Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:

Revision 01 of 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'

When Hamming Meets Euclid: the Approximability of Geometric TSP and MST

LUCA TREVISAN*

November 15, 1996

Abstract

We prove that the Traveling Salesperson Problem (MIN TSP), the Minimum Steiner Tree Problem (MIN ST), the Minimum k-Steiner Tree problem (MIN k-ST) and the k-Center Problem (MIN k-CENTER) are Max SNP-hard (and thus NP-hard to approximate within some constant $r > 1$) even if all cities (respectively, points) lie in the geometric space $\mathcal{R}^n$ (n is the number of cities/points) and distances are computed with respect to the l_1 (rectilinear) metric. The MIN TSP and MIN k-CENTER hardness results also hold for any l_p metric, including the *Euclidean* ℓ_2 metric, and (under randomized reductions) also in $\mathcal{R}^{\log n}$ for the Euclidean metric. Arora's approximation scheme for Euclidean MIN TSP in $\mathcal{R}^d$ runs in time $n^{\tilde{O}(\log^{d-2} n)/\epsilon^{d-1}}$ and achieves approximation $(1+\epsilon)$; our result implies that this running time cannot be improved to $n^{d/\epsilon}$ unless NP has subexponential randomized algorithms. We also prove, as an intermediate step, the hardness of approximating the above problems in *Hamming spaces*. The only previous hardness results involved metrics where all distances are 1 or 2.

1 Introduction

Given a metric space and a set U of points into it, the Traveling Salesperson Problem (MIN TSP) is to find a closed tour of shortest total length visiting each point exactly once, while the Minimum Steiner Tree Problem (MIN ST) is to find the minimum cost tree connecting all the points of U; the tree can possibly contain points not in U, that are called "Steiner points".

Both problems are among the most classical and most widely studied ones in Combinatorial Optimization, Operations Research and Computer Science during the past few decades, and before. Important special cases arise when the metric space is $\mathcal{R}^k$ and the distance is computed according to the ℓ_1 norm (the *rectilinear* case) or the ℓ_2 norm (the *Euclidean* case).

We establish the first non-approximability results for this class of problems. As an intermediate step, we prove that they are hard to approximate also in *Hamming spaces*. The Hamming versions of MIN TSP and MIN ST seem to have never been considered before. Our main contributions are: (i) the identification of this class of metric spaces as the "right" one to prove hardness in more natural geometric spaces, and (ii) the derivation of combinatorial results that could have some independent interest.

Our techniques prove hardness of approximation for other problems, including the Minimum k-Center Problem studied by Hochbaum and Shmoys [HS85], and all the problems mentioned in Arora's paper [Aro96] on approximation schemes for geometric problems.

We now state and discuss our results for MIN TSP and MIN ST.

*Dept. of Computer Science, University of Geneva, Rue General-Dufour 24, CH 1211 Geneva, Switzerland. Email: trevisan@cui.unige.ch. Work partly done at the University of Rome "La Sapienza".

The Traveling Salesperson Problem

Interest in the MIN TSP started during the 1930's. In 1966, the (already) long-standing failure of developing an efficient algorithm for the MIN TSP led Edmonds [Edm66] to conjecture that the problem is not in P: this is sometimes referred to as the first statement of the $\mathsf{P} \neq \mathsf{NP}$ conjecture. See the book of Lawler et al. [LLKS85] for a very complete survey on MIN TSP. Here we will only review the results that are relevant for the present paper. The MIN TSP is NP-hard even if the cities are restricted to lie in $\mathcal{R}^2$ and the distances are computed according to the ℓ_2 norm [GGJ76, Pap77]. Due to such a negative result, research concentrated on developing good heuristics. 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. A 3/2-approximation algorithm that works for any metric space is due to Christofides [Chr76]. In twenty years of research no improvement of this bound had 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 MIN TSP 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 MIN TSP cannot be approximated within a factor $(1+\epsilon)$ in polynomial time, unless $\mathsf{P} = \mathsf{NP}$. The complexity of approximating MIN TSP in the case of geometric metrics remained a major open question. In his PhD thesis, Arora noted that proving the Max SNP-hardness of Euclidean MIN TSP in $\mathcal{R}^2$ should be very difficult, but that this could perhaps be done in $\mathcal{R}^{k(n)}$ for sufficiently large $k(n)$ ([Aro94, Chapter 9]). The relevance of non-approximability results for geometric MIN TSP was also stated in the open questions section of a survey by Arora and Lund [AL96]. In [GKP95], Grigni, Koutsopias and Papadimitriou proved that the restriction of the MIN 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 Euclidean MIN TSP has a PTAS in $\mathcal{R}^2$. They again posed the question of determining the approximability of the problem for higher dimensions. In a very recent breakthrough, Arora [Aro96] developed a PTAS for the MIN TSP in $\mathcal{R}^2$ under any ℓ_p metric. Such an algorithm also works in higher dimensional spaces and, in particular, it runs in time $n^{\tilde{O}(\log^{k(n)-2} n)/\epsilon^{k(n)-1}}$ in $\mathcal{R}^{k(n)}$. 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 Euclidean MIN TSP in $\mathcal{R}^n$ or if, at least, it was possible to have the running time being singly exponential in $k(n)$, e.g. $n^{O(k(n)/\epsilon)}$.

Our Results. In this paper we essentially answer negatively to both questions. We prove that MIN TSP in $\mathcal{R}^n$ is Max SNP-hard using any ℓ_p metric (thus, unless $\mathsf{P} = \mathsf{NP}$, there cannot be a PTAS for these problems). Furthermore, we show that $(1+\epsilon_1)$-approximating Euclidean MIN TSP in $\mathcal{R}^{\log n}$ is NP-hard under randomized reductions, for a proper constant ϵ_1. The latter result implies that there cannot be an algorithm that finds $(1+\epsilon)$-approximate solutions for Euclidean MIN TSP in $\mathcal{R}^k$ running in time $n^{O(k/\epsilon)}$ for any $\epsilon > 0$, unless $\mathsf{NP} \subseteq \mathsf{RQP}$, 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 n-dimensional case is proved by means of a reduction from the version of the metric MIN TSP that was shown to be Max SNP-hard in [PY93]. The reduction uses a mapping (see Lemma 5) of the metric spaces of [PY93] into Hamming spaces and the observation (see Proposition 3) that, for elements of $\{0,1\}^n$ a "gap" in the Hamming distance is preserved if distances are computed according to a ℓ_p metric. Our mapping of the metric spaces of [PY91] into Hamming spaces is *not* an *approximate isometry*, that is, it does *not* preserve distances up to negligible distorsion. We also suspect that such kind of mapping would be provably impossible. Instead, our mapping introduces a fairly high (yet constant) distorsion, but satisfies an additional condition that makes the mapping be an *L-reduction* [PY91]. Our mapping, combined with a reduction by Kariv and Hakimi [KH76], gives also non-approximability results for the Minimum k-Center Problem. The Minimum k-Cities Traveling Salesman Problem (MIN k-TSP) and the Minimum Degree-Restricted Steiner Tree Problem (two problems mentioned in Arora's paper [Aro96] on approximation schemes for geometric problems) are generalizations of the MIN TSP. The hardness results that we prove for MIN TSP clearly extend to them.

The Minimum Steiner Tree Problem

The origins of the MIN ST problem seem to be even more remote than the MIN TSP's ones: the case when $|U| = 3$ and the metric space is $\mathcal{R}^2$ with the ℓ_2 norm has been studied by the Italian mathematician Torricelli (a student of Galilei's) in 17th century. Reportedly, Gauss had an interest to this problem as well. Recent results about this problem are similar to the ones for MIN TSP: exact optimization is NP-hard in $\mathcal{R}^2$ both in the Rectilinear (ℓ_1) case [GJ77] and in the Euclidean (ℓ_2) case [GGJ77]. Constant-factor approximation is achievable in any metric space (the best factor should be 1.644 due to Karpinski and Zelikovsky [KZ95]), in general metric spaces the problem is Max SNP-hard [BP89], Arora's algorithm achieves approximation $(1+\epsilon)$ in $\mathcal{R}^k$ in time $n^{\tilde{O}(\log^{k(n)-2} n)/\epsilon^{k(n)-1}}$. No non-approximability result was known for geometric versions of the problem.

Our Results. We prove the Max SNP-hardness of the problem in $\mathcal{R}^n$ under the ℓ_1 norm. As a preliminary step, we prove the hardness of the problem restricted to Hamming spaces. The latter hardness is proved via a reduction from the Minimum Vertex Cover problem (MIN VC) restricted to triangle-free graphs of maximum degree 3. The Max SNP-hardness of this very restricted version of MIN VC is proved in this paper and could be used as a starting point for other non-approximability results. The reduction from MIN VC to Hamming MIN ST uses a combinatorial result (Claim 17) stating that for an instance where all points have weight[1] 2 or 0, if a technical condition is satisfied, there exists an optimum solution where all Steiner points have weight 1. We remark that there exists an instance of Hamming Steiner Tree where all the points have weight 3 or 0 and such that an optimum solution must contain a Steiner point of weight at least 4. Thus, our combinatorial result cannot be generalized too much. Reducing from Hamming Steiner Tree to Rectilinear Steiner Tree requires another combinatorial result (Theorem 19): for an instance where all the points are in $\{0,1\}^n \subset \mathcal{R}^n$, there exists an optimum solution where all the Steiner points lie in $\{0,1\}^n$. We prove this fact using the *integrality property* of Min-CUT linear programming relaxations. Our non-approximability result extends to MIN k-ST, the variation where one is also given an integer k and the goal is to find a minimum Steiner tree among the ones involving at most k Steiner points.

Discussion

We give the first non-approximability results for geometric versions of network optimization problems. For Euclidean MIN TSP, there is little room for improvement of our results, as well as there

[1]For a vector $u \in \{0,1\}^n$, its weight is defined as the number of non-zero coefficients, e.g. the weight of $(0,1,1,0,1)$ is three.

is little room for improving Arora's algorithm. If we believe that NP has not sub-exponential algorithms, then the best possible running time for an approximation scheme for Euclidean Min TSP is of the form $2^{2^d/\epsilon}\mathrm{poly}(n)$; alternatively, our non-approximability result could be extended to $\mathcal{R}^{\log/\log\log n}$. Much more consistent improvements are possible for Min ST, however our results at least state very clearly that the number of dimensions *does matter* in the running time of an approximation scheme for these geometric problems.

We feel that one important contribution of this paper is the recognition of Hamming spaces as a class of metric spaces that seem to retain most of the hardness of general metrics while having a nice combinatorial structure. We believe that other non-approximability results could be established using Hamming spaces as intermediate steps. We also think that it should be worth trying to improve Christofides algorithm in Hamming spaces. While the well-behaved structure of Hamming spaces should not make this task impossible, it is likely that such an improved algorithm could give useful ideas for more general cases.

