A Probabilistic Analysis of Christofides' Algorithm ⋆

Markus Bläser[1] and Konstatinos Panagiotou[2] and B. V. Raghavendra Rao[1]

[1] Department of Computer Science, Saarland University
{mblaeser,bvrr}@cs.uni-saarland.de
[2] Department of Mathematics, University of Munich
kpanagio@math.lmu.de

Abstract. Christofides' algorithm is a well known approximation algorithm for the metric travelling salesman problem. As a first step towards obtaining an average case analysis of Christofides' algorithm, we provide a probabilistic analysis for the stochastic version of the algorithm for the Euclidean traveling salesman problem, where the input consists of n randomly chosen points in $[0,1]^d$. Our main result provides bounds for the length of the computed tour that hold almost surely. We also provide an experimental evaluation of Christofides's algorithm.

1 Introduction

The Traveling Salesman Problem, TSP for short, is a well-known NP-hard combinatorial optimization problem. For general edge weights, it is even NP-hard to find any sub-exponential approximation, see e.g. [10]. One natural restriction is the case where the edge weights fulfill the triangle inequality. The problem remains NP-hard (and APX-hard) for this restriction as well, but constant factor approximation algorithms are well known, like Christofides' algorithm [6] or the tree doubling algorithm [7]. Euclidean Traveling Salesman Problem (ETSP for short) is the restriction of metric TSP, where the vertices of the graph are points in $\mathbb{R}^d$ and the edge weights are the Euclidean distances between them. ETSP is also NP-hard to compute exactly, however efficient approximation schemes are known [1, 13].

There has been a lot of interest to understand the asymptotic behavior of Euclidean combinatorial optimization problems, in particular ETSP. In their seminal paper [4], Beardwood, Halton and Hammersley performed a probabilistic analysis, where they showed the following remarkable result:

Theorem 1. *Let $d \geq 2$. Let $U_1, \ldots, U_n$ be n independent uniformly distributed random points over $[0,1]^d$. There exists a constant $\gamma_{\mathsf{ETSP}} = \gamma_{\mathsf{ETSP}}(d) > 0$ such that almost surely*

$$\lim_{n\to\infty} \frac{\mathsf{ETSP}(U_1, \ldots, U_n)}{n^{(d-1)/d}} = \gamma_{\mathsf{ETSP}}.$$

⋆ This project is partially funded by the DFG grant BL 511/7-1. The work was performed while the second author was affiliated with Max Planck Institute for Informatics in Saarbrücken.

In words, the authors of [4] showed that when we draw n points uniformly at random from $[0,1]^d$, the cost of an optimal tour is almost surely asymptotically equal to $\gamma_{\mathsf{ETSP}}\, n^{(d-1)/d}$. Later on, this result was generalized to many other problems on Euclidean spaces, like the minimum spanning tree problem [17]. In particular, Steele [16] provided a general framework that provides similar results for all Euclidean functionals that are sub-additive (see Definition 1).

Motivated by the result of Beardwood *et al.*, Karp [12] gave a partitioning heuristic for ETSP that runs in polynomial time and asymptotically outputs an optimal tour with probability 1 over points uniformly sampled from $[0,1]^d$. Starting with Karp's work, there has been a lot of interest in the probabilistic analysis of heuristic algorithms for Euclidean optimization problems. For instance, Avis, Davis, and Steele [2] showed complete convergence for the greedy algorithm for the Euclidean minimum matching problem and Goemans and Bertsimas [9] provided the almost sure asymptotics for the Held-Karp relaxation of ETSP. For an overview of further results, we refer to [8, 18, 19, 3].

In this paper we study Christofides' algorithm, a popular heuristic for metric TSP. It starts with computing a minimum spanning tree and then a minimum matching on the odd-degree vertices. In the resulting graph, every node has even degree and therefore, the graph is Eulerian. We obtain a TSP tour by taking shortcuts in the Eulerian tour. For arbitrary metrics, Christofides' algorithm achieves a 3/2-approximation.

