A PTAS for Euclidean Maximum Scatter TSP

László Kozma* Tobias Mömke*†

Abstract. We study the problem of finding a tour of n points in $\mathbb{R}^d$ in which *every edge is long*. More precisely, we wish to find a tour that maximizes the length of the shortest edge in the tour. The problem is known as Maximum Scatter TSP, and it was introduced by Arkin et al. (SODA 1997), motivated by applications in manufacturing and medical imaging. Arkin et al. gave a 2-approximation for the metric version of the problem and showed that this is the best possible ratio achievable in polynomial time (assuming $\mathsf{P} \neq \mathsf{NP}$). They raised the question of whether one can obtain a better approximation ratio in the planar Euclidean case. We answer this question in the affirmative in a more general setting, by giving a polynomial-time approximation scheme (PTAS) for Maximum Scatter TSP in an arbitrary fixed-dimensional Euclidean space.

1 Introduction

Let $P = \{p_1, \ldots, p_n\}$ be a set of points in $\mathbb{R}^d$. A tour T of P is a sequence $T = (p_{i_1}, \ldots, p_{i_n})$, where $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$. The *scatter* of tour T is the minimum distance between neighboring points of T, i. e., $\min\{d(p_{i_1}, p_{i_2}), \ldots, d(p_{i_{n-1}}, p_{i_n}), d(p_{i_n}, p_{i_1})\}$. The Maximum Scatter Travelling Salesman Problem (MSTSP) asks for a tour of P with maximum scatter. We study this problem in the *geometric* setting where the distance function d is the Euclidean distance between points.

Arkin et al. [1] initiated the study of MSTSP in 1997, motivated by problems in manufacturing (riveting) and medical imaging. They gave a simple 2-approximation algorithm for the more general metric problem (where distances are only required to satisfy the triangle inequality). They also showed that for the metric variant, the approximation ratio of 2 is optimal (assuming $\mathsf{P} \neq \mathsf{NP}$). It was left open whether a better approximation ratio can be obtained in polynomial time if the problem has more geometric structure (e. g., if distances are Euclidean). Arkin et al. raise this question for the planar case (see also [6] and [13, p. 681]).

It is natural to expect that geometric structure should lead to stronger approximation-guarantees. The same phenomenon has been observed for the standard TSP problem: for metric TSP the best known approximation ratio is 1.5 (Christofides [5]), with a current lower bound of $\frac{123}{122}$ (Karpinski et al. [9]), whereas for Euclidean TSP Arora [2] and Mitchell [11] independently obtained polynomial-time approximation schemes (PTAS). Similarly, the Euclidean MaxTSP problem (where the goal is to maximize the *total* length of the tour) admits a PTAS (Barvinok [3]), but the metric version is currently known to admit only a $\frac{7}{8}$-approximation (Kowalik and Mucha [10]).

In this paper we answer the open question about planar MSTSP, by giving a polynomial-time $(1-\epsilon)$-approximation, for arbitrary fixed $\epsilon > 0$. In fact, we present a PTAS for MSTSP in arbitrary fixed-dimensional Euclidean spaces. Since MSTSP is known to be strongly NP-complete in dimensions 3 and above [7], our result settles the classical complexity status of the problem in these dimensions. We show the following result.

Theorem 1 *Let P be a set of n points in $\mathbb{R}^d$. A tour of P whose scatter is at least a $(1-\epsilon)$ factor of the MSTSP optimum can be found in time $O\left(n^{(100d/\epsilon^2)^d}\right)$.*

Further related work. TSP is one of the cornerstones of combinatorial optimization and several variants have been considered in the literature (we refer to [8] for a survey). Minimizing variants are more common, but there exist natural settings in which tours with long edges are desirable. This is the case in certain manufacturing operations where nearby elements in a sequence are required to be geometrically well-separated in order to avoid interferences [1].

