all edges in these cycles to ϵ; all other edges in G are given very large weight. For one special edge in the first cycle, we change the weight to 1. This keeps the first cycle k-optimal but now its weight is many times that of the second cycle.

The graph G has n vertices denoted $1, 2, \ldots, n$. Its edge weights are

1. $wt(1, 2) = 1$
2. $wt(i, i + 1) = \epsilon$ for all $i > 1$, and $wt(n, 1) = \epsilon$.
3. $wt(k + 3, 2k + 4) = \epsilon$.
4. $wt(j, 2k + 4 - j) = \epsilon$ for all $1 \leq j \leq k$.
5. All other edges have weight kn.

In general, T is $1, 2k+3, 2k+2, 2,\quad 3, 2k+1, 2k, 4,\quad 5, 2k-1, 2k-2, 6,\quad \ldots,\quad k-1, k+5, k+4, k,\quad k+1, k+2, k+3,\quad 2k+4, 2k+5, 2k+6, 2k+7, \ldots, n-2, n-1, n$. This tour has weight $n\epsilon$. For example, when $k = 8$ and $n = 26$, the optimal tour T is $1, 19, 18, 2, 3, 17, 16, 4, 5, 15, 14, 6, 7, 13, 12, 8, 9, 10, 11, 20, 21, 22, 23, 24, 25, 26$.

The tour T' is $1, 2, 3, \ldots, n$ with weight $1 + (n - 1)\epsilon$. If we set $\epsilon = 1/(Mn)$ the performance ratio will exceed M as desired. We still have to verify that T' is k-optimal. This is straightforward and left to the reader. ■

Plesnik [18] showed that for a graph with n vertices, the worst-case performance ratio of 2-opt could be as bad as $\frac{\sqrt{n}}{2\sqrt{2}}$, and conjectured that the worst-case performance ratio for 3-opt is 2. We show that Plesnik's bound for 2-opt is tight up to a constant factor by proving an upper bound of $4\sqrt{n}$ on the performance ratio of 2-opt. We also disprove his conjecture for 3-opt by proving lower bounds that approach infinity as n goes to infinity for k-opt for all k; a lower bound for 2-opt then follows as a special case.

The upper bounds in this section and in section 5 use techniques similar to those of [4, 20].

Let M be any arbitrary metric space with a distance function d. Let V be a set of points in M, and let $n = |V|$. Let $OPT(V)$ be an optimal tour on V and let $T(V)$ be *any* tour on V which is locally optimal with respect to the 2-opt algorithm.

We first state a simple fact which follows from the triangle inequality.

Fact 3.2. $V' \subseteq V \Rightarrow wt(OPT(V')) \leq wt(OPT(V))$.

Lemma 3.3. For any $k \in \{1, 2, \ldots, n\}$, let $E_k = \{\text{edges } e \in T(V) | wt(e) > \frac{2 \cdot wt(OPT(V))}{\sqrt{k}}\}$. Then $|E_k| < k$.

Proof: Suppose otherwise; so for some k, $r = |E_k| \geq k$. Orient the edges of $T(V)$ in a consistent manner, i.e., so that the directed edges form a directed Hamiltonian cycle. Consider the directed edges (with the above orientation) of E_k, $(t_1, h_1), (t_2, h_2), \ldots, (t_r, h_r)$, where the t_i's are the tails and the h_i's are the heads of these directed edges.

We first see that not too many tails can be clustered very closely together. Consider any sphere of radius $\frac{wt(OPT(V))}{\sqrt{k}}$ around some point in the metric space. We show that the number of tails (of edges from E_k) in this sphere is less than $\sqrt{k}$.

Suppose otherwise, so that the tails $t_{i_1}, t_{i_2}, \ldots, t_{i_p}$ all lie in the sphere for some $p \geq \sqrt{k}$. Let $h_{i_1}, h_{i_2}, \ldots, h_{i_p}$ be the corresponding heads. For any $u \neq v$, $d(t_{i_u}, t_{i_v}) \leq \frac{2 \cdot wt(OPT(V))}{\sqrt{k}}$, since t_{i_u} and t_{i_v}

lie in the sphere. This implies that $d(h_{i_u}, h_{i_v}) \geq \frac{2 \cdot wt(OPT(V))}{\sqrt{k}}$, since otherwise we get a shorter valid tour $(T(V) \cup \{(t_{i_u}, t_{i_v}), (h_{i_u}, h_{i_v})\}) - \{(t_{i_u}, h_{i_u}), (t_{i_v}, h_{i_v})\}$ with a 2-change operation. But since, by supposition, we have $p \geq \sqrt{k}$ heads and these heads are pairwise at a distance at least $\frac{2 \cdot wt(OPT(V))}{\sqrt{k}}$ apart, the optimal tour on these heads is of length at least $2 \cdot wt(OPT(V))$, which contradicts Fact 3.2.

Now we show that a large number of tails have to be at a large distance apart. Pick any arbitrary tail t_i and consider the sphere (of radius $\frac{wt(OPT(V))}{\sqrt{k}}$) centered around t_i. "Kill" all the tails within this sphere. By the above argument, fewer than $\sqrt{k}$ tails can have been killed. Now pick any remaining "live" tail and kill all tails in the sphere centered at this tail. Repeat this process until all tails have been killed. Since there are at least k tails and in a single iteration we kill fewer than $\sqrt{k}$ tails, this process can be repeated more than $\sqrt{k}$ times. Clearly, the tails at the center of the spheres are at a distance greater than $\frac{wt(OPT(V))}{\sqrt{k}}$ apart from each other, and there are greater than $\sqrt{k}$ of them, therefore the optimal tour on the tails is of length greater than $wt(OPT(V))$, which contradicts Fact 3.2. ∎

Theorem 3.4. $\frac{wt(T(V))}{wt(OPT(V))} \leq 4\sqrt{n}$.

Proof: Note that Lemma 3.3 implies that the weight of the kth largest edge is at most $\frac{2 \cdot wt(OPT(V))}{\sqrt{k}}$. Hence

$$\begin{aligned} wt(T(V)) &= \sum_{k=1}^{n} wt(k^{th} \text{ largest edge}) \\ &\leq \sum_{k=1}^{n} \frac{2 \cdot wt(OPT(V))}{\sqrt{k}} \\ &= 2 \cdot wt(OPT(V)) \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \\ &\leq 2 \cdot wt(OPT(V)) \int_{x=0}^{n} \frac{1}{\sqrt{x}} \\ &= 4 \cdot wt(OPT(V)) \cdot \sqrt{n}. \blacksquare \end{aligned}$$

3.1 Lower Bounds for 2-Opt and k-Opt

We prove a lower bound on the performance ratio of k-opt; Plesnik's lower bound on the performance ratio of 2-opt [18] follows as a special case.

Theorem 3.5. For any $k \geq 2$, for infinitely many values of n, there exists a complete weighted n-node graph $G_{k,n}$ with positive edge weights satisfying the triangle inequality, and a k-optimal tour $T_{k,n}$ of $G_{k,n}$, such that $\frac{wt(T_{k,n})}{wt(OPT(G_{k,n}))} \geq \frac{1}{4} \cdot n^{\frac{1}{2k}}$ if $k \geq 3$, and $\frac{wt(T_{k,n})}{wt(OPT(G_{k,n}))} \geq \frac{1}{4}\sqrt{n}$ if $k = 2$.

Define the *girth* of a graph as the number of edges in its smallest cycle, provided it is not a forest.

Lemma 3.6. Suppose there exists a connected unweighted graph $G_{k,n,m}$ with n vertices and m edges, having girth at least $2k$, in which every vertex has even degree. Then there is an m-vertex complete weighted graph G_1 (with positive edge weights satisfying the triangle inequality) and a k-optimal tour T of G_1 such that $\frac{wt(T)}{wt(OPT(G_1))} \geq \frac{m}{2n}$.

Proof: Assume we are given $G_{k,n,m} = G = (V, E)$. Since G is connected and every vertex has even degree, G has an Eulerian tour ET.

Using G and ET, we construct a complete weighted graph $G_1 = (V_1, E_1)$ and a tour T for G_1. Let $V(G) = \{x_1, x_2, \ldots, x_n\}$. We think of each vertex x_i in G as a "supervertex" corresponding to $deg_G(x_i)/2$ vertices in G_1, so

$$V_1 = \{x_{1,1}, \ldots, x_{1,deg_G(x_1)/2}, \ldots, x_{n,1}, \ldots, x_{n,deg_G(x_n)/2}\}.$$

The number of vertices in G_1 is $(deg_G(x_1) + \cdots + deg_G(x_n))/2 = m$.

Let $d_G(x_i, x_j)$ be the length of the shortest path from x_i to x_j in G. The edge weights of G_1 are as follows:

1. $\forall i, s, t, s \neq t,\ wt(x_{i,s}, x_{i,t}) = \epsilon$, where $\epsilon = \frac{1}{n^2}$.
2. $\forall i, j, s, t, i \neq j,\ wt(x_{i,s}, x_{j,t}) = d_G(x_i, x_j)$.

By inspection, it is easy to see that the edge weights of G_1 satisfy the triangle inequality.

The tour T on G_1 is constructed as follows. Suppose that the rth vertex of the Eulerian tour ET of G is vertex x_i. Suppose that this is the lth time ET has entered and exited vertex x_i, $1 \leq l \leq deg(x_i)/2$. Then the rth vertex of tour T of G_1 is $x_{i,l}$. Since ET enters and exits each vertex x_i of G exactly $deg_G(x_i)/2$ times and there are precisely $deg_G(x_i)/2$ vertices in each supervertex, this procedure gives us a tour T. Note that for all $\{x_i, x_j\} \in E$ there is a unique pair s, t such that $\{x_{i,s}, x_{j,t}\} \in T$.

Since the weight of the minimum spanning tree of G_1 is at most $(n-1) + (n \cdot n \cdot \epsilon) = n$, and edge weights satisfy the triangle inequality, $wt(OPT(G_1)) \leq 2n$. In the tour T, there are m edges each of weight 1, and so $wt(T) = m$. Hence we get $\frac{wt(T)}{wt(OPT(G_1))} \geq \frac{m}{2n}$, so all we need to prove Lemma 3.6 is

Claim 3.7. T is k-optimal.

Proof of claim 3.7: If not, then there is a tour T' of G_1 which is obtained from T by a k'-change operation, $k' \leq k$, such that $wt(T') < wt(T)$.

A *closed walk* is a walk which begins and ends at the same vertex, repeated edges and vertices allowed. A *simple closed walk* is a closed walk with no repeated edges.

Claim 3.8. Viewing T and T' as sets of edges, there are sets $C \subseteq T - T'$ and $C' \subseteq T' - T$ such that $C \cup C'$ is the edge set of a simple closed walk, $|C| = |C'| \leq k$, $wt(C) > wt(C')$, and every vertex in V_1 is incident to the same number (0, 1, or 2) of edges of C as C'.

Proof: Let Δ denote symmetric difference. Since for all $v \in V_1$, $deg_{T\Delta T'} v$ is either 0, 2, or 4, $T\Delta T'$ can be partitioned into a collection of vertex-disjoint simple closed walks $P_1, P_2, \ldots, P_s$. Further, since $wt(T) > wt(T')$, at least one of the P_i, say, P_1, has to satisfy $wt(P_1 \cap T) > wt(P_1 \cap T')$. It is easy to verify that $C = P_1 \cap T$ and $C' = P_1 \cap T'$ have the desired properties. ■

Let C, C' be as in Claim 3.8. Let $G_2 = (V, E_2)$ be a weighted multigraph with the following edges. Between every pair of vertices there will be one edge of positive integral weight and zero or one edge of weight -1. Specifically, between x_i and x_j, $i \neq j$, there is an edge in E_2 of weight $d_G(x_i, x_j)$, which is a positive integer. For that x_i and x_j, if there are s, t such that $\{x_{i,s}, x_{j,t}\} \in C$, then ($s$ and t are unique and) in addition to the edge of positive weight between x_i and x_j, there is an edge in E_2 between x_i and x_j of weight $-d_G(x_i, x_j)$. A crucial fact is that $-d_G(x_i, x_j) = -1$ in this case, because every edge in T is of weight 1. Let us denote the set of edges of positive weight in E_2 as D' and let us denote the set of edges of weight -1 as D. Each edge in C gives rise to exactly one edge in D, so $|D| = |C| \leq k$.

Note that there is an obvious correspondence between G_1 and G_2: the vertices inside a supervertex in G_1 are merged into a single vertex in G_2, with the intra-supervertex edges in G_1 "disappearing."

Edges from C', like arbitrary edges of G_1, are either of weight ϵ or of positive integral weight. An edge in C' of positive integral weight is an edge $\{x_{i,s}, x_{j,t}\}$ for some $i \neq j$ having weight $d_G(x_i, x_j)$ and is said to correspond to the edge in D' between x_i and x_j of weight $d_G(x_i, x_j)$. An edge in C' of weight ϵ is said to correspond to nothing. An edge in C is of unit length and is an edge $\{x_{i,s}, x_{j,t}\}$ such that $d_G(x_i, x_j) = 1$, i.e., $\{x_i, x_j\} \in E$. Such an edge is said to correspond to the edge in D (of weight -1) between x_i and x_j. (Several edges of C' may correspond to the same edge in D'. However, different edges in C correspond to different edges in D.) With this correspondence, the simple closed walk in G_1 which uses each edge in $C \cup C'$ exactly once corresponds to a closed walk P in G_2. (Edges of weight ϵ are not needed and do not appear.) P need not be simple since edges in D' may have several "preimages" in C'. Since each edge in $C \cup C'$ is traversed exactly once and different edges in C correspond to different edges in D, it follows that no edges in D are traversed twice. Each, in fact, is traversed exactly once by P. The weight of P is at most $wt(C') - wt(C) < 0$.