Despite its worst case approximation ratio of 3/2, Christofides' algorithm is known to perform better in practice. Analyzing Christofides' algorithm on random point sets has been an open problem, as posed by Frieze and Yukich in 2002 [8]. In this work, we introduce the functional CHR as the sum of cost of a minimum spanning tree and a minimum matching on the odd-degree vertices of the minimum spanning tree (see Section 3). Clearly, CHR serves as a worst case upper bound on the cost of the tour computed by Christofides' algorithm. We prove:

Theorem 2. *Let $d \geq 2$. Let $U_1, \ldots, U_n$ be n i.i.d. uniformly distributed random points over $[0,1]^d$. There exists a constant $\gamma_{\mathsf{CHR}} = \gamma_{\mathsf{CHR}}(d) > 0$ such that almost surely*

$$\lim_{n\to\infty} \frac{\mathsf{CHR}(U_1, \ldots, U_n)}{n^{(d-1)/d}} = \gamma_{\mathsf{CHR}}.$$

As a main ingredient in our proof, we show that the functional CHR satisfies a weak form of geometric sub-additivity (see Definition 2). Then we show that the techniques developed by Steele [16] can be extended to weakly sub-additive functionals.

Note that this result for Christofides' algorithm is not a consequence of the results known for Euclidean minimum spanning trees and Euclidean minimum matching, for the matching computed by the Christofides' algorithm depends on the odd degree vertices in the minimum spanning tree. Moreover, it is not clear if Christofides' functional is sub-additive, and thus it is not possible to apply known methods. However, we show that Christofides' functional fulfills a weaker property, which is still strong enough to obtain the desired results. A related approach to overcome the limitations of the classical methods was taken by Baltz *et al.* [3], who studied a routing problem with multiple depots.

Before we proceed with the proofs we comment on the values of the constants γ_{ETSP} and γ_{CHR} in Theorems 1 and 2. It is well-known that the worst-case approximation ratio of Christofides' algorithm is $3/2$. On the other hand, the theorems above imply that the approximation ratio is at most $\gamma_{\mathsf{CHR}}/\gamma_{\mathsf{TSP}}$ on almost all pointsets. It is an intriguing question whether this ratio is $< 3/2$. However, a numerical or analytical evaluation seems from a current perspective very difficult: although numerous efforts have been made in the past, see e.g. [14, 5, 11] and many references therein, the constant γ_{ETSP} is not known exactly. Similarly, as in all previous proofs regarding the asymptotic behavior of Euclidean functionals, our proof does not provide any way of computing the value of γ_{CHR}. On the positive side, our experimental evaluation on random points shows that γ_{CHR} is strictly below 1.5, even without shortcutting (see Section 5).

2 Preliminaries

Most of the notations used here are from [19]. In this paper the distance between two points in $[0,1]^d$ is always the Euclidean distance.

Euclidean Functionals Let $d > 1$ be a fixed dimension. An *Euclidean functional* in dimension d is a function f that maps any finite point set $X \subset \mathbb{R}^d$ to a positive real number $f(X)$. We use the following Euclidean functionals in the paper:

- $\mathsf{TSP}(X)$ = total edge weight of a minimum Euclidean traveling salesman tour of X.
- $\mathsf{MM}(X)$ = total edge weight of a minimum weight Euclidean perfect matching of X. (If $|X|$ is odd, then one point will be left unmatched.)
- $\mathsf{MST}(X)$ = total edge weight of a minimum Euclidean spanning tree of X.

Following standard notation, for an Euclidean functional f and a hypercube $\mathcal{H} \subset \mathbb{R}^d$, $f(.,\mathcal{H})$ will denote the restriction of f, when the points are restricted to $\mathcal{H}$. In other words, for any point set X we have that $f(X,\mathcal{H}) = f(X \cap \mathcal{H})$. We define certain properties of Euclidean functionals that will be used throughout.

An Euclidean functional f is *monotone* if for every $F \subseteq G$, $f(F,\mathcal{H}) \leq f(G,\mathcal{H})$. Moreover, f is said to be *homogeneous* if

$$\forall\, \alpha > 0: \quad f(\alpha\, F, \alpha\, \mathcal{H}) = \alpha \cdot f(F,\mathcal{H}),$$

where $\alpha\, X = \{\alpha\, x : x \in X\}$ for any $X \subseteq \mathbb{R}^d$. We will say that f is *translation invariant* if

$$\forall\, a \in \mathbb{R}^d: \quad f(F,\mathcal{H}) = f(F + a, \mathcal{H} + a)$$

where $X + a = \{x + a : x \in X\}$ for $X \subseteq \mathbb{R}^d$. We say that f admits a *growth bound*, if there is a constant $C > 0$ such that

$$f(F,[0,1]^d) \leq C|F|^{(d-1)/d}, \tag{1}$$

Finally, f is called *smooth* if there is a constant C such that for all $F, G \subset [0,1]^d$,

