Wojciech RAFAJŁOWICZ, Ewa SKUBALSKA-RAFAJŁOWICZ
Politechnika Wrocławska
Katedra Informatyki Technicznej

ZANURZENIE METRYCZNEGO PROBLEMU KOMIWOJAŻERA W PRZESTRZENIACH EUKLIDESOWYCH

Streszczenie. Przedstawiamy problem zanurzenia metrycznego problemu komiwojażera w przestrzeniach euklidesowych. W pracy przedstawiono algorytm zanurzania problemu przedstawionego w postaci macierzy odległości. Efektem działania algorytmu są położenia punktów w przestrzeni Euklidesowej. Dzięki takim zanurzeniom uzyskuje się możliwość zastosowania znanych, bardzie efektywnych algorytmów dla problemu TSP. Algorytm ten został przetestowany numerycznie na przykładach ze zbioru TSPLIB.

EMBEDDING OF THE METRIC tsp INTO eUCLIDEAN SPACES

Summary. This paper presents a fast incremental algorithm for embedding the metric TSP data sets in Euclidean spaces. The algorithm input is in a form of a distance matrix. The result are positions in the Euclidean space. The application of the proposed method allows for using more efficient methods for the Euclidean TSP problem. Proposed method was tested on TSPLIB – standard set of benchmarks for TSP.

1. Introduction

The traveling salesman problem (TSP) is one of most well known problems in combinatorial optimization [7], [11]. The problem can be formulated as follows. Given a complete undirected graph $G = (V, E)$ with vertex set V and set of edges E, with non-negative edge costs $d : \ E \to R_+$, the objective is to find a Hamiltonian cycle in G of minimum cost. In the general TSP formulation there are no restrictions on the cost function. But, in general, TSP cannot be approximated in polynomial time (unless P = NP). In order to find approximate solutions for TSP, one should require that instances of the problem have costs that satisfy the triangle inequality ($d_{ij} \leqslant d_{ik} + d_{kj},\ i, j, k \in V$). Such problem is known as the Metric TSP (MTSP) problem. The Euclidean TSP (ETSP) is a special case of the Metric TSP. The vertex set V is consider as $|V| = n$ points in $\mathcal{R}^m$, where m is fixed. The graph is complete and the Euclidean distance is used as a cost function.

In this paper, we assume that only distance matrix $D = (d_{ij}),\ i, j \in V$ is given. Our goal is to check if the matrix an Euclidean distance matrix (EDM) [4], [5], to determine the minimum dimension m of the adequate Euclidean space and give the location of vertices V in this space. That is, possibly to determine the function $\psi : V \rightarrow \mathcal{R}^m$. It is obvious that

Our motivations are twofolds:
1) Sanjeev Arora has found a Polynomial Time Approximation Scheme (PTAS) for Euclidean TSP [2]. An algorithm admits the PTAS scheme, if for any fixed error parameter $\varepsilon > 1$, the running time is bounded by a polynomial in n and the costs of solution (the computed tour) does not exceed $(1 + 1/\varepsilon)OPT$, where OPT stands for the costs of an optimal tour. Furthermore, it is known that there exists a constant c > 1, for which c–approximation is NP-hard, if the TSP is metric [7], [6].
2) Even if the Metric TSP is not an ETSP , a low-dimensional finite vector space representing a part of the graph, or a decomposition of the problem onto several independent partial low-dimensional Euclidean embeddings could be a source of new approximate solutions.

The paper is organized as follows: firstly the motivation in form of a metric TSP problem is given, then the problem of reconstruction euclidean distance matrix is given, an algorithm for this reconstruction is proposed. Finally an account of numerical experiments is presented as well as some propositions for further research.