Let $G_3 = (V, E_3)$ be a weighted multigraph obtained by replacing each edge from G by two edges, one of weight $+1$ and one of weight -1. An edge of G_2 of positive integral weight is an edge between some x_i and x_j with $i \neq j$. Such an edge has weight $d_G(x_i, x_j)$ and is said to correspond to some fixed (shortest) path in G_3 between x_i and x_j consisting of $d_G(x_i, x_j)$ edges of weight $+1$. An edge of G_2 of weight -1 between, say, x_i and x_j, is said to correspond to the identical edge in G_3. (There will be many more negative edges in G_3 than there are in G_2, since C is small.) With this correspondence, the closed walk P in G_2 corresponds to a closed walk W in G_3 of the same weight. Edges of weight $+1$ may be traversed many times, but no edge of weight -1 can be traversed even twice, since no edge of weight -1 is traversed twice by P. Let the edges of weight $+1$ in W be edges $c_1, c_2, \ldots, c_r$ occurring in W $m_1, m_2, \ldots, m_r$ times, respectively. The number of edges of weight $+1$ in W, including multiplicities of course, equals $\sum_{j=1}^{r} m_j$. Since $wt(W) = wt(P) < 0$ and $wt(W) = (\sum_{j=1}^{r} m_j) - |D|$, we have $m_1 + m_2 + \cdots + m_r < |D|$. Also, the number of edges in W is $(\sum_{j=1}^{r} m_j) + |D| < 2|D| \leq 2k$.

One of the following must be true:

- For every edge in W of weight -1, there is another edge in W with the same endpoints. But since W never has two negative edges with the same endpoints, this other edge must have weight $+1$. We infer that $wt(W) \geq 0$, a contradiction.
- There is some edge in W of weight -1 such that there is no other edge in W with the same endpoints. But since W is a closed walk, this implies that there is some set of edges $S \subseteq W$ such that S is the edge set of a simple cycle. Since $S \subseteq W$ and W has fewer than $2k$ edges, S also has fewer than $2k$ edges. But then there is a cycle in G corresponding to S, and this cycle has fewer than $2k$ edges since S has fewer than $2k$ edges, which is a contradiction since the girth of G is at least $2k$. ■

Lemma 3.9. For all $k \geq 2$, for infinitely many n the graphs $G_{k,n,m}$ of Lemma 3.6 exist with $\frac{m}{2n} \geq \frac{m^{\frac{1}{2k}}}{4}$.

Proof: In order to prove that these graphs exist for infinitely many n, it suffices to show that for any n_0, there exists such a graph $G_{k,n,m}$ with $n > n_0$.

We first present an extremal graph-theoretic lemma from [5] [Theorem 1.4′, Chapter III].

Fact 3.10. Let q, δ, g be positive integers such that $q \geq \frac{(\delta-1)^{g-1}-1}{\delta-2}$. Then there exists a δ-regular graph having $2q$ vertices and girth at least g.

Let $p \geq n_0$ be a positive integer. Let $q = (2p)^{2k-1}$, $\delta = 2p$ and $g = 2k$. The parameters q, δ, g satisfy the hypothesis of Fact 3.10; let G' be the graph from Fact 3.10. G' has $2q$ vertices, girth at least $2k$, and is $(2p)$-regular. Let G be the largest connected component of G'. We claim that G has the desired properties.

Clearly G is connected, every vertex has even degree, and the girth is at least $2k$. Let $n = |V(G)|$. Since $p \geq n_0$ and G is $2p$-regular, we get $n > 2p > n_0$. Let $m = |E(G)|$. Since G is $2p$-regular, $m = pn \leq p(2q) = 2p(2p)^{2k-1} = (2p)^{2k}$, which implies that $\frac{m}{2n} = \frac{p}{2} \geq \frac{m^{\frac{1}{2k}}}{4}$. This completes the proof of Lemma 3.9. ■

Lemma 3.11. For infinitely many n the graphs $G_{2,n,m}$ of Lemma 3.6 exist with $\frac{m}{2n} = \frac{\sqrt{m}}{4}$.

Proof: We will prove the result for all values of n which are multiples of 4. Let $p = n/4$. Let $G = K_{2p,2p}$. G is connected, every vertex has even degree and G has no cycles of length 3. G has exactly $m = 4p^2$ edges so $m/n = p = \sqrt{m}/2$. ■

Theorem 3.5 now follows from Lemmas 3.6, 3.9 and 3.11. ■

4 Bounds on Performance Ratios for Geometric Graphs

In the previous section we found that the triangle inequality by itself ensures a $\Theta(\sqrt{n})$ worst-case performance ratio. Now we put stronger conditions on the distances, requiring them to be induced by a norm on $\mathbb{R}^m$, and show the worst-case performance ratio is between $c \log n / \log \log n$ and $O(\log n)$.

4.1 The Upper Bound

We find an upper bound on the performance ratio of any 2-optimal tour for geometric graphs, under any norm and in any dimension. A large portion of this subsection is based on concepts presented in [6].

We begin by stating a well-known property about norms and introducing a few definitions. Consider any positive integer $d \geq 2$ and any norm N on $\mathbb{R}^m$. Let $d_N(x, y)$ denote the distance between x and y in the metric generated by N. Let $d(x, y)$ denote the Euclidean distance between x and y. By the well-known comparability of norms [15, page 132], there exist $l_N, u_N > 0$ such that for every x and y,

$$l_N \cdot d_N(x, y) \leq d(x, y) \leq u_N \cdot d_N(x, y). \tag{1}$$

In this section we use the concept of angles. As usual, angles are defined by the inner product and the Euclidean metric. The angle between a and b in $\mathbb{R}^m$ is

$$\arccos \frac{a \cdot b}{||a||_2 ||b||_2},$$

which we take to be in the interval $[0, \pi]$.

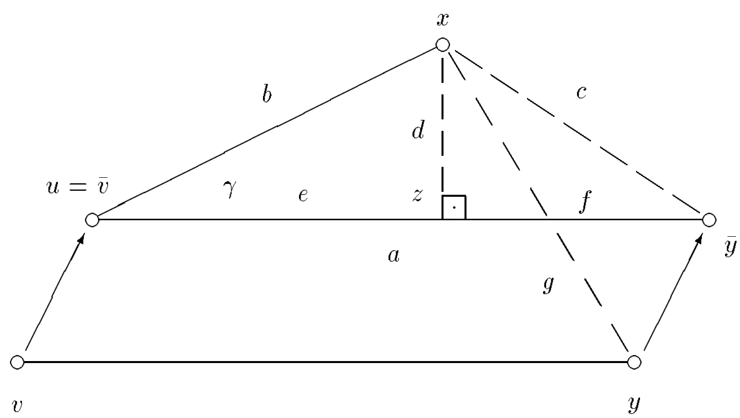

Figure 1: Illustration for Lemma 4.1.

Consider any norm N on $\mathbb{R}^m$. For l_N and u_N satisfying (1), let $\theta_N = \arctan(\frac{l_N}{4u_N})$. Define the angle between directed line segments $\overrightarrow{uv}$ and $\overrightarrow{wx}$ to be the angle between the vectors $v - u$ and $x - w$ ($\overrightarrow{ab}$ denotes a line segment directed from a to b). Two directed line segments $\overrightarrow{ux}$ and $\overrightarrow{vy}$ are said to be *similar-directional* (with respect to N, l_N and u_N) if the angle between them is at most θ_N. More intuitively, similar-directional means that the two directed line segments point in almost the same direction, since the angle θ_N is small. Note that for $\theta_N < \pi/2$, if $\overrightarrow{ux}$ and $\overrightarrow{vy}$ are similar-directional, then $\overrightarrow{ux}$ and $\overrightarrow{yv}$ are *not* similar-directional. N, l_N and u_N will be implicit when we write *similar-directional* instead of *similar-directional with respect to N, l_N and u_N*.

Let V be a finite nonempty set of points in $\mathbb{R}^m$ with norm N. Let G be the geometric graph induced by V. Let T' be a 2-optimal tour of G with (directed) edge set E' (2-optimality is with respect to distances on the metric induced by N).

Build a set $\boldsymbol{E}'$ of directed line segments in $\mathbb{R}^m$ corresponding to E' as follows. Suppose the tour $T' = (v_1, v_2, \ldots, v_n, v_1)$. Then $\boldsymbol{E}' = \{\overrightarrow{v_1v_2}, \overrightarrow{v_2v_3}, \ldots, \overrightarrow{v_{n-1}v_n}, \overrightarrow{v_nv_1}\}$. Every vertex is the tail of exactly one line segment in $\boldsymbol{E}'$ and the head of exactly one line segment in $\boldsymbol{E}'$.

We now present an important technical lemma. Intuitively, this lemma says that if there are two similar-directional directed line segments $\overrightarrow{uw}$ and $\overrightarrow{vx}$ in $\boldsymbol{E}'$, then u and v must be separated by a distance greater than half the length of the shorter segment. Consequently, the originating points of two similar-directional segments cannot be too close together.

Lemma 4.1. Let $G = (V, E)$ be a geometric graph in $\mathbb{R}^m$ under norm N. Let T' be a 2-optimal tour of G having (directed) edge set E'. Let $\overrightarrow{ux}$ and $\overrightarrow{vy}$ be any two similar-directional segments in $\boldsymbol{E}'$. If $d_N(u, x) \le d_N(v, y)$, then $d_N(u, v) > \frac{1}{2} \cdot d_N(u, x)$.

Proof. Let $l_N, u_N > 0$ be the constants defined in (1) and $\theta_N = \arctan \frac{l_N}{4u_N}$. Let γ be the angle between directed segments $\overrightarrow{ux}$ and $\overrightarrow{vy}$. To prove the lemma we assume that

$$d_N(u, x) \le d_N(v, y), \qquad \gamma \le \theta_N, \qquad \text{and} \qquad d_N(u, v) \le \frac{1}{2} \cdot d_N(u, x) \tag{2}$$

and we derive a contradiction.

Note that since $l_N \le u_N$, $\tan \theta_N = \frac{l_N}{4u_N} < 1$. So, if $\gamma \ge \pi/4$, $\gamma > \theta_N$. Thus, we may assume that

$\gamma < \pi/4$.

Consider the configuration obtained by translating $\overrightarrow{vy}$ in space such that v coincides with u. Let $\bar{v} = u$ be the translate of v, and let $\bar{y}$ be the translate of y. Points $u = \bar{v}, x, \bar{y}$ lie in a 2-dimensional plane. The situation is illustrated by Figures 1 and 2 .

Throughout this proof we use primed lower case letters, $a', b', c', \ldots$, to denote distances in the N metric, while unprimed lower case letters denote distances in the Euclidean metric. For example, if $a' = d_N(x, y)$, then $a = d(x, y)$, and vice-versa.

Let $a' = d_N(v, y)$, $b' = d_N(u, x)$, $c' = d_N(x, \bar{y})$, and $g' = d_N(x, y)$. (Recall that $a = d(v, y)$, $b = d(u, x)$, $c = d(x, \bar{y})$, and $g = d(x, y)$ are the corresponding Euclidean distances.) Using this notation, (2) implies that

$$b' \leq a' \qquad \text{and} \qquad d_N(u, v) \leq \frac{1}{2} b'. \tag{3}$$

Claim: $g' \geq a'$.

Proof. Suppose otherwise. We first see that $|\{u, x, v, y\}| = 4$. Clearly, $u \neq x$ and $v \neq y$. Clearly $u \neq v$, since otherwise u is the tail of two line segments in E', $\overrightarrow{ux}$ and $\overrightarrow{uy}$. Similarly, $x \neq y$. If $u = y$, then $d_N(u, v) = d_N(v, y)$ and $d_N(u, v) \leq \frac{1}{2} d_N(u, x) \leq \frac{1}{2} d_N(v, y)$, which is a contradiction. If $v = x$, then $d_N(u, v) = d_N(u, x)$ and $d_N(u, v) \leq \frac{1}{2} d_N(u, x)$, which is a contradiction. Hence, $|\{u, x, v, y\}| = 4$.

By assumption, $d_N(u, v) \leq \frac{1}{2} d_N(u, x) < d_N(u, x)$, and $g' < a'$, so $d_N(u, v) + d_N(x, y) < d_N(u, x) + d_N(v, y)$. Also, (u, v) is not in E', because if it was, either $\overrightarrow{uv} \in \boldsymbol{E}'$ or $\overrightarrow{vu} \in \boldsymbol{E}'$. But if $\overrightarrow{uv} \in \boldsymbol{E}'$ then the vertex u is the tail of two line segments in E', namely $\overrightarrow{ux}$ and $\overrightarrow{uv}$, and if $\overrightarrow{vu} \in \boldsymbol{E}'$ then the vertex v is the tail of two line segments in E', namely $\overrightarrow{vy}$ and $\overrightarrow{vu}$. Similarly, $(x, y) \notin E'$. But, now we can interchange two edges from the tour T', (u, x) and (v, y), with the two edges (v, u) and (y, x) which are not in the tour, to get a smaller valid tour, which contradicts the 2-optimality of T'. ∎

We now consider two cases: $a \geq b$ and $a < b$.

The case in which $a \geq b$ is illustrated in Figure 1, where z is the orthogonal projection of x onto the segment $\bar{v}\bar{y}$, $d' = d_N(x, z)$, $e' = d_N(\bar{v}, z)$, and $f' = d_N(z, \bar{y})$. Since $\gamma < \pi/4$ and $a \geq b$, z does belong to the segment $\bar{v}\bar{y}$.

The case in which $a < b$ is illustrated in Figure 2, where z is the orthogonal projection of $\bar{y}$ onto the segment ux, $d' = d_N(\bar{y}, z)$, $e' = d_N(u, z)$, and $f' = d_N(z, x)$. Since $\gamma < \pi/4$ and $a < b$, z does belong to the segment ux.

Case 1: $a \geq b$ (a and b are Euclidean distances). See Figure 1.