$$|f(F \cup G, [0,1]^d) - f(F, [0,1]^d)| \leq C|G|^{(d-1)/d}. \tag{2}$$

A further important property of Euclidean functionals is *sub-additivity.*

Definition 1 (Sub-additivity). *Let $Q_1, \ldots, Q_{m^d}$ be a partition of $[0,1]^d$ into equal-sized sub-cubes of edge m^{-1} each. Then f is sub-additive if there is a $C = C(d) \geq 0$ such that for $m \in \mathbb{Z}^+$,*

$$f(F, [0,1]^d) \leq \sum_{i=1}^{m^d} f(F, Q_i) + Cm^{d-1}. \tag{3}$$

Sub-additivity is the one of the most important properties used in several studies of Euclidean functionals, in particular $\mathsf{TSP}, \mathsf{MM}$, and MST, see [8, 19] for an excellent survey.

Proposition 1. *The Euclidean functionals* $\mathsf{TSP}, \mathsf{MM}$*, and* MST *are homogeneous, translation invariant, smooth, sub-additive, and admit a growth bound.*

The Christofides' functional is not sub-additive. However, we define a weaker property that turns out to be sufficient for our analysis of the Christofides' functional.

Definition 2 (Weak Sub-additivity). *Let $Q_1, \ldots, Q_{m^d}$ be a partition of the unit box $[0,1]^d$ into equal-sized sub-cubes of edge m^{-1} each. Then, f is said to be weakly sub-additive, if there are constants $C = C(d), C' = C'(d) \geq 0$ and $\epsilon = \epsilon(d) > 0$ such that for $m \in \mathbb{Z}^+$,*

$$f(F, [0,1]^d) \leq \sum_{i=1}^{m^d} f(F, Q_i) + Cn^{((d-1)/d)-\epsilon} m^{\epsilon} + C'm^{d-1}. \tag{4}$$

We will use the following (folklore) facts about Euclidean minimum matching. We give a simple proof for completeness.

Lemma 1. *Let $A, B \subset [0,1]^d$ be two finite, disjoint sets of points with even cardinality. Then*

$$\mathsf{MM}(B) \leq \mathsf{MM}(A \cup B) + \mathsf{MM}(A)$$

Proof. Let M be a minimum matching of $A \cup B$ and M_1 be that of A. Consider the graph $G = (A \cup B, M_1 \cup M_2)$. For every $u \in B$, there is a unique path P_u originating at u. Moreover, P_u ends at some $u' \in B$, and all the remaining vertices in P are from A. This gives a matching for B of the required cost.

Lemma 2 (Folklore). *Let T be a MST of n points. Then the cost of minimum matching on the odd-degree vertices of T is bounded by* $\mathsf{MST}(T)$*.*

Further Notation If S is any collection of edges in a graph, and v is a vertex then $\Delta_S(v)$ denotes the degree of v in the sub-graph induced by S. Moreover, we will write $\|S\|$ for the sum of the lengths of the edges in S.

3 Proof of the Main Result

In this section we present the main steps that are needed to achieve the proof of Theorem 2. We begin with defining the Euclidean functional given by Christofides algorithm. For any point set $F \subset [0,1]^d$ let $\mathsf{CHR}(F)$ denote the cost of a minimum spanning tree T of the points in F plus the cost of a minimum matching of the odd-degree points in T. In symbols, we write

$$\mathsf{CHR}(F) \stackrel{\triangle}{=} \mathsf{MST}(F) + \mathsf{OM}(F)$$

where $\mathsf{OM}(F)$ denotes the cost of a minimum matching on the odd degree vertices in the minimum spanning tree. When F has more than one minimum spanning trees, $\mathsf{CHR}(F)$ is defined as the minimum over all such trees. However, we will not consider such exceptional cases in our analysis, since the spanning tree of a random point set is unique.

Note that strictly speaking, the functional CHR defined above does not measure the length of the tour obtained by Christofides' algorithm, as we ignore shortcuts. This is done in order to have more structure in the functional CHR, even though it weakens the analysis a bit.

The first step in our proof is to establish the following lemma about the structure of the functional CHR. It is the main contribution of our paper, and it is proved in Section 4.

Lemma 3. *The Euclidean functional* CHR *is homogeneous, translation invariant, smooth, and admits a growth bound. Moreover, it is weakly sub-additive.*

With this fact at hand, we proceed with proving a general result that determines the asymptotic value of the expectation of a weakly sub-additive Euclidean functional. This result, together with the proof, are generalizations of the corresponding theorems that hold for sub-additive functionals only, and thus they can be applied to a wider class of functions. Due to space limitations, the proof is omitted.

