Submatrices of Non-Tree-Realizable Distance Matrices

J. M. S. Simões-Pereira*
Department of Computer Science
Hunter College
and
Graduate Center
of the City University of New York
New York, New York 10021

and

Christina M. Zamfirescu†
Department of Computer Science
Hunter College
of the City University of New York
New York, New York 10021

Submitted by Richard A. Brualdi

ABSTRACT

We deal with distance matrices of real (this means, not necessarily integer) numbers. It is known that a distance matrix D of order n is tree-realizable if and only if all its principal submatrices of order 4 are tree-realizable. We discuss bounds for the number, denoted $Q_j(D)$, of non-tree-realizable principal submatrices of order $j \geqslant 4$ of a non-tree-realizable distance matrix D of order $n \geqslant j$, and we construct some distance matrices which meet extremal conditions on $Q_j(D)$. Our starting point is a proof that a non-tree-realizable distance matrix of order 5 has at least two non-tree-realizable principal submatrices of order 4. Optimal realizations (by graphs with circuits) of distance matrices which are not tree-realizable are not yet as well known as optimal realizations which are trees. Using as starting point the optimal realization of the (arbitrary) distance matrix of order 4, we investigate optimal realizations of non-tree-realizable distance matrices with the minimum number of non-tree-realizable principal submatrices of order 4.

---

*The first-named author discussed this material in seminars given in the departments of Mathematical Sciences of the Universities of Coimbra and Aveiro (Portugal) under the NATO Senior Scientists Programme, Fellowship number 5.2.03B.172.

†The second-named author was partly supported by a grant from the City University of New York PSC/CUNY Research Award Program.

 0024-3795/82/030001 + 17$02.75

1. INTRODUCTION

As in [5], we define a distance matrix of order n as a symmetric, nonnegative, square matrix D with entries d_{ij} such that, for $i, j, k = 1,\ldots,n$, $d_{ii} = 0$ and $d_{ij} \leqslant d_{ik} + d_{kj}$.

Distance matrices are the most natural tool to describe finite metric spaces. They have been studied in several papers [1–11]. (Generalized distance matrices with negative entries have been studied in [8] but will not be considered here.) Moreover, realizations of distance matrices have a wide range of applications: perhaps the most recently found application is in a biological model for the analysis and synthesis of evolutionary trees [4].

Let $G = (W, E)$ be a graph, W and E its vertex and edge sets, respectively; let $V \subseteq W$ and $|V| = n$. Consider a function $f\colon E \to R^+$, where R^+ is the set of the nonnegative real numbers. We say that f assigns a length or weight to each edge of G. For any $i, j \in W$, we define the distance $d(i, j)$ between i and j as the minimum value among sums $\Sigma f(e)$ taken over any path between i and j. We say that G realizes D if and only if, for some V, $d(i, j) = d_{ij}$ for $i, j = 1,\ldots,n$. If G is a realization of D, we call the vertices in V external and those in $W - V$ internal. Trivially, G can be required to have no internal vertex of degree less than three. A realization is called optimal when the sum $\Sigma f(e)$ taken over E is minimal among all realizations of D.

For later reference, we recall the following known results:

THEOREM A [5, 11]. *If D has a tree realization, then this realization is optimal.*

THEOREM B [10]. *The matrix D has a tree realization if and only if its principal submatrices of order 4 have tree realizations.*

THEOREM C [5]. *If G is an optimal realization of D, then to any vertex $p \in W$ of degree higher than one we can associate two vertices $i, j \neq p$ such that $d_{ij} = d(i, p) + d(p, j)$.*

THEOREM D [5, 9]. *No optimal realization contains a triangle.*

THEOREM E [2, 6]. *A matrix of order 4 is tree-realizable if and only if, among the three sums $d_{12} + d_{34}$, $d_{13} + d_{24}$, $d_{14} + d_{23}$, two are equal and not smaller than the third one.*

In the next two statements, $D_i(a)$ is a matrix obtained from D by subtracting the nonnegative number a from all entries in the ith row or the

ith column of D except for d_{ii}. For $p, r = 1, \ldots, n$ and $p, r \neq i$, we set $a_0^i = \min\{(d_{pi} + d_{ir} - d_{pr})/2\}$.