Using (3) and the triangle inequality several times, we obtain

$$g' \leq c' + d_N(y, \bar{y}) = c' + d_N(u, v) \leq c' + \frac{1}{2} b' \leq d' + f' + \frac{1}{2} b'$$

implying

$$d' \geq g' - \frac{1}{2} b' - f' \geq a' - f' - \frac{1}{2} b' = e' - \frac{1}{2} b' \geq e' - \frac{1}{2}(d' + e')$$

implying

$$d'(1 + \frac{1}{2}) \geq e'(1 - \frac{1}{2}).$$

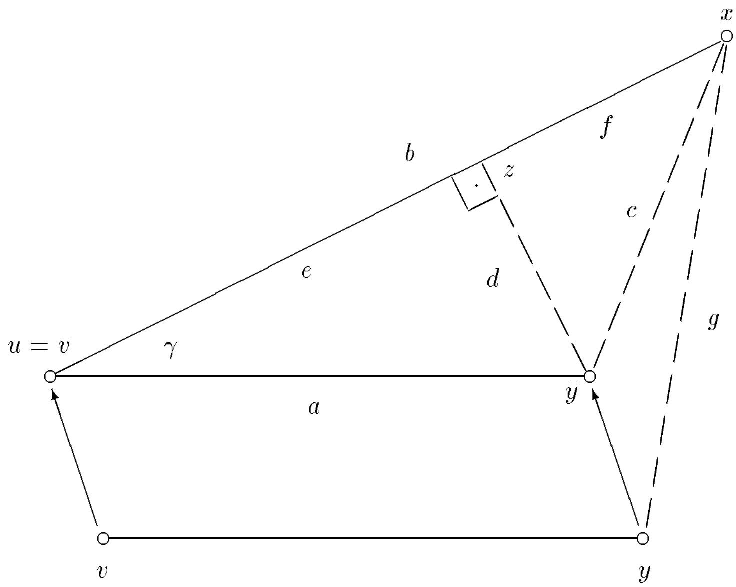

Figure 2: Illustration for Lemma 4.1.

Using (1) we have that

$$\frac{1-\frac{1}{2}}{1+\frac{1}{2}} \le \frac{d'}{e'} \le \frac{u_N d}{l_N e} = \frac{u_N}{l_N}\tan\gamma,$$

implying

$$\tan\gamma \ge \frac{l_N(1-\frac{1}{2})}{u_N(1+\frac{1}{2})} = \frac{l_N}{3u_N} > \frac{l_N}{4u_N} = \tan\theta_N.$$

Since $\tan\gamma > \tan\theta_N$, we have that $\gamma > \theta_N$, a contradiction.

Case 2: $a < b$ (a and b are Euclidean distances). See Figure 2.

Using (3) and the triangle inequality several times, we obtain

$$g' \le c' + d_N(y,\bar{y}) = c' + d_N(u,v) \le c' + \frac{1}{2}b' \le d' + f' + \frac{1}{2}b'$$

implying

$$d' \ge g' - \frac{1}{2}b' - f' \ge b' - f' - \frac{1}{2}b' = e' - \frac{1}{2}b' \ge e' - \frac{1}{2}a' \ge e' - \frac{1}{2}(d'+e')$$

(the second inequality follows from $g' \ge a' \ge b'$) implying

$$d'(1+\frac{1}{2}) \ge e'(1-\frac{1}{2}).$$

As in Case 1, we obtain $\gamma > \theta_N$, a contradiction.

This completes the proof of Lemma 4.1. ∎

We now analyze the weight of the tour T'. In $\mathbb{R}^m$, for any angle $\alpha > 0$, consider a cover of $\mathbb{R}^m$ by some finite number $B(d,\alpha)$ of circular (overlapping) cones, all having the same origin P, such that two distinct points different from P in the same cone form, at P, an angle at most α. We use in Theorem 4.2 the well-known fact that $B(d,\alpha)$ is finite for every $\alpha > 0$ and every d. This *covering problem* has been extensively studied.

We mention the following upper bound due to Rogers [19]:

$$B(m, \alpha) \text{ is } O\left(m^{3/2}\left(\log \frac{m}{\sin(\alpha/2)}\right)\left(\frac{1}{\sin(\alpha/2)}\right)^m\right).$$

Theorem 4.2. Fix m and a norm N on $\mathbb{R}^m$. Let $G = (V, E)$ be an n-vertex geometric graph in $\mathbb{R}^m$ under norm N. Let OPT be the weight of the optimal tour on G. Let T' be any 2-optimal tour of G. Then the weight of T' is $O(\log n) \cdot OPT$. (The constant implicit in the big O depends on m and N.)

Proof. Let $\theta_N = \arctan(\frac{l_N}{4u_N})$, where the constants $l_N, u_N > 0$ are according to (1). At some arbitrary point P in $\mathbb{R}^m$, we cover the space by a constant number of circular cones $C_1, C_2, \ldots, C_{B(m,\theta_N)}$, such that every two line segments containing P and lying within the same cone subtend, at P, an angle at most θ_N. As noted, $B(m, \theta_N)$ depends only on d and N.

Call these the *original cones*. Construct $B(m, \theta_N)$ congruent cones around each of the n vertices of G by translating each original cone so that its origin shifts from P to that vertex. Hence, corresponding to each vertex of G are $B(m, \theta_N)$ cones, one cone corresponding to each of the original cones $C_1, C_2, \ldots, C_{B(m,\theta_N)}$. Let C_{ji} be the cone with its origin at vertex j that is a translate of original cone C_i.

Let E' be the edge set of the 2-optimal tour T'. Let $\boldsymbol{E}'$ be the set of directed line segments corresponding to E'. Let $\boldsymbol{E}'_i$ be the set of directed line segments in $\boldsymbol{E}'$ that appear in $\cup_j C_{ji}$. We claim that the sum of the weights of the segments in $\boldsymbol{E}'_i$ is bounded by $O(\log n) \cdot wt(OPT)$, for $1 \leq i \leq B(m, \theta_N)$. Since the sets $\boldsymbol{E}'_i$ cover the set $\boldsymbol{E}'$ and the number of cones is a constant, proving this claim is enough to prove the lemma.

Clearly all the directed line segments in $\boldsymbol{E}'_i$ are similar-directional. Hence by Lemma 4.1, if $\overrightarrow{u_1v_1}$ and $\overrightarrow{u_2v_2}$ are two directed line segments in $\boldsymbol{E}'_i$, and if the former one is shorter, then $d_N(u_1, u_2) > \frac{1}{2} d_N(u_1, v_1)$.

Let T be an an optimal tour on G, so $wt(T) = OPT$. We are now going to account for the length of the line segments in $\boldsymbol{E}'_i$ using the length of the edges in the T. Consider a walk along the the edges of T starting from an arbitrary vertex, and ending at the same vertex. As we walk along the path we encounter the originating points of the line segments in $\boldsymbol{E}'_i$. Let the order of the line segments encountered from $\boldsymbol{E}'_i$ be $\overrightarrow{e_1}, \ldots, \overrightarrow{e_q}$.

We claim that there exist $\lfloor q/2 \rfloor$ line segments in $\boldsymbol{E}'_i$ with total length at most $2 \cdot OPT$. Consider a pair of consecutive directed line segments $\overrightarrow{e_j}, \overrightarrow{e_{j+1}}$. By Lemma 4.1, the distance between the originating points of $\overrightarrow{e_j}$ and $\overrightarrow{e_{j+1}}$ is longer than $\frac{1}{2}$ times the length of the shorter segment. Hence the distance along T between the originating points of $\overrightarrow{e_j}$ and $\overrightarrow{e_{j+1}}$ is also longer than $\frac{1}{2}$ times the length of the shorter segment. We may charge the length of the shorter line segment to the length of the section of T between the originating points of $\overrightarrow{e_j}$ and $\overrightarrow{e_{j+1}}$. The charge is at most two times the length of the path. Taking the $\lfloor q/2 \rfloor$ disjoint consecutive pair of segments, $\{\overrightarrow{e_1}, \overrightarrow{e_2}\}, \{\overrightarrow{e_3}, \overrightarrow{e_4}\}, \ldots$, and charging the shorter line segment of each pair to the corresponding path in T between the originating points, each section of the path T is charged at most once. Thus, the total length of the shorter line segments of the chosen pairs is at most $2 \cdot OPT$.

Now we consider only edges in $\boldsymbol{E}'_i$ that have not been charged yet, and repeat the same process. After $O(\log n)$ steps, each edge in $\boldsymbol{E}'_i$ is charged, giving the bound of $O((\log n) \cdot OPT)$ for the total weight of the edges in $\boldsymbol{E}'_i$.

Since $B(m, \theta_N)$ is constant (dependent only on m and N), we conclude that $wt(\boldsymbol{E}')$ is $O((\log n) \cdot OPT)$, and therefore $O(\log n) \cdot OPT$. ■

4.2 A Lower Bound for 2-Opt Under L_2

Theorem 4.3. There exists a constant $c > 0$ such that for infinitely many values of n, there exists a n-node graph G_n embedded in the Euclidean plane under the L_2 metric and a 2-optimal tour T_n of G_n such that $\frac{wt(T_n)}{wt(OPT(G_n))} \geq c \cdot \frac{\log n}{\log\log n}$.

We will prove the result for those values of n which satisfy $n = 2(1+p^2+p^4+p^6+\cdots+p^{2p})+2p+p^{2p}+1$ for any positive odd integer $p \geq 3$. Note that $p > c' \frac{\log n}{\log\log n}$, for some $c' > 0$.

We exhibit a set of n vertices V (all lying on the $n \times n$ grid in the Euclidean plane) such that $wt(OPT(V))$ is at most $18 \cdot p^{2p}$. We then construct a 2-optimal tour T on V of weight at least $2p \cdot p^{2p} \geq 2c' \frac{\log n}{\log\log n} \cdot p^{2p}$. Hence, we will get that $\frac{wt(T)}{wt(OPT(V))} \geq \frac{c'}{9} \cdot \frac{\log n}{\log\log n}$.

Our construction is a modification of a construction due to Bentley and Saxe [3] and Alon and Azar [1] and is shown in Figure 3. We construct V in three parts, V_1, V_2 and V_3. The vertices in V_1 are in $p+1$ layers, where each layer is a set of equally spaced points on a horizontal line of length p^{2p}. The coordinates of the points in level i, $0 \leq i \leq p$, are (ja_i, b_i) where $a_i = p^{2p-2i}$ and $0 \leq j \leq p^{2p}/a_i$, and b_i will be defined later. Thus $a_0 = p^{2p}$, $a_1 = p^{2p-2}$, ..., and $a_p = 1$. Hence in layer 0 there are only two points, in layer 1 there are p^2+1, in layer i there are $p^{2i}+1 = \frac{p^{2p}}{a_i}+1$ points, up to layer p which contains $p^{2p}+1$ points. Let $b_0 = 0$. The vertical distance between layer i and layer $i+1$ (i.e., $b_{i+1} - b_i$) is $c_i = p^{2p-1-2i}$, for all i. Note that $p \cdot a_{i+1} = p \cdot p^{2p-2i-2} = c_i = \frac{p^{2p-2i}}{p} = \frac{a_i}{p}$.

V_2 is a copy of V_1 shifted to the right. For every vertex in V_1 with coordinates (e, f), there is a vertex in V_2 with coordinates $(e + 2p^{2p}, f)$. These are the only vertices in V_2.

Finally, we fill in the gaps in the topmost layer to get V_3. Since $a_p = 1$, let $V_3 = \{(j, b_p) | p^{2p} < j < 2 \cdot p^{2p}\}$. The set of all the vertices is $V = V_1 \cup V_2 \cup V_3$. Note that $|V| = n$.

Claim 4.4. $wt(OPT(V)) \leq 18p^{2p}$.

Proof: Since $wt(OPT(V))$ is no more than twice the weight of the optimal spanning tree, it suffices to show that there is a spanning tree of weight at most $9p^{2p}$. Consider the spanning tree built as follows: for every point in every layer, other than the bottom layer, draw a vertical line to the point directly above it in the next higher layer. Also draw the horizontal line in the topmost layer (layer number p). The total length of this tree is at most

$$3p^{2p} + 2\sum_{i=0}^{p-1} c_i(\frac{p^{2p}}{a_i} + 1) \leq 3(p^{2p}(1 + \sum_{i=0}^{p-1} 2c_i/a_i)) = 3(p^{2p}(1 + p\frac{2}{p})) = 9p^{2p}. ■$$

Define the tour T on V to be as shown in the Figure 3. Note that since $p+1$ is even, we can always construct this tour.

Claim 4.5. $wt(T(V)) > 2p \cdot p^{2p}$.

Proof: Consider just the horizontal edges in T. Each layer has horizontal edges whose combined weight is at least $2p^{2p}$ and there are $p+1$ layers. ■

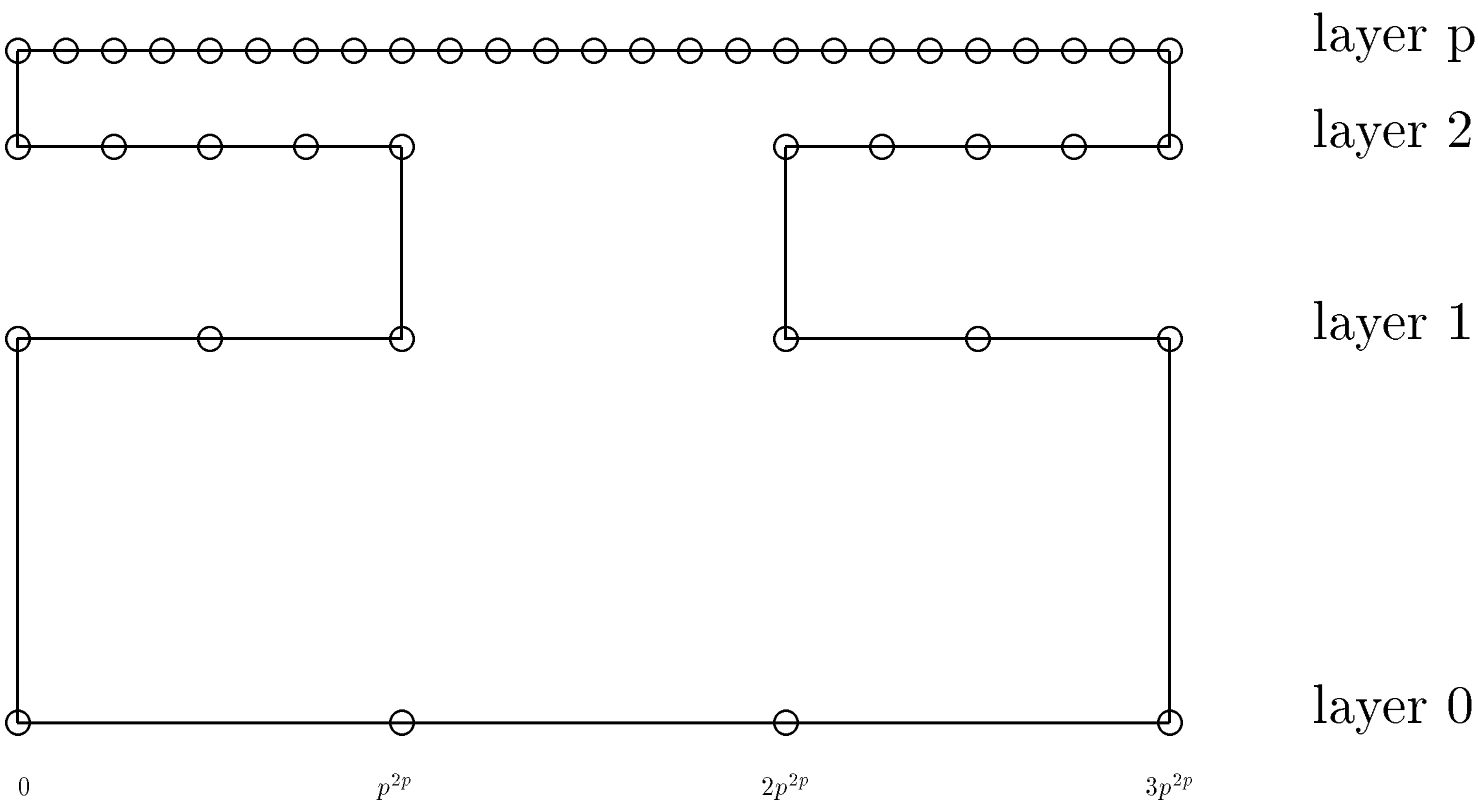

Figure 3: The tour T

Claim 4.6. T is 2-optimal.