Theorem 3. *Let $d \geq 2$. Let f be a smooth, weakly sub-additive Euclidean functional that admits a growth bound. There is a $\gamma_f = \gamma_f(d)$ such that if $U_1, \ldots, U_n$ are uniform i.i.d over $[0,1]^d$, then*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[f(U_1,\ldots,U_n)\right]}{n^{(d-1)/d}} = \gamma_f.$$

Together with Lemma 3, the above theorem implies that there is a $\gamma_{\mathsf{CHR}} \geq 0$ such that

$$\mathbb{E}\left[\mathsf{CHR}(U_1,\ldots,U_n)\right] = (1+o(1)) \cdot \gamma_{\mathsf{CHR}} \cdot n^{(d-1)/d}. \quad (5)$$

However, as $\gamma_{\mathsf{MST}} > 0$, see [8], we also obtain that $\gamma_{\mathsf{CHR}} > 0$.

Note that the above collection of arguments almost shows Theorem 2. To complete the proof it remains to show that $\mathsf{CHR}(U_1,\ldots,U_n)$ is typically very close to its expected value. This is performed by the next well-known result, which follows immediately from the arguments exposed in [15].

Theorem 4. *Let f be a homogeneous, translation invariant and smooth Euclidean functional. Suppose that there is a $\gamma_f = \gamma_f(d)$ such that*

$$\lim_{n\to\infty} \frac{\mathbb{E}[f(U_1,\ldots,U_n)]}{n^{(d-1)/d}} = \gamma_f.$$

Then there is a $C = C(d) > 0$ such that for sufficiently large n

$$\mathbb{P}\left[|f(U_1,\ldots,U_n) - \mathbb{E}[f(U_1,\ldots,U_n)]| > t\right] \le \exp\left\{-C\frac{t^{2d/(d-1)}}{n}\right\}.$$

Theorem 2 is then a direct consequence of (5) together with Lemma 3 and the above result, which we apply with, say, $t = n^{2(d-1)/3d} = o(n^{(d-1)/d})$.

4 Christofides' functional

This section is devoted to the proof of Lemma 3. We start with proving all the properties except for the weak sub-additivity of it.

Lemma 4. CHR *is a homogeneous, translation invariant, smooth Euclidean functional that admits a growth bound.*

Proof. As translation or scaling does not change the relative distances between the points, CHR is homogeneous and translation invariant. The growth bound can be obtained by that of MST and minimum matching, see [8, 19].

We now argue that CHR is also smooth. Let $D(d)$ denote the bound on the maximum degree in any Euclidean minimum spanning tree of a d-dimensional pointset. Note that $D(d)$ depends only on d. Let $F, G \subset [0,1]^d$ be any finite sets of points. Let T be a minimum spanning tree of F and $O \subseteq F$ denote the set of odd-degree points in T. Let T' be the minimum spanning tree of $F \cup G$, obtained by iteratively adding points from G, one at a time, and then adding/removing necessary edges to/from T. Let us examine the first step in the above procedure. Let $v \in G$, and let T_1 be a minimum spanning tree of $F \cup \{v\}$. Then by the degree bound on the Euclidean minimum spanning tree, v can have at most $D(d)$ incident edges in any MST of $F \cup \{v\}$. For each such edge there is at most one edge in T that has to be removed to ensure the acyclicity of T_1. So, each edge incident to v can affect the degrees of at most 3 points. Thus, the degrees of at most $3D(d)$ points in T_1 are different from that in T, and we infer that in total at most $3D(d)|G|$ points will have their degrees in T' different from that in T. Let O' denote the set of odd degree vertices in T'. The above discussion guarantees that

$$||O| - |O'|| \le 3D(d)|G|.$$

As $O \cap O'$ and $O \setminus O'$ form a partition of O, we have:

$$\mathsf{MM}(O) \le \mathsf{MM}(O \cap O') + \mathsf{MM}(O \setminus O') + t_1, \tag{6}$$

where t_1 is the cost of a single edge if $|O \cap O'|$ is odd and zero otherwise. By Lemma 1 with $B = O \cap O'$, $A = O' \setminus (O \cap O')$ (hence $A \cup B = O'$),

$$\mathsf{MM}(O \cap O') \leq \mathsf{MM}(O') + \mathsf{MM}\left(O' \setminus (O \cap O')\right).$$