MSTSP (a.k.a. max min TSP) appears similar to Bottleneck TSP (a.k.a. min max TSP), a problem known to be NP-complete already in the planar Euclidean case [8]. For metric Bottleneck TSP, 2-approximation is the best possible [12], and we are not aware of stronger approximation-results for geometric variants. Despite the similarity between MSTSP and Bottleneck TSP, it is unclear whether any techniques can be transferred from one problem to the other.

Open question. Our current work does not address the complexity status of solving MSTSP *exactly* in the planar Euclidean case. It remains open whether this problem is NP-hard (the situation is the same for MaxTSP). We note that this question has a natural equivalent formulation: is checking for existence of a Hamiltonian cycle NP-complete in complements of unit disk graphs?

*Department of Computer Science, Saarland University, kozma@cs.uni-saarland.de, moemke@cs.uni-saarland.de

†Partially funded by Deutsche Forschungsgemeinschaft grant BL511/10-1 and MO 2889/1-1.

2 The PTAS (Proof of Theorem 1)

Consider a set P of n points in $\mathbb{R}^d$, a threshold value ℓ, and a precision parameter $\epsilon > 0$. Given these inputs, a PTAS for the MSTSP problem is an algorithm with running time polynomial in n, required to return "yes" if a tour of P exists with scatter (i. e., shortest length) at least ℓ. The algorithm is required to return "no" if there is no tour of P with scatter at least $\ell(1-\epsilon)$, and is otherwise allowed to return "yes" or "no" arbitrarily.

Such a PTAS is an approximation algorithm for the *decision version* of the MSTSP problem. Observe that the optimum value of the MSTSP problem can only take one of $\binom{n}{2}$ possible values (the distances between points in P). Thus, a binary search over the possible values turns a PTAS of the above kind into a PTAS for the optimization problem. In the following, we focus on the decision problem. Before proceeding to the algorithm, we present some structural observations upon which the algorithm relies.

Given a point set $P \in \mathbb{R}^d$, let G_P be a graph with vertex set $V(G_P) = P$ and edge set $E(G_P) = \{\{x, y\} \mid x, y \in P \ \wedge \ d(x, y) \geq \ell\}$. In words, G_P contains all edges with length at least ℓ. The MSTSP decision problem asks whether G_P contains a Hamiltonian cycle. The following result is well-known (see e. g., [4]), and is also used by Arkin et al.

Lemma 2 (Dirac's theorem) *A graph G with n vertices has a Hamiltonian cycle if the degree of every vertex in G is at least $\frac{n}{2}$. Furthermore, in such a case, a Hamiltonian cycle can be found in $O(n^2)$ time.*

Observe that if the condition of Lemma 2 holds for G_P, then we are done. If that is not the case, then there is a vertex in G_P, whose degree is less than $\frac{n}{2}$. In other words, there is a point $p \in P$, such that $|B_p \cap P| > \frac{n}{2}$, where B_p is the open ball of radius ℓ with center p. Let us fix p to be such a point, and let B'_p be the open ball of radius 2ℓ with center p. We show that the optimal solution can be assumed to have a certain structure in relation to B_p and B'_p.

Lemma 3 *Suppose a tour T of P with scatter at least ℓ exists. Then there exists a tour T' of P with scatter at least ℓ, such that for every pair $x, y \in P$ of neighboring points in T', at least one of x and y is contained in B'_p.*

Proof. Suppose this is not the case. Since B_p contains more than half of the points in P, it must contain at least one edge of T entirely. Let $\{z, t\}$ be such an edge. Since both x and y are outside of B'_p we have $d(x, t), d(x, z), d(y, t), d(y, z) \geq \ell$. Thus, we can replace the edges $\{x, y\}$ and $\{z, t\}$ in T, with either $\{x, z\}$ and $\{y, t\}$, or $\{x, t\}$ and $\{y, z\}$, depending on the ordering of the points in T. We obtain another tour with scatter at least ℓ, that no longer contains the edge $\{x, y\}$. We proceed in the same way until we have removed all edges with both endpoints outside of B'_p. See Fig. 1 for an illustration. ◻