We first present some simple notation. For any point A, let A_x, A_y be its x and y coordinates. We use AB to refer both to the edge (line segment) and its length. $(AB)_x$ is the length of the projection of AB onto the x-axis, i.e., $(AB)_x = |A_x - B_x|$. We define $(AB)_y$ similarly. Note that $AB \geq (AB)_x, (AB)_y$. We say that two edges AB and CD, which are either both vertical or both horizontal, *overlap* if the following holds: let the projection of AB onto the infinite line containing CD be $A'B'$. Then $A'B' \cap CD$ consists of more than a single point.

We next state and prove a simple geometric lemma.

Lemma 4.7. Let EF and GH be horizontal line segments in the Euclidean plane, $G_x \leq E_x < F_x \leq H_x$. Let $EF = 1$, $GH = q^2$, $q \geq 1$, so $G_x \leq E_x \leq H_x - 1$. Let the vertical distance between EF and GH be z. If $z \geq q$, then $\min\{EG + FH, EH + FG\} \geq EF + GH$.

Proof: Let $z \geq q$. Clearly, $EH + FG > EG + FH$ so all we need to prove is $EG + FH \geq EF + GH$. For $0 \leq a \leq q^2 - 1$, define $f(a) = \sqrt{a^2 + q^2} + \sqrt{(q^2 - 1 - a)^2 + q^2}$. Let $a = E_x - G_x$. Then $H_x - F_x = (q^2 - 1) - a$. $EG + FH = \sqrt{a^2 + z^2} + \sqrt{(q^2 - 1 - a)^2 + z^2} \geq \sqrt{a^2 + q^2} + \sqrt{(q^2 - 1 - a)^2 + q^2} = f(a)$. Since $EF + GH = q^2 + 1$, in order to show that $EG + FH \geq EF + GH$, it suffices to show that for $0 \leq a \leq q^2 - 1$, $f(a) \geq q^2 + 1$.

We will show that the minimum value of $f(a)$ in the interval $[0, q^2 - 1]$ occurs at $a = (q^2 - 1)/2$.

This suffices since $f(\frac{q^2-1}{2}) = q^2+1$.

$$
\begin{aligned}
f'(a) &= \frac{a}{\sqrt{a^2+q^2}} - \frac{(q^2-1-a)}{\sqrt{(q^2-1-a)^2+q^2}} \\
&= \frac{1}{\sqrt{1+(\frac{q}{a})^2}} - \frac{1}{\sqrt{1+(\frac{q}{q^2-1-a})^2}}.
\end{aligned}
$$

In the interval $[0, \frac{q^2-1}{2})$, $a < q^2-1-a$, and hence $f'(a) < 0$. In the interval $(\frac{q^2-1}{2}, q^2-1]$, $a > q^2-1-a$, and hence $f'(a) > 0$. Hence the minimum value of $f(a)$ in the interval $[0, q^2-1]$ occurs at $a = (q^2-1)/2$. ■

Proof of Claim 4.6 :

Suppose otherwise. So, in a single 2-change operation, from T we can get another tour T' such that $wt(T') < wt(T)$. Label the four vertices involved as A, B, C, D so that $E(T) - E(T') = \{AB, CD\}$, $E(T') - E(T) = \{AC, BD\}$ and $AC + BD < AB + CD$. Note that the vertices A, B, C, D have to be distinct, since in a single 2-change operation we cannot replace two edges out of one vertex.

Since all edges in $E(T)$ are either horizontal or vertical, there are three cases:

Case 1: AB and CD are both vertical edges. If AB and CD overlap, then they are the two vertical edges which face each other. But then $(AC)_x + (BD)_x \geq 2p^{2p} \geq AB + CD$. If AB and CD don't overlap, assume without loss of generality that $A_y, B_y \leq C_y < D_y$. If $A_y < B_y$ then $(AC)_y \geq AB$ and $(BD)_y \geq CD$. If $A_y > B_y$ then $(BD)_y \geq AB + CD$.

Case 2: One of AB or CD is horizontal and the other is vertical. Assume without loss of generality that AB is horizontal and CD is a vertical. By construction, and since A, B, C, D are all distinct, exactly one of the following has to be true:

Subcase (i): Either $C_y = A_y = B_y$ or $D_y = A_y = B_y$. Then, by construction, $A_x, B_x < C_x = D_x$ or $A_x, B_x > C_x = D_x$. If $C_y = A_y$ then $(BD)_y = CD$ and, by construction, $(AC)_x \geq AB$. Similarly, if $D_y = B_y$ then $(AC)_y = CD$ and $(BD)_x \geq AB$.

Subcase (ii): $C_y, D_y > A_y = B_y$. Since AB is horizontal ($A_x \neq B_x$) and CD is vertical ($C_x = D_x$) either $C_x \neq A_x$ or $D_x \neq B_X$. If $C_x \neq A_x$ then, by construction, $(AC)_x \geq AB$ and, by construction, $(BD)_y > CD$. Similarly, if $D_x \neq B_x$ then $(BD)_x \geq AB$ and $(AC)_y > CD$.

Subcase (iii): $C_y, D_y < A_y = B_y$. If $C_y < D_y$ then $(AC)_y = (AD)_y + CD$, and, by construction, $(AD)_y > AB$, implying $(AC)_y > AB + CD$. Similarly, if $D_y < C_y$ then $(BD)_y > AB + CD$.

Case 3: AB and CD are both horizontal edges.

Subcase (i): AB and CD are non-overlapping. Assume without loss of generality that $A_x, B_x \leq C_x < D_x$. If $A_x < B_x$ then $(AC)_x \geq AB$ and $(BD)_x \geq CD$. If $B_x < A_x$ then $(BD)_x \geq AB + CD$.

Subcase (ii): AB and CD are overlapping. Assume without loss of generality that AB is the smaller, higher edge and that $C_x \leq D_x$ so $C_x \leq A_x, B_x \leq D_x$. Suppose AB is l levels above CD, $l \geq 1$. Then $CD = p^{2l} \cdot AB$. The difference in height between them is $AB \cdot (p + p^3 + \cdots + p^{2l-1}) \geq AB \cdot p^l$. Scaling all three quantities so that $AB = 1$, we see that the hypotheses of Lemma 4.7 are satisfied, and hence $AC + BD \geq AB + CD$. ■

5 Bounds on the Length of 2-Optimal Tours in the Unit Hypercube

In this section we show that for every m and every norm on $\mathbb{R}^m$ there is a $O(n^{1-1/m})$ upper bound on the length of any 2-optimal tour on n points in the m-dimensional unit hypercube. (The constant implicit in the big O depends on m and the norm.)

Notation: an arc is an ordered pair (h, t). The distance between h and t is denoted $d(h, t)$, and the (directed) line segment between them is denoted $\overline{ht}$. The *orientation* of an arc (h, t) is the (Euclidean) unit-length vector $(h-t)/||h-t||_2$. The difference between two orientations r and s is the angle between them as defined in section 4, i.e. $\arccos(r \cdot s)$. Thus the orientations of (h, t) and (t, h) differ by π.

Initially we work with the Euclidean norm.

Theorem 5.1. For any dimension $m \geq 2$ there exists a constant c_m such that all Euclidean 2-optimal tours on all configurations of n points in the unit hypercube have Euclidean length less than $c_m n^{1-\frac{1}{m}}$.

For any $\epsilon > 0$ define the *long arcs* in T as those of length at least ϵ. For each long arc (h, t), define the *heart* of the arc as the interior of the hypercylinder of radius $\epsilon/8$ and length $d(h, t)/2$, with the height oriented parallel to ht, and with the center of the hypercylinder at the midpoint of segment $\overline{ht}$. In the 2-D case the heart is a rectangle of width $\epsilon/4$.

Formally, suppose without loss of generality (h, t) is a vertical arc of length k with endpoints $t = (0, 0, \ldots, 0)$ and $h = (0, 0, \ldots, 0, k)$. Let $B = \{x \in \mathbb{R}^m |\ ||x|| < \epsilon/8; x_m = 0\}$, so B is the interior of a $m-1$-dimensional ball of radius $\epsilon/8$ in the subspace $x_m = 0$ of $\mathbb{R}^m$. Let M be the interior of the middle half of the segment $\overline{ht}$, $M = \{(0, 0, \ldots, 0, \lambda) | k/4 < \lambda < 3k/4\}$. Then the *heart* $H((h, t))$ is the set sum $B + M$.

We say that arc (h_1, t_1) *attacks* arc (h_2, t_2) iff the line segment $h_1 t_1$ intersects the heart of (h_2, t_2). Note: attacking is not a symmetric relation.

At times it will be convenient to refer to the heart of a segment, or to say that a line segment attacks another, even if the segment's endpoints are not tour points. The intended meaning is obvious.

Let $\theta = \arccos\ (31/32)$. A *family* of arcs is any collection of long arcs in $\mathbb{R}^m$ whose orientations differ pairwise by at most θ.

Lemma 5.2. No arc attacks another arc in the same family.

Proof. We prove the lemma in three steps. First, it suffices to consider the case where arcs have length exactly ϵ. Second, if two arcs are parallel and one attacks the other, then they violate 2-optimality by more than $\epsilon/4$. Third, if two arcs are oriented within θ and one attacks the other, then they can be made parallel while still attacking, without changing things by more than $\epsilon/4$.

Step 1 begins with a simple geometric definition.

Definition 1. If the line segment $\overline{HT}$ contains the line segment $\overline{ht}$, and their orientations are consistent (so $d(H, h) \leq d(H, t)$), then we say $\overline{HT}$ is an *extension* of $\overline{ht}$.

For any four points h_1, t_1, h_2, t_2, define the function

$$G(h^1, t^1, h^2, t^2) = d(h^1, t^1) + d(h^2, t^2) - d(h^1, h^2) - d(t^1, t^2).$$

The function G measures the decrease in tour length if arcs (h^1, t^1) and (h^2, t^2) were removed in a 2-change operation. The two arcs cannot both be in a 2-optimal tour if G is strictly positive. (If G is positive, we can swap out arcs (h^1, t^1), (h^2, t^2) and swap in arcs (h^1, h^2), (t^1, t^2). The tour remains connected, because the original tour included h^1 just before t^1, and h^2 just before t^2, in that order.) The following sub-lemma implies that if a pair of arcs has positive G value, then so does any pair of extensions of these arcs.

Lemma 5.3. Suppose $\overline{H^iT^i}$ is an extension of $\overline{h^it^i}$ for $i = 1, 2$. Then $G(H^1, T^1, H^2, T^2) \geq G(h^1, t^1, h^2, t^2)$. The result holds in $\mathbb{R}^m$ with respect to any norm.

Proof. Observe how G changes as the shorter segments are stretched by extending the endpoints in turn. Because H^1, h^1, and t^1 are collinear and any norm scales, $d(H^1, t^1) = d(H^1, h^1) + d(h^1, t^1)$. By the triangle inequality $d(H^1, h^2) - d(h^1, h^2) \leq d(H^1, h^1) = d(H^1, t^1) - d(h^1, t^1)$. Thus extending h^1 to H^1 cannot decrease G. By a symmetric argument the other components of G are nondecreasing as the segments are extended and the claim follows. ∎

Suppose long arc (H^1, T^1) attacks long arc (H^2, T^2). Obviously the segment $\overline{H^1T^1}$ is an extension of some segment $\overline{h^1t^1}$ that has length ϵ and that also attacks (H^2, T^2). Now consider all segments that are of length ϵ and can be extended to $\overline{H^2T^2}$. The union of the hearts of these segments is a hypercylinder of radius $\epsilon/8$ and length $d(H^2, T^2) - \epsilon/2 \geq d(H^2, T^2)/2$, with the same center as the heart of (H^2, T^2); so it contains the heart of (H^2, T^2). Therefore (at least) one of these segments is attacked by $\overline{h^1t^1}$. Denote the attacked segment by $\overline{h^2t^2}$.

By lemma 5.3, if $G(h^1, t^1, h^2, t^2) > 0$ then $G(H^1, T^1, H^2, T^2) > 0$. To prove our lemma it therefore suffices to consider the case $d(h^1, t^1) = d(h^2, t^2) = \epsilon$. This completes step one of the proof.