2 Preliminaries

We denote by $\mathcal{R}$ the set of real numbers. For an integer n we denote by $[n]$ the set $\{1, \ldots, n\}$. For a vector $\vec{a} \in \mathcal{R}^n$ and an index $i \in [n]$, we denote by $\vec{a}[i]$ the i-th coordinate of $\vec{a}$, Given an instance x of an optimization problem A, we will denote by $\mathsf{opt}_A(x)$ the cost of an optimum solution for x, we will also typically omit the subscript. For a feasible solution y (usually a tour or a tree) of an instance x of an optimization problem A, we denote its cost by $\mathsf{cost}_A(x, y)$ or, more often, as $\mathsf{cost}(y)$. In this paper we will use the notions of L-reduction and Max SNP-hardness. Max SNP is a class of constant-factor approximable optimization problems that includes Max 3SAT, we refer the reader to [PY91] for the formal definition.

Definition 1 (L-reduction) *An optimization problem A us said to be L-reducible to an optimization problem B if two constants α and β and two polynomial-time computable functions f and g exist such that*

1. *For an instance x of A, $x' = f(x)$ is an instance of B, and it holds $\mathsf{opt}_B(x') \leq \alpha \mathsf{opt}_A(x)$.*
2. *For an instance x of A, and a solution y' feasible for $x' = f(x)$, $y = g(x, y')$ is a feasible solution for x and it holds $|\mathsf{opt}_A(x) - \mathsf{cost}_A(x, y)| \leq \beta |\mathsf{opt}_B(x') - \mathsf{cost}_B(x', y')|$.*

We say that an optimization problem is Max SNP-hard if all Max SNP-problems are L-reducible to it. From [ALM+92] it follows that if a problem A is Max SNP-hard, then a constant $\epsilon > 0$ exists such that $(1 + \epsilon)$-approximating A is NP-hard.

A function $d : U \times U \to \mathcal{R}$ 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 2 ($(1,2) - B$ metrics) *A metric $d : U \times U \to \mathcal{R}$ 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 MIN TSP is Max SNP-hard even when restricted to $(1,2) - B_0$ metrics.

For an integer $p \geq 1$, the ℓ_p norm in $\mathcal{R}^n$ is defined as $||(u_1, \ldots, u_n)||_p = (\sum_{i=1}^n |u_i|^p)^{(1/p)}$. The distance induced by the ℓ_p norm is defined as $d_p(\vec{u}, \vec{v}) = ||\vec{u} - \vec{v}||_p$. For a positive integer n, we denote by d_H^n the Hamming metric in $\{0,1\}^n$ (we will usually omit the superscripts). We will make some use of the following fact.

Proposition 3 *Let $\vec{u}, \vec{v} \in \{0,1\}^n \subseteq \mathcal{R}^n$. Then $d_p(\vec{u}, \vec{v}) = d_H(\vec{u}, \vec{v})^{1/p}$.*

Before starting with the presentation of our results, we make the following important caveat.

Remark 4 *In some of the proofs of this paper we implicitly make 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 or Steiner Tree into an instance where coordinates are positive integers whose value is $O(n^2)$.*

3 MIN TSP and MIN k-CENTER

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 5 *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 simple corollary is required in the proof of our hardness result.

Corollary 6 *Let $p \geq 1$ be fixed. 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 $f : U \to \mathcal{R}^{3B|U|/2}$ such that for any $u, v \in U$, $d_p(f(u), f(v)) = 1$ if $d(u,v) = 1$ and $d_p(f(u), f(v)) = 1 + \delta$ if $d(u,v) = 2$. Such an embedding is computable in time polynomial in $|U|$.*

PROOF: Map $U = \{u_1, \ldots, u_n\}$ into a set $U' = \{\vec{u}'_1, \ldots, \vec{u}'_n\}$ as in Lemma 5. From Proposition 3 we have that for any i and j, if $d(u_i, u_j) = 1$ then $d_p(\vec{u}'_i, \vec{u}'_j) = (2(B-1))^{1/p}$, and if $d(u_i, u_j) = 2$ then $d_p(\vec{u}'_i, \vec{u}'_j) = (2B)^{1/p}$. If we divide each coordinate of the points $\vec{u}'_i$ by $(2(B-1))^{1/p}$, we obtain a set of points in $\mathcal{R}^{3Bn/2}$ whose distances satisfy the hypothesis of the corollary, with $\delta = (2B/2(B-1))^{1/p} - 1$. The entire process can be done in time polynomial in $|U|$. □

The main result of this section is now only a matter of standard calculations.

Theorem 7 *For any fixed $p \geq 1$, the* MIN TSP *is* Max SNP*-hard when restricted to the ℓ_p metric in $\mathcal{R}^{O(n)}$ (n is the number of cities).*

PROOF: For some constant B_0, the MIN 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 TSP in $\mathcal{R}^{O(n)}$. Let $x = (U, d)$ be an instance of the MIN TSP, where $U = \{u_1, \ldots, u_n\}$ and d is a $(1,2) - B_0$ metric. We map the cities into $\mathcal{R}^{3B_0 n/2}$ as in Corollary 6, thus obtaining an instance x' of MIN TSP in $\mathcal{R}^{3B_0 n/2}$. It is easy to see that, for any tour π,

$$\mathsf{cost}(x', \pi) = n + \delta(\mathsf{cost}(x, \pi) - n) = n(1-\delta) + \delta \mathsf{cost}(x, \pi) .$$

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

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

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

Remark 8 *Given an instance of* MIN TSP *with n points, if one adds n^c more points, all of them being at distance $1/O(n^{c+1})$ from some point of the instance, this perturbs the optimum in a negligible way. We can use this simple observation to scale down our hardness result to $\mathcal{R}^n$ (adding $3B_0 n/2 - 1$ points), or even to $\mathcal{R}^{n^\delta}$ for fixed $\delta > 0$ (adding $O(n^{1/\delta})$ points).*

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

Corollary 9 *For any positive integer $p \geq 1$, a constant $\epsilon^{(p)} > 0$ such that approximating the* Min TSP *in $\mathcal{R}^n$ within $(1+\epsilon^{(p)})$ is* NP*-hard.*

The following lemma, which was stated in [LLR95, Theorem 3.1] and was implicit in [JL84][2], will be used to give a non-approximability result for Euclidean Min TSP in $\mathcal{R}^{\log n}$.

Lemma 10 ([JL84]) *There exists a constant $\mu > 0$ such that the following holds. Let U be a set of n points into $\mathcal{R}^n$ and let $\gamma > 0$. Then there exists an embedding f of U into $\mathcal{R}^{\mu \log n/\gamma^2}$ such that for any $\vec{u}, \vec{v} \in U$, $(1-\gamma)d_2(\vec{u},\vec{v}) \leq d_2(f(\vec{u}), f(\vec{v})) \leq (1+\gamma)d_2(\vec{u},\vec{v})$. Such an embedding is computable by a randomized polynomial time algorithm.*

We can now prove the non-approximability result for $\mathcal{R}^{\log n}$.

Theorem 11 *There exist a constant $\epsilon_1 > 0$ such that $(1+\epsilon_1)$-approximating the Euclidean* Min TSP *in $\mathcal{R}^{\log n}$ is* NP*-hard under randomized reductions (and thus infeasible, unless* RP = NP*).*

Proof:[Of Theorem 11] Let $\epsilon^{(2)}$ be the constant of Corollary 9 for the Euclidean case. Fix constants c, γ and ϵ_1 such that

$$(1+\epsilon_1)(1+\gamma)/(1-\gamma) \leq 1+\epsilon^{(2)} \text{ and } c = \mu/\gamma^2.$$

Assume we have a $(1+\epsilon_1)$-approximate algorithm for Euclidean Min TSP in $\mathcal{R}^{\log n}$. Then the same algorithm can also be adapted to work in $\mathcal{R}^{\log n}$ (use a preprocessing phase that adds a bunch of n^c new dummy points, see Remark 8). Given an instance x with points in $\mathcal{R}^n$, we can map it into an instance x' with points in $\mathcal{R}^{c \log n}$ using Lemma 10 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

$$(1-\gamma)\mathsf{cost}(x,\pi) \leq \mathsf{cost}(x',\pi) \leq (1+\gamma)\mathsf{cost}(x,\pi)$$

and thus

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

If a tour π is $(1+\epsilon_1)$-approximate for x', then

$$\frac{\mathsf{opt}(x)}{m(x,\pi)} \leq \frac{\mathsf{opt}(x')/(1-\gamma)}{m(x',\pi)/(1+\gamma)} \leq \frac{1+\gamma}{1-\gamma}(1+\epsilon_1) \leq 1+\epsilon^{(2)}$$

and so π is $(1+\epsilon^{(2)})$-approximate for x. Since finding $(1+\epsilon^{(2)})$-approximate solutions in $\mathcal{R}^n$ is NP-hard, the above randomized reduction yields the NP-hardness (under randomized reductions) of $(1+\epsilon_1)$-approximating the Euclidean Min TSP in $\mathcal{R}^{\log n}$. □

The previous theorem could be improved in two aspects: use a deterministic reduction (so that hardness would be established under the nicer P $\neq$ NP condition instead that under the RP $\neq$ NP condition) and prove hardness for other ℓ_p metrics. Both improvements boil down to proving the following conjecture.

Conjecture 1 *For any fixed $p \in \mathcal{Z}^+$ and $\gamma > 0$, a polynomial-time computable function $f : \mathcal{R}^n \rightarrow \mathcal{R}^{(\mathrm{poly}(1/\gamma))\log n}$ exists such that, for any two vectors $\vec{u}, \vec{v} \in \mathcal{R}^n$*

$$(1-\gamma)d_p(\vec{u},\vec{v}) \leq d_p(f(\vec{u}), f(\vec{v})) \leq (1+\gamma)d_p(\vec{u},\vec{v}) \ .$$

[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].

The k-Center Problem In the MIN k-CENTER problem one is given a set of points U in a metric space and an integer k. The goal is to select a set S of k points (the *centers*) so that it is minimized the maximum distance between a point of $U - S$ and the closest center. The problem has been studied in several papers, including [KH76, HS85].