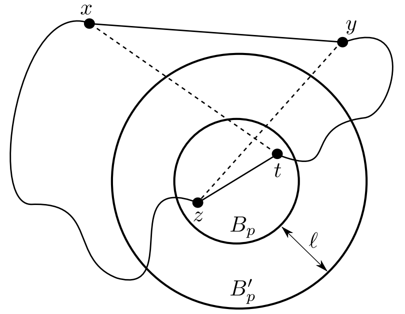

Fig. 1: Illustration of Lemma 3. The dashed edges can replace {x,y} and {z,t} in the optimal tour.

The next ingredient of the algorithm is a coarsening of the input, by rounding points in P to points of a grid. Let $\mathbb{G}_\delta$ be a δ-scaling of the d-dimensional unit grid, i. e., $\mathbb{G}_\delta = \{\delta(n_1, \ldots, n_d) \mid n_1, \ldots, n_d \in \mathbb{Z}\}$, for an arbitrary $\delta > 0$. Let f_δ (or simply f) be the mapping from $\mathbb{R}^d$ to $\mathbb{G}_\delta$ that maps each point to its nearest grid point (breaking ties arbitrarily). The following properties result from basic geometric considerations.

Lemma 4 *With f and δ as defined earlier, we have:*
(i) $d(x, y) \geq d(f(x), y) - \delta\sqrt{d}/2$ for all $x, y \in \mathbb{R}^d$,
(ii) $|B \cap \mathbb{G}_\delta| \leq (2\ell/\delta + 1)^d$ for every open ball B of radius ℓ.

Observe that f maps the graph G_P to a multigraph H_P defined as follows. Let $V(H_P) = \{v \mid v = f(x), x \in P\}$, i. e., the set of grid points with at least one mapped point of P, and let $E(H_P) = \{\{u, v\} \mid u = f(x), v = f(y), \{x, y\} \in E(G_P)\}$, i. e., the pairs of grid points to which edges of G_P are mapped. We also maintain multiplicities on edges of H_P, i. e., we keep track of how many edges of G_P are mapped to each edge of H_P.

A tour T of P (i. e., a Hamiltonian cycle of G_P) is mapped by f to an *Eulerian tour* of H_P. It is not hard to see that given this Eulerian tour of H_P, a tour of P can be recovered (by replacing multiple occurrences of a grid point with the points in P that are mapped to it). Moreover, the scatter of the recovered tour is not far from that of T (by Lemma 4(i)). However, the edges of H_P are not available to us, and thus it seems prohibitively expensive to guess a correct Eulerian tour on the vertices of H_P (there may be $\Omega(n)$ vertices). The key insight of the algorithm is that it is sufficient to consider the portion of H_P that falls inside B'_p.

At a high level, the strategy to obtain an approximation is the following (see Fig. 2 for an illustration). Assuming that the optimal tour has the property from Lemma 3, it consists of edges inside B'_p, and "hops" of two consecutive edges, connecting a point outside B'_p with two points inside B'_p. We replace such hops with virtual edges, both of whose endpoints are in B'_p. The