For step two, suppose a segment attacks a parallel segment, and that both segments have length ϵ. Our aim is to show G exceeds $\epsilon/4$.

Consider the 2-dimensional (affine) subspace (plane) spanned by the four endpoints of the two parallel segments. Now the intersections of the hypercylindrical hearts of the segments with this plane are precisely the 2-dimensional rectangular hearts of segments in the 2-D case. Therefore, for the remainder of step 2 we work in 2-D.

Also without loss of generality (but assuming Euclidean distances) take the second (attacked) segment to be vertical, with x and y coordinates $h^2_x = t^2_x = t^2_y = 0$, and $h^2_y = \epsilon$. We can further take the x and y coordinates of the first (attacking) segment's endpoints as all nonnegative. Geometrically we are placing the attacking segment above and to the right of the other. The attacking segment is parallel, hence vertical, so $h^1_y \geq \epsilon$. Since the first segment attacks the second, we must have $h^1_x = t^1_x < \epsilon/8$, and also $t^1_y < 3\epsilon/4$, whence $h^1_y < 7\epsilon/4$ (since the first segment has length ϵ). (The inequalities are strict because the heart is the interior of the rectangle.) Now

$$d(t^1, t^2) \leq |t^1_y - t^2_y| + |t^1_x - t^2_x| = t^1_y + t^1_x < t^1_y + \epsilon/8.$$

Similarly $d(h^1, h^2) < h^1_y - h^2_y + \epsilon/8 = h^1_y - 7\epsilon/8$. Thus

$$G(h^1, t^1, h^2, t^2) > 2\epsilon - t^1_y - \epsilon/8 - h^1_y + 7\epsilon/8 > \epsilon(2 - 3/4 - 1/8 - 7/4 + 7/8) = \epsilon/4.$$

This completes the second step of the proof of the lemma.

For the third step, suppose two segments of length ϵ have orientation differing by at most θ. Without loss of generality assume that the first attacks the second.

Let $p \in \overline{h^1t^1} \cap H((h^2,t^2))$ be a point of intersection of the first arc and the second arc's heart. Hold $\overline{h^1t^1}$ tacked at p and rotate it to be parallel with (h^2,t^2). Call the new segment $\overline{\tilde{h}^1,\tilde{t}^1}$.

The new segment attacks the second arc because of p. Thus the new segment and the second segment meet the conditions of part two. Therefore,

$$G(\tilde{h}^1,\tilde{t}^1,h^2,t^2) > \epsilon/4.$$

We next claim that when the attacking arc is rotated, the value of G does not change by more than $\epsilon/4$, i.e.,

$$|G(h^1,t^1,h^2,t^2) - G(\tilde{h}^1,\tilde{t}^1,h^2,t^2)| \leq \epsilon/4.$$

Observe that $d(h^1,t^1) = d(\tilde{h}^1,\tilde{t}^1)$ by construction. Therefore the value of G changes for only two reasons: $d(t^1,t^2)$ is replaced by $d(\tilde{t}^1,t^2)$; and $d(h^1,h^2)$ is replaced by $d(\tilde{h}^1,h^2)$. Recall that we set the value of θ so $\cos(\theta) = 31/32$. The idea is that for small rotations $\tilde{t}^1$ cannot be too far from t^1.

By the law of cosines, at any angle $\zeta \leq \theta$,

$$d(t^1,\tilde{t}^1)^2 = d(t^1,p)^2 + d(\tilde{t}^1,p)^2 - 2d(t^1,p)d(\tilde{t}^1,p)\cos(\zeta)$$

$$= 2d(t^1,p)^2(1-\cos(\zeta)) \leq 2d(t^1,p)^2(1-\cos(\theta)) = 2d(t^1,p)^2(1-\frac{31}{32}) = d(t^1,p)^2/16.$$

Therefore $d(t^1,\tilde{t}^1) \leq d(t^1,p)/4$.

By properties of similar triangles, we also have $d(h^1,\tilde{h}^1) \leq d(h^1,p)/4$. Putting all this together with the triangle inequality, we find that rotation changes G by at most $d(p,t^1)/4 + d(p,h^1)/4 = \epsilon/4$. Therefore

$$G(h^1,t^1,h^2,t^2) > 0.$$

This completes the third and final step of the proof of lemma 5.2. ■

Definition 2. The *soul* of an arc is the hypercylinder defined exactly as the heart of the arc but with radius and length half that of the heart.

The soul therefore has radius $\epsilon/16$.

Lemma 5.4. If the souls of two long arcs intersect, then they attack each other.

Proof. Let p be a point of intersection of the souls of (h^1,t^1) and (h^2,t^2). Let p^i denote the nearest point on segment $\overline{h^it^i}$ to p. Then p^i is in the soul of (h^i,t^i) and $d(p,p^i) < \epsilon/16$, for $i = 1,2$. By the triangle inequality we have $d(p^1,p^2) < \epsilon/8$.

On the other hand, the heart of (h^1,t^1) has radius $\epsilon/8$, and it extends $d(h^1,t^1)/8 \geq \epsilon/8$ beyond the soul along the arc in both directions as well. Therefore, for any point of the arc that is in the soul, its open ball of radius $\epsilon/8$ (the set of points at distance less than $\epsilon/8$) is completely contained in the heart.

Taking p^1 as this point, it follows that p^2 is in the heart of (h^1,t^1). Therefore (h^2,t^2) attacks (h^1,t^1). By symmetry, (h^1,t^1) attacks (h^2,t^2). ■

Fix the dimension m. Any soul is contained in the slightly-larger-than-unit hypercube of side length $1+2\epsilon/16$ and volume at most $k_1 = (9/8)^m$ (if $\epsilon \leq 1$). By Lemmas 5.2 and 5.4 the sum of the volumes of the souls in a family F is at most k_1.

Now, the volume of the soul of arc (h,t) is $k_2\epsilon^{m-1}d(h,t)$ for some constant k_2. So

$$k_2\epsilon^{m-1}\sum_{(h,t)\in F} d(h,t) \le k_1 \Longrightarrow \sum_{(h,t)\in F} d(h,t) \le k_1\epsilon^{1-m}/k_2.$$

This bounds the sum of the lengths of long arcs in a single family.

For every possible orientation $||u||_2 = 1$ define a corresponding set $F_u = \{v | v\cdot u > \cos(\theta/2)\}$. Notice that for any T, the set F_u induces a family of arcs of T. The set of all F_u is an open cover of the compact unit sphere $\{u|\, ||u||_2 = 1\}$. Extract a finite subcover of cardinality k_3. Note that k_3 is independent of T and n.

For all T the subcover provides a finite collection of families whose union is the set of all long arcs in the tour T. Therefore, the sum of the lengths of all long arcs in T is bounded by $k_3[k_1\epsilon^{1-m}/k_2]$.

The total length of all short arcs is obviously bounded by $n\epsilon$. Choose $\epsilon = n^{-\frac{1}{m}}$. The sum of lengths of all arcs in T is less than $n\epsilon + \frac{k_1k_3}{k_2}\epsilon^{1-m} = n^{1-\frac{1}{m}} + \frac{k_1k_3}{k_2}n^{-\frac{1-m}{m}} = c_m n^{1-\frac{1}{m}}$ and Theorem 5.1 is proved. ■

For the 2-D case we can use a region larger than the soul, and change a few other details to get an explicit bound on tour length.

Corollary 5.5. In the two-dimensional Euclidean case, the length of any 2-optimal tour is less than $8\sqrt{26}\sqrt{n} + 234$.

Proof.

Define the *left heart* of a large arc as that half of the arc's heart which lies strictly to the left the arc, when walking from tail to head.

Lemma 5.6. The left hearts of arcs in a family are all mutually disjoint.

Proof. Suppose two long arcs (h^1,t^1) and (h^2,t^2) have intersecting left hearts. Suppose further that the two arcs' orientations differ by η where $|\eta| \le \pi/3 = \arccos 1/2$. We show that one of the arcs attacks the other, and the result follows from Lemma 5.2.

Let q be a point of intersection of the left hearts. Drop a perpendicular from q to segment $\overline{h^it^i}$ at intersection point $q^i, i = 1,2$. Consider without loss of generality the case $d(q,q^1) \le d(q,q^2)$. We prove in this case arc (h^1,t^1) attacks (h^2,t^2).

Extend the line segment $\overline{qq^2}$ to point q^{22} so that q^2 is the midpoint of the other points: $q^2 = (q+q^{22})/2$. Observe that the entire segment $\overline{qq^{22}}$ is within the heart (not necessarily the left heart) of (h^2,t^2). Therefore, all we have to do to prove (h^2,t^2) is attacked, is to verify that the segment $\overline{h^1t^1}$ intersects this segment $\overline{qq^{22}}$.

Let p denote the point of intersection of the (infinite) lines h^1t^1 and qq^{22}. The intersection p is sure to exist because $\eta < \pi/2$. First we show that p is "between" q and q^{22}, or simply that $d(q,p) \le d(q,q^{22})$. Since $\eta \le \pi/3$, we have

$$\frac{d(q,q^1)}{d(q,p)} = \cos\eta \ge \cos\pi/3 = 1/2.$$

Hence $d(q,p) \le 2d(q,q^1) \le 2d(q,q^2) = d(q,q^{22})$ as desired.

Second we show that p is between h^1 and t^1. Now,

$$\frac{d(q^1, p)}{d(q, q^1)} = \tan \eta \leq \tan \pi/3 = \sqrt{3}.$$

Also, q is in the left heart of (h^1, t^1) and q^1 lies on that arc. Therefore $d(q, q^1) \leq \epsilon/8$. Finally, recall that the arc (h^1, t^1) is long, whence $d(h^1, t^1) \geq \epsilon$. Putting these inequalities together, we find that q^1 and p are near each other:

$$d(q^1, p) \leq \sqrt{3} d(q, q^1) \leq \epsilon\sqrt{3}/8 < \epsilon/4 \leq d(h^1, t^1)/4.$$

Since q^1 is in inner half of the arc, the above implies that p lies in the arc (h^1, t^1). So p is the desired point of intersection of the two segments, (h^2, t^2) is attacked, and the lemma is proved. ∎

Any left heart is contained in the square of side $1 + \epsilon/4$ and area at most $1 + 9\epsilon/16$ (since $\epsilon \leq 1$). By Lemma 5.6 the sum of the areas of the left hearts in a family F is at most $1 + 9\epsilon/16$.

The area of the left heart of arc (h, t) is $\epsilon d(h, t)/16$. So

$$(\epsilon/16) \sum_{(h,t) \in F} d(h, t) \leq 1 + 9\epsilon/16 \Longrightarrow \sum_{(h,t) \in F} d(h, t) \leq 16/\epsilon + 9.$$

The number of families needed to cover the circle is $\lceil \frac{2\pi}{\arccos{(31/32)}} \rceil = 26$. Therefore, the sum of lengths of all long arcs in T is bounded by $26(16/\epsilon + 9) = 416/\epsilon + 234$.

Choose $\epsilon = \sqrt{416}/\sqrt{n}$. The sum of lengths of all arcs in T is less than $\sqrt{416}\sqrt{n}$ (for the short arcs) plus $\sqrt{416}\sqrt{n} + 234 = 8\sqrt{26}\sqrt{n} + 234$ and corollary 5.5 is proved. ∎

It is also easy to generalize theorem 5.1 to arbitrary norms on $\mathbb{R}^m$. The trick is to do most of the work with respect to the Euclidean norm and switch norms later.

Theorem 5.7. For any dimension $m \geq 2$ and any norm N on $\mathbb{R}^m$ there exists a constant $c_{m,N}$ such that all 2-optimal tours on all configurations of n points in the unit hypercube have length less than $c_{m,N} n^{1-\frac{1}{m}}$.

The proof is a modification of the proof of theorem 5.1.

Proof. Choose m and N. As in section 4.1, by the comparability of norms, there exists a constant $K_N \geq 1$ bounding the ratio (and reciprocal of the ratio) between the N-induced distance and the Euclidean distance between any pair of distinct points in $\mathbb{R}^m$. Define G with respect to norm N. Define *long arc* as one whose Euclidean length is at least ϵ. Define *heart* as before, except with radius $\epsilon/(8K_N^2)$. Define *attack* as before. Define *family* as before, except choose θ so that $\cos\theta = 1 - \frac{1}{128K_N^4}$.

Lemma 5.8. No arc attacks another arc in the same family.

Proof. Sub-lemma 5.3 has already been proved for the general case.

For step 2, our aim is to prove G exceeds $\frac{\epsilon}{4K_N}$, if $\overline{h^1t^1}$ and $\overline{h^2t^2}$ are parallel, both have length ϵ, and $\overline{h^1t^1}$ attacks $\overline{h^2t^2}$. Let $d()$ denote Euclidean distance; let $d_N()$ denote distance induced by N. As before it suffices to work in 2-D. Let pt and ph denote the respective projections of t^2 and h^2 on the line through h^1, t^1. Then, since $d(h^1, t^1) = \epsilon$ and $d(t^1, pt) < \frac{3}{4}\epsilon$,

$$d(h^1, t^1) - d(t^1, pt) > \epsilon/4.$$

Now, because all these points are collinear, and all norms scale, we have

$$d_N(h^1, t^1) - d_N(t^1, pt) > \epsilon/(4K_N).$$

By the triangle inequality, $d_N(t^1, t^2) \leq d_N(t^1, pt) + d_N(pt, t^2)$. By comparability, $d_N(pt, t^2) \leq K_N d(pt, t^2) < K_N \cdot \epsilon/(8K_N^2) = \epsilon/(8K_N)$.

Putting these inequalities together,

$$d_N(h^1, t^1) - d_N(t^1, t^2) \geq [d_N(h^1, t^1) - d_N(t^1, pt)] - d_N(pt, t^2) > \epsilon/(4K_N) - \epsilon/(8K_N) = \epsilon/(8K_N).$$

By a similar argument, $d_N(h^2, t^2) - d_N(h^1, h^2) > \epsilon/(8K_N)$. Hence $G > \epsilon/(4K_N)$ in the norm N.