THEOREM F [5]. *$D_i(a)$ is a distance matrix if and only if $a \leqslant a_0^i$.*

THEOREM G [5]. *An optimal realization G of D can be obtained from an optimal realization $G_i(a)$ of $D_i(a)$, where $a \leqslant a_0^i$, by the following construction: add a vertex i' to $G_i(a)$, and add an edge of length a linking i' to the vertex i of $G_i(a)$. Moreover, if $G_i(a)$ is the unique optimal realization of $D_i(a)$, then G is the unique optimal realization of D.*

The operation which leads from G to $G_i(a)$ or from D to $D_i(a)$ is called a compactification by (the amount) a. A remark is that we may have $a_0^i = d_{ij}$ for some $j \neq i$. In such a case, by the minimality of a_0^i, we obtain, for every $r \neq i$, $d_{ij} + d_{ir} - d_{jr} \geqslant d_{ij} + d_{ij} - d_{jj} = 2d_{ij}$, hence $d_{ir} \geqslant d_{ij} + d_{jr}$; by the definition of a distance matrix, this means that, for every $r \neq i$, $d_{ir} = d_{ij} + d_{jr}$. We then say that i is a pendant index. Compactification by a_0^i yields a matrix with $d_{ij} = 0$ and $d_{ip} = d_{jp}$ for all p.

Given a distance matrix, for each i which is not a pendant index compactify the matrix by a_0^i. The matrix we obtain will be called fully compactified.

An entry d_{ij} of D is called basic if no $k \neq i, j$ exists such that $d_{ij} = d_{ik} + d_{kj}$; otherwise d_{ij} is nonbasic.

If D is fully compactified of order n with no pendant i, then, for each k, there is a pair p, r such that $d_{kp} + d_{kr} = d_{pr}$. As immediate consequences, D has at least n basic entries and at least 2 nonbasic entries. The following matrix, for example, has 2 nonbasic entries:

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

2. PRELIMINARY REMARKS

Our first result follows immediately from Theorem E and the definitions.

THEOREM 1. *Let D and $D_i(a)$ be distance matrices. Then either they are both tree-realizable or they are both non-tree-realizable.*

Our next result is the following theorem.

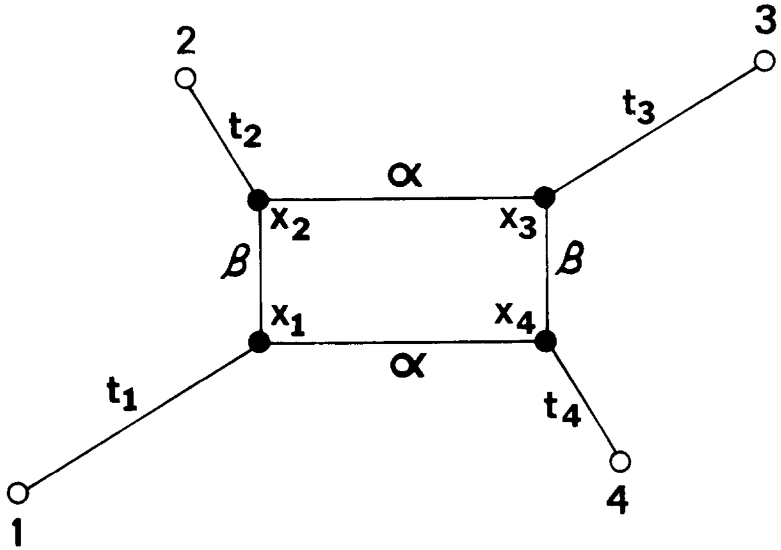

FIG. 1

THEOREM 2. *The optimal realization (Figure 1) of a non-tree-realizable distance matrix of order 4 is a rectangle of vertices x_1, x_2, x_3, x_4 plus four pendant edges at its vertices with lengths $t_1, t_2, t_3, t_4 \geqslant 0$.*