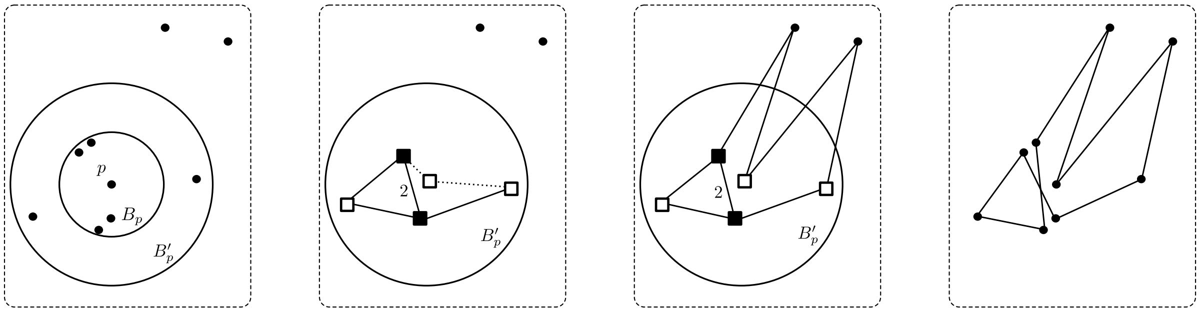

Fig. 2: Illustration of the algorithm. (i) Input point set with open balls B_p and B'_p with center p and radii ℓ and 2ℓ respectively. (ii) Points inside B'_p mapped to grid points (shown as squares), and "guessed" edges. Filled squares indicate grid points to which more than one point is mapped. Dotted lines indicate virtual edges, and the number indicates the multiplicity of an edge (omitted if 1). (iii) Virtual edges matched to points outside of B'_p and extended to hops, resulting in a multigraph. (iv) An Eulerian tour of the multigraph, expanded to a tour on the initial point set.

resulting tour is entirely in B'_p, and we can "guess" its image under f. This is now feasible, since the number of grid points involved is bounded by Lemma 4(ii). We also guess the multiplicities of all edges, i. e., how many original edges have been mapped to each edge, and how many edges are virtual.

We then disambiguate the virtual edges, i. e., we find a suitable midpoint outside of B'_p for each hop. This is achieved by solving a perfect matching problem. We obtain a multigraph on which we find an Eulerian tour. Finally, from the Eulerian tour we recover a tour of P. The distortion in distances due to rounding (i. e., the approximation ratio) is controlled by the choice of the grid resolution δ.

We only focus on answering whether a tour with the required scatter value exists. It will be clear that *constructing* such a tour can be achieved with minor changes to the algorithm. More details follow.

Algorithm. INPUT: a set P of n points in $\mathbb{R}^d$, a threshold ℓ, and a precision parameter $\epsilon > 0$.

1. Set $\delta = \epsilon\ell/(2\sqrt{d})$, and let $\ell' = \ell(1 - \epsilon/2)$.
2. Find $p \in P$ such that $|B_p \cap P| > \frac{n}{2}$, where B_p and B'_p are the open balls with center p of radius ℓ and 2ℓ. If no such p exists, output YES.
3. Let $f\colon P \to \mathbb{G}_\delta$ map points to their nearest grid point. Compute the set $C = \{f(x) \mid x \in (P \cap B'_p)\}$, and for each $v \in C$, compute the sets $f^{-1}(v) = \{x \mid f(x) = v\}$.
4. Let $m, v : \binom{C}{2} \to \mathbb{N}$. For all $\{u, v\} \subseteq C$, guess $m(\{u, v\})$ and $v(\{u, v\})$, such that
 (i) $m(\{u, v\}) = 0$ if $d(u, v) < \ell'$, and
 (ii) for all $v \in C$:
 $$\sum_{u \in C \setminus \{v\}} \big(m(\{u, v\}) + v(\{u, v\})\big) = 2|f^{-1}(v)|.$$
5. Construct a bipartite graph B as follows:
 - for each $\{u, v\} \subseteq C$, add $v(\{u, v\})$ vertices labeled $\{u, v\}$ to left vertex set $L(B)$.
 - for each $x \in P \setminus B'_p$ add a vertex labeled x to the right vertex set $R(B)$.
 - add an edge $(\{u, v\}, x)$ between $\{u, v\} \in L(B)$ and $x \in R(B)$ to $E(B)$ iff $d(u, x), d(v, x) \geq \ell'$.
6. Find a perfect matching M of B; if there is none, output NO.
7. Construct a multigraph H as follows:
 - let $V(H) = C \cup (P \setminus B'_p)$.
 - for all $\{u, v\} \subseteq C$ add $m(\{u, v\})$ copies of the edge $\{u, v\}$ to $E(H)$.
 - for all $(\{u, v\}, x) \in M$ add the edges $\{u, x\}$ and $\{v, x\}$ to $E(H)$.
8. Find an Eulerian tour Q of H; if there is none, output NO.
9. Transform Q into a tour T of P, by replacing multiple occurrences of every point $v \in C$, with the points in $f^{-1}(v)$ in arbitrary order.
10. Output YES.