Substituting in (6), and using that $|O \setminus O'|, |O' \setminus (O \cap O')| \leq 3D(d)|G|$ and $\mathsf{MM}(X) = O(|X|^{(d-1)/d})$ for any pointset $X \subset [0,1]^d$

$$\begin{aligned}\mathsf{MM}(O) &\leq \mathsf{MM}(O') + \mathsf{MM}(O \setminus O') + \mathsf{MM}\left(O' \setminus (O \cap O')\right) + t_1 \\ &\leq \mathsf{MM}(O') + 3(3D(d)|G|)^{(d-1)/d}\end{aligned}$$

By interchanging the roles of O and O' in the above argument, we can show similarly that

$$\mathsf{MM}(O') \leq \mathsf{MM}(O) + \mathsf{MM}(O' \setminus O) + \mathsf{MM}(O \setminus O') + t_2$$

where t_2 is the cost of a single edge in $[0,1]^d$. Hence,

$$\mathsf{MM}(O') \leq \mathsf{MM}(O) + 3(3D(d)|G|)^{(d-1)/d}$$

Thus we have shown that there is a $C = C(d) > 0$ such that

$$\mathsf{OM}(F) - C|G|^{(d-1)/d} \leq \mathsf{OM}(F \cup G) \leq \mathsf{OM}(F) + C|G|^{(d-1)/d}. \tag{7}$$

By the definition of CHR and the triangle inequality, we have

$$|\mathsf{CHR}(F) - \mathsf{CHR}(F \cup G)| \leq |\mathsf{MST}(F) - \mathsf{MST}(F \cup G)| + |\mathsf{OM}(F) - \mathsf{OM}(F \cup G)|.$$

As MST is a smooth functional, $|\mathsf{MST}(F) - \mathsf{MST}(F \cup G)| = O(|G|^{(d-1)/d})$. Combining this with (7) then proves the claim.

We cannot show that CHR is sub-additive. However, we show that it satisfies a weaker form of subadditivity, which, however, is sufficient for our purposes.

Lemma 5. CHR *is weakly sub-additive.*

Before proving Lemma 5, we prove some of the structural properties of Euclidean minimum spanning trees and minimum matchings.

Notation We use the following notation in Lemmas 6 and 7. Let T be a minimum spanning tree of a finite point set $F \subset [0,1]^d$. Let $Q_1, \ldots Q_{m^d}$ be the partitioning of $[0,1]^d$ into sub-cubes side length m^{-1} each. An edge $e = (u,v)$ in T is called a *boundary edge*, if $u \in Q_i$ and $v \in Q_j$, where $i \neq j$. A boundary edge (u,v) of T is called *short*, if $Q_i \cap Q_j \neq \emptyset$. We shall say that Q_i and Q_j are adjacent in this case. Every boundary edge that is not short will be denoted as *long*. A point $v \in F$ is said to be a *boundary point*, if it is incident to at least one of the boundary edges of T. Let B denote the set of boundary points of T that are incident on short edges. Let $B_i = B \cap Q_i$. Let $r \leq m^{-1}$ be a parameter to be chosen later. Let Q_i and Q_j be two adjacent sub-cubes, and $\mathcal{B}_{i,j}$ denote the boundary between them (i.e, a sub-cube in dimension at most $d-1$). Let $C_1, \ldots, C_t$ denote the partitioning of $\mathcal{B}_{i,j}$ into sub-cubes of side length r each. As $\sum_{k=1}^{t} \mathsf{Vol}(C_i) = \mathsf{Vol}(\mathcal{B}_{i,j}) \leq m^{-(d-1)}$, we have

$$t \leq m^{-(d-1)}/r^{d-1}. \tag{8}$$

For $1 \leq k \leq t$, let $\mathcal{C}_{i,k}$ (resp. $\mathcal{C}_{j,k}$) denote the hyper-rectangle in Q_i (resp. Q_j) with C_k as one of its base face.

Lemmas 6 and 7 provide some structural properties of B. The proof of the next statement is not very difficult and can be found in the Appendix.