For step 3, the reader can verify that θ was chosen so that rotating a line segment of length ϵ through an angle θ cannot move the endpoints by more than $\epsilon/(8K_N^2)$, *with respect to Euclidean distance.* Then the same rotation does not change endpoint distances in the norm N by more than $\epsilon/(8K_N)$. By the triangle inequality, G does not change by more than $\epsilon/(4K_N)$. Therefore G remains strictly positive in the norm N. This proves lemma 5.8. ■

Define the soul as having half the length and radius of the smaller heart used in this proof. The rest of the proof of theorem 5.1 goes through without change, *in the Euclidean metric.* At the end we have a bound of the form $cn^{1-\frac{1}{m}}$ on the *Euclidean* length of T (although T is 2-optimal with respect to the norm N). Multiplying this bound by K_N gives a bound for the norm N and the proof is complete. ■

6 Expected Value of Performance Ratio in the Unit Hypercube

In this section we combine Theorem 5.7 with well-known distributional properties of optimal tour lengths to show that the expected performance ratio is bounded by a constant.

Let S_n be any set of n points in the d-dimensional unit hypercube $[0,1]^d \subset \mathbb{R}^d$. Let $OPT(S_n)$ be an optimal tour (under norm N) on S_n, and let $T(S_n)$ be a 2-optimal tour (under norm N). Let I_n be n points picked i.i.d. from the d-dimensional unit hypercube under the uniform distribution.

As an immediate corollary to Theorem 5.7 and a lower bound of $\Omega(n^{\frac{d-1}{d}})$ on $E[OPT(I_n)]$ [9], we infer that there exists a constant $\gamma_{d,N}$ such that $\frac{E[wt(T(I_n))]}{E[OPT(I_n)]} \leq \gamma_{d,N}$ for all n; the ratio of the expected values is bounded. We can also show that $\frac{wt(T(I_n))}{wt(OPT(I_n))}$ is $O(1)$ with high probability and that the expected value of this ratio is $O(1)$.

The following is easily obtainable from Lemma 3 in [9, p. 190]:

Fact 6.1. There exist constants $F_N > 0$ and $0 < \rho < 1$ such that for all $n > 1$,

$$P\left[wt(OPT(I_n)) \leq F_N \cdot n^{\frac{d-1}{d}}\right] \leq \rho^n.$$

From this and Theorem 5.7 we get

Theorem 6.2.

$$P\left[wt(T(I_n)) \geq \frac{c_{m,N}}{F_N} \cdot wt(OPT(I_n))\right] \leq \rho^n.$$

Corollary 6.3. For all m and all norms N on $\mathbb{R}^m$ there exists a constant $c'_{m,N}$ such that

$$E\left[\frac{wt(T(I_n))}{wt(OPT(I_n))}\right] \leq c'_{m,N},$$

where $T(I_n)$ (respectively $OPT(I_n)$) is the length of the longest 2-optimal tour (respectively the shortest tour) on the points I_n with respect to N.

Proof: We first note that for any set of points S_n, $\frac{wt(T(S_n))}{wt(OPT(S_n))} \leq n$; this follows since if the diameter (under norm N) of S_n is D, then $wt(T(S_n)) \leq nD$ and $wt(OPT(S_n)) \geq D$.

Let n_0 be such that $\forall n \geq n_0$, $n\rho^n \leq 1$. Let $\delta_N = \max_{2 \leq n \leq n_0} E\left[\frac{wt(T(I_n))}{wt(OPT(I_n))}\right]$. Now consider $n \geq n_0$.

$$\begin{aligned}
& E\left[\frac{wt(T(I_n))}{wt(OPT(I_n))}\right] \\
= \quad & P[wt(T(I_n)) < \frac{c_{m,N}}{F_N} \cdot wt(OPT(I_n))] \\
& \cdot E\left[\frac{wt(T(I_n))}{wt(OPT(I_n))} \middle| \, wt(T(I_n)) < \frac{c_{m,N}}{F_N} \cdot wt(OPT(I_n))\right] \\
+ \quad & P[wt(T(I_n)) \geq \frac{c_{m,N}}{F_N} \cdot wt(OPT(I_n))] \\
& \cdot E\left[\frac{wt(T(I_n))}{wt(OPT(I_n))} \middle| \, wt(T(I_n)) \geq \frac{c_{m,N}}{F_N} \cdot wt(OPT(I_n))\right] \\
< \quad & \frac{c_{m,N}}{F_N} + n\rho^n \\
\leq \quad & \frac{c_{m,N}}{F_N} + 1.
\end{aligned}$$

Taking $c'_{m,N} = \max\{\delta_N, \frac{c_{m,N}}{F_N} + 1\}$, we are done. ■

7 Expected Running Time of 2-Opt

This section gives polynomial upper bounds on the average number of iterations of 2-opt under the L_2 and L_1 norms.

7.1 The L_2 metric

In the first subsection, we prove that the average number of iterations done by the 2-opt local-improvement algorithm on n random points in the Euclidean unit square is $O(n^{10} \log n)$. Prior to this paper, no polynomial upper bound on the expected time was known. However, W. Kern proved a related result [10]:

Theorem 7.1. There is a c such that the probability that 2-opt does more than n^{16} iterations is at most c/n.

Kern's proof allows the possibility that 2-opt does exponentially many iterations with probability $\Omega(1/n)$. Kern himself writes, "Our approach does not seem to yield interesting results about average running times." We will prove that the expected time is polynomial and we will rely heavily on Kern's lemmas in doing so.

The basic idea of Kern's proof is to show that with probability at least $1 - c/n$, every iteration decreases the cost by at least $\epsilon(n) > 0$; the initial tour being of length at most $\sqrt{2}n$, the number of iterations can then not exceed $\frac{\sqrt{2}n}{\epsilon(n)}$.

To prove that the expected number of iterations is polynomial, we need the following definitions and lemma from [10].

Definition 3. Given points $P, Q, R, S \in [0,1]^2$, define $G(P,Q,R,S) = [d(P,Q) + d(R,S)] - [d(P,R) + d(Q,S)]$.

Definition 4. Given three points P, Q, R in the unit square and $\epsilon > 0$, define $B_\epsilon(P,Q,R)$ to be the set of points S in the unit square such that $|G(P,Q,R,S)| \leq \epsilon$.

Lemma 7.2. [10] There is a $K \geq 1$ with the following property. For any three points P, Q, R in the unit square with $P \neq Q$, the area of $B_\epsilon(P,Q,R)$ (which is the conditional probability that $|G(P,Q,R,S)| \leq \epsilon$, given P, Q, R) is bounded above by $K\sqrt{\epsilon}/d(P,Q)$.

Let $X_1, X_2, ..., X_n$ be points chosen independently and uniformly at random from the unit square.

Definition 5. If i, j, k, l are distinct elements of $\{1, 2, ..., n\}$, define $F(i,j,k,l) = G(X_i, X_j, X_k, X_l)$.

Definition 6. Define $Q = \{(i,j,k,l,i',j',k',l')$ such that $i,j,k,l,i',j',k',l' \in \{1,2,...,n\}$, $|\{i,j,k,l\}| = |\{i',j',k',l'\}| = 4$, and $\{i,j,k,l\} \neq \{i',j',k',l'\}\}$.

Definition 7. Given n random points $X_1, X_2, ..., X_n$ in the unit square, define

$$\hat{F} = \min\{F(i',j',k',l') | (i,j,k,l,i',j',k',l') \in Q \text{ and } 0 < F(i,j,k,l) \leq F(i',j',k',l')\}.$$

Definition 8. Let $N = \min\{n!, 1/\hat{F}\}$.

We now give a very rough road map of the proof that the expected number of iterations done by 2-opt is $O(n^{10} \log n)$. It is not hard to see that any two consecutive improving 2-changes involve distinct 4-sets of vertices. By the definition of $\hat{F}$, 2-opt must decrease the cost of the tour by at least $\hat{F}$ in any two consecutive iterations. The cost of the initial tour being at most $\sqrt{2}n$, the number of iterations cannot exceed $2\sqrt{2}n/\hat{F}$. Clearly the number of iterations never exceeds $n!$. Thus the number of iterations done is at most $\min\{(2n\sqrt{2}) \cdot n!, 2n\sqrt{2}/\hat{F}\} = (2n\sqrt{2})N$. Our goal is therefore to bound $E[N]$.

Notice that if $\hat{F} \in [\frac{\epsilon}{2}, \epsilon)$, then N, which is at least $\frac{1}{2n\sqrt{2}}$ times the number of iterations, is bounded by $\frac{2}{\epsilon}$. The chance that $\hat{F}$ is in this interval is bounded by $Cn^8\epsilon$, since $P[\hat{F} \leq \epsilon] \leq Cn^8\epsilon$ (this is Lemma 7.5). Hence the contribution to the expected value of N due to $[\frac{\epsilon}{2}, \epsilon)$ is bounded by $2Cn^8$. Since $N = n!$ if $\hat{F} < \frac{1}{n!}$ and $\hat{F} \leq 2$ always, we need consider only $\lg n! + O(1)$ intervals, each of which contributes at most $2Cn^8$ to $E[N]$. Thus $E[N]$ is $O(n^8 \lg n!)$, and the expected number of iterations is $O(n^{10} \log n)$.

Now we continue with the proof.

Lemma 7.3. Let $\epsilon > 0$. Choose points $X_1, X_2, ..., X_n$ in the unit square independently and uniformly at random. Let $(i,j,k,l,i',j',k',l') \in Q$. Suppose that $l \notin \{i',j',k',l'\}$ and that $l' \notin \{i,j,k,l\}$. Then

1. If $\{i,j\} \neq \{i',j'\}$, then $P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon] \leq 64\pi^2K^2\epsilon$.
2. If $\{i,j\} = \{i',j'\}$, then $P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon] \leq 14\pi K^2\epsilon \cdot \lg\frac{1}{\epsilon}$ if $\epsilon \leq \frac{1}{2}$.

Proof. From Lemma 7.2 we have

$$P[X_l \in B_\epsilon(X_i, X_j, X_k)|X_i, X_j, X_k] = P[|F(i,j,k,l)| \leq \epsilon|X_i, X_j, X_k] \leq \frac{K\sqrt{\epsilon}}{d(X_i, X_j)}.$$

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon|X_i, X_j, X_k, X_{i'}, X_{j'}, X_{k'}]$$

(notice that l and l' must be distinct from each other and from i, j, k, i', j', k', though the latter six need not be distinct)

$$= P[X_l \in B_\epsilon(X_i, X_j, X_k), X_{l'} \in B_\epsilon(X_{i'}, X_{j'}, X_{k'})|X_i, X_j, X_k, X_{i'}, X_{j'}, X_{k'}] \leq \frac{K\sqrt{\epsilon}}{d(X_i, X_j)} \frac{K\sqrt{\epsilon}}{d(X_{i'}, X_{j'})}$$

$$= \frac{K^2\epsilon}{d(X_i, X_j)d(X_{i'}, X_{j'})}.$$

Therefore

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon|X_i, X_j, X_k, X_{i'}, X_{j'}, X_{k'}] \leq \frac{K^2\epsilon}{d(X_i, X_j)d(X_{i'}, X_{j'})}$$

and therefore

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon|X_i, X_j, X_{i'}, X_{j'}] \leq \frac{K^2\epsilon}{d(X_i, X_j)d(X_{i'}, X_{j'})}. \tag{4}$$

If $\{i,j\} \neq \{i',j'\}$, then

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon]$$

$$= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon|d(X_i, X_j) \in [2^{-r}, 2^{-r+1}), d(X_{i'}, X_{j'}) \in [2^{-s}, 2^{-s+1})]$$

$$\cdot P[d(X_i, X_j) \in [2^{-r}, 2^{-r+1}), d(X_{i'}, X_{j'}) \in [2^{-s}, 2^{-s+1})]$$

Now

$$P[d(X_i, X_j) \in [2^{-r}, 2^{-r+1}), d(X_{i'}, X_{j'}) \in [2^{-s}, 2^{-s+1})] \leq P[d(X_i, X_j) \leq 2^{-r+1}, d(X_{i'}, X_{j'}) \leq 2^{-s+1}]$$

and

$$P[d(X_i, X_j) \leq 2^{-r+1}, d(X_{i'}, X_{j'}) \leq 2^{-s+1}] \leq \pi(2^{-r+1})^2\pi(2^{-s+1})^2$$

since $\{i,j\} \neq \{i',j'\}$. So

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon]$$

$$\leq \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon|d(X_i, X_j) \in [2^{-r}, 2^{-r+1}), d(X_{i'}, X_{j'}) \in [2^{-s}, 2^{-s+1})]$$

$$\cdot P[d(X_i, X_j) \in [0, 2^{-r+1}), d(X_{i'}, X_{j'}) \in [0, 2^{-s+1})]$$

$$\leq \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{K^2\epsilon}{2^{-r}2^{-s}}\pi(2^{-r+1})^2\pi(2^{-s+1})^2$$

$$= 16\pi^2 \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} K^2\epsilon 2^r 2^s 2^{-2r} 2^{-2s}$$

$$= 16\pi^2 K^2 \epsilon \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} 2^{-r} 2^{-s}$$

$$= 16\pi^2 K^2 \epsilon (\sum_{r=0}^{\infty} 2^{-r})(\sum_{s=0}^{\infty} 2^{-s})$$

$$= 64\pi^2 K^2 \epsilon.$$

If instead $\{i, j\} = \{i', j'\}$, we have

$$P[X_l \in B_\epsilon(X_i, X_j, X_k) | X_i, X_j, X_k] = P[|F(i,j,k,l)| \leq \epsilon | X_i, X_j, X_k] \leq \frac{K\sqrt{\epsilon}}{d(X_i, X_j)}.$$

Equation (4) tells us that

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon | X_i, X_j] \leq \frac{K^2 \epsilon}{d(X_i, X_j)^2}.$$

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon]$$

$$= \sum_{s=0}^{\infty} P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon | d(X_i, X_j) \in [2^{-s}, 2^{-s+1})] P[d(X_i, X_j) \in [2^{-s}, 2^{-s+1})]$$

Since

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon | d(X_i, X_j) \in [2^{-s}, 2^{-s+1})] \leq \min\left\{\frac{K^2\epsilon}{(2^{-s})^2}, 1\right\}$$

$$= \min\{K^2 \epsilon 2^{2s}, 1\},$$

we have

$$P[|F(i,j,k,l)| \leq \epsilon, |F(i',j',k',l')| \leq \epsilon] \leq \sum_{s=0}^{\infty} P[d(X_i, X_j) \in [2^{-s}, 2^{-s+1})] \cdot \min\{K^2 \epsilon 2^{2s}, 1\}$$

$$\leq \sum_{s=0}^{\infty} P[d(X_i, X_j) \leq 2^{-s+1}] \cdot \min\{K^2 \epsilon 2^{2s}, 1\}$$

$$\leq \sum_{s=0}^{\infty} 4\pi 2^{-2s} \cdot \min\{K^2 \epsilon 2^{2s}, 1\}$$

Now $K^2 \epsilon 2^{2s} < 1$ if and only if $s < \frac{1}{2} \lg \frac{1}{K^2 \epsilon}$. So the quantity is at most

$$\sum_{s=0}^{\lfloor \frac{1}{2} \lg \frac{1}{K^2\epsilon} \rfloor} 4\pi 2^{-2s} K^2 \epsilon 2^{2s} + \sum_{s=\lceil \frac{1}{2} \lg \frac{1}{K^2\epsilon} \rceil}^{\infty} 4\pi 2^{-2s} \cdot 1$$

$$\leq 4\pi K^2 \epsilon (1 + \frac{1}{2} \lg \frac{1}{K^2 \epsilon}) + 8\pi K^2 \epsilon.$$

If $\epsilon \leq \frac{1}{2}$, then since $K \geq 1$ we have

$$4\pi K^2 \epsilon (1 + \frac{1}{2} \lg \frac{1}{K^2 \epsilon}) + 8\pi K^2 \epsilon \leq 14\pi K^2 \epsilon \cdot \lg \frac{1}{\epsilon}. \blacksquare$$

Lemma 7.4. Let $\epsilon > 0$. Let $(i, j, k, l, i', j', k', l') \in Q$.

