Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:
ECCC TR96-046
FTP: ftp.eccc.uni-trier.de:/pub/eccc/
WWW: http://www.eccc.uni-trier.de/eccc/
Email: ftpmail@ftp.eccc.uni-trier.de with subject 'help eccc'

On the Approximability of the Multi-dimensional Euclidean TSP

Luca Trevisan*

August 1996

Abstract

We consider the Traveling Salesperson Problem (TSP) restricted to Euclidean spaces of dimension at most $k(n)$, where n is the number of cities. We are interested in the relation between the asymptotic growth of $k(n)$ and the approximability of the problem. We show that the problem is Max SNP-hard when $k(n) = n - 1$. Thus, for a certain constant $\epsilon_1 > 0$, the $(n-1)$-dimensional Euclidean TSP cannot be approximated within a factor $(1+\epsilon_1)$ in polynomial time, unless P = NP. Using a previous result about embedding of Euclidean metrics in low-dimensional spaces, we can also prove that constants c and ϵ_2 exist such that approximating the $c \log n$-dimensional Euclidean TSP within $(1+\epsilon_2)$ is NP-hard under randomized reductions (and thus not solvable in P, unless RP = NP). This contrasts with the recent result by Arora [Aro96] who presented a polynomial time approximation scheme for the bidimensional Euclidean TSP. Additionally, we note extensions of our result to other geometric metrics.

1 Introduction

The Traveling Salesperson Problem (TSP) is defined as follows: given a set of "cities" $U = \{u_1, \ldots, u_n\}$ and a "distance function" $d : U \times U \to \mathbf{R}$, find a permutation π of $\{1, \ldots, n\}$ (also called "tour") that minimizes the "total length" (or "cost") $m((U,d),\pi) = d(u_{\pi[n]}, u_{\pi[1]}) + \sum_{i=1}^{n-1} d(u_{\pi[i]}, u_{\pi[i+1]})$.

Interest in such problem started during the 1930's. In 1966, the (already) long-standing failure of developing an efficient algorithm for the TSP led Edmonds [Edm66] to conjecture that the problem is not in P: this is sometimes referred to as the first statement of the P $\neq$ NP conjecture. Indeed, the general version of the TSP was proved NP-hard in the original Karp paper [Kar72]. More recently it was proved that the TSP remains NP-complete even if the cities are restricted to lie in the bidimensional plane and the distances are computed according to the Euclidean metric[1] [GGJ76, Pap77]. Due to those negative results, research concentrated on developing good heuristics. There is an extensive literature on this field, we will only review the results that are relevant for the present paper (see the book of Lawler et al. [LLKS85] for a very complete survey). Recall that an r-approximate algorithm ($r > 1$) is a polynomial-time heuristic that is guaranteed to deliver a tour whose cost is at most r times the optimum cost. Assuming P $\neq$ NP, in the general case, for any $r > 1$ there cannot be an r-approximate algorithm [SG76], however if the distance function

*Dept. of Computer Science, University of Rome "La Sapienza", Via Salaria 113, 00198 Rome, Italy. Email: trevisan@dsi.uniroma1.it

[1] In general, the $k(n)$-dimensional Euclidean TSP ($k(n)$ETSP) is the restriction of the TSP to the case in which cities lie in $\mathbf{R}^{k(n)}$ and the distances are computed according to the ℓ_2 norm.