Proof. Let D^* be non-tree-realizable. For $i = 1,2,3,4$, let $t_i = \min\{(d^*_{ia} + d^*_{ib} - d^*_{ab})/2\}$ over all pairs a, b. [By the remark in the Introduction, we may consider $a \neq b$; if $t_i = (d^*_{ia} + d^*_{ia} - d^*_{aa})/2 = d^*_{ia}$, then, by that remark and Theorem E, D^* is tree-realizable.] Let D be obtained from D^* by successive compactifications by the amounts t_1, t_2, t_3, t_4. By Theorem 1, D is non-tree-realizable. Moreover, for each i, there exist a, b such that

$$d_{ai} + d_{ib} = d_{ab}. \tag{0}$$

For $i = 1$, suppose, without loss of generality, that

$$d_{21} + d_{14} = d_{24}. \tag{1}$$

It follows from (1) that the pair a, b which satisfies (0) for $i = 2$ can not be 1,4. It can not be 3,4 either. In fact, if $d_{32} + d_{24} = d_{34}$, then, by (1), we obtain

$$d_{32}+d_{21}+d_{14}=d_{34}, \tag{2}$$

which yields $d_{43}>d_{23}$ and $d_{43}>d_{32}+d_{21}\geqslant d_{13}$; hence

$$d_{42}+d_{43}-d_{23}>0 \tag{3}$$

and

$$d_{41}+d_{43}-d_{13}>0. \tag{4}$$

From (1) it follows that $d_{42}>d_{21}$; hence

$$d_{42}+d_{41}-d_{21}>0. \tag{5}$$

The inequalities (3), (4), and (5) contradict the hypothesis that D is compactified. The pair a,b which satisfies (0) for $i=2$ is therefore 1,3. We have thus

$$d_{32}+d_{21}=d_{31}. \tag{6}$$

Note that the equality (6) is like (1) with its subscripts or indices increased by 1 mod 4. A similar argument with all subscripts increased by 1 mod 4 proves that (6) implies

$$d_{43}+d_{32}=d_{42} \tag{7}$$

and (7) in turn implies

$$d_{14}+d_{43}=d_{13}. \tag{8}$$

Moreover, it follows from (1), (6), (7), and (8) that $d_{14}=d_{23}$ and $d_{12}=d_{34}$.

Any optimal realization of D must contain the paths represented as sides of the rectangle in Figure 2. Among such paths, no two adjacent ones can have an internal vertex in common: this would contradict one of the equalities (1), (6), (7), or (8). Any two nonadjacent paths must also be disjoint. In fact, suppose the paths joining the pairs 1,2 and 4,3 have common internal vertices, denoted by x_i. Since the paths joining 2,3 and 4,1 are pairwise disjoint from them, we may choose an internal vertex x_0 joined by single