1. If $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \le 1$, then $P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon] \le 64\pi^2 K^2 \epsilon$.
2. If $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \ge 2$ and $\epsilon \le \frac{1}{2}$, then $P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon] \le 64\pi^2 K^2 \epsilon \cdot \lg \frac{1}{\epsilon}$.

Proof. By the symmetry in F, we have $F(a,b,c,d) = F(c,d,a,b) = F(b,a,d,c)$. This means that it is possible to move any one of the indices into the final position without changing the value of $F(a,b,c,d)$. Formally, if $x \in \{a,b,c,d\}$, then there is a permutation (a',b',c',d') of $\{a,b,c,d\}$ such that $d' = x$ and $F(a',b',c',d') = F(a,b,c,d)$ always.

Since $\{i,j,k,l\} \ne \{i',j',k',l'\}$, we can find an x in $\{i,j,k,l\} - \{i',j',k',l'\}$ and a y in $\{i',j',k',l'\} - \{i,j,k,l\}$. By moving x and y to the last position, without loss of generality we may assume that $l \notin \{i',j',k',l'\}$ and $l' \notin \{i,j,k,l\}$.

Now we invoke Lemma 7.3. If $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \le 1$, clearly $\{i,j\} \ne \{i',j'\}$. Lemma 7.3 implies that $P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon] \le 64\pi^2 K^2 \epsilon$.

If $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \ge 2$, then possibly $\{i,j\} = \{i',j'\}$ and possibly not. In the former case, $P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon] \le 14\pi K^2 \epsilon \cdot \lg \frac{1}{\epsilon} \le 64\pi^2 K^2 \epsilon \cdot \lg \frac{1}{\epsilon}$ (if $\epsilon \le \frac{1}{2}$). In the latter, $P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon] \le 64\pi^2 K^2 \epsilon$. ■

Recall the definition of $\hat{F}$:

$$\hat{F} = \min\{F(i',j',k',l') | (i,j,k,l,i',j',k',l') \in Q \text{ and } 0 < F(i,j,k,l) \le F(i',j',k',l')\}.$$

Lemma 7.5. Let $n \ge 8$ and let $C = 9280\pi^2 K^2$. Let $2^{-n^2} \le \epsilon \le \frac{1}{2}$. Then $P[\hat{F} \le \epsilon] \le Cn^8 \epsilon$.

In $P[\hat{F} \le \epsilon] \le Cn^8\epsilon$, the n^8 comes from the fact that Q in the definition of $\hat{F}$ has at most n^8 8-tuples. The ϵ comes from Lemma 7.2, which has a $\sqrt{\epsilon}$. Very roughly, because $\hat{F}$ involves two 4-tuples, we can replace the $\sqrt{\epsilon}$ in Lemma 7.2 by an ϵ.

Proof. In this terminology, $\hat{F} \le \epsilon$ if and only if there is an 8-tuple $(i,j,k,l,i',j',k',l') \in Q$ such that $0 < F(i,j,k,l) \le \epsilon$ and $0 < F(i',j',k',l') \le \epsilon$.

$$P[\hat{F} \le \epsilon] \le \sum P[|F(i,j,k,l)| \le \epsilon, |F(i',j',k',l')| \le \epsilon],$$

where the summation is over Q. There are at most n^8 8-tuples $(i,j,k,l,i',j',k',l') \in Q$ such that $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \le 1$. There are at most $12^2 n^6$ 8-tuples such that $|\{i,j,k,l\} \cap \{i',j',k',l'\}| \ge 2$. By Lemma 7.4, if $0 < \epsilon \le \frac{1}{2}$, then

$$P[\hat{F} \le \epsilon] \le n^8(64\pi^2 K^2 \epsilon) + 144n^6(64\pi^2 K^2 \epsilon \lg \frac{1}{\epsilon}).$$

Since $2^{-n^2} \le \epsilon \le \frac{1}{2}$, $\frac{1}{\epsilon} \le 2^{n^2}$ and $\lg \frac{1}{\epsilon} \le n^2$. So the expression above is no more than $n^8(64\pi^2 K^2)\epsilon + 144n^6(64\pi^2 K^2 n^2)\epsilon = n^8(9280\pi^2 K^2)\epsilon$. Let $C = 9280\pi^2 K^2$. Therefore $P[\hat{F} \le \epsilon] \le Cn^8\epsilon$ if $2^{-n^2} \le \epsilon \le \frac{1}{2}$ and $n \ge 8$. ■

Recall that $N = \min\{n!, 1/\hat{F}\}$.

Lemma 7.6. $E[N] \le 4Cn^9 \lg n$ if $n \ge 8$.

Proof. We have

$$E[N] \le \sum_{r=-1}^{\lceil \lg n! \rceil} P[\hat{F} \in [2^{-r}, 2^{-r+1})] \cdot 2^r + P[\hat{F} < \frac{1}{n!}] \cdot n!$$

$$\leq \sum_{r=-1}^{\lceil \lg n! \rceil} P[\hat{F} \leq 2^{-r+1}]2^r + P[\hat{F} < \frac{1}{n!}]n!.$$

So

$$
\begin{aligned}
E[N] &\leq \sum_{r=-1}^{1} P[\hat{F} \leq 2^{-r+1}]2^r + \sum_{r=2}^{\lceil \lg n! \rceil} P[\hat{F} \leq 2^{-r+1}]2^r + P[\hat{F} < \frac{1}{n!}]n! \\
&\leq (\frac{1}{2} + 1 + 2) + \sum_{r=2}^{\lceil \lg n! \rceil} (Cn^8 2^{-r+1})2^r + Cn^8 \frac{1}{n!} n! \\
&= 3.5 + 2Cn^8(\lceil \lg n! \rceil - 1) + Cn^8 \\
&\leq 3.5 + 2Cn^8(n \lg n) \\
&\leq 4Cn^9 \lg n
\end{aligned}
$$

since we are assuming $n \geq 8$. ■

Theorem 7.7. The average number of 2-changes made by algorithm 2-opt when run on n random points in the Euclidean unit square is at most $(8C\sqrt{2})n^{10} \lg n$, for any $n \geq 8$.

Proof. If a 2-change is made, replacing edges (x_i, x_j) and (x_k, x_l) by (x_i, x_k) and (x_j, x_l), then the same four vertices $\{i, j, k, l\}$ cannot be used in an improving 2-change in the next iteration. Therefore, any two consecutive improving 2-changes involve distinct 4-tuples of vertices.

By the definition of $\hat{F}$, 2-opt must decrease the cost of the tour by at least $\hat{F}$ in any two consecutive iterations. The cost of the initial tour being at most $\sqrt{2}n$, the number of iterations cannot exceed $2\sqrt{2}n/\hat{F}$. Clearly the number of iterations never exceeds $n!$.

The number of iterations done is at most

$$
\begin{aligned}
&\min\{(2n\sqrt{2}) \cdot n!, \frac{2n\sqrt{2}}{\hat{F}}\} \\
&= (2n\sqrt{2}) \cdot \min\{n!, \frac{1}{\hat{F}}\} \\
&= (2n\sqrt{2})N.
\end{aligned}
$$

So the average number of iterations is at most $(2n\sqrt{2})E[N] \leq (8C\sqrt{2})n^{10} \lg n$ for all $n \geq 8$. ■

7.2 The L_1 metric

In this subsection we bound the expected number of iterations of 2-opt under the L_1 norm. In contrast with Theorem 7.7 the proof is somewhat nonconstructive, but it is fairly short and the polynomial is of lower order.

Theorem 7.8. Let n points be independently sampled from the uniform distribution in the m-dimensional unit hypercube. Let 2-opting be performed with respect to the L_1 norm. Then the expected number of iterations required is $O(n^6 \log n)$.

Proof. Let I^m denote the unit hypercube. Suppose v^1, v^2, v^3, v^4 are four points sampled independently from the uniform distribution on I^m. Let Z denote the random variable equal to $G(v^1, v^2, v^3, v^4) = ||v^1 - v^2||_1 + ||v^3 - v^4||_1 - ||v^1 - v^3||_1 - ||v^2 - v^4||_1$. Note that distances are computed according to the L_1 norm.

We study the distribution of Z. For clarity we focus on the case $m = 2$, and use parenthetical remarks to extend the proof to the m-dimensional case.

The four points in I^2 are defined by eight $(4m)$ random variables, denoted $x_i, y_i : i = 1, \ldots, 4$, drawn independently from the uniform distribution on $[0, 1]$. Let $g(x_1, \ldots, x_4) = |x_1 - x_2| + |x_3 - x_4| - |x_1 - x_3| - |x_2 - x_4|$. Then Z is the sum of the independent and identically distributed variables X and Y where $X = g(x_1, x_2, x_3, x_4)$ and $Y = g(y_1, y_2, y_3, y_4)$. The key is to understand the distribution of X, because $X + Y$ will have a distribution found as the convolution of two i.i.d. variables with this distribution. (In m dimensions Z is the m-fold convolution of i.i.d. variables with this distribution.)

Lemma 7.9. With probability $1/3$ the variable $X = 0$; with probability $2/3$ the variable X is distributed according to a continuous density function $\bar{h}$ on $[2, 2]$.

Proof. Consider the conditional distribution of X, conditioned on the event $\pi_1 = \{x_1 \geq x_2 \geq x_3 \geq x_4\}$. Notice that $X = 2x_3 - 2x_2$ under this condition. Now if we take four samples i.i.d. from the uniform distribution, and let W equal the difference between the 3rd and 2nd largest, then obviously W has a continuous density function. Therefore X has continuous conditional density function conditioned on event π_1. We denote the conditional probability density by h_{π_1}.

By symmetry, for seven other π's we can make the same argument, and for each we get a continuous conditional probability density function h_{π_i}. For eight additional π_i the conditional distribution is like $-W$, i.e., twice the difference between the 2nd and 3rd largest.

For the last eight π_i, the conditional distribution of X is degenerate with all its mass at zero. This occurs when the projections of the arcs on the x-axis overlap and have opposite orientations.

The unconditional density function of X is

$$h = \sum_{i=1}^{24} \frac{1}{24} h_{\pi_i}.$$

Therefore X has a hybrid distribution. It has a mass point of weight $8/24 = 1/3$ at 0. The remaining $2/3$ of the mass has continuous density because each of the 16 contributing h_{π_i}'s is continuous. ∎

When the distribution of X is convolved with itself to get Z, the result is a hybrid distribution: $Z = 0$ with probability $1/9$; Z is distributed according to a continuous density, denoted $\bar{h}^2$, with probability $8/9$. This is because X and Y are independent. (In higher dimensions each additional convolution is of two independent hybrid distributions, each containing one mass point at zero, and continuous everywhere else, and these properties are obviously preserved in the sum of the distributions.) In m dimensions, the probability is $(1/3)^m$ that a given random 4-tuple of points has $Z = 0$; Z is distributed according to a continuous density, denoted $\bar{h}^m$, with probability $1 - (1/3)^m$.

A local improvement algorithm will not make any of the 2-changes corresponding to 4-tuples where $Z = 0$. The algorithm will be fast if there are no very small positive Z values.

Now consider the density function $\bar{h}^m$. It is continuous everywhere on $[-2m, 2m]$ and is symmetric around 0. Since a continuous function on a compact set attains its maximum, $\bar{h}^m$ has a maximum M.

This implies that

$$P[0 < |Z| \leq \epsilon] \leq 2M\epsilon \quad \forall \epsilon > 0.$$

Now let S denote a sample of n points on I^m. Let Δ denote the smallest of all non-zero absolute differences in tour length from 2-changes:

$$\Delta = \min_{\substack{v^1, v^2, v^3, v^4 \in S \\ G(v^1, v^2, v^3, v^4) \neq 0}} |G(v^1, v^2, v^3, v^4)|$$

(where the minimum is over distinct points). Since the minimum is taken over fewer than n^4 4-tuples, we get

$$P[\Delta < \epsilon] < n^4(2M\epsilon) = cn^4\epsilon.$$

By exactly the same argument as lemma 7.6 with an n^4 instead of an n^8 term, we have $E[\min\{n!, 1/\Delta\}] \leq 4cn^5 \log n$. The cost of the initial tour being at most $2n$, the expected number of iterations is at most $(2n)4cn^5 \log n$, which is $O(n^6 \log n)$. ■

8 Extending Lueker's Construction

In 1976, G. Lueker [16] constructed, for each $n \geq 4$, a family of n-node weighted cliques $< G_n >$ on vertex set $\{x_0, x_1, x_2, ..., x_{n-1}\}$ with the property that the naive 2-opt algorithm can do at least $2^{\lfloor n/2 \rfloor - 2}$ iterations. Although the only thing we need from his construction is that each of the tours includes the edge $\{x_0, x_{n-1}\}$, we give Lueker's (unpublished) construction here.