Theorem 12 *For any fixed $p \geq 1$, it is* NP*-hard to approximate* MIN k-CENTER *within any factor smaller than $(1.5)^{1/p}$ when the points are in $\mathcal{R}^n$ and distances are computed according to the ℓ_p norm.*

PROOF: [Of Theorem 12] Kariv and Hakimi [KH76] give a simple reduction from the Dominating Set problem to the k-center problem. Given a graph G, consider a metric space and a set U of points such that each point of U corresponds to a node of the graph and such that points corresponding to adjacent vertices are at distance d_1, while the others are at distance d_2 (where $d_1 < d_2$). It is easy to prove that, if the graph admits a dominating set of size k, the optimum of the k-Center instance (U, k) is d_1; otherwise the optimum is d_2. The dominating set problem is NP-hard even for graphs of maximum degree 3 [GJ79]. Such graphs can be mapped into an ℓ_p metric space so that adjacent vertices are at distance $4^{1/p}$ and non-adjacent vertices are at distance $6^{1/p}$. Thus, the k-center problem cannot be approximated within a factor smaller than $(1.5)^{1/p}$ in ℓ_p normed $\mathcal{R}^n$ spaces. □

4 The MIN ST Problem

The hardness of approximating MIN ST will be established with a longish chain of reductions. The starting point is the following hardness result, that may have a little independent interest. Recall that in the Minimum Vertex Cover (MIN VC) one is given a graph $G = (V, E)$ and looks for the smallest set $C \subseteq V$ such that C contains at least one endpoint of any edge in E.

Theorem 13 *The* MIN VC *problem is* Max SNP*-hard even when restricted to triangle-free graphs with maximum degree 3 (we call this restriction* MIN TF VC-3*).*

PROOF: [Of Theorem 13] The MAX 2SAT problem is Max SNP-hard even when restricted to instances where each variable occurs in at most 3 clauses (apply to MAX 2SAT the reduction from MAX 3SAT to MAX 3SAT-3 described in [Pap94]). One can assume without loss of generality that the 3 occurrences of each variable are either one positive occurrence and two negative occurrences, or vice versa. We reduce MAX 2SAT-3 to MIN VC using the reduction of [PY91]: we create a graph with a node for any occurrence of any literal, putting an edge between two nodes if they represent literals that occur in the same clause or if they are one the complement of the other. See [PY91] for the proof that this is an L-reduction. The obtained graph has maximum degree 3: each literal is adjacent to the fellow literal occurring in the same clause and to the (at most) two occurrences of its complement. Also, the graph is triangle-free: let l_1, l_2 and l_3 be any three occurrences of literals. Since clauses contain only two literals, from pigeonhole principle it follows that one of the three occurrences (say, l_1) does not occur in the same clause with any of other two. Then, if l_1, l_2 and l_3 form a triangle it follows that l_2 and l_3 are both the complement of l_1. Being adjacent, they also have to occur in the same clause, but this is a contraddiction since the literals occurring in a clause have to be different. □

We note in passing that, as a corollary, we obtain that the MAX INDEPENDENT SET problem is Max SNP-hard in the same, very restricted class of graphs. We now move to the restriction of MIN ST to Hamming spaces.

Notation: For a pair of indices $i,j \in [n]$ we define $\vec{a}^n_{i,j} \in \{0,1\}^n$ as the n-dimensional boolean vector all whose coordinates are zero but the i-th and the the j-th, e.g. $\vec{a}^5_{1,4} = (1,0,0,1,0)$. Similarly, we let $\vec{a}^n_i$ be the the vector in $\{0,1\}^n$ whose only non-zero coordinate is the i-th, e.g. $\vec{a}^4_3 = (0,0,1,0)$. For a vector $\vec{a} \in \{0,1\}^n$ and indices $i,j \in [n]$, we let $\mathsf{red}_{i,j}(\vec{a}) \in \{0,1\}^n$ be the vector defined as follows

$$\mathsf{red}_{i,j}(\vec{a})[h] = \begin{cases} \vec{a}[h] & \text{if } h \neq i \wedge h \neq j \\ 0 & \text{if } (h = i \vee h = j) \wedge \vec{a}[i] = \vec{a}[j] = 1 \end{cases}$$

In other words, $\mathsf{red}_{i,j}(\vec{a})$ is equal to $\vec{a}$ unless a has a one in the i-th and the j-th coordinate. In this latter case, the i-th and the j-th coordinate of $\mathsf{red}_{i,j}(\vec{a})$ are set to zero. For example $\mathsf{red}_{1,3}(0,1,1,1) = (0,1,1,1)$, while $\mathsf{red}_{2,3}(0,1,1,1) = (0,0,0,1)$. We will make use of the following simple combinatorial lemma.

Lemma 14 *For any $\vec{a},\vec{b} \in \{0,1\}^n$, for any $i,j \in [n]$, $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) \leq d_H(\vec{a},\vec{b})$.*

Proof: Without loss of generality, assume $i = 1$ and $j = 2$. The lemma trivially holds when both $\vec{a} = \mathsf{red}_{1,2}(\vec{a})$ and $\vec{b} = \mathsf{red}_{1,2}(\vec{b})$. Assume $\vec{a} \neq \mathsf{red}_{1,2}(\vec{a})$ (the other case is perfectly symmetric), that is, $\vec{a} = (1,1,\vec{a}')$ (with $\vec{a}' \in \{0,1\}^{n-2}$). There are four cases to be considered:

- If $\vec{b} = (1,1,\vec{b}')$, then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = d_H((0,0,\vec{a}'),(0,0,\vec{b}')) = d_H(\vec{a}',\vec{b}') = d_H(\vec{a},\vec{b})$.
- If $\vec{b} = (0,1,\vec{b}')$ or $b = (1,0,\vec{b}')$ then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = 1 + d_H(\vec{a}',\vec{b}') = d_H(\vec{a},\vec{b})$.
- If $\vec{b} = (0,0,\vec{b}')$, then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = d_H(\vec{a}',\vec{b}') = d_H(\vec{a},\vec{b}) - 2$.

□

Theorem 15 *The* Min ST *problem is* Max SNP*-hard when restricted to Hamming spaces.*

Proof: We give an L-reduction from Min TF VC-3. Let $G = (V,E)$ be a triangle-free graph of maximum degree 3, assume $V = [n]$ and let $m = |E|$. We define an instance of Hamming Min ST as follows: the number of dimensions is n and the set of points is

$$U = \{\vec{0}\} \cup \{\vec{a}^n_{ij} : \{i,j\} \in E\}$$

where $\vec{0}$ is the vector with all zero entries.

Claim 16 *Given a vertex cover $C \subseteq V$ in G it is possible to find a Steiner tree for U of cost $m+C$.*

Proof:[Of Claim 16] Let $S = \{\vec{a}^n_i : i \in C\}$. Consider the graph whose vertex set is $S \cup U$ and such that two vertices are adjacent iff their Hamming distance is one. We claim that this graph is connected: indeed all the nodes of S are clearly adjacent to $\vec{0}$; furthermore any node in U is adjacent to some node in S (since C is a vertex cover), thus all the nodes are connected to $\vec{0}$. Since the graph is connected it admits a spanning tree, that is also a Steiner tree for U. All the edges of such Steiner tree have cost 1, and there are $|C| + m$ of them (because the tree has $|S| + |U| = |C| + m + 1$ nodes), so the claim follows. □

From the above claim it follows that $\mathsf{opt}(U) \leq m + \mathsf{opt}(G) \leq 4\mathsf{opt}(G)$, and we have established the first condition of the L-reducibility. As usual, the other condition is more difficult to prove.

Claim 17 *Given a Steiner tree T for U it is possible to find in polynomial time another Steiner tree T' such that: (i) $\mathsf{cost}(T') \leq \mathsf{cost}(T)$ and (ii) all the edes of T' have cost one and all the Steiner nodes of T' are weight-one vectors.*

PROOF:[Of Claim 17] We first make sure that all edges have cost 1: any edge of cost $d > 1$ is broken into a length-d path using $d-1$ additional Steiner nodes. Let S be the new set of Steiner vertices. We now reduce the number of non-zero coordinates of Steiner vertices. For any $\{i,j\} \notin E$ we map each point $\vec{a} \in S \cup U$ into $\mathsf{red}_{i,j}(\vec{a})$; this mapping only changes Steiner points (by definition of $\mathsf{red}_{i,j}$, definition of U, and the fact that $\{i,j\} \notin E$). From Lemma 14 we also have that any phase does not increase the cost of the tree. At the end of this set of transformations, we run a "clean-up" phase that does the following: if some transformation has collapsed one node onto another, we take only one node (if a Steiner node is collapsed onto a node in U we clearly take the node in U). If the transformation creates cycles, we break them (e.g. finding a spanning tree of the final graph), and, again, this does not increase the cost. It remains to see that, after this process, no Steiner node can have more than one non-zero coordinate. If a Steiner node has some set of k non-zero coordinates, then they must correspond to a clique in G (otherwise, at some phase, some of them would have been changed by the application of the red operator): since G is triangle-free, $k \leq 2$, but if $k = 2$ then the Steiner node would be equal to a node of U, and thus would have been removed in the clean-up phase. It follows that $k = 1$. □

From the above claim, the following one follows quite easily.

Claim 18 *Given a Steiner tree T for U it is possible to find in polynomial time a vertex cover C for G such that $|C| \leq \mathsf{cost}(T) - m$*

PROOF: [Of Claim 18] We first find the Steiner tree T' as in the previous claim. Then, if we let S be the set of Steiner vertices of T', we have that $\mathsf{cost}(T) \geq \mathsf{cost}(T') = |S| + |U| - 1 = |S| + m$. Let $C = \{i : \vec{a}_i \in S\}$; it is easy to see that $|C| = |S| \leq \mathsf{cost}(T) - m$ and that C is a vertex cover in G. The latter fact follows from the fact that for any edge $\{i,j\} \in E$, the vector $\vec{a}_{i,j}$ belongs to U; since all the edges of T' have cost one, and T' does not contain weight-3 vectors, then either $\vec{a}_i \in S$ or $\vec{a}_j \in S$ (otherwise $\vec{a}_{i,j}$ would be an isolated node in T' contraddicting the fact that T' be connected). □

If T is any Steiner tree of U, the vertex cover C for G computed according the previous claim satisfies

$$\mathsf{cost}(C) - \mathsf{opt}(G) \leq (\mathsf{cost}(T) - m) - (\mathsf{opt}(U) - m) = \mathsf{cost}(T) - \mathsf{opt}(U)$$

and so also the second condition of the L-reduction is satisfied. □

If the following conjecture holds, then we can reduce MIN VC-B to Hamming MIN ST (without imposing the triangle-free restriction).