2. The Metric TSP

The TSP problem is the MTSP if and only if:

- $d_{ij} \leqslant 0,\ i, j \in V,$
- $d_{ij} = 0$ if and only if $i = j$
- $d_{ij} \leqslant d_{ik} + d_{kj},\ i, j, k \in V$

There is a special case of the Metric TSP called the Euclidean TSP. In the Euclidean TSP, the weights of edges corresponds to the Euclidean distances between their endpoints in the Euclidean space. Thus, there exists the function $\psi : V \rightarrow \mathcal{R}^m$, such that

$$d_{ij} = ||\psi(i) - psi(j)||,\ i, j \in V,$$

where $|| \cdot ||$ is the Euclidean norm in $\mathcal{R}^m$. For the sake of simplicity of the notation we will assume that $|V| = n + 1$ and $i, j \in \{0, \ldots, n\}$.

3. The Euclidean Distance Matrix Problem

Euclidean distance matrices (EDMs) have appeared during the last thirty years, motivated by applications to the multidimensional scaling problem and molecular conformation problems in Biology [8]. These applications focus on the construction or reconstruction of sets of points such that the distances between these points are as close as possible to a given set of inter-point distances

Define $a_{ij} = d_{ij}^2$ as a squared distance between nodes $i \in V$ and $j \in V$. Matrix $A = \{a_{ij}\}$ is the distance matrix if and only if all elements on the diagonal of A are zero, the matrix is symmetric, i.e., $a_{ij} = a_{ji}$, $a_{ij} \geqslant 0$ and (by the triangle inequality)

$$\sqrt{a_{ij}} \leqslant \sqrt{a_{ik}} + \sqrt{a_{kj}}.$$

Theorem 1. ([8], [10], [14]) *A necessary and sufficient condition for the isometric embeddability of a finite metric set* (S, d) *of* $n + 1$ *elements in an Euclidean space* $\mathcal{R}^n$ *is that the following statement be true:*
The matrix $[\frac{1}{2}(a_{0i} + a_{0j} - a_{ij})]_{i,j=1,\ldots,n}$ *is positive definite.*

There are also known a conditions of embeddability based on the Cayley-Menger derterminants [8], [15], but discussion of that approach is out of scope of our paper.

It is well known that a complete graph embedded in R^m is rigid in R^m [1]. An embedding is locally unique [13], i.e., we say that $p : V \to \mathcal{R}^m$ and $q : V \to \mathcal{R}^m$ are congruent, if

$$||p_i - p_j|| = ||q_i - q_j||$$

for all pairs $i, j \in V$. As a consequence, if a MTSP is the Euclidean Problem, any solution $\psi : V \to \mathcal{R}^m$ is locally unique.

4. An Incremental Algorithm for Embedding MTSP in the Euclidean Space

In this section we provide an computionally efficient algorithm for embedding the MTSP in the Euclidean Space. We assume that the true dimension is not known. The algorithm provides the dimension number of the embedding and the locally unique embedding of V when the TSP is the ETSP. In such a case the embedding does not depend on the vertex chosen as a starting point.

We assume that metric distance matrix $D = [d_{ij}]_{ij,0,\ldots,n}$ and squared distance matrix $A = [d_{ij}^2]_{ij,0,\ldots,n}$ are given. We begin with a set of three vertices, let say, vertices labeled by $0, 1$ and 2. It is obvious that these vertices can be embedded in $\mathcal{R}^2$ and this embedding is locally unique. Vertex-representing points form a triangle and the shape of this triangle is unique. So, without loss of the generality we can locate vertex v_0 in $x_0 = (0, 0)$, v_1 in $x_1 = (d_{01}, 0)$ and v_2 in $x_2 = (x_{21}, x_{22})$, where x_2 coordinates are obtained by solving the following system of two equations:

$$(x_{21} - x_{01})^2 + (x_{22} - x_{02})^2 = a_{02},$$

$$(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 = a_{12}.$$

The system simplifies to

$$x_{21}^2 + x_{22}^2 = a_{02},\ (x_{21} - d_{01})^2 + x_{22}^2 = a_{12}.$$

Further, substracting the first equation from the second one, we arrive at:

$$-2d_{01}x_{21} + a_{01} = a_{12}.$$

Thus, $x_{21} = -0.5(a_{12} - a_{01})/d_{01}$, and consecutively

$$x_{22}^2 = a_{02} - \frac{(a_{12} - a_{01})^2}{4a_{01}}.$$

Due to the triangle inequality,

$$a_{02} - \frac{(a_{12} - a_{01})^2}{4a_{01}} \geqslant 0.$$

Thus, there exist at most two solutions of the system with x_{22} being a positive or a negative real number. When points x_1, x_2, x_3 are colinear, $x_{22} = 0$ and the embedding dimension is $m = 1$. As a consequence, all coordinates $x_{\dot{2}}$ can be neglected and could be removed.