satisfies the triangle inequality, then it possible to achieve a 3/2-approximation in polynomial time [Chr76]. In twenty years of research no improvement of this bound has been found, even in the restricted case of geometric metrics. In the late 1980's, the emergence of the theory of Max SNP-hardness [PY91] gave a means of possibly understanding this lack of results. Indeed, Papadimitriou and Yannakakis [PY93] proved that the TSP problem is Max SNP-hard even when restricted to metric spaces (as we shall see later, the result also holds for a particularly restricted class of metric spaces), and thus a constant $\epsilon > 0$ exists such that metric TSP cannot be approximated within a factor $(1+\epsilon)$ in polynomial time, unless P = NP. The current estimation of such a constant is not very good (of the order of 10^{-6}), however such a result gives at least some *qualitative* insigth into the approximability properties of the problem. The complexity of approximating TSP in the case of geometric metrics remained open. In his PhD thesis, Arora noted that proving the Max SNP-hardness of 2ETSP should be very difficult, but that the $k(n)$ETSP for sufficiently large $k(n)$ could perhaps be proved Max SNP-hard with relatively simple reductions ([Aro94, Chapter 9]). In [GKP95], Grigni et al. proved that the restriction of the TSP to shortest paths metrics of planar graphs can be approximated within $(1+\epsilon)$ in time $n^{O(1/\epsilon)}$. Such an approximation algorithm is called a *Polynomial Time Approximation Scheme (PTAS)*. This result led Grigni et al. [GKP95] to conjecture that 2ETSP has a PTAS. In a very recent breakthrough, Arora [Aro96] developed a PTAS for the 2ETSP. Such an algorithm can be adapted to work in higher dimensional spaces and, in particular, it runs in time $n^{\tilde{O}(\log^{k(n)-2} n)/\epsilon^{k(n)-1}}$ for instances of $k(n)$ETSP. Note that the dependence of the running time on $k(n)$ is doubly exponential. In a preliminary version of [Aro96] Arora asked if it was possible to develop a PTAS for $k(n)$ETSP for arbitrary k, or if, at least, it was possible to have the running time being singly exponential in k, e.g. $n^{O(k(n)/\epsilon)}$.

In this paper we essentially answer negatively to both questions. We prove that $(n-1)$ETSP is Max SNP-hard (thus, unless P = NP, there cannot be a PTAS for $(n-1)$ETSP). Furthermore, we show that $(1+\epsilon_2)$-approximating $(c\log n)$ETSP is NP-hard under randomized reductions, for proper constants c and ϵ_2. The latter result implies that there cannot be an algorithm that finds $(1+\epsilon)$-approximate solutions for $k(n)$ETSP running in time $n^{O(k(n)/\epsilon)}$ for any $\epsilon > 0$, unless NP $\subseteq$ RQP, where RQP is the class of problems solvable by randomized algorithms with one-sided error and running time $n^{O(\log n)}$.

The Max SNP-hardness of the general case is proved by means of a reduction from the version of the metric TSP that was shown to be Max SNP-hard in [PY93]. The reduction uses a mapping of the metric spaces of [PY93] into Hamming spaces. Our result extends to other geometric metrics.

2 Preliminaries

We denote by $\mathbf{R}$ the set of real numbers. Given an instance $x = (U, d)$ of the TSP, we will denote by $\mathsf{opt}(x)$ the cost of an optimum solution for x.

In this paper we will use the notions of L-reduction and Max SNP-hardness; we refer the reader to [PY91] for definitions.

Recall that a function $d : U \times U \to \mathbf{R}$ is a *metric* if it is non-negative, if $d(u,v) = 0$ iff $u = v$, if it is symmetric (i.e. $d(u,v) = d(v,u)$ for any $u, v \in U$), and it satisfies the *triangle inequality* (i.e. $d(u,v) \leq d(u,z) + d(z,v)$ for any $u, v, z \in U$).

Definition 1 ($(1,2)-B$ metrics) *A metric $d : U \times U \to \mathbf{R}$ is a $(1,2)-B$ metric if it satisfies the following properties:*

1. *For any $u, v \in U$, $u \neq v$, $d(u,v) \in \{1,2\}$.*

2. For any u, at most B elements of U are at distance 1 from u.

Papadimitriou and Yannakakis [PY93] have shown that a constant $B_0 > 0$ exists such that the TSP problem is Max SNP-hard even when restricted to $(1,2) - B_0$ metrics.

For any positive integer n, we denote by d_H^n the Hamming metric in $\{0,1\}^n$ and by d_E^n the Euclidean metric in $\mathbf{R}^n$. We will usually omit the superscripts. We will make use of the following well known fact.

Proposition 2 *Let $u, v \in \{0,1\}^n \subseteq \mathbf{R}^n$. Then $d_E(u,v) = \sqrt{d_H(u,v)}$.*

3 The Reductions

Let us begin with a lemma relating $(1,2) - B$ metrics and Hamming metrics. The lemma gives a "distance preserving" embedding of $(1,2) - B$ metric spaces into Hamming spaces.

Lemma 3 *Let U be a finite set and d be a $(1,2)-B$ metric over U. Then there exists an embedding $f : U \to \{0,1\}^{3B|U|/2}$ such that for any $u, v \in U$,*

1. *$d_H(f(u), f(v)) = 2B$ if $d(u,v) = 2$, and*
2. *$d_H(f(u), f(v)) = 2(B-1)$ if $d(u,v) = 1$.*

Such an embedding is computable in time polynomial in $|U|$.

