

A Note on Completing Quasi-distance and Distance Matrices

Yulin Zhang*, A. Veloso da Costa*, J. M. S. Simões-Pereira†

Dedicated to Charles R. Johnson

Abstract: *We give a necessary and sufficient condition for the existence of a quasi-distance matrix where some positive off-diagonal entries have been prescribed. Moreover, we give an algorithm for obtaining such a matrix. We analyze also the case where distances instead of quasi-distances are considered.*

Keywords: Quasi-distance matrix; Distance matrix; Completing a matrix.

MR Classification Subjects: 15A83, 15B99

1. Introduction

A *distance matrix* is a square matrix D whose entries are non-negative real numbers which satisfy the following three conditions, the third one being called the triangle inequality:

$$\forall i : d_{i,i} = 0; \quad \forall i, k : d_{i,k} = d_{k,i}; \quad \forall i, j, k : d_{i,k} \leq d_{i,j} + d_{j,k}.$$

We assume that the entries $d_{i,k}$ are positive for $i \neq k$, as is usually done; finally, we point out that, if we drop the symmetry in this definition, then D will be called a *quasi-distance matrix*.

Distance and quasi-distance matrices have been widely investigated and have applications in several areas, such as biology, psychology, archeology, engineering, computer science (see [1, 2, 3, 4, 5, 6] and respective lists of references). An interesting application is their use in a biologic model to reconstruct phylogenetic trees from matrices whose entries represent certain genetic distances among biological species (see [2]). Recently, in [7], we gave an alternative to deal with distances and quasi-distances using a so-called *tower matrix*; the tower matrix gives us more information than the classical distance and quasi-distance matrices.

In what follows, we prove a necessary and sufficient condition for the following problem to have a solution:

Consider an $n \times n$ matrix W of real numbers with a main diagonal of zeros and with a few more positive entries prescribed. Choose the remaining entries so that the matrix M we obtain is a quasi-distance matrix.

2. Main result

As usual, we may associate to the entries $m_{i,j}$ of a matrix M the arcs (i, j) of a digraph D so that these arcs have lengths, denoted $d_{i,j}$, equal to the entries of the matrix. And we define the length of a path as the sum of the lengths of its arcs. We prove:

Theorem 1: *The necessary and sufficient condition for the existence of a quasi-distance matrix M with a given set of positive entries and a main*

diagonal of zeros is that the digraph D associated to the given entries has no sub-digraph formed by a directed path P from i to j plus a single arc (i, j) whose length is greater than the length of the path.

Proof: The existence of a directed path from i to j together with a single arc (i, j) whose length is greater than the length of the path is therefore a *forbidden configuration*: such configuration may be seen as a cycle with all arcs oriented in the same sense except one of them which we call the *inverted arc*.

For the proof of the theorem, the necessity is an immediate consequence of the triangle inequality in the definitions of distance and quasi-distance matrix.

Regarding the sufficiency, the reasoning is longer. We'll show that, if there is no forbidden configuration in a digraph D where not all pairs of vertices are linked by an arc, then it is always possible to add an arc (i, j) to D with a conveniently chosen length in such a way that the digraph obtained, call it $D_{i,j}$, has no forbidden configuration either. Once we have proved the possibility of adding one arc, we'll argue that, when we have all pairs of vertices linked by arcs, without forbidden configurations, the complete graph trivially satisfies quasi-distance conditions. Notice that a failure of a triangular inequality would correspond to a forbidden configuration with three arcs.

Now, to choose the lengths to be given to the new arcs, there are two rules:

R1) If there is any path from i to j , then the length we give to the new arc (i, j) will be equal to the length of the shortest path from i to j ;

R2) If there is no path from i to j , then the length we give to the new arc (i, j) will be the length L of the longest arc already existing in the digraph.

Since, by hypothesis, D had no forbidden configuration, if $D_{i,j}$ has one, then the arc (i, j) belongs to it; in fact, the arc (i, j) may be the inverted arc of this forbidden configuration or not. We distinguish:

a) If (i, j) is the *inverted* arc of a forbidden configuration, then this means that there is a path in D whose length is shorter than the length of the arc (i, j) but this contradicts the choice we made for the length of the arc (i, j) .

b) If (i, j) is not the *inverted* arc of the forbidden configuration it belongs to, then we have two cases to consider: in the case where we gave to (i, j) the length of the shortest path from i to j , replace (i, j) by such path and we obtain a forbidden configuration which is entirely formed by arcs of D against the hypothesis (notice that the forbidden configuration may have repeated vertices but when this happens it contains a smaller elementary forbidden configuration); in the case where we gave to (i, j) the length L , this value L cannot be smaller than the length of the inverted arc of the cycle they belong to, which means that the path from i to j which includes the arc of length L cannot be shorter than the length of the arc (i, j) , which in turn means that this cycle is not a forbidden configuration.