Conjecture 2 *Let $U \subset \{0,1\}^n$ be an instance of Hamming* MIN ST *such that $\vec{0} \in U$ and all vectors of U have weight at most 2. Then there exists an optimum solution where all the Steiner nodes have weight at most 2.*

Janos Körner proposed a further generalization: if U is contained in the Hamming sphere centered in some $\vec{u} \in U$ and of radius k, then there exists an optimum solution all whose Steiner nodes lie in the same sphere. This seemed to be a reasonable combinatorial analog of the fact that if the points are in $\mathcal{R}^k$ and distances are computed according to the Euclidean metric, the Steiner points of an optimum solution will be in the *convex hull* of the points of the instance. Lately, Janos refuted the generalized conjecture. The instance $U = \{(0,0,0,0),(0,1,1,1),(1,0,1,1),(1,1,0,1),(1,1,1,0)\}$ refutes the generalized conjecture even for $k = 3$. An optimum solution of cost 7 uses the Steiner node $(1,1,1,1)$. Computational experiments show that any solution without $(1,1,1,1)$ has cost at least 8.

To approach the MIN ST in ℓ_1 normed spaces we use a reduction from the Hamming case. Note that for points in $\{0,1\}^n$ the ℓ_1 distance equals the Hamming distance. However, the reduction is non-trivial since $\mathcal{R}^n$ contains so many points that are not in $\{0,1\}^n$ and we have to argue that having much more choice for the Steiner nodes does not make the problem easier. The Rectilinear MIN ST problem looks very much like a *relaxation* of the Hamming MIN ST problem; our reduction makes use of a *rounding scheme* proving that the relaxation does not change the optimum.

Theorem 19 *Let $U \subseteq \{0,1\}^n \subset \mathcal{R}^n$ be an instance of Rectilinear* MIN ST *all whose points are in the Boolean cube. Let T be a feasible solution for U. Then it is possible to find in polynomial time (in the size of T) another solution T' such that $\mathsf{cost}(T') \leq \mathsf{cost}(T)$ and all the Steiner nodes of T' are in $\{0,1\}^n$.*

Before proving the theorem, we note the following relevant consequence.

Corollary 20 *For any instance $U \subseteq \{0,1\}^n$ of Rectilinear* MIN ST*, an optimum solution exists all whose Steiner points are in $\{0,1\}^n$.*

We now prove Theorem 19.

PROOF:[Of Theorem 19] Let $S = \{\vec{s}_1, \ldots, \vec{s}_m\}$ be the set of Steiner points of T, and let E be the set of edges of T. For any $\vec{s}_j \in S$ we will find a new point $\vec{s}'_j \in \{0,1\}^n$, so that if we let T' be the tree obtained from T by substituting the $\vec{s}$ points with the corresponding $\vec{s}'$ points, the cost of T' is not greater than the cost of T. The latter statement is equivalent to

$$\sum_{(\vec{s}_j,\vec{u})\in E,\vec{u}\in U} ||\vec{s}_j - \vec{u}||_1 + \sum_{(\vec{s}_j,\vec{s}_h)\in E} ||\vec{s}_j - \vec{s}_h||_1 \geq \sum_{(\vec{s}'_j,\vec{u})\in E,\vec{u}\in U} ||\vec{s}'_j - \vec{u}||_1 + \sum_{(\vec{s}'_j,\vec{s}'_h)\in E} ||\vec{s}'_j - \vec{s}'_h||_1$$

We will indeed prove something stronger, namely, that for any $i \in [n]$ it holds

$$\sum_{(\vec{s}_j,\vec{u})\in E\ ,\ \vec{u}\in U} |\vec{s}_j[i] - \vec{u}[i]| + \sum_{(\vec{s}_j,\vec{s}_h)\in E} |\vec{s}_j[i] - \vec{s}_h[i]| \geq \sum_{(\vec{s}'_j,\vec{u})\in E\ ,\ \vec{u}\in U} |\vec{s}_j[i] - \vec{u}[i]| + \sum_{(\vec{s}'_j,\vec{s}'_h)\in E} |\vec{s}'_j[i] - \vec{s}'_h[i]| \tag{1}$$

Let $i \in [n]$ be fixed, we now see how to find values of $\vec{s}'_1[i], \ldots, \vec{s}'_m[i] \in \{0,1\}$ such that (1) holds. We express as a linear program the problem of finding values of $\vec{s}'_1[i], \ldots, \vec{s}'_m[i]$ that minimize the right-hand side of (1). For any $j \in [m]$ we have a variable x_j (representing the value to be given to $\vec{s}'_j[i]$) and for any edge $e = (\vec{a}, \vec{b})$ such that at least one endpoint is in S we have a variable y_e, representing the lenght $|\vec{a}[i] - \vec{b}[i]|$. The linear program is as follows

$$
\begin{array}{lll}
\min & \sum_e y_e & \\
\text{Subject to} & & \\
& y_e \geq x_j - x_h & \forall e = (\vec{s}_j, \vec{s}_h) \in E \\
& y_e \geq x_h - x_j & \forall e = (\vec{s}_j, \vec{s}_h) \in E \\
& y_e \geq x_j & \forall e = (\vec{s}_j, \vec{u}_h) \in E \text{ such that } \vec{u}_h[i] = 0 \\
& y_e \geq 1 - x_j & \forall e = (\vec{s}_j, \vec{u}_h) \in E \text{ such that } \vec{u}_h[i] = 1 \\
& x_j \geq 0 & \\
& y_e \geq 0 &
\end{array}
\qquad \text{(LP)}
$$

Setting $x_j = s_j[i]$ and setting $y_{(\vec{a},\vec{b})} = |\vec{a}[i] - \vec{b}[i]|$ yields a feasible solution, and its cost is the left-hand side of (1). Let $(\vec{x}^*, \vec{y}^*)$ be an optimum solution for (LP). From the previous observation we have that setting $\vec{s}_j'[i] = x_j^*$ we satisfy (1). It remains to be seen that (LP) has an optimum solution where all variables take value from $\{0,1\}$. This follows from the fact that (LP) is the linear programming relaxation of an undirected Min-CUT problem, where all the $\vec{u}$ such that $\vec{u}[i] = 0$ (respectively, $\vec{u}[i] = 1$) are identifed with the source (respectively, the sink), each $\vec{s}_j$ is a node, and the edges are like in T. It is well known (see e.g. [PS82]) that a Min-CUT linear programming relaxation has optimum 0/1 solutions, and that such a solution can be found in polynomial time. □

Remark 21 *There seems to be no natural analog of Theorem 19 in other norms. Even in $\mathcal{R}^2$, using the Eulcidean metric, we have that the optimum solution of the instance $\{(0,0),(1,0),(0,1)\}$ must use a Steiner point not in $\{0,1\}^2$.*

Theorem 22 *Rectilinear* MIN ST *is* Max SNP*-hard.*

PROOF: We reduce from Hamming MIN ST. The reduction leaves the instance unchanged. For an instance $U \subseteq \{0,1\}^n$, we let $\mathsf{opt}_H(U)$ (respectively, $\mathsf{opt}_R(U)$) be the cost of an optimum solution for U, when seen as an instance of Hamming MIN ST (respectively, of Rectilinear MIN ST). Clearly, we have that $\mathsf{opt}_R(U) \leq \mathsf{opt}_H(U)$. Given a solution T for U, we find a solution T' as in Theorem 19. Since in $\{0,1\}^n$ the distance induced by the ℓ_1 norm equals the Hamming distance, we have that $\mathsf{cost}_H(T') = \mathsf{cost}_R(T') \leq \mathsf{cost}_R(T)$. We have an L-reduction with $\alpha = \beta = 1$. □

It is easy to see that the Rectilinear MIN k-ST and the Hamming MIN k-ST are Max SNP-hard as well. The instances produced by the above reductions have the property that an optimum solution has at most n Steiner nodes. So the reduction works unchanged for MIN k-ST.

5 Open questions

We don't know how to extend our non-approximability result for MIN ST to the Euclidean case. Arora [Aro96] notes that, by inspecting the way his algorithm works, it is possible to claim that, for any instance of Euclidean MIN ST, there exists a near-optimal solution where the Steiner points lie in some well-specified positions (either at "portals" or in positions chosen at the bottom of the recursion). This observation could perhaps be a starting point.

We don't have explicit estimations of the constants to within which it is hard to approximate geometric MIN TSP and rectilinear MIN ST. The constant for MIN TSP should be only slightly smaller than the corresponding constant for the $(1,2) - B$ case (estimated around $1 + 10^{-5}$). The constant for MIN ST is more likely to be around $1 + 10^{-4}$. Finding much stronger estimations (comparable to the 3/2 bound of Christofides and the 1.644 bound of Karpinski and Zelikovsky) is an open and challenging question. It appears that MIN TSP and MIN ST lack the nice logical definability that allows to prove very strong non-approximability results for MAX CUT and MAX 3SAT using so-called "gadget reductions" [BGS95, TSSW96, Hås96].

Acknowledgments

I'm grateful to Sanjeev Arora for proving Theorem 11 and to Madhu Sudan for proving Lemma 14. I thank Laszlo Babai and Janos Körner for very interesting and useful discussions. Ray Greenlaw carefully read a preliminary versions of this paper and gave several helpful remarks.

References

[AL96] S. Arora and C. Lund. Hardness of approximations. In *Approximation Algorithms for NP-hard Problems*. PWS Publishing, 1996.

[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.

[BGS95] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP's and non-approximability – towards tight results (3rd version). Technical Report TR95-24, Electronic Colloquium on Computational Complexity, 1995. Preliminary version in *Proc. of FOCS'95.*

[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.

[GGJ77] M.R. Garey, R.L. Graham, and D.S. Johnson. The complexity of computing Steiner minimal trees. *SIAM Journal of Applied Mathematics*, 34:477–495, 1977.

[GJ77] M.R. Garey and D.S. Johnson. The rectilinear Steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-Completeness*. Freeman, 1979.

[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.

[Hås96] J. Håstad. Getting optimal in-approximability results. Manuscript, September 1996.

[HS85] D.S. Hochbaum and D.B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184, 1985.

[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 paces in low dimensional normed spaces. In *Geometric Aspects of Functional Analysis*, pages 177–184. LNM 1267, Springer-Verlag, 1987.

[KH76] O. Kariv and S.L. Hakimi. An algorithmic approach to network location problem – Part I: the p-centers. Unpublished manuscript, cited in [GJ79], 1976.

[KZ95] M. Karpinski and A. Zelikovsky. New approximation algorithms for the Steiner tree problems. Technical Report TR95-030, Electronic Colloquium on Computational Complexity, 1995.

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

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. Preliminary Version in *Proc. of FOCS'94*.

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

[Pap94] C.H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[PS82] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prenitce-Hall, 1982.

[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.

[TSSW96] L. Trevisan, G.B. Sorkin, M. Sudan, and D.P. Williamson. Gadgets, approximation, and linear programming. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, 1996.

Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:

Revision 01 of 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'

When Hamming Meets Euclid: the Approximability of Geometric TSP and MST

LUCA TREVISAN*

November 15, 1996

Abstract

We prove that the Traveling Salesperson Problem (MIN TSP), the Minimum Steiner Tree Problem (MIN ST), the Minimum k-Steiner Tree problem (MIN k-ST) and the k-Center Problem (MIN k-CENTER) are Max SNP-hard (and thus NP-hard to approximate within some constant $r > 1$) even if all cities (respectively, points) lie in the geometric space $\mathcal{R}^n$ (n is the number of cities/points) and distances are computed with respect to the l_1 (rectilinear) metric. The MIN TSP and MIN k-CENTER hardness results also hold for any l_p metric, including the *Euclidean* ℓ_2 metric, and (under randomized reductions) also in $\mathcal{R}^{\log n}$ for the Euclidean metric. Arora's approximation scheme for Euclidean MIN TSP in $\mathcal{R}^d$ runs in time $n^{\tilde{O}(\log^{d-2} n)/\epsilon^{d-1}}$ and achieves approximation $(1+\epsilon)$; our result implies that this running time cannot be improved to $n^{d/\epsilon}$ unless NP has subexponential randomized algorithms. We also prove, as an intermediate step, the hardness of approximating the above problems in *Hamming spaces*. The only previous hardness results involved metrics where all distances are 1 or 2.

1 Introduction

Given a metric space and a set U of points into it, the Traveling Salesperson Problem (MIN TSP) is to find a closed tour of shortest total length visiting each point exactly once, while the Minimum Steiner Tree Problem (MIN ST) is to find the minimum cost tree connecting all the points of U; the tree can possibly contain points not in U, that are called "Steiner points".

Both problems are among the most classical and most widely studied ones in Combinatorial Optimization, Operations Research and Computer Science during the past few decades, and before. Important special cases arise when the metric space is $\mathcal{R}^k$ and the distance is computed according to the ℓ_1 norm (the *rectilinear* case) or the ℓ_2 norm (the *Euclidean* case).

We establish the first non-approximability results for this class of problems. As an intermediate step, we prove that they are hard to approximate also in *Hamming spaces*. The Hamming versions of MIN TSP and MIN ST seem to have never been considered before. Our main contributions are: (i) the identification of this class of metric spaces as the "right" one to prove hardness in more natural geometric spaces, and (ii) the derivation of combinatorial results that could have some independent interest.

Our techniques prove hardness of approximation for other problems, including the Minimum k-Center Problem studied by Hochbaum and Shmoys [HS85], and all the problems mentioned in Arora's paper [Aro96] on approximation schemes for geometric problems.

We now state and discuss our results for MIN TSP and MIN ST.

*Dept. of Computer Science, University of Geneva, Rue General-Dufour 24, CH 1211 Geneva, Switzerland. Email: trevisan@cui.unige.ch. Work partly done at the University of Rome "La Sapienza".

The Traveling Salesperson Problem

Interest in the MIN TSP started during the 1930's. In 1966, the (already) long-standing failure of developing an efficient algorithm for the MIN TSP led Edmonds [Edm66] to conjecture that the problem is not in P: this is sometimes referred to as the first statement of the $\mathsf{P} \neq \mathsf{NP}$ conjecture. See the book of Lawler et al. [LLKS85] for a very complete survey on MIN TSP. Here we will only review the results that are relevant for the present paper. The MIN TSP is NP-hard even if the cities are restricted to lie in $\mathcal{R}^2$ and the distances are computed according to the ℓ_2 norm [GGJ76, Pap77]. Due to such a negative result, research concentrated on developing good heuristics. 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. A 3/2-approximation algorithm that works for any metric space is due to Christofides [Chr76]. In twenty years of research no improvement of this bound had 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 MIN TSP 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 MIN TSP cannot be approximated within a factor $(1+\epsilon)$ in polynomial time, unless $\mathsf{P} = \mathsf{NP}$. The complexity of approximating MIN TSP in the case of geometric metrics remained a major open question. In his PhD thesis, Arora noted that proving the Max SNP-hardness of Euclidean MIN TSP in $\mathcal{R}^2$ should be very difficult, but that this could perhaps be done in $\mathcal{R}^{k(n)}$ for sufficiently large $k(n)$ ([Aro94, Chapter 9]). The relevance of non-approximability results for geometric MIN TSP was also stated in the open questions section of a survey by Arora and Lund [AL96]. In [GKP95], Grigni, Koutsopias and Papadimitriou proved that the restriction of the MIN 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 Euclidean MIN TSP has a PTAS in $\mathcal{R}^2$. They again posed the question of determining the approximability of the problem for higher dimensions. In a very recent breakthrough, Arora [Aro96] developed a PTAS for the MIN TSP in $\mathcal{R}^2$ under any ℓ_p metric. Such an algorithm also works in higher dimensional spaces and, in particular, it runs in time $n^{\tilde{O}(\log^{k(n)-2} n)/\epsilon^{k(n)-1}}$ in $\mathcal{R}^{k(n)}$. 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 Euclidean MIN TSP in $\mathcal{R}^n$ or if, at least, it was possible to have the running time being singly exponential in $k(n)$, e.g. $n^{O(k(n)/\epsilon)}$.

Our Results. In this paper we essentially answer negatively to both questions. We prove that MIN TSP in $\mathcal{R}^n$ is Max SNP-hard using any ℓ_p metric (thus, unless $\mathsf{P} = \mathsf{NP}$, there cannot be a PTAS for these problems). Furthermore, we show that $(1+\epsilon_1)$-approximating Euclidean MIN TSP in $\mathcal{R}^{\log n}$ is NP-hard under randomized reductions, for a proper constant ϵ_1. The latter result implies that there cannot be an algorithm that finds $(1+\epsilon)$-approximate solutions for Euclidean MIN TSP in $\mathcal{R}^k$ running in time $n^{O(k/\epsilon)}$ for any $\epsilon > 0$, unless $\mathsf{NP} \subseteq \mathsf{RQP}$, 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 n-dimensional case is proved by means of a reduction from the version of the metric MIN TSP that was shown to be Max SNP-hard in [PY93]. The reduction uses a mapping (see Lemma 5) of the metric spaces of [PY93] into Hamming spaces and the observation (see Proposition 3) that, for elements of $\{0,1\}^n$ a "gap" in the Hamming distance is preserved if distances are computed according to a ℓ_p metric. Our mapping of the metric spaces of [PY91] into Hamming spaces is *not* an *approximate isometry*, that is, it does *not* preserve distances up to negligible distorsion. We also suspect that such kind of mapping would be provably impossible. Instead, our mapping introduces a fairly high (yet constant) distorsion, but satisfies an additional condition that makes the mapping be an *L-reduction* [PY91]. Our mapping, combined with a reduction by Kariv and Hakimi [KH76], gives also non-approximability results for the Minimum k-Center Problem. The Minimum k-Cities Traveling Salesman Problem (MIN k-TSP) and the Minimum Degree-Restricted Steiner Tree Problem (two problems mentioned in Arora's paper [Aro96] on approximation schemes for geometric problems) are generalizations of the MIN TSP. The hardness results that we prove for MIN TSP clearly extend to them.

The Minimum Steiner Tree Problem

The origins of the MIN ST problem seem to be even more remote than the MIN TSP's ones: the case when $|U| = 3$ and the metric space is $\mathcal{R}^2$ with the ℓ_2 norm has been studied by the Italian mathematician Torricelli (a student of Galilei's) in 17th century. Reportedly, Gauss had an interest to this problem as well. Recent results about this problem are similar to the ones for MIN TSP: exact optimization is NP-hard in $\mathcal{R}^2$ both in the Rectilinear (ℓ_1) case [GJ77] and in the Euclidean (ℓ_2) case [GGJ77]. Constant-factor approximation is achievable in any metric space (the best factor should be 1.644 due to Karpinski and Zelikovsky [KZ95]), in general metric spaces the problem is Max SNP-hard [BP89], Arora's algorithm achieves approximation $(1+\epsilon)$ in $\mathcal{R}^k$ in time $n^{\tilde{O}(\log^{k(n)-2} n)/\epsilon^{k(n)-1}}$. No non-approximability result was known for geometric versions of the problem.

Our Results. We prove the Max SNP-hardness of the problem in $\mathcal{R}^n$ under the ℓ_1 norm. As a preliminary step, we prove the hardness of the problem restricted to Hamming spaces. The latter hardness is proved via a reduction from the Minimum Vertex Cover problem (MIN VC) restricted to triangle-free graphs of maximum degree 3. The Max SNP-hardness of this very restricted version of MIN VC is proved in this paper and could be used as a starting point for other non-approximability results. The reduction from MIN VC to Hamming MIN ST uses a combinatorial result (Claim 17) stating that for an instance where all points have weight[1] 2 or 0, if a technical condition is satisfied, there exists an optimum solution where all Steiner points have weight 1. We remark that there exists an instance of Hamming Steiner Tree where all the points have weight 3 or 0 and such that an optimum solution must contain a Steiner point of weight at least 4. Thus, our combinatorial result cannot be generalized too much. Reducing from Hamming Steiner Tree to Rectilinear Steiner Tree requires another combinatorial result (Theorem 19): for an instance where all the points are in $\{0,1\}^n \subset \mathcal{R}^n$, there exists an optimum solution where all the Steiner points lie in $\{0,1\}^n$. We prove this fact using the *integrality property* of Min-CUT linear programming relaxations. Our non-approximability result extends to MIN k-ST, the variation where one is also given an integer k and the goal is to find a minimum Steiner tree among the ones involving at most k Steiner points.

Discussion

We give the first non-approximability results for geometric versions of network optimization problems. For Euclidean MIN TSP, there is little room for improvement of our results, as well as there

[1]For a vector $u \in \{0,1\}^n$, its weight is defined as the number of non-zero coefficients, e.g. the weight of $(0,1,1,0,1)$ is three.

is little room for improving Arora's algorithm. If we believe that NP has not sub-exponential algorithms, then the best possible running time for an approximation scheme for Euclidean MIN TSP is of the form $2^{2^d/\epsilon}\text{poly}(n)$; alternatively, our non-approximability result could be extended to $\mathcal{R}^{\log/\log\log n}$. Much more consistent improvements are possible for MIN ST, however our results at least state very clearly that the number of dimensions *does matter* in the running time of an approximation scheme for these geometric problems.

We feel that one important contribution of this paper is the recognition of Hamming spaces as a class of metric spaces that seem to retain most of the hardness of general metrics while having a nice combinatorial structure. We believe that other non-approximability results could be established using Hamming spaces as intermediate steps. We also think that it should be worth trying to improve Christofides algorithm in Hamming spaces. While the well-behaved structure of Hamming spaces should not make this task impossible, it is likely that such an improved algorithm could give useful ideas for more general cases.