Note. The "guessing" in step 4 should be thought of as a loop over all possible values of m and v satisfying the requirements. The overall output is NO if the output is NO for all possible values of step 4.

Correctness. We prove two claims which together imply that the algorithm is a PTAS for MSTSP: (1) if the algorithm outputs YES, then there is a tour of P with scatter at least $\ell(1 - \epsilon)$, and (2) if there is a tour of P with scatter at least ℓ, then the algorithm outputs YES.

(1) If we obtain YES in step 2, then we have a tour with scatter at least ℓ by Lemma 2. Suppose that the algorithm returns YES in step 10. This means that steps 5–9 were successful with the values of m and v chosen in step 4, and T is a tour of P. Consider an arbitrary edge $\{x, y\}$ of T. By step 9, there is a corresponding edge $\{u, v\}$ in the Eulerian tour Q of H. By construction of H in step 7, either (a) $u, v \in C$, or (b) $(\{u, w\}, v) \in M$ or $(\{v, w\}, u) \in M$, for some grid point $w \in C$.

In case (a) by condition (i) of step 4, we have $d(u, v) \geq \ell'$. Since $\{u, v\} = \{f(x), f(y)\}$, from Lemma 4(i) we obtain $d(x, y) \geq \ell' - \delta\sqrt{d} = \ell(1 - \epsilon)$.

In case (b) by the construction of B in step 5, we have $d(u, v) \geq \ell'$. Since $\{u, v\}$ equals either $\{f(x), y\}$ or $\{x, f(y)\}$, from Lemma 4(i) we obtain $d(x, y) \geq \ell' - \delta\sqrt{d}/2 \geq \ell(1 - \epsilon)$.

(2) Assume now that a tour T of P with scatter at least ℓ exists, and that the solution is not trivially found in step 1. Assume also w. l. o. g. that T has the special structure described in Lemma 3, i. e., it consists of hops and of edges entirely inside B'_p. Consider an edge $\{x, y\}$ of T, such that $x, y \in B'_p$. Then, after step 3, $f(x), f(y) \in C$ holds, and we say that $\{x, y\}$ *maps* to $\{f(x), f(y)\}$. Consider now a hop of T, i. e., two consecutive edges $\{x, w\}$ and $\{w, y\}$, such that $x, y \in B'_p$ and $w \in P \setminus B'_p$. Then, after step 3, $f(x), f(y) \in C$ holds, and we say that the hop $\{x, w, y\}$ *virtually maps* to $\{f(x), f(y)\}$.

Consider now the values m and v guessed in step 4, and let $m^*(\{u, v\})$ be the number of edges in T that map to $\{u, v\}$, and let $v^*(\{u, v\})$ be the number of hops in T that virtually map to $\{u, v\}$. Since every point in T has degree 2, it follows that the number of edges and hops mapped to an edge incident to some $u \in C$ is twice the number of points in P mapped to u. Furthermore, for all edges $\{x, y\} \subseteq B'_p$ of T, we have $d(f(x), f(y)) \geq \ell - \delta\sqrt{d} = \ell'$ (by Lemma 4(i)). Therefore, guessing the correct values $m = m^*$ and $v = v^*$ is consistent with the conditions in step 4.