Proof: Let $U = \{u_1, \ldots, u_n\}$. Recall that a $(1,2) - B$ metric (U,d) can be represented as an undirected graph $G = (U,E)$, where $\{u,v\} \in E$ iff $d(u,v) = 1$ (see [PY93]). Let $E = \{e_1, \ldots, e_m\}$. An edge e is said to be *incident* on a vertex u if u is one of the endpoints of e. The *degree* of a vertex u (denoted by $deg(u)$) is the number of edges incident on u. Note that any vertex in G has degree at most B, and thus $m \leq Bn/2$. The embedding f of U into $\{0,1\}^{3Bn/2}$ is defined as follows: for any $i = 1, \ldots, n$, for any $j = 1, \ldots, 3Bn/2$, the j-th coordinate of $f(u_i)$ is

$$f(u_i)[j] = \begin{cases} 1 & \text{if } (i-1)B + 1 \leq j \leq iB - deg(u_i), \\ 1 & \text{if } nB + 1 \leq j \leq m + nB \text{ and } e_{j-nB} \text{ is incident on } u_i \text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

We first note that, by construction, $f(u_i)$ has $B - deg(u_i)$ nonzero coordinates among the first nB ones, and $deg(u_i)$ nonzero coordinates among those between the $(nB+1)$-th and the $(nB+m)$-th. All of the other coordinates are zero. It follows that *for any $u \in U$, $f(u)$ has exactly B nonzero coordinates*, and thus, the Hamming distance between any two different points in $f(U)$ is at most $2B$. More specifically, the Hamming distance between $f(u_i)$ and $f(u_j)$ $(i \neq j)$ is equal to $2(B - a_{ij})$, where a_{ij} is the number of indices of coordinates such that both $f(u_i)$ and $f(u_j)$ are equal to one. Since $f(u_i)$ and $f(u_j)$ cannot have a one in the same position in any of the first nB coordinates, it follows that a_{ij} is equal to the number of indices h, $1 \leq h \leq m$, such that $f(u_i)[nB+h] = f(u_j)[nB+h] = 1$. It is not hard to see that $f(u_i)[nB+h] = f(u_j)[nB+h] = 1$ if and only if $\{u_i, u_j\} = e_h$, and thus, a_{ij} can only be either 0 or 1, and it can be 1 if and only if $d(u_i, u_j) = 1$. Clearly the embedding can be computed in time $O(B|U|^2)$: since B is constant, this is polynomial in $|U|$. □

The following corollary is required in the proof of our main theorem.

Corollary 4 *Let U be a finite set and d be a $(1,2)-B$ metric over U. Then there exist a constant δ (depending on B) and an embedding $f : U \to \mathbf{R}^{|U|-1}$ such that for any $u, v \in U$, $d_E(f(u), f(v)) = 1$ if $d(u,v) = 1$ and $d_E(f(u), f(v)) = 1 + \delta$ if $d(u,v) = 2$. Such an embedding is computable in time polynomial in $|U|$.*

PROOF: Map $U = \{u_1, \dots, u_n\}$ into a set $U' = \{u'_1, \dots, u'_n\}$ as in Lemma 3. From Proposition 2 we have that for any i and j, if $d(u_i, u_j) = 1$ then $d_E(u'_i, u'_j) = \sqrt{2(B-1)}$, and if $d(u_i, u_j) = 2$ then $d_E(u'_i, u'_j) = \sqrt{2B}$. If we divide each coordinate of the points u'_i by $\sqrt{2(B-1)}$, we obtain a set of points in $\mathbf{R}^{3Bn/2}$ whose distances satisfy the hypothesis of the corollary, with $\delta = (\sqrt{2B}/\sqrt{2(B-1)}) - 1$. Now, note that those n points lie in some $(n-1)$-dimensional affine subspace and therefore can be mapped in polynomial time into $\mathbf{R}^{n-1}$ preserving the Euclidean distance. The entire process can be done in time polynomial in $|U|$. □

We are now ready to prove our main result.