Adding a new vector to the Euclidean space leads to the following problem:

Problem 1. *Let's assume that we have given coordinates of* $r + 1$ *points representing an embedding of* $V_r \subset V$, $|V_r| = r + 1$ *in a Euclidean space, i.e.,* $x_0, x_1, \ldots x_r \in R^s$. s *is dimension of the Euclidean space,* $s \leqslant r$. *It is clear, that we have:*

$$||x_i - x_j||^2 = a_{ij},\ i,\ j \in 0, \ldots, r.$$

Find a vector representing a new vertex, let say, $v_{r+1} \in V - V_r$.

It is clear that the dimension of the Euclidean embedding space could be larger. Namely, $m = s + 1$ or $m = s$.

Our goal is to find

$$x_{r+1} = (x_{r+1,1}, x_{r+1,2}, \ldots, x_{r+1,s+1})$$

such that

$$||x_i - x_{r+1}||^2 = a_{i,r+1},\ i = 0, 1, \ldots, r, \tag{1}$$

where dimension of all points $X_r = [x_i]_{i=0,1,\ldots,r}$ is expanded to $s + 1$, and the $(s + 1)$-th coordinates of all vectors in X_r are set to zero. If $x_{r+1} \in \mathcal{R}^{s+1}$ exists and $x_{r+1,s+1} = 0$ the system of $r+1$ points lays in s dimensional Euclidean space, ie., $x_1, x_2, \ldots x_r, x_{r+1} \in \mathcal{R}^s$. All last coordinates could be removed or neglected. When $x_{r+1,s+1} \neq 0$, the new dimension is $s + 1$. The lack of a real solution means that the metric d is not a Euclidean metric. Observe, that system of equations (1) is equivalent to the system consisting of

$$\sum_{j=1}^{s+1} x_{r+1,j}^2 = a_{0,r+1}, \tag{2}$$

and, after substraction, of the system of r linear equations

$$\sum_{j=1}^{\min[i,s]} x_{i,j} x_{r+1,j} = \frac{1}{2}[a_{0,i} - a_{i,r+1}],\ i = 1, \ldots, r. \tag{3}$$

System (3) can be overderminated (when $s < r$) but consistent. Thus, the solution of (3) can be obtained using ordinary least square method. The first s coordinates of x_{r+1} are given by:

$$\mathcal{X}^T \mathcal{X} x_{r+1}(1 : s)^T = \mathcal{X}^T b, \tag{4}$$

where $x_{r+1}(1 : s) = (x_{r+1,1}, \ldots, x_{r+1,s})$, $b = \frac{1}{2}[a_{0,i} - a_{i,r+1}]_{i=1,\ldots,r}^T$ and $\mathcal{X} = [x_{ij}]_{i=1:r,j=1:s}^T \in \mathcal{R}^{s\times r}$. The computational complexity of solving (3) is $O(r^3)$. Thus, the computational complexity of embedding V, is $O(n^4)$ [16]. Furthermore, using triangle-like structure of $\mathcal{X}$, it is possible to get complexities $O(r^2)$ and $O(n^3)$, respectively. Nevertheless, the more complex approach allows us better control numerical errors.

There are known conditions which guarantee that the real solution of (2)-(3) exists [4], [9], [8], [15], but verifying these conditions has similar computational complexity as a direct solving of the problem.

Table 1

Dependence of dimensionality on required accuracy ϵ.