2 Preliminaries

We denote by $\mathcal{R}$ the set of real numbers. For an integer n we denote by $[n]$ the set $\{1, \ldots, n\}$. For a vector $\vec{a} \in \mathcal{R}^n$ and an index $i \in [n]$, we denote by $\vec{a}[i]$ the i-th coordinate of $\vec{a}$, Given an instance x of an optimization problem A, we will denote by $\mathsf{opt}_A(x)$ the cost of an optimum solution for x, we will also typically omit the subscript. For a feasible solution y (usually a tour or a tree) of an instance x of an optimization problem A, we denote its cost by $\mathsf{cost}_A(x, y)$ or, more often, as $\mathsf{cost}(y)$. In this paper we will use the notions of L-reduction and $\mathsf{Max\ SNP}$-hardness. $\mathsf{Max\ SNP}$ is a class of constant-factor approximable optimization problems that includes MAX 3SAT, we refer the reader to [PY91] for the formal definition.

Definition 1 (L-reduction) *An optimization problem A us said to be L-reducible to an optimization problem B if two constants α and β and two polynomial-time computable functions f and g exist such that*

1. *For an instance x of A, $x' = f(x)$ is an instance of B, and it holds $\mathsf{opt}_B(x') \leq \alpha\mathsf{opt}_A(x)$.*
2. *For an instance x of A, and a solution y' feasible for $x' = f(x)$, $y = g(x, y')$ is a feasible solution for x and it holds $|\mathsf{opt}_A(x) - \mathsf{cost}_A(x, y)| \leq \beta|\mathsf{opt}_B(x') - \mathsf{cost}_B(x', y')|$.*

We say that an optimization problem is $\mathsf{Max\ SNP}$-hard if all $\mathsf{Max\ SNP}$-problems are L-reducible to it. From [ALM+92] it follows that if a problem A is $\mathsf{Max\ SNP}$-hard, then a constant $\epsilon > 0$ exists such that $(1 + \epsilon)$-approximating A is NP-hard.

A function $d : U \times U \to \mathcal{R}$ 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 2 ($(1, 2) - B$ metrics) *A metric $d : U \times U \to \mathcal{R}$ 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 MIN TSP is Max SNP-hard even when restricted to $(1,2) - B_0$ metrics.

For an integer $p \geq 1$, the ℓ_p norm in $\mathcal{R}^n$ is defined as $||(u_1, \ldots, u_n)||_p = (\sum_{i=1}^n |u_i|^p)^{(1/p)}$. The distance induced by the ℓ_p norm is defined as $d_p(\vec{u}, \vec{v}) = ||\vec{u} - \vec{v}||_p$. For a positive integer n, we denote by d_H^n the Hamming metric in $\{0,1\}^n$ (we will usually omit the superscripts). We will make some use of the following fact.

Proposition 3 *Let $\vec{u}, \vec{v} \in \{0,1\}^n \subseteq \mathcal{R}^n$. Then $d_p(\vec{u}, \vec{v}) = d_H(\vec{u}, \vec{v})^{1/p}$.*

Before starting with the presentation of our results, we make the following important caveat.

Remark 4 *In some of the proofs of this paper we implicitly make 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 or Steiner Tree into an instance where coordinates are positive integers whose value is $O(n^2)$.*

3 MIN TSP and MIN k-CENTER

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 5 *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 simple corollary is required in the proof of our hardness result.

Corollary 6 *Let $p \geq 1$ be fixed. 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 $f : U \to \mathcal{R}^{3B|U|/2}$ such that for any $u, v \in U$, $d_p(f(u), f(v)) = 1$ if $d(u,v) = 1$ and $d_p(f(u), f(v)) = 1 + \delta$ if $d(u,v) = 2$. Such an embedding is computable in time polynomial in $|U|$.*

Proof: Map $U = \{u_1, \ldots, u_n\}$ into a set $U' = \{\vec{u}'_1, \ldots, \vec{u}'_n\}$ as in Lemma 5. From Proposition 3 we have that for any i and j, if $d(u_i, u_j) = 1$ then $d_p(\vec{u}'_i, \vec{u}'_j) = (2(B-1))^{1/p}$, and if $d(u_i, u_j) = 2$ then $d_p(\vec{u}'_i, \vec{u}'_j) = (2B)^{1/p}$. If we divide each coordinate of the points $\vec{u}'_i$ by $(2(B-1))^{1/p}$, we obtain a set of points in $\mathcal{R}^{3Bn/2}$ whose distances satisfy the hypothesis of the corollary, with $\delta = (2B/2(B-1))^{1/p} - 1$. The entire process can be done in time polynomial in $|U|$. □

The main result of this section is now only a matter of standard calculations.

Theorem 7 *For any fixed $p \geq 1$, the* Min TSP *is* Max SNP*-hard when restricted to the ℓ_p metric in $\mathcal{R}^{O(n)}$ (n is the number of cities).*

Proof: For some constant B_0, the Min 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 TSP in $\mathcal{R}^{O(n)}$. Let $x = (U, d)$ be an instance of the Min TSP, where $U = \{u_1, \ldots, u_n\}$ and d is a $(1,2) - B_0$ metric. We map the cities into $\mathcal{R}^{3B_0 n/2}$ as in Corollary 6, thus obtaining an instance x' of Min TSP in $\mathcal{R}^{3B_0 n/2}$. It is easy to see that, for any tour π,

$$\mathsf{cost}(x', \pi) = n + \delta(\mathsf{cost}(x, \pi) - n) = n(1 - \delta) + \delta \mathsf{cost}(x, \pi) \ .$$

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

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

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

Remark 8 *Given an instance of* Min TSP *with n points, if one adds n^c more points, all of them being at distance $1/O(n^{c+1})$ from some point of the instance, this perturbs the optimum in a negligible way. We can use this simple observation to scale down our hardness result to $\mathcal{R}^n$ (adding $3B_0 n/2 - 1$ points), or even to $\mathcal{R}^{n^\delta}$ for fixed $\delta > 0$ (adding $O(n^{1/\delta})$ points).*

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

Corollary 9 *For any positive integer $p \geq 1$, a constant $\epsilon^{(p)} > 0$ such that approximating the* Min TSP *in $\mathcal{R}^n$ within $(1+\epsilon^{(p)})$ is* NP*-hard.*

The following lemma, which was stated in [LLR95, Theorem 3.1] and was implicit in [JL84][2], will be used to give a non-approximability result for Euclidean Min TSP in $\mathcal{R}^{\log n}$.

Lemma 10 ([JL84]) *There exists a constant $\mu > 0$ such that the following holds. Let U be a set of n points into $\mathcal{R}^n$ and let $\gamma > 0$. Then there exists an embedding f of U into $\mathcal{R}^{\mu \log n/\gamma^2}$ such that for any $\vec{u}, \vec{v} \in U$, $(1-\gamma)d_2(\vec{u},\vec{v}) \leq d_2(f(\vec{u}), f(\vec{v})) \leq (1+\gamma)d_2(\vec{u},\vec{v})$. Such an embedding is computable by a randomized polynomial time algorithm.*

We can now prove the non-approximability result for $\mathcal{R}^{\log n}$.

Theorem 11 *There exist a constant $\epsilon_1 > 0$ such that $(1+\epsilon_1)$-approximating the Euclidean* Min TSP *in $\mathcal{R}^{\log n}$ is* NP*-hard under randomized reductions (and thus infeasible, unless* RP = NP*).*

Proof:[Of Theorem 11] Let $\epsilon^{(2)}$ be the constant of Corollary 9 for the Euclidean case. Fix constants c, γ and ϵ_1 such that

$$(1+\epsilon_1)(1+\gamma)/(1-\gamma) \leq 1+\epsilon^{(2)} \text{ and } c = \mu/\gamma^2.$$

Assume we have a $(1+\epsilon_1)$-approximate algorithm for Euclidean Min TSP in $\mathcal{R}^{\log n}$. Then the same algorithm can also be adapted to work in $\mathcal{R}^{\log n}$ (use a preprocessing phase that adds a bunch of n^c new dummy points, see Remark 8). Given an instance x with points in $\mathcal{R}^n$, we can map it into an instance x' with points in $\mathcal{R}^{c \log n}$ using Lemma 10 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

$$(1-\gamma)\mathsf{cost}(x,\pi) \leq \mathsf{cost}(x',\pi) \leq (1+\gamma)\mathsf{cost}(x,\pi)$$

and thus

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

If a tour π is $(1+\epsilon_1)$-approximate for x', then

$$\frac{\mathsf{opt}(x)}{m(x,\pi)} \leq \frac{\mathsf{opt}(x')/(1-\gamma)}{m(x',\pi)/(1+\gamma)} \leq \frac{1+\gamma}{1-\gamma}(1+\epsilon_1) \leq 1+\epsilon^{(2)}$$

and so π is $(1+\epsilon^{(2)})$-approximate for x. Since finding $(1+\epsilon^{(2)})$-approximate solutions in $\mathcal{R}^n$ is NP-hard, the above randomized reduction yields the NP-hardness (under randomized reductions) of $(1+\epsilon_1)$-approximating the Euclidean Min TSP in $\mathcal{R}^{\log n}$. □

The previous theorem could be improved in two aspects: use a deterministic reduction (so that hardness would be established under the nicer P $\neq$ NP condition instead that under the RP $\neq$ NP condition) and prove hardness for other ℓ_p metrics. Both improvements boil down to proving the following conjecture.

Conjecture 1 *For any fixed $p \in \mathcal{Z}^+$ and $\gamma > 0$, a polynomial-time computable function $f : \mathcal{R}^n \rightarrow \mathcal{R}^{(\mathrm{poly}(1/\gamma))\log n}$ exists such that, for any two vectors $\vec{u}, \vec{v} \in \mathcal{R}^n$*

$$(1-\gamma)d_p(\vec{u},\vec{v}) \leq d_p(f(\vec{u}), f(\vec{v})) \leq (1+\gamma)d_p(\vec{u},\vec{v}) \ .$$

[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].

The k-Center Problem In the Min k-Center problem one is given a set of points U in a metric space and an integer k. The goal is to select a set S of k points (the *centers*) so that it is minimized the maximum distance between a point of $U - S$ and the closest center. The problem has been studied in several papers, including [KH76, HS85].