Theorem 5 *$(n-1)$ETSP is Max SNP-hard.*

PROOF: For some constant B_0, the TSP is Max SNP-hard when restricted to $(1,2) - B_0$ metrics [PY93]. We shall now describe an L-reduction from the $(1,2) - B_0$ metric TSP to the $(n-1)$ETSP. Let $x = (U, d)$ be an instance of the TSP, where $U = \{u_1, \dots, u_n\}$ and d is a $(1,2) - B_0$ metric. We map the cities into $\mathbf{R}^{n-1}$ as in Corollary 4, thus obtaining an instance x' of $(n-1)$ETSP. It is easy to see that, for any tour π,

$$m(x', \pi) = n + \delta(m(x, \pi) - n) = n(1-\delta) + \delta m(x, \pi) \ .$$

This implies

$$\mathsf{opt}(x') = n(1-\delta) + \delta \mathsf{opt}(x) \leq \mathsf{opt}(x)$$

(since $0 < \delta \leq 1$, and $\mathsf{opt}(x) \geq n$) and that

$$m(x, \pi) - \mathsf{opt}(x) = \frac{1}{\delta}(m(x', \pi) - \mathsf{opt}(x'))$$

Thus, we have an L-reduction with $\alpha = 1$ and $\beta = 1/\delta$. □

Combining the above theorem with the results of Arora et al. [ALM+92], we have the following non-approximability result for the Euclidean TSP.

Corollary 6 *There exists a constant $\epsilon_1 > 0$ such that approximating the $(n-1)$ETSP within $(1+\epsilon_1)$ is NP-hard.*

By straightforward modifications of the proofs of Corollary 4 and Theorem 5, obtain the following result.

Theorem 7 *Let $\{d^k\}_{k \geq 1}$ be a family of metrics over $\mathbf{R}^k$ such that the distance between two points in $\{0,1\}^k$ is a monotone increasing function of their Hamming distance (and is independent of k). Then the TSP in $(\mathbf{R}^{3Bn/2}, d^{3Bn/2})$ is Max SNP-hard.*

Examples of such kinds of metrics are the metrics based on ℓ_p norms for any fixed $p \geq 1$.

The following lemma, which was stated in [LLR94, Theorem 3.4] and was implicit in [JL84][2], will be used to give a non-approximability result for $O(\log n)$ETSP.

[2]Reference [JL84] is not easy to collect. The results of [JL84] are also presented in [JLS87], and an alternative (and simpler) proof is given in [FM88]

Lemma 8 ([JL84]) *There exists a constant* $\mu > 0$ *such that the following holds. Let* U *be a set of* n *points into* $\mathbf{R}^n$ *and let* $\gamma > 0$*. Then there exists an embedding* f *of* U *into* $\mathbf{R}^{\mu \log n/\gamma^2}$ *such that for any* $u, v \in U$*,* $(1-\gamma)d_E(u,v) \leq d_E(f(u), f(v)) \leq (1+\gamma)d_E(u,v)$*. Such an embedding is computable by a randomized polynomial time algorithm.*

We can now prove the non-approximability result for $O(\log n)$ETSP.

Theorem 9 *There exist constants* ϵ_2 *and* c *such that* $(1+\epsilon_2)$*-approximating the* $(c \log n)$*ETSP is* NP*-hard under randomized reductions (and thus infeasible, unless* RP = NP*).*

PROOF: Let ϵ_1 be the constant of Corollary 6. Fix constants c, γ and ϵ_2 such that

$$0 < \epsilon_2 < \epsilon_1, \quad (1+\epsilon_2)(1+\gamma)/(1-\gamma) \leq 1+\epsilon_1, \text{ and } c = \mu/\gamma^2.$$

Assume we have an r-approximate algorithm for $(c \log n)$ETSP. Given an instance x of nETSP, we can map it into an instance x' of $(c \log n)$ETSP using Lemma 8 with parameter γ. Since the cost of a solution is the sum of the distances between certain pairs of cities, it immediately follows that for any tour π we have

$$(1-\gamma)m(x,\pi) \leq m(x',\pi) \leq (1+\gamma)m(x,\pi)$$

and thus

$$\mathsf{opt}(x) \leq \mathsf{opt}(x')/(1-\gamma) \ .$$

If a tour π is $(1+\epsilon_2)$-approximate for x', then

$$\frac{\mathsf{opt}(x)}{m(x,\pi)} \leq \frac{\mathsf{opt}(x')/(1-\gamma)}{m(x',\pi)/(1+\gamma)} \leq \frac{1+\gamma}{1-\gamma}(1+\epsilon_2) \leq 1+\epsilon_1$$

and so π is $(1+\epsilon_1)$-approximate for x. Since finding $(1+\epsilon_1)$-approximate solutions for nETSP is NP-hard, the above randomized reduction yields the NP-hardness (under randomized reductions) of $(1+\epsilon_2)$-approximating the $(c \log n)$ETSP. □