ϵ	1.0e-03	5.0e-03	1.0e-02	5.0e-02	1.0e-01	5.0e-01	1.0e+00	5.0e+00
d	7	7	6	3	2	2	1	1

5. Experiments

The well known and widely used set of benchmarks for TSP problems is called TSPLIB, originally published in [12] . The optimal solutions for those problems are known (at http://www.iwr.
uni-heidelberg.de/groups/comopt/software/TSPLIB95). The set consists of 111 different problem in 6 general types. Of these we are interested in the following groups:

- symmetric problems with euclidean 2D distances,
- symmetric problems with explicit weights in form of distance matrices.

The implementation of algorithm described in previous section was done in Python programming language, linear algebra methods came from `numpy.linalg` package. The implemented algorithm made use of least-square method of solving the overdefined linear system (3) in order to reduce possible numerical error. Even with LSQ method, the decision whether to add additional dimension require some decision point. We had chosen

$$x_{r+1}(s+1) \leqslant \varepsilon,$$

where ε is dependent on scale (in the sense of distances not number of elements) of the specific problem. The easiest choice is

$$\varepsilon = \epsilon \cdot \max_{ij} a_{ij}$$

As a method of verification of the concept and its accuracy the 61 problems of type `EUC_2D` (two dimensional problems where coordinates of the cities are known and the distance is calculated as simple euclidean norm) were converted to distance matrices. We know the exact dimensionality is 2. By this reversed method we can easily verify accuracy of the result.

In the fig. 1 we can see the results of the reconstruction in the comparison to original data. The resulting image is accurate with respect rotation and mirroring. The general accuracy was around $\epsilon = 10^{-5}$.

All experiments with data of known dimensionality had shown that the choice of ϵ (which generates ε) is crucial to accuracy of reconstruction. If we are too generous with this parameter the resulting dimension might be correct but some points will be put into space with smaller then required dimension and then inaccuracy would occur.

Overall the correct results for all 61 problem were achieved with $\epsilon = 10^{-5}$.

After confirming the feasibility of the idea we had proceed to problems where dimension is not known. Only distances between points (towns) are known. The good example is a `bayg29` problem coming from distances between Bavarian towns. As this region is located in highlands and mountains we can expect at least 3 dimension. The actual distances differ from simple 3D euclidean distances due to requirements of a road construction in such difficult terrain.