Lemma 6. *With the notations above, suppose that $\overline{AB}$ and $\overline{CD}$ are two boundary edges such that $A, D \in \mathcal{C}_{i,k}$ and $B, C \in \mathcal{C}_{j,k'}$ for some $1 \leq k, k' \leq t$. Then, at least one point each from $\{A, D\}$ and $\{B, C\}$ is at distance at most $\sqrt{d}r$ to $\mathcal{B}_{i,j}$.*

Corollary 1. *Let Q_i and Q_j be two adjacent sub-cubes. Then there are at most t^2 boundary edges between points in Q_i and Q_j of length more than $2\sqrt{d}r$.*

Proof. By Lemma 6, for a sub-rectangle $\mathcal{C}_{i,k}$, there are at most t boundary points in Q_j at a distance of at least $\sqrt{d}r$ from the boundary. As there are t such rectangles $\mathcal{C}_{i,k}$, we get the desired bound.

In Lemma 7 below, we bound the cost of a minimum spanning tree or a minimum matching for the points in B_i. (See Appendix for a proof.)

Lemma 7. *Suppose $n_i = |Q_i \cap F| \geq 1$. There is an $\epsilon = \epsilon(d) > 0$ such that the cost of a minimum matching or a minimum spanning tree of any subset of points in B_i is $O(m^{-1} n_i^{((d-1)/d)-\epsilon})$.*

We also bound the total edge length of long boundary edges in T. (See Appendix for a proof.)

Lemma 8. *The total length of all long boundary edges in T is $O(m^{d-1})$.*

Proof (of Lemma 5). Let T be a minimum spanning tree of F. Let O denote the set of odd degree points in T. Let T_i denote the restriction of T to Q_i obtained by removing the edges incident to points outside Q_i. Let T_i' denote a minimum spanning tree for $F_i = F \cap Q_i$ obtained by adding necessary edges to T_i. Let $S_i = E(T_i') \setminus E(T_i)$. Let O_i' denote the set of odd degree points in T_i'. Let $O_i = O \cap Q_i$. By definition, $\mathsf{CHR}(F, [0,1]^d) = \|T\| + \mathsf{MM}(O)$. We need to prove

$$\mathsf{CHR}(F, [0,1]^d) \leq \sum_{i=1}^{m^d} \mathsf{CHR}(F_i, Q_i) + O(n^{((d-1)/d)-\epsilon} m^{\epsilon} + m^{d-1}). \tag{9}$$

for some $\epsilon = \epsilon(d) > 0$. Applying the geometric sub-additivity of the Euclidean minimum spanning tree functional, and that of Euclidean minimum matching,

$$\mathsf{CHR}(F, [0,1]^d) \leq \sum_{i=1}^{m^d} \|T_i'\| + \sum_{i=1}^{m^d} \mathsf{MM}(O_i) + O(m^{d-1}). \tag{10}$$

By the definition of CHR, we have $\mathsf{CHR}(F_i) = \|T_i'\| + \mathsf{MM}(O_i')$. Thus it is sufficient to bound $\mathsf{MM}(O_i)$ in terms of $\mathsf{MM}(O_i')$. This is performed by the next claim.

Claim 5 $\mathsf{MM}(O_i) \le \mathsf{MM}(O'_i) + \|S_i\| + O(m^{-1} n_i^{((d-1)/d)-\epsilon} + m^{-1})$.

Applying the above claim on (10),

$$\mathsf{CHR}(F, [0,1]^d) \le \sum_{i=1}^{m^d} \Big(\|T'_i\| + \mathsf{MM}(O'_i) + \|S_i\| + C' m^{-1} n_i^{((d-1)/d)-\epsilon} + O(m^{-1}) \Big)$$

The crucial observation is that it is sufficient to replace $\|T'_i\|$ by $\|T_i\|$ with a small additive term in the above sum, since $\|T'_i\| = \|T_i\| + \|S_i\|$. This can be done using Lemma 7. We prove

$$\|T'_i\| \le \|T_i\| + O(m^{-1} n_i^{((d-1)/d)-\epsilon} + m^{-1})$$

To see this, note first that if T_i is connected, then $\|S_i\| = 0$, hence assume that T_i is not connected. Then at least one point in each of the connected components of T_i is a boundary point. So, T_i plus a spanning tree of all boundary points in F_i and a single edge connecting them gives a spanning tree τ_i of F_i, and hence $\|T'_i\| \le \|\tau_i\|$. The boundary points in F_i can be partitioned into B_i, and the remaining boundary points of F_i that are incident on long boundary points. By Lemma 7, we have $\mathsf{MST}(B_i) \le O(m^{-1} n_i^{((d-1)/d)-\epsilon}) + O(m^{-1})$. The boundary points that are incident on long boundary points can be connected arbitrarily to each other, as their total length is at most $O(m^{d-1})$ by Lemma 8. Thus,

$$\|T'_i\| \le \|\tau_i\| \le \|T_i\| + \mathsf{MST}(B_i) + O(m^{-1}) + \mathsf{MST}(\text{long boundary points in } F_i)$$