Remark 10 *In all proofs of this section we implicitly made the (unrealistic) assumption that arbitrary real numbers can appear in an instance and that arithmetic operations (including squared roots) can be computed over them in constant time. However, our results still hold if we instead assume that numbers are rounded and stored in a floating point notation using* $O(\log n)$ *bits. This fact follows from a minor modification of the argument used in [Aro96] to reduce a general instance of Euclidean TSP to an instance where coordinates are positive integers whose value is* $O(n^2)$*.*

4 Open Questions

A deterministic version of Lemma 8 (besides being a *per se* interesting result) would imply the NP-hardness of $(1+\epsilon_2)$-approximating $c \log n$ETSP (we can only prove infeasibility under the hypothesis that RP $\neq$ NP). Proving an analogous of Lemma 8 for any ℓ_p norm would imply the hardness of approximating the TSP problem in $O(\log n)$ dimensional ℓ_p spaces.

The Steiner Tree problem is Max SNP-hard when restricted to metric spaces where all distance are 1 or 2 [BP89] (but the reduction of [BP89] does not give $(1,2) - B$ metrics). In the case of Euclidean spaces, Arora [Aro96] gives approximation schemes with the same efficiency of the approximation schemes for Euclidean TSP. We conjecture that the n-dimensional Euclidean Steiner Tree problem is Max SNP-hard.

Acknowledgments

I'm grateful to Sanjeev Arora for proving Theorem 9 and to Ray Greenlaw and Madhu Sudan for helpful comments on preliminary versions of this paper.

References

[ALM+92] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. In *Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science*, pages 14–23, 1992.

[Aro94] S. Arora. *Probabilistic Checking of Proofs and Hardness of Approximation Problems.* PhD thesis, University of California at Berkeley, 1994.

[Aro96] S. Arora. Polynomial time approximation schemes for euclidean TSP and other geometric problems. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, 1996. To appear.

[BP89] M. Bern and P. Plassmann. The steiner tree problem with edge lengths 1 and 2. *Information Processing Letters*, 32:171–176, 1989.

[Chr76] N. Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, Carnegie-Mellon University, 1976.

[Edm66] J. Edmonds. Optimum branchings. *Journal of Research National Bureau of Standards, Part B*, 17B(4):233–240, 1966.

[FM88] P. Frankl and H. Maehara. The Johnson-Linderstrauss lemma and the spericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44:355–362, 1988.

[GGJ76] M.R. Garey, R.L. Graham, and D.S. Johnson. Some NP-complete geometric problems. In *Proceedings of the 8th ACM Symposium on Theory of Computing*, pages 10–22, 1976.

[GKP95] M. Grigni, E. Koutsoupias, and C.H. Papadimitriou. An approximation scheme for planar graph TSP. In *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pages 640–645, 1995.

[JL84] W.B. Johnson and J. Lindestrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[JLS87] W.B. Johnson, J. Lindestrauss, and G. Schechtman. On Lipschitz embedding of finite metric spaces in low dimensional normed spaces. In *Geometric Aspects of Functional Analysis*, pages 177–184. LNM 1267, Springer, 1987.

[Kar72] R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[LLKS85] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys. *The traveling salesman problem.* John Wiley, 1985.

[LLR94] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. In *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pages 577–591, 1994.

[Pap77] C.H. Papadimitriou. Euclidean TSP is NP-complete. *Theoretical Computer Science*, 4:237–244, 1977.

[PY91] C. H. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43:425–440, 1991. Preliminary version in *Proc. of STOC'88.*

[PY93] C.H. Papadimitriou and M. Yannakakis. The travelling salesman problem with distances one and two. *Mathematics of Operations Research*, 18:1–11, 1993.

[SG76] S. Sahni and T. Gonzalez. P-complete approximation problems. *Journal of the ACM*, 23:555–565, 1976.