Theorem 12 *For any fixed $p \geq 1$, it is* NP*-hard to approximate* Min k-Center *within any factor smaller than $(1.5)^{1/p}$ when the points are in $\mathcal{R}^n$ and distances are computed according to the ℓ_p norm.*

Proof: [Of Theorem 12] Kariv and Hakimi [KH76] give a simple reduction from the Dominating Set problem to the k-center problem. Given a graph G, consider a metric space and a set U of points such that each point of U corresponds to a node of the graph and such that points corresponding to adjacent vertices are at distance d_1, while the others are at distance d_2 (where $d_1 < d_2$). It is easy to prove that, if the graph admits a dominating set of size k, the optimum of the k-Center instance (U, k) is d_1; otherwise the optimum is d_2. The dominating set problem is NP-hard even for graphs of maximum degree 3 [GJ79]. Such graphs can be mapped into an ℓ_p metric space so that adjacent vertices are at distance $4^{1/p}$ and non-adjacent vertices are at distance $6^{1/p}$. Thus, the k-center problem cannot be approximated within a factor smaller than $(1.5)^{1/p}$ in ℓ_p normed $\mathcal{R}^n$ spaces. □

4 The Min ST Problem

The hardness of approximating Min ST will be established with a longish chain of reductions. The starting point is the following hardness result, that may have a little independent interest. Recall that in the Minimum Vertex Cover (Min VC) one is given a graph $G = (V, E)$ and looks for the smallest set $C \subseteq V$ such that C contains at least one endpoint of any edge in E.

Theorem 13 *The* Min VC *problem is* Max SNP*-hard even when restricted to triangle-free graphs with maximum degree 3 (we call this restriction* Min TF VC-3*).*

Proof: [Of Theorem 13] The Max 2SAT problem is Max SNP-hard even when restricted to instances where each variable occurs in at most 3 clauses (apply to Max 2SAT the reduction from Max 3SAT to Max 3SAT-3 described in [Pap94]). One can assume without loss of generality that the 3 occurrences of each variable are either one positive occurrence and two negative occurrences, or vice versa. We reduce Max 2SAT-3 to Min VC using the reduction of [PY91]: we create a graph with a node for any occurrence of any literal, putting an edge between two nodes if they represent literals that occur in the same clause or if they are one the complement of the other. See [PY91] for the proof that this is an L-reduction. The obtained graph has maximum degree 3: each literal is adjacent to the fellow literal occurring in the same clause and to the (at most) two occurrences of its complement. Also, the graph is triangle-free: let l_1, l_2 and l_3 be any three occurrences of literals. Since clauses contain only two literals, from pigeonhole principle it follows that one of the three occurrences (say, l_1) does not occur in the same clause with any of other two. Then, if l_1, l_2 and l_3 form a triangle it follows that l_2 and l_3 are both the complement of l_1. Being adjacent, they also have to occur in the same clause, but this is a contraddiction since the literals occurring in a clause have to be different. □

We note in passing that, as a corollary, we obtain that the Max Independent Set problem is Max SNP-hard in the same, very restricted class of graphs. We now move to the restriction of Min ST to Hamming spaces.

Notation: For a pair of indices $i, j \in [n]$ we define $\vec{a}^n_{i,j} \in \{0,1\}^n$ as the n-dimensional boolean vector all whose coordinates are zero but the i-th and the the j-th, e.g. $\vec{a}^5_{1,4} = (1,0,0,1,0)$. Similarly, we let $\vec{a}^n_i$ be the the vector in $\{0,1\}^n$ whose only non-zero coordinate is the i-th, e.g. $\vec{a}^4_3 = (0,0,1,0)$. For a vector $\vec{a} \in \{0,1\}^n$ and indices $i, j \in [n]$, we let $\mathsf{red}_{i,j}(\vec{a}) \in \{0,1\}^n$ be the vector defined as follows

$$\mathsf{red}_{i,j}(\vec{a})[h] = \begin{cases} \vec{a}[h] & \text{if } h \neq i \wedge h \neq j \\ 0 & \text{if } (h = i \vee h = j) \wedge \vec{a}[i] = \vec{a}[j] = 1 \end{cases}$$

In other words, $\mathsf{red}_{i,j}(\vec{a})$ is equal to $\vec{a}$ unless a has a one in the i-th and the j-th coordinate. In this latter case, the i-th and the j-th coordinate of $\mathsf{red}_{i,j}(\vec{a})$ are set to zero. For example $\mathsf{red}_{1,3}(0,1,1,1) = (0,1,1,1)$, while $\mathsf{red}_{2,3}(0,1,1,1) = (0,0,0,1)$. We will make use of the following simple combinatorial lemma.

Lemma 14 *For any $\vec{a}, \vec{b} \in \{0,1\}^n$, for any $i, j \in [n]$, $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) \leq d_H(\vec{a}, \vec{b})$.*

Proof: Without loss of generality, assume $i = 1$ and $j = 2$. The lemma trivially holds when both $\vec{a} = \mathsf{red}_{1,2}(\vec{a})$ and $\vec{b} = \mathsf{red}_{1,2}(\vec{b})$. Assume $\vec{a} \neq \mathsf{red}_{1,2}(\vec{a})$ (the other case is perfectly symmetric), that is, $\vec{a} = (1,1,\vec{a}')$ (with $\vec{a}' \in \{0,1\}^{n-2}$). There are four cases to be considered:

- If $\vec{b} = (1,1,\vec{b}')$, then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = d_H((0,0,\vec{a}'),(0,0,\vec{b}')) = d_H(\vec{a}', \vec{b}') = d_H(\vec{a}, \vec{b})$.
- If $\vec{b} = (0,1,\vec{b}')$ or $b = (1,0,\vec{b}')$ then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = 1 + d_H(\vec{a}', \vec{b}') = d_H(\vec{a}, \vec{b})$.
- If $\vec{b} = (0,0,\vec{b}')$, then $d_H(\mathsf{red}_{i,j}(\vec{a}), \mathsf{red}_{i,j}(\vec{b})) = d_H(\vec{a}', \vec{b}') = d_H(\vec{a}, \vec{b}) - 2$.

□

Theorem 15 *The* Min ST *problem is* Max SNP*-hard when restricted to Hamming spaces.*

Proof: We give an L-reduction from Min TF VC-3. Let $G = (V, E)$ be a triangle-free graph of maximum degree 3, assume $V = [n]$ and let $m = |E|$. We define an instance of Hamming Min ST as follows: the number of dimensions is n and the set of points is

$$U = \{\vec{0}\} \cup \{\vec{a}^n_{ij} : \{i,j\} \in E\}$$

where $\vec{0}$ is the vector with all zero entries.

Claim 16 *Given a vertex cover $C \subseteq V$ in G it is possible to find a Steiner tree for U of cost $m + C$.*

Proof:[Of Claim 16] Let $S = \{\vec{a}^n_i : i \in C\}$. Consider the graph whose vertex set is $S \cup U$ and such that two vertices are adjacent iff their Hamming distance is one. We claim that this graph is connected: indeed all the nodes of S are clearly adjacent to $\vec{0}$; furthermore any node in U is adjacent to some node in S (since C is a vertex cover), thus all the nodes are connected to $\vec{0}$. Since the graph is connected it admits a spanning tree, that is also a Steiner tree for U. All the edges of such Steiner tree have cost 1, and there are $|C| + m$ of them (because the tree has $|S| + |U| = |C| + m + 1$ nodes), so the claim follows. □

From the above claim it follows that $\mathsf{opt}(U) \leq m + \mathsf{opt}(G) \leq 4\mathsf{opt}(G)$, and we have established the first condition of the L-reducibility. As usual, the other condition is more difficult to prove.

Claim 17 *Given a Steiner tree T for U it is possible to find in polynomial time another Steiner tree T' such that: (i) $\mathsf{cost}(T') \leq \mathsf{cost}(T)$ and (ii) all the edes of T' have cost one and all the Steiner nodes of T' are weight-one vectors.*

PROOF:[Of Claim 17] We first make sure that all edges have cost 1: any edge of cost $d > 1$ is broken into a length-d path using $d - 1$ additional Steiner nodes. Let S be the new set of Steiner vertices. We now reduce the number of non-zero coordinates of Steiner vertices. For any $\{i, j\} \notin E$ we map each point $\vec{a} \in S \cup U$ into $\mathsf{red}_{i,j}(\vec{a})$; this mapping only changes Steiner points (by definition of $\mathsf{red}_{i,j}$, definition of U, and the fact that $\{i, j\} \notin E$). From Lemma 14 we also have that any phase does not increase the cost of the tree. At the end of this set of transformations, we run a "clean-up" phase that does the following: if some transformation has collapsed one node onto another, we take only one node (if a Steiner node is collapsed onto a node in U we clearly take the node in U). If the transformation creates cycles, we break them (e.g. finding a spanning tree of the final graph), and, again, this does not increase the cost. It remains to see that, after this process, no Steiner node can have more than one non-zero coordinate. If a Steiner node has some set of k non-zero coordinates, then they must correspond to a clique in G (otherwise, at some phase, some of them would have been changed by the application of the red operator): since G is triangle-free, $k \leq 2$, but if $k = 2$ then the Steiner node would be equal to a node of U, and thus would have been removed in the clean-up phase. It follows that $k = 1$. □

From the above claim, the following one follows quite easily.

Claim 18 *Given a Steiner tree T for U it is possible to find in polynomial time a vertex cover C for G such that $|C| \leq \mathsf{cost}(T) - m$*

PROOF: [Of Claim 18] We first find the Steiner tree T' as in the previous claim. Then, if we let S be the set of Steiner vertices of T', we have that $\mathsf{cost}(T) \geq \mathsf{cost}(T') = |S| + |U| - 1 = |S| + m$. Let $C = \{i : \vec{a}_i \in S\}$; it is easy to see that $|C| = |S| \leq \mathsf{cost}(T) - m$ and that C is a vertex cover in G. The latter fact follows from the fact that for any edge $\{i, j\} \in E$, the vector $\vec{a}_{i,j}$ belongs to U; since all the edges of T' have cost one, and T' does not contain weight-3 vectors, then either $\vec{a}_i \in S$ or $\vec{a}_j \in S$ (otherwise $\vec{a}_{i,j}$ would be an isolated node in T' contraddicting the fact that T' be connected). □

If T is any Steiner tree of U, the vertex cover C for G computed according the previous claim satisfies

$$\mathsf{cost}(C) - \mathsf{opt}(G) \leq (\mathsf{cost}(T) - m) - (\mathsf{opt}(U) - m) = \mathsf{cost}(T) - \mathsf{opt}(U)$$

and so also the second condition of the L-reduction is satisfied. □

If the following conjecture holds, then we can reduce MIN VC-B to Hamming MIN ST (without imposing the triangle-free restriction).