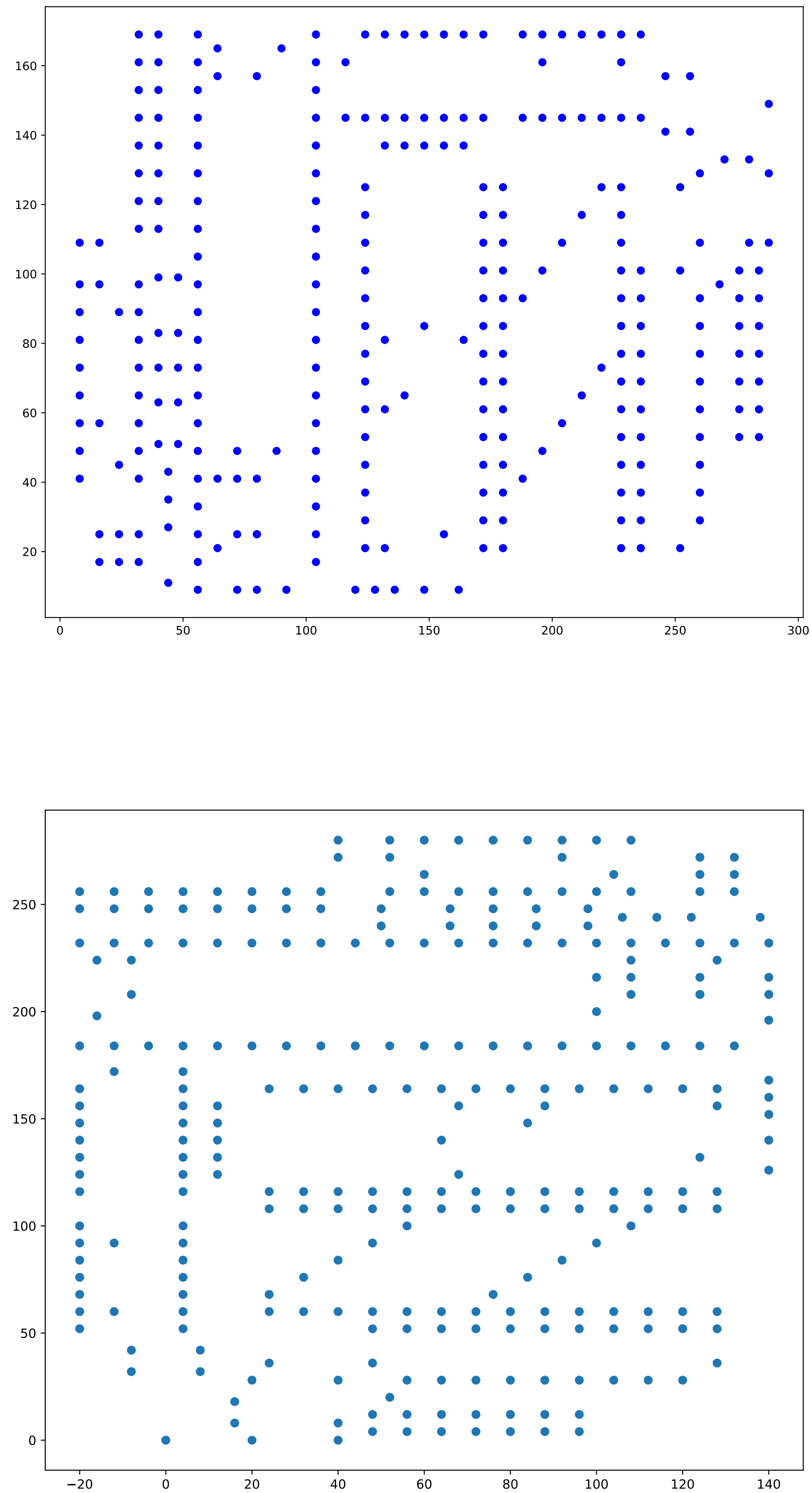

Fig. 1. Original (first) and reconstructed (second) layout of `a280` problem from tsplib