Here is Lueker's construction. For each $0 \leq i < j \leq n-1$, define the weight w_{ij} of edge $\{x_i, x_j\}$ as follows.

1. j is even.
   (a) If i=0, then $w_{ij} = 2^{2j}$.
   (b) If i is positive and odd, then $w_{ij} = 2^{2i+3}$.
   (c) If i is positive and even, then $w_{ij} = 2^{2i}$.
2. j is odd.
   (a) If i=0, then $w_{ij} = 2^{2j+3}$.
   (b) If i is positive and odd, then $w_{ij} = 2^{2i+4}$.
   (c) If i is positive and even, then $w_{ij} = 2^{2i+3}$.

Let $a_k =< x_1, x_2, x_3, ..., x_{2k} >$ and let a'_k be its reverse. Lueker proves the following theorem.

Theorem 8.1. Let T be a tour of G_n that contains the block $< x_0, a_k, x_{2k+1} >$. Then there is a sequence of at least 2^{k-1} improving 2-changes which lead to the replacement of this block by $< x_0, a'_k, x_{2k+1} >$ (and which leaves the rest of the tour unchanged).

We refer the reader to [16] for a proof. The proof converts the string $< x_0, a_k, x_{2k+1}, x_{2k+2}, x_{2k+3} >$ to $< x_0, a'_k, x_{2k+1}, x_{2k+2}, x_{2k+3} >$ in at least 2^{k-1} steps by induction, then to $< x_0, a'_k, x_{2k+2}, x_{2k+1}, x_{2k+3} >$ in one step, then to $< x_0, x_{2k+1}, x_{2k+2}, a_k, x_{2k+3} >$ in one step, then to $< x_0, x_{2k+1}, x_{2k+2}, a'_k, x_{2k+3} >$ in at least 2^{k-1} steps by a clever use of the inductive assertion, and last to $< x_0, x_{2k+2}, x_{2k+1}, a'_k, x_{2k+3} >$. Notice that, as claimed, edge $\{x_0, x_{n-1}\}$ is in all tours.

It is now our job to extend the result to all $k \geq 3$. Since $k \geq 4$ is easy, we do that case first.

Theorem 8.2. For any $k \geq 4$, for any $N \geq 2k$, there is an N-vertex weighted graph on which there exists a sequence of at least $2^{\lfloor N/2 \rfloor - k}$ improving k-changes.

Proof. The idea is to take G_n and add $2(k-2)$ new vertices at distance one from each other and from the original vertices. Any 2-change in the original proof can be converted to a k-change in the new graph, by flipping two original edges and $k-2$ new ones.

Let $l = k - 2$ and add to G_n $2l$ new vertices:

$$s, t, y_1, z_1, y_2, z_2, ..., y_{l-1}, z_{l-1}.$$

The distance between any pair of these vertices, as well as that between these new vertices and the n old ones, is 1. Let

$$A = \{\{y_1, z_1\}, \{y_2, z_2\}, \{y_3, z_3\}, ..., \{y_{l-1}, z_{l-1}\}\},$$

let

$$L = \{\{s, y_1\}, \{z_1, y_2\}, \{z_2, y_3\}, ..., \{z_{l-1}, t\}\},$$

and let

$$R = \{\{s, z_1\}, \{y_1, z_2\}, \{y_2, z_3\}, ..., \{y_{l-1}, t\}\}.$$

The key fact is that $A \cup L$ is the edge set of a Hamiltonian path from s to t among the new vertices, $A \cup R$ is the edge set of a Hamiltonian path from s to t among the new vertices, and A, L, R are pairwise disjoint. Furthermore, $A \cup L$ and $A \cup R$ differ in exactly l edges.

Now let S, T be the edge sets of any two tours of G_n that both contain edge $\{x_0, x_{n-1}\}$ and that share exactly $n-2$ edges. Let $S' = (S - \{x_0, x_{n-1}\}) \cup (\{\{x_0, s\}, \{x_{n-1}, t\}\} \cup A \cup L)$ and let $T' = (T - \{x_0, x_{n-1}\}) \cup (\{\{x_0, s\}, \{x_{n-1}, t\}\} \cup A \cup R)$. S' induces a tour of the new graph, and T' induces a tour of the new graph as well, and S' and T' differ in exactly $2 + l = k$ edges. Since all the new edges have weight 1, if moving from S to T on the original graph is a cost-decreasing 2-change, then moving from S' to T' in the new graph is a cost-decreasing k-change. (In the next step, we can interchange the roles of the y's and z's, since they are symmetric.)

Lueker's proof gives a sequence of at least $2^{\lfloor n/2 \rfloor - 2}$ cost-decreasing 2-changes on an n-vertex graph (if $n \geq 4$). The new graph we built has $N = n + 2l = n + 2(k-2) = n + 2k - 4$ vertices, and there is a sequence of $2^{\lfloor n/2 \rfloor - 2}$ cost-decreasing k-changes. In terms of the number N of vertices in the new graph, the number of k-changes is at least $2^{\lfloor N/2 \rfloor - k}$. ∎

Now we tackle $k = 3$.

Theorem 8.3. For each even $N \geq 8$, there is an N-vertex weighted graph on which there exists a sequence of at least $2^{\lfloor N/4 \rfloor - 1}$ improving 3-changes.

Proof. We use w_{ij} to denote the weight of $\{x_i, x_j\}$ in G_n. Let $w_{ii} = 0$ for all i. Take two copies of G_n. One copy has vertex set $V = \{x_0, ..., x_{n-1}\}$; the other has vertex set $V' = \{x'_0, ..., x'_{n-1}\}$. The weight of edge $\{x'_i, x'_j\}$, $i \neq j$, and that of $\{x_i, x'_j\}$ equals w_{ij}. Call the new $2n$-node graph H_n.

Let S be the edge set of any tour in G_n containing edge $\{x_0, x_{n-1}\}$. There is an obvious associated tour of H_n: Let $A = S - \{\{x_0, x_{n-1}\}\}$. Let $A^* = \{\{x'_i, x'_j\} | \{x_i, x_j\} \in A\}$. Then the associated tour contains edges $S' = \{\{x_0, x'_{n-1}\}, \{x'_0, x_{n-1}\}\} \cup A \cup A^*$. It is not hard to see that S' is the edge set of a tour in H_n.

We will prove the following. Let S be any tour of G_n containing $\{x_0, x_{n-1}\}$, and let S' be the associated tour of H_n. Let T be a tour of G_n obtained from S by one cost-decreasing 2-change. Then the tour T' of H_n associated with T can be obtained from S' via *two* cost-decreasing 3-changes. (We are implicitly identifying tours with their edge sets.)

Consider the tour S as directed from x_0 to x_{n-1}. Suppose that the 2-change involves swapping out the edges $\{x_i, x_j\}, \{x_k, x_l\}$, where, in S, the vertex x_i is the first among the four visited by S, x_j is second, x_l is third and x_k, fourth. Since the 2-change results in a tour T, it must replace the two missing edges by $\{x_i, x_l\}, \{x_k, x_j\}$. (If they were replaced by $\{x_i, x_k\}, \{x_l, x_j\}$, then the new "tour" would be disconnected.)

Let S' and T' be the tours of H_n associated with S and T, respectively. S' and T' differ only in that to get from S' to T' one drops the four edges $\{x_i, x_j\}$, $\{x'_i, x'_j\}$, $\{x_k, x_l\}$ $\{x'_k, x'_l\}$, and adds $\{x_i, x_l\}$, $\{x'_i, x'_l\}$, and $\{x_k, x_j\}$, $\{x'_k, x'_j\}$.

We know that switching from S to T decreased the cost; thus $(w_{ij} + w_{kl}) - (w_{il} + w_{kj}) > 0$. In S', we have the four edges $\{x_i, x_j\}$, $\{x_k, x_l\}$, $\{x'_i, x'_j\}$, $\{x'_k, x'_l\}$. We leave the last one unchanged and we change the first three to $\{x_i, x_l\}$, $\{x_k, x'_j\}$, $\{x'_i, x_j\}$. It is easy to see that the new "tour" is indeed a tour, so this is a valid 3-change provided that we have decreased the cost. The decrease in cost is

$$(w_{ij} + w_{kl} + w_{ij}) - (w_{il} + w_{kj} + w_{ij}) = (w_{ij} + w_{kl}) - (w_{il} + w_{kj}),$$

which we know to be positive. The next 3-change leaves edge $\{x_i, x_l\}$ unchanged. It flips $\{x_k, x'_j\}, \{x'_i, x_j\}, \{x'_k, x'_l\}$ to $\{x_k, x_j\}$, $\{x'_i, x'_l\}$, $\{x'_k, x'_j\}$. The decrease in cost is

$$(w_{kj} + w_{kl} + w_{ij}) - (w_{il} + w_{kj} + w_{kj}) = (w_{ij} + w_{kl}) - (w_{il} + w_{kj}),$$

which we know to be positive. The resulting edge set contains $\{x_k, x_j\}$, $\{x'_k, x'_j\}$, $\{x_i, x_l\}$, $\{x'_i, x'_l\}$, and is otherwise the same as S', so it is T', and therefore a tour.

Lueker's proof gave $2^{\lfloor n/2 \rfloor - 2}$ 2-changes on an n-vertex graph, $n \geq 4$. We have $N = 2n$, and we make two 3-changes for each original 2-change. In terms of N, we have $2^{\lfloor N/4 \rfloor - 1}$ 3-changes, if $N \geq 8$ is even. ■

9 Open Problems

- One of the best TSP algorithms in actual experiments [2] is the Lin-Kernighan algorithm [14], a local search algorithm with a more complex neighborhood structure. Since a Lin-Kernighan optimal tour is also 2-optimal, all the upper bounds on the performance ratio of 2-opt also hold for Lin-Kernighan. Can one do better for Lin-Kernighan?

- Our lower bounds on the performance ratio of k-opt are obtained by showing that there is some k-optimal tour of large weight. Suppose we start with a random tour and then deterministically make improving k-changes. Can we get better performance guarantees?
- Can Lueker's results be extended to the Euclidean plane, i.e., is there a graph in the Euclidean plane for which there exists an exponential number of improving 2-changes?
- Can Theorem 3.4 be generalized to any k-opt algorithm, i.e., for arbitrary metric spaces can one prove that as k increases, the performance guarantee of the k-opt algorithm improves?

10 Acknowledgments

We thank Jim Dai, Bob Foley, Tony Hayter, Mark Krentel, Kirk Pruhs, David Shmoys, Mike Todd, Sundar Vishwanathan and Mihalis Yannakakis for helpful discussions on the material of the paper. and George Lueker for kindly permitting us to include his construction in section 8.

References

[1] N. Alon and Y. Azar, "On-line Steiner Trees in the Euclidean Plane," *8th Annual ACM Symposium on Computational Geometry* (1992), 337-343.

[2] J. L. Bentley, Invited talk at the *1st SIAM Symposium on Discrete Algorithms* (1990).

[3] J. Bentley and J. Saxe, "An Analysis of Two Heuristics for the Euclidean Traveling Salesman Problem," *Eighteenth Annual Allerton Conference on Communication, Control and Computing* (1980), 41-49.

[4] M. Bern, and D. Eppstein, "Worst-Case Bounds for Subadditive Geometric Graphs," *9th Annual ACM Symposium on Computational Geometry* (1993), 183–188.

[5] B. Bollobás, *Extremal Graph Theory.* Academic Press, London, 1978.

[6] B. Chandra, G. Das, G. Narasimhan, and J. Soares, "New Sparseness Results on Graph Spanners," *8th Annual ACM Symposium on Computational Geometry* (1992), 192-201.

[7] L. K. Grover, "Local Search and the Local Structure of NP-complete Problems," *Operations Research Letters* **12** (1992), 235–243.

[8] D. J. Johnson, C. H. Papadimitriou, M. Yannakakis, "How Easy is Local Search?", *JCSS* **37** (1988), 79-100.

[9] R. M. Karp and J. M. Steele, "Probabilistic Analysis of Heuristics," *The Traveling Salesman Problem*, eds. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, D. B. Shmoys. John Wiley and Sons, 1985, 181-205.

[10] W. Kern, "A Probabilistic Analysis of the Switching Algorithm for the Euclidean TSP," *Mathematical Programming* **44** (1989), 213-219.

[11] M. W. Krentel, "Structure in Locally Optimal Solutions," *30th IEEE Symposium on the Foundations of Computer Science* (1989), 216-221.

[12] M. W. Krentel, private communication.

[13] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, D. B. Shmoys, *The Traveling Salesman Problem.* John Wiley and Sons, 1985.

[14] S. Lin and B. W. Kernighan, "An Effective Heuristic for the Traveling Salesman Problem," *Operations Research* **21** (1973), 489-516.

[15] L. Loomis and S. Sternberg, *Advanced Calculus*, Addison-Wesley, 1968.

[16] G. Lueker, manuscript, Princeton University, Princeton, NJ, 1975.

[17] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[18] J. Plesnik, "Bad Examples of the Metric Traveling Salesman Problem for the 2-Change Heuristic," *Acta Mathematica Universitatis Comenianae* (1986), 203-207.

[19] C. A. Rogers, "Covering a Sphere With Spheres," *Mathematika* **10** (1963), 157-164.

[20] T. L. Snyder and J. M. Steele, "A Priori Inequalities for the Euclidean Traveling Salesman," *8th Annual ACM Symposium on Computational Geometry* (1992), 344-349.