This completes the *Proof*.

This result has an analogue for distances instead of quasi-distances. It is immediate to obtain its proof which is similar to the proof of Theorem 1. We state:

Theorem 2: *The necessary and sufficient condition for the existence of a distance matrix M with a given set of positive entries above the main diagonal is that the graph G associated to the given entries has no sub-graph formed by a cycle where there is an edge whose length is greater than the sum of the lengths of the remaining edges of the cycle.*

3. A remark

The general results we have just proved show us that, for instance, when the digraph associated to the given entries is a directed (or even a non-directed) path, completing the matrix is always possible. The same can be said when the digraph is a star; or when it becomes a tree after we disregard the orientation of the arcs.

4. A numerical example

Start with the following matrix

$$W = \begin{bmatrix} 0 & + & 3 & 5 \\ + & 0 & + & 2 \\ + & 1 & 0 & + \\ + & + & + & 0 \end{bmatrix}$$

whose associated digraph is in Figure 1:

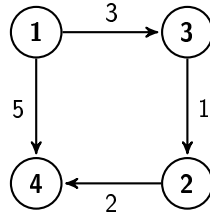


Figure 1: The digraph associated to the matrix W .

To complete the matrix, our rules yield the remaining entries:

$$d_{1,2} = 4, d_{2,1} = 5, d_{2,3} = 8, d_{3,1} = 6, d_{3,4} = 3, d_{4,1} = 8, d_{4,2} = 12, d_{4,3} = 11.$$

With these values, we get the following quasi-distance matrix M :

$$M = \begin{bmatrix} 0 & 4 & 3 & 5 \\ 5 & 0 & 8 & 2 \\ 6 & 1 & 0 & 3 \\ 8 & 12 & 11 & 0 \end{bmatrix}$$

This solution is not unique. Keep following our rules but assign the values in a different sequence, for instance,

$$d_{4,1} = 5, d_{4,2} = 9, d_{4,3} = 8, d_{3,1} = 8, d_{3,4} = 3, d_{2,1} = 7, d_{2,3} = 10, d_{1,2} = 4$$

We get a different quasi-distance matrix, namely

$$M' = \begin{bmatrix} 0 & 4 & 3 & 5 \\ 7 & 0 & 10 & 2 \\ 8 & 1 & 0 & 3 \\ 5 & 9 & 8 & 0 \end{bmatrix}$$

References

- [1] Agnese Baldisserrri and Elena Rubei: *Distance matrices of some positive weighted graphs*, Australasian J. of Combinatorics 70-2 (2018), 185-201.
- [2] A. Dress, K. T. Huber, J. Koolen, V. Moulton and A. Spillner: *Basic Phylogenetic Combinatorics*, Cambridge University Press, Cambridge, 2012.
- [3] Jeremy P. Spinrad: *Efficient Graph Representations*, Fields Institute Monographs, American Mathematical Society, 2003.
- [4] Leslie Hogben: *Handbook of Linear Algebra*, Chapman and Hall/CRC Press, 2007.
- [5] Salton N. and M. Nei: *The neighbor-joining method: a new method for reconstructing phylogenetic trees*, Molecular Biology and Evolution 4-4, (July 1987), 406-425.
- [6] J. M. S. Simões-Pereira: *Matemática Discreta: Grafos, Redes, Aplicações*(in Portuguese), Editora Luz da Vida, Coimbra, Portugal, 2009.
- [7] Yulin Zhang, A. Veloso da Costa and J. M. S. Simões-Pereira: *The tower matrix, an alternative to deal with distances and quasi-distances*, Journal of Multiple-Valued Logic and Soft Computing 32 (2019), 449-454.

*Dept. of Math., University of Minho, Braga, Portugal.

E-mails: zhang@math.uminho.pt, aveloso@math.uminho.pt. Research by these authors was supported by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia, within Project UID/MAT/00013/2013.

†Dept. of Math., University of Coimbra, Coimbra, Portugal. E-mail: siper@mat.uc.pt. Research supported by the Divine Cities Foundation.