Conjecture 2 *Let $U \subset \{0, 1\}^n$ be an instance of Hamming* MIN ST *such that $\vec{0} \in U$ and all vectors of U have weight at most 2. Then there exists an optimum solution where all the Steiner nodes have weight at most 2.*

Janos Körner proposed a further generalization: if U is contained in the Hamming sphere centered in some $\vec{u} \in U$ and of radius k, then there exists an optimum solution all whose Steiner nodes lie in the same sphere. This seemed to be a reasonable combinatorial analog of the fact that if the points are in $\mathcal{R}^k$ and distances are computed according to the Euclidean metric, the Steiner points of an optimum solution will be in the *convex hull* of the points of the instance. Lately, Janos refuted the generalized conjecture. The instance $U = \{(0,0,0,0),(0,1,1,1),(1,0,1,1),(1,1,0,1),(1,1,1,0)\}$ refutes the generalized conjecture even for $k = 3$. An optimum solution of cost 7 uses the Steiner node $(1,1,1,1)$. Computational experiments show that any solution without $(1,1,1,1)$ has cost at least 8.

To approach the MIN ST in ℓ_1 normed spaces we use a reduction from the Hamming case. Note that for points in $\{0,1\}^n$ the ℓ_1 distance equals the Hamming distance. However, the reduction is non-trivial since $\mathcal{R}^n$ contains so many points that are not in $\{0,1\}^n$ and we have to argue that having much more choice for the Steiner nodes does not make the problem easier. The Rectilinear MIN ST problem looks very much like a *relaxation* of the Hamming MIN ST problem; our reduction makes use of a *rounding scheme* proving that the relaxation does not change the optimum.

Theorem 19 *Let $U \subseteq \{0,1\}^n \subset \mathcal{R}^n$ be an instance of Rectilinear* MIN ST *all whose points are in the Boolean cube. Let T be a feasible solution for U. Then it is possible to find in polynomial time (in the size of T) another solution T' such that $\mathsf{cost}(T') \leq \mathsf{cost}(T)$ and all the Steiner nodes of T' are in $\{0,1\}^n$.*

Before proving the theorem, we note the following relevant consequence.

Corollary 20 *For any instance $U \subseteq \{0,1\}^n$ of Rectilinear* MIN ST, *an optimum solution exists all whose Steiner points are in $\{0,1\}^n$.*

We now prove Theorem 19.

PROOF:[Of Theorem 19] Let $S = \{\vec{s}_1, \ldots, \vec{s}_m\}$ be the set of Steiner points of T, and let E be the set of edges of T. For any $\vec{s}_j \in S$ we will find a new point $\vec{s}'_j \in \{0,1\}^n$, so that if we let T' be the tree obtained from T by substituting the $\vec{s}$ points with the corresponding $\vec{s}'$ points, the cost of T' is not greater than the cost of T. The latter statement is equivalent to

$$\sum_{(\vec{s}_j,\vec{u})\in E, \vec{u}\in U} ||\vec{s}_j - \vec{u}||_1 + \sum_{(\vec{s}_j,\vec{s}_h)\in E} ||\vec{s}_j - \vec{s}_h||_1 \geq \sum_{(\vec{s}'_j,\vec{u})\in E, \vec{u}\in U} ||\vec{s}'_j - \vec{u}||_1 + \sum_{(\vec{s}'_j,\vec{s}'_h)\in E} ||\vec{s}'_j - \vec{s}'_h||_1$$

We will indeed prove something stronger, namely, that for any $i \in [n]$ it holds

$$\sum_{(\vec{s}_j,\vec{u})\in E\ ,\ \vec{u}\in U} |\vec{s}_j[i] - \vec{u}[i]| + \sum_{(\vec{s}_j,\vec{s}_h)\in E} |\vec{s}_j[i] - \vec{s}_h[i]| \geq \sum_{(\vec{s}'_j,\vec{u})\in E\ ,\ \vec{u}\in U} |\vec{s}_j[i] - \vec{u}[i]| + \sum_{(\vec{s}'_j,\vec{s}'_h)\in E} |\vec{s}'_j[i] - \vec{s}'_h[i]| \tag{1}$$

Let $i \in [n]$ be fixed, we now see how to find values of $\vec{s}'_1[i], \ldots, \vec{s}'_m[i] \in \{0,1\}$ such that (1) holds. We express as a linear program the problem of finding values of $\vec{s}'_1[i], \ldots, \vec{s}'_m[i]$ that minimize the right-hand side of (1). For any $j \in [m]$ we have a variable x_j (representing the value to be given to $\vec{s}'_j[i]$) and for any edge $e = (\vec{a}, \vec{b})$ such that at least one endpoint is in S we have a variable y_e, representing the lenght $|\vec{a}[i] - \vec{b}[i]|$. The linear program is as follows

$$
\begin{array}{lll}
\min & \sum_e y_e & \\
\text{Subject to} & & \\
& y_e \geq x_j - x_h & \forall e = (\vec{s}_j, \vec{s}_h) \in E \\
& y_e \geq x_h - x_j & \forall e = (\vec{s}_j, \vec{s}_h) \in E \\
& y_e \geq x_j & \forall e = (\vec{s}_j, \vec{u}_h) \in E \text{ such that } \vec{u}_h[i] = 0 \\
& y_e \geq 1 - x_j & \forall e = (\vec{s}_j, \vec{u}_h) \in E \text{ such that } \vec{u}_h[i] = 1 \\
& x_j \geq 0 & \\
& y_e \geq 0 &
\end{array}
\tag{LP}
$$

Setting $x_j = s_j[i]$ and setting $y_{(\vec{a},\vec{b})} = |\vec{a}[i] - \vec{b}[i]|$ yields a feasible solution, and its cost is the left-hand side of (1). Let $(\vec{x}^*, \vec{y}^*)$ be an optimum solution for (LP). From the previous observation we have that setting $\vec{s}'_j[i] = x^*_j$ we satisfy (1). It remains to be seen that (LP) has an optimum solution where all variables take value from $\{0,1\}$. This follows from the fact that (LP) is the linear programming relaxation of an undirected Min-CUT problem, where all the $\vec{u}$ such that $\vec{u}[i] = 0$ (respectively, $\vec{u}[i] = 1$) are identifed with the source (respectively, the sink), each $\vec{s}_j$ is a node, and the edges are like in T. It is well known (see e.g. [PS82]) that a Min-CUT linear programming relaxation has optimum 0/1 solutions, and that such a solution can be found in polynomial time. □

Remark 21 *There seems to be no natural analog of Theorem 19 in other norms. Even in* $\mathcal{R}^2$*, using the Eulcidean metric, we have that the optimum solution of the instance* $\{(0,0),(1,0),(0,1)\}$ *must use a Steiner point not in* $\{0,1\}^2$.

Theorem 22 *Rectilinear* MIN ST *is* Max SNP*-hard.*

PROOF: We reduce from Hamming MIN ST. The reduction leaves the instance unchanged. For an instance $U \subseteq \{0,1\}^n$, we let $\mathsf{opt}_H(U)$ (respectively, $\mathsf{opt}_R(U)$) be the cost of an optimum solution for U, when seen as an instance of Hamming MIN ST (respectively, of Rectilinear MIN ST). Clearly, we have that $\mathsf{opt}_R(U) \leq \mathsf{opt}_H(U)$. Given a solution T for U, we find a solution T' as in Theorem 19. Since in $\{0,1\}^n$ the distance induced by the ℓ_1 norm equals the Hamming distance, we have that $\mathsf{cost}_H(T') = \mathsf{cost}_R(T') \leq \mathsf{cost}_R(T)$. We have an L-reduction with $\alpha = \beta = 1$. □

It is easy to see that the Rectilinear MIN k-ST and the Hamming MIN k-ST are Max SNP-hard as well. The instances produced by the above reductions have the property that an optimum solution has at most n Steiner nodes. So the reduction works unchanged for MIN k-ST.

5 Open questions

We don't know how to extend our non-approximability result for MIN ST to the Euclidean case. Arora [Aro96] notes that, by inspecting the way his algorithm works, it is possible to claim that, for any instance of Euclidean MIN ST, there exists a near-optimal solution where the Steiner points lie in some well-specified positions (either at "portals" or in positions chosen at the bottom of the recursion). This observation could perhaps be a starting point.

We don't have explicit estimations of the constants to within which it is hard to approximate geometric MIN TSP and rectilinear MIN ST. The constant for MIN TSP should be only slightly smaller than the corresponding constant for the $(1,2) - B$ case (estimated around $1 + 10^{-5}$). The constant for MIN ST is more likely to be around $1 + 10^{-4}$. Finding much stronger estimations (comparable to the 3/2 bound of Christofides and the 1.644 bound of Karpinski and Zelikovsky) is an open and challenging question. It appears that MIN TSP and MIN ST lack the nice logical definability that allows to prove very strong non-approximability results for MAX CUT and MAX 3SAT using so-called "gadget reductions" [BGS95, TSSW96, Hås96].

Acknowledgments

I'm grateful to Sanjeev Arora for proving Theorem 11 and to Madhu Sudan for proving Lemma 14. I thank Laszlo Babai and Janos Körner for very interesting and useful discussions. Ray Greenlaw carefully read a preliminary versions of this paper and gave several helpful remarks.

References

[AL96] S. Arora and C. Lund. Hardness of approximations. In *Approximation Algorithms for NP-hard Problems*. PWS Publishing, 1996.

[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.

[BGS95] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP's and non-approximability – towards tight results (3rd version). Technical Report TR95-24, Electronic Colloquium on Computational Complexity, 1995. Preliminary version in *Proc. of FOCS'95.*

[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.

[GGJ77] M.R. Garey, R.L. Graham, and D.S. Johnson. The complexity of computing Steiner minimal trees. *SIAM Journal of Applied Mathematics*, 34:477–495, 1977.

[GJ77] M.R. Garey and D.S. Johnson. The rectilinear Steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-Completeness.* Freeman, 1979.

[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.

[Hås96] J. Håstad. Getting optimal in-approximability results. Manuscript, September 1996.

[HS85] D.S. Hochbaum and D.B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184, 1985.

[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 paces in low dimensional normed spaces. In *Geometric Aspects of Functional Analysis*, pages 177–184. LNM 1267, Springer-Verlag, 1987.

[KH76] O. Kariv and S.L. Hakimi. An algorithmic approach to network location problem – Part I: the p-centers. Unpublished manuscript, cited in [GJ79], 1976.

[KZ95] M. Karpinski and A. Zelikovsky. New approximation algorithms for the Steiner tree problems. Technical Report TR95-030, Electronic Colloquium on Computational Complexity, 1995.

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

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. Preliminary Version in *Proc. of FOCS'94.*

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

[Pap94] C.H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[PS82] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prenitce-Hall, 1982.

[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.

[TSSW96] L. Trevisan, G.B. Sorkin, M. Sudan, and D.P. Williamson. Gadgets, approximation, and linear programming. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, 1996.