Let $\{x_1, w_1, y_1\}, \ldots, \{x_k, w_k, y_k\}$ denote all the hops in T, where $w_i \in P \setminus B'_p$, for all i. Let $u_i = f(x_i)$, and $v_i = f(y_i)$, and let us call $M(T) = \big\{ \big(\{u_i, v_i\}, w_i\big) \mid i = 1, \ldots, k \big\}$ the *hop-matching* of T. Observe that $M(T)$ is a valid perfect matching for the graph B constructed in step 5, therefore, step 6 will succeed. We cannot, however, guarantee that $M(T)$ will be revovered in step 6. Observe that any other perfect matching M of B corresponds to a shuffling of the points w_i in B, and thus it is a hop-matching of a tour T' in which the points w_i have been correspondingly shuffled. T' differs from T only in its hops, and by construction of B in step 5, we see that T' must have a scatter at least $\ell' - \delta\sqrt{d}/2$.

It can be seen easily that the edges of T' are mapped to an Eulerian tour of the multigraph H constructed in step 7, and thus, step 8 succeeds. Again, we cannot guarantee that the recovered Eulerian tour is the same as the one to which T' maps. Any Eulerian tour of H, however, must respect the edge-multiplicities of H, which in turn are determined by the number of points that map to each vertex of H. Therefore, step 9 must succeed, and the output is Yes.

Note. The restrictions on m and v in step 4 can be strengthened, resulting in a smaller number of iterations (and thus better running time). For instance, since each virtual edge corresponds to a hop via a point outside of B'_p, we could require the values of v to sum to $|P \setminus B'_p|$. We ignore such technicalities, as they do not affect the correctness of the algorithm – in the case of wrong values, we get the No output in some of the later steps.

Running time. The cost of steps 1–3 is dominated by the cost of the loop starting in step 4. We observe that by Lemma 4(ii), $|C| \leq (9.8\sqrt{d}/\epsilon)^d$. Steps 5 and 6 amount to finding a perfect matching, and steps 7 and 8 amount to finding an Eulerian tour, both in a graph with $O(n)$ vertices. As for step 4, observe that the values of the functions m and v over all pairs in C sum to $|P \cap B'_p| \leq n$, so we need to consider at most $\binom{n}{|C|^2}$ ways of distributing a value of at most n into the integer values of m and v. Multiplying, and using a standard bound on the binomial, we obtain that the running time is at most $O(n^{|C|^2+3}) = O\left(n^{(100d/\epsilon^2)^d}\right)$.

This concludes the proof of Theorem 1.

References

[1] E. M. Arkin, Y. Chiang, J. S. B. Mitchell, S. Skiena, and T. Yang. On the maximum scatter TSP. In *SODA 1997*.

[2] S. Arora. Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, 1998.

[3] A. I. Barvinok. Two algorithmic results for the traveling salesman problem. *Math. Oper. Res.*, 21(1):65–84, 1996.

[4] B. Bollobás. *Modern Graph Theory.* Springer, 1998.

[5] N. Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Carnegie Mellon University, 1976.

[6] E. D. Demaine, J. S. B. Mitchell, and J. O'Rourke. The open problems project, 2012.

[7] S. P. Fekete. Simplicity and hardness of the maximum traveling salesman problem under geometric distances. In *SODA, 1999.*

[8] G. Gutin, A. Punnen, A. Barvinok, E. K. Gimadi, and A. I. Serdyukov. *The Traveling Salesman Problem and Its Variations.* Springer, 2002.

[9] M. Karpinski, M. Lampis, and R. Schmied. New inapproximability bounds for TSP. *J. Comput. Syst. Sci.*, 81(8):1665–1677, 2015.

[10] L. Kowalik and M. Mucha. Deterministic - approximation for the metric maximum TSP. *Theoretical Comp. Sc.*, 410(4749):5000 – 5009, 2009.

[11] J. S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric tsp, k-mst, and related problems. *SIAM J. Comput*, 28:402–408, 1996.

[12] R. Parker and R. L. Rardin. Guaranteed performance heuristics for the bottleneck travelling salesman problem. *Oper. Res. Lett.*, 2(6):269–272, Mar. 1984.

[13] J.-R. Sack and J. Urrutia, eds. *Handbook of Computational Geometry.* North-Holland Publishing Co., 2000.