Thus there is a constant $C_1 = C_1(d) \ge 0$ such that,

$$\begin{aligned} \|T\| &\le \sum_{i=1}^{m^d} \|T_i\| + C_1 \cdot (n_i^{((d-1)/d)-\epsilon} + m^{-1}) + \mathsf{MST}(\text{long boundary points in } F_i) \\ &\le \sum_{i=1}^{m^d} \big(\|T_i\| + C_1 \cdot (n_i^{((d-1)/d)-\epsilon} + m^{-1}) \big) + O(m^{d-1}). \end{aligned}$$

Hence, there is a $C'' = C''(d) \ge 0$ such that

$$\begin{aligned} \mathsf{CHR}(F, [0,1]^d) &= \|T\| + \mathsf{MM}(O) \\ &\le \sum_{i=1}^{m^d} \Big(\|T_i\| + C_1 \cdot (m^{-1} n_i^{((d-1)/d)-\epsilon} + m^{-1}) \Big) \\ &\quad + \sum_{i=1}^{m^d} \Big(\mathsf{MM}(O'_i) + \|S_i\| + C' m^{-1} n_i^{((d-1)/d)-\epsilon} + O(m^{-1}) \Big) \\ &= \sum_{i=1}^{m^d} \Big(\|T_i\| + \|S_i\| + \mathsf{MM}(O'_i) + C'' m^{-1} n_i^{((d-1)/d)-\epsilon} + O(m^{-1}) \Big) \end{aligned}$$

As $\|T_i'\| = \|T_i\| + \|S_i\|$ and $\mathsf{CHR}(F, Q_i) = \|T_i'\| + \mathsf{MM}(O_i')$, we have that

$$\begin{aligned}\mathsf{CHR}(F,[0,1]^d) &\le \sum_{i=1}^{m^d} \Big\{ \mathsf{CHR}(F_i, Q_i) + C''(m^{-1} n_i^{((d-1)/d)-\epsilon} + m^{-1}) \Big\} \\ &\le \sum_{i=1}^{m^d} \mathsf{CHR}(F_i, Q_i) + C_2 \big(\sum_i n_i\big)^{((d-1)/d)-\epsilon} m^{\epsilon} \\ &\quad \text{(Hölder's inequality)} \\ &= \sum_{i=1}^{m^d} \mathsf{CHR}(F_i, Q_i) + C_2 n^{((d-1)/d)-\epsilon} m^{\epsilon}.\end{aligned}$$

To complete the proof of Lemma 5, we need to prove Claim 5. Due to space limitations, the proof is omitted and can be found in the full version of the paper.

5 Experimental Evaluation

In this section we present simulation results that shed some light on the actual values of the constants γ_{ETSP} and γ_{CHR}. In particular, we provide experimental evidence that the value of Christofides' functional is *strictly* less than 3/2 times the length of an optimal TSP tour through n random points.

Our experimental setup is as follows. Let $n_i = 500\,i$, where $1 \le i \le 20$. For any i in the given range, we generated independently 100 sets of n_i uniformly distributed random points in $[0,1]^2$, and computed the average and the standard deviation of three parameters: i) the size of a minimum spanning tree (MST), ii) the size of a minimum matching on the odd degree vertices (OM) of the minimum spanning tree, and iii) the ratio (MST + OM) / MST. Note that the latter is an upper bound for the approximation ratio of Christofides' algorithm, since the length of a minimum spanning tree is a lower bound for the length of a TSP tour.

The results of the experiments are summarized in Figures 1 and 2, and lead to the following conclusions. Observe that the ratio (MST + OM) / MST stabilizes quickly around approximately 1.3347, and the standard deviation becomes small very quickly. In other words, even if we do not perform any shortcutting, the approximation ratio stays well below the worst-case value 3/2.

References

1. S. Arora. Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, 1998.
2. D. Avis, B. Davis, and J. M. Steele. Probabilistic analysis of a greedy heuristic for euclidean matching. *Probability in the Engineering and Informational Sciences*, 2(02):143–156, 1988.