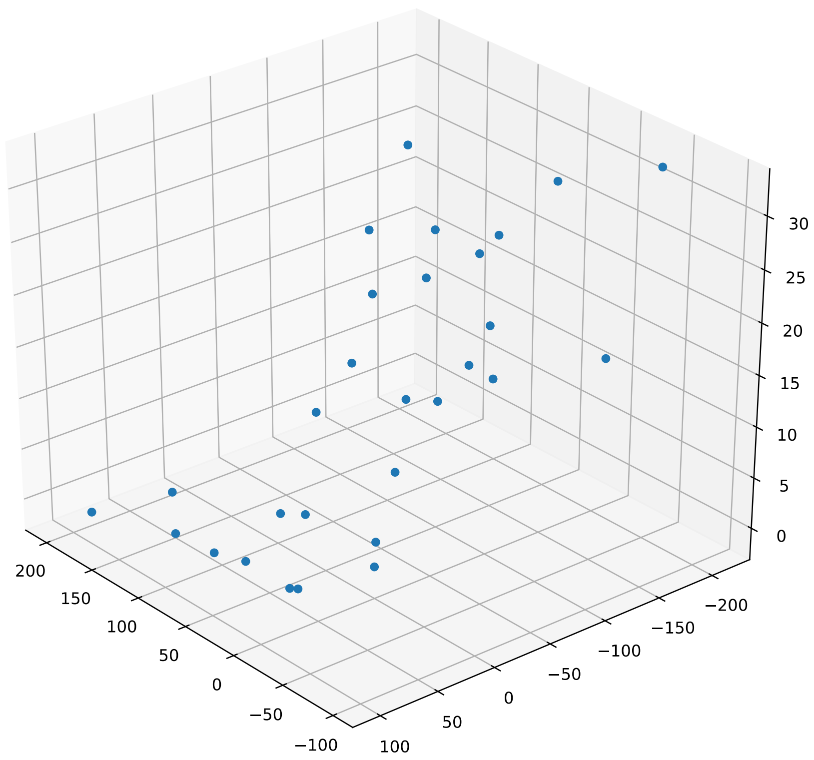

Fig. 2. A 3D reconstruction of `bayg29` distances

The results of reconstruction can be seen in fig. 2. The resulting dimensionality depends on required accuracy, that dependence can be seen in tab. 1.

Overall the numerical experiments had shown the proposed algorithm can provide good results with carefully chosen accuracy.

6. Comments and conclusions

The proposed method was extensively tested with problems from well-know TSPLIB. The results were promising. Careful selection of accuracy parameter can provide exact solution or approximation with reduced dimension.

The proposed algorithm provides the dimension number of the embedding and a partial embedding when the TSP is not Euclidean as a whole. In such a case the partial embedding depends on the vertex chosen as a starting point. The problem of partial embedding of the MTSP is combinatorial in nature and requires exponential-time complexity to be solved. In our opinion it is a new open problem which is worth further research. This direction of research is also important in the case of the ETSP problem. In $\mathcal{R}^m$, the PTAS has space and time complexities of $\mathcal{O}\left(n(\log n)^{(\mathcal{O}(\sqrt{m}\varepsilon))^{(m-1)}}\right)$, but the dimension should be relatively small in comparison to n (i.e., $\mathcal{O}(\log \log n)$) in order to get the polynomial running time [3].

REFERENCES

1. Asimow L., Roth B.: The rigidity of graphs. Trans. Amer. Math. Soc., 245, 1978, p. 279–289.
2. Arora S.: Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. Journal of the ACM , 45(5), 1998, p. 753–782.
3. Arora S.: Approximation schemes for NP-hard geometric optimization problems: a survey. Math. Program., Ser. B 97. 2003, p. 43–69.
4. Blumenthal L.: Theory and Applications of Distance Geometry, Oxford University Press, Oxford, 1953.
5. Gower J.: Euclidean distance geometry. Math. Sci., 7 , 1982, p. 1–14.
6. Lampis M.: Improved inapproximability for TSP, in Approximation, Randomization, and Combinatorial Optimization, vol. 7408 of LNCS, Springer, 2012, p. 243–253.
7. Lawler E., Lenstra J., Rinnooy Kan A.: The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization. Wiley, New York, 1985.
8. Liberti L., Lavor C., Nelson Maculan N., Mucherino A.: Euclidean Distance Geometry and Applications. SIAM Review, 56(1), 2014, p. 3–69.
9. Menger K.: New foundation of Euclidean geometry. Amer. J. Math., 53, 1931, p. 721–745.
10. Mitchell J.: Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric tsp, k-mst, and related problems. SIAM Journal on Computing, 28(4), 1999, p. 1298–1309.
11. Papadimitriou CH.:The Euclidean traveling salesman problem is NP-complete. Theoretical Computer Science 4, 1978, p. 237–244.
12. G. Reinelt.: TSPLIB – A Traveling Salesman Problem Library. ORSA Journal on Computing, 3(4), 1991, p. 376–384.
13. Saxe J.B.: Embeddability of weighted graphs in k-space is strongly NP-hard. Proc. of the 17th Allerton Conference in Communications, Control, and Computing, 1979, p. 480–489.
14. Schoenberg, I. J.: Metric spaces and positive definite functions. Transactions of the American Mathematical Society, 44, 1938, p. 522–536.
15. M. Sippl and H. Scheraga, Cayley-Menger coordinates. Proc. Natl. Acad. Sci. USA, 83, 1986, p. 2283–2287.
16. Stroeker, R. J.: On the sum of consecutive cubes being a perfect square. Compositio Mathematica. 97(1-2), 1995, p. 295–307