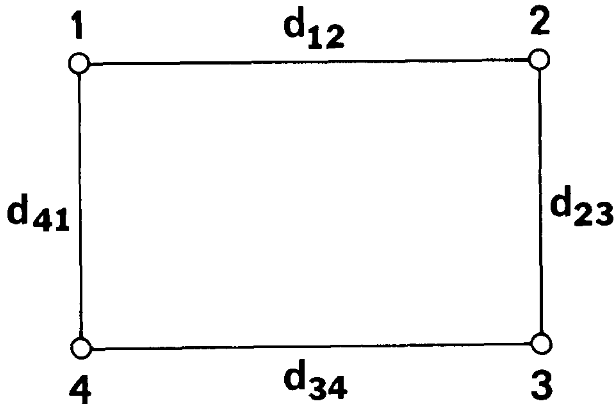

FIG. 2

edges to 2 and 3 (see Figure 3). By (6) we have

$$d(1, x_0) + d(x_0, 3) \geqslant d(1, x_0) + d(x_0, 2) + d_{23};$$

hence

$$d(x_0, 3) \geqslant d(x_0, 2) + d_{23}.$$

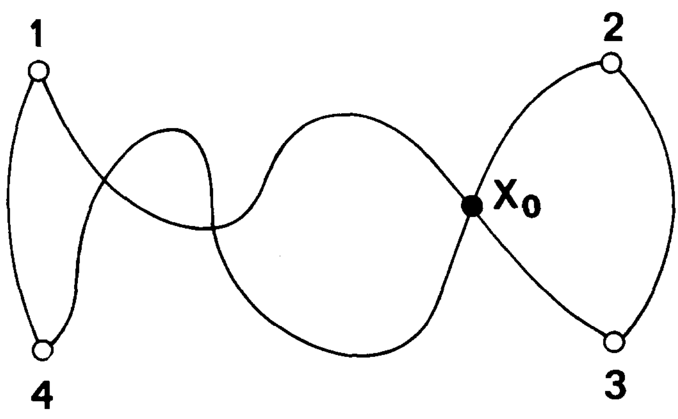

FIG. 3

This would imply

$$d_{43} = d(4, x_0) + d(x_0, 3) \geqslant d(4, x_0) + d(x_0, 2) + d_{23} \geqslant d_{42} + d_{23}$$

and hence $d_{43} = d_{42} + d_{23}$, which contradicts (7). Due to Theorem D, the graph of Figure 2 is therefore the optimal realization of D. By Theorem G, the proof of Theorem 2 is completed. ■

3. NON-TREE-REALIZABLE DISTANCE MATRICES OF ORDER 5

From now on, $\langle\{i_1,\dots,i_m\}\rangle$ denotes a matrix D of order m; if we consider D as a principal submatrix of another matrix, say $\langle\{1,\dots,n\}\rangle$, then the entries of D are those in the rows and columns $i_1,\dots,i_m$ of $\langle\{1,\dots,n\}\rangle$. It is natural to denote the external vertices of any realization of D by $i_1,\dots,i_m$. We also say that $i_1,\dots,i_m$ are the indices of D.

It follows from Theorem E that, if $\langle\{i, j, k, h\}\rangle$ is non-tree-realizable, then one of the sums $d_{ij} + d_{kh}$, $d_{ik} + d_{jh}$, $d_{ih} + d_{jk}$ is strictly greater than the other two. If, say, $d_{ij} + d_{kh}$ is the largest sum, then we call indices i, j (and indices k, h too) opposite.

THEOREM 3. *A non-tree-realizable distance matrix D of order 5 has at least two non-tree-realizable principal submatrices of order 4.*

Proof. By Theorem 1, we may consider only compactified matrices.

First suppose that no pendant index exists.

With no loss of generality, let $\langle\{1,2,3,4\}\rangle$ be non-tree-realizable, and d_{51} and d_{52} basic. These basic entries are distances between 5 and two indices of $\langle\{1,2,3,4\}\rangle$ which may be opposite (case 1) or nonopposite (case 2) in $\langle\{1,2,3,4\}\rangle$.

Case 1. The assumption means that

$$d_{12} + d_{34} > d_{13} + d_{24}, \tag{9}$$

$$d_{12} + d_{34} > d_{14} + d_{23}. \tag{10}$$

If the equality $d_{54} + d_{43} = d_{53}$ holds, then add d_{54} to both sides of (9) and (10) and use this equality; by Theorem E, $\langle\{1,2,3,5\}\rangle$ is non-tree-realizable.

Suppose now that

$$d_{54}+d_{43}>d_{53}. \tag{11}$$

Since d_{52} and d_{51} are basic, we have also

$$d_{54}+d_{42}>d_{52}, \tag{12}$$

$$d_{54}+d_{41}>d_{51}. \tag{13}$$

Moreover,

$$d_{34}+d_{42}>d_{32}. \tag{14}$$

To prove (14) note that adding d_{14} to both sides of $d_{32}=d_{34}+d_{42}$ yields $d_{32}+d_{14}=d_{34}+d_{42}+d_{14}\geqslant d_{34}+d_{12}$, which contradicts (10). Similarly,

$$d_{34}+d_{41}>d_{31}. \tag{15}$$

Since D is compactified, the inequalities (11), (12), (13), (14), (15) imply

$$d_{24}+d_{41}=d_{21}. \tag{16}$$

Add now d_{41} and d_{42} to both sides of (12) and (13), respectively, and use (16); by Theorem E, $\langle\{1,2,4,5\}\rangle$ is non-tree-realizable.