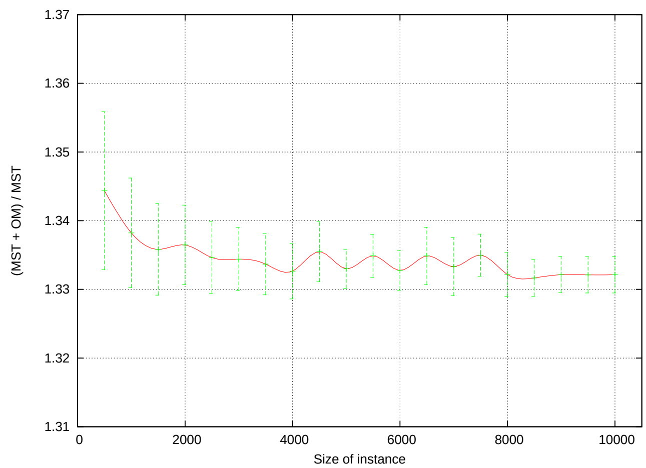

Fig. 1. An experimental upper bound for the value of the Christofides' functional divided by the length of a optimum TSP tour.

3. A. Baltz, D. P. Dubhashi, A. Srivastav, L. Tansini, and S. Werth. Probabilistic analysis for a multiple depot vehicle routing problem. *Random Struct. Algorithms*, 30(1-2):206–225, 2007.
4. J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Proc. Cambridge Philos. Soc.*, 55:299–327, 1959.
5. D. J. Bertsimas and G. van Ryzin. An asymptotic determination of the minimum spanning tree and minimum matching constants in geometrical probability. *Operations Research Letters*, 9(4):223 – 231, 1990.
6. N. Christofides. Worst-case analysis of a new heuristic for the traveling salesman problem. Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.
7. V. Deineko and A. Tiskin. Fast minimum-weight double-tree shortcutting for metric tsp: Is the best one good enough? *J. Exp. Algorithmics*, 14:4.6–4.16, 2010.
8. A. M. Frieze and J. E. Yukich. Probabilistic analysis of the traveling salesman problem. In G. Gutin and A. Punnen, editors, *The traveling salesman problem and its variations*, pages 257–308. Kluwer Academic Publisher, 2002.
9. M. X. Goemans and D. J. Bertsimas. Probabilistic analysis of the held and karp lower bound for the euclidean traveling salesman problem. *Mathematics of Operations Research*, 16(1):72–89, 1991.
10. G. Gutin and A. P. Punnen, editors. *The traveling salesman problem and its variations*, volume 12 of *Combinatorial Optimization*. Kluwer Academic Publishers, 2002.
11. D. S. Johnson, L. A. McGeoch, and E. E. Rothberg. Asymptotic experimental analysis for the held-karp traveling salesman bound. SODA '96, pages 341–350, 1996.

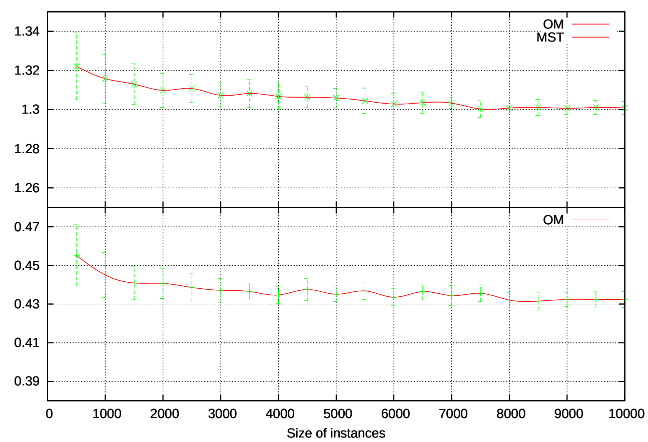

Fig. 2. The length of MST and OM.

12. R. M. Karp. Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane. *Math. of Operat. Research*, 2(3):209–224, 1977.
13. J. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.*, 28(4):1298–1309, 1999.
14. W. T. Rhee. On the travelling salesperson problem in many dimensions. *Random Struct. Algorithms*, 3(3):227–234, 1992.
15. W. T. Rhee. A matching problem and subadditive euclidean functionals. *Ann. Appl. Probab.*, 3(3):794–801, 1993.
16. J. M. Steele. Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.*, 9(3):365–376, 1981.
17. J. M. Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.*, 16:1767–1787, 1988.
18. J. M. Steele. *Probability Theory and Combinatorial Optimization*, volume 69 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1997.
19. J. E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*, volume 1675 of *Lecture Notes in Mathematics*. Springer-Verlag, 1998.