Case 2. The assumption means that

$$d_{13}+d_{24}>d_{12}+d_{34}, \tag{17}$$

$$d_{13}+d_{24}>d_{14}+d_{32}. \tag{18}$$

We assume also that d_{53} and d_{54} are nonbasic; otherwise the situation would be equivalent to case 1.

If the equality $d_{53}=d_{51}+d_{13}$ holds, then add d_{51} to both sides of (17) and (18) and use this equality; by Theorem E, $\langle\{2,3,4,5\}\rangle$ is non-tree-realizable. Similarly, if $d_{54}=d_{52}+d_{24}$, then $\langle\{1,3,4,5\}\rangle$ is non-tree-realizable.

Suppose now that

$$d_{51}+d_{13}>d_{53} \tag{19}$$

and

$$d_{52}+d_{24}>d_{54}. \tag{20}$$

Since d_{53} and d_{54} are nonbasic, among the four equalities

$$d_{53}=d_{52}+d_{23}, \tag{21}$$

$$d_{53}=d_{54}+d_{43}, \tag{21'}$$

$$d_{54}=d_{51}+d_{14}, \tag{22}$$

$$d_{54}=d_{53}+d_{34} \tag{22'}$$

we must have (21) or (21′) and (22) or (22′). Obviously, (21′) and (22′) can not occur simultaneously. Further, (21′) and (22) yield $d_{53}=d_{51}+d_{14}+d_{43} \geqslant d_{51}+d_{13}$, which contradicts (19). Similarly, (21) and (22′) yield $d_{54}=d_{52}+d_{23}+d_{34} \geqslant d_{52}+d_{24}$, which contradicts (20). Thus, (21) and (22) hold. We claim now that

$$d_{14}+d_{43}=d_{13}. \tag{23}$$

In fact, since d_{51} and d_{52} are basic, $d_{14}+d_{45}>d_{15}$ and $d_{24}+d_{45}>d_{25}$. Moreover, by (22) and (19), $d_{34}+d_{45}=d_{34}+d_{41}+d_{15} \geqslant d_{31}+d_{15}>d_{35}$. Finally, since $d_{14}+d_{42}=d_{12}$ would imply $d_{14}+d_{42}+d_{34}=d_{12}+d_{34}$ and therefore $d_{13}+d_{42} \leqslant d_{12}+d_{34}$, which contradicts (17), we have also $d_{14}+d_{42}>d_{12}$; and a similar argument shows that $d_{34}+d_{42}>d_{32}$. This proves our claim: otherwise D would not be compactified.

Add now d_{13} to both sides of (22). Using (19) and (23), we obtain respectively

$$d_{13}+d_{54}=d_{13}+d_{51}+d_{14}>d_{53}+d_{14},$$

$$d_{13}+d_{54}=d_{13}+d_{51}+d_{14}=d_{14}+d_{43}+d_{51}+d_{14}>d_{51}+d_{43}.$$

By Theorem E, $\langle\{1,3,4,5\}\rangle$ is non-tree-realizable.

To complete the proof of the theorem, suppose that the compactified matrix D has a pendant index, say 5. Let j be such that $d_{5r}=d_{5j}+d_{jr}$ for all $r \neq 5$. Without loss of generality, let $j=4$. By Theorem E, any $\langle\{i,k,4,5\}\rangle$ is tree-realizable: among the sums $d_{54}+d_{ik}$, $d_{5i}+d_{k4}=d_{54}+d_{4i}+d_{k4}$, d_{5k}

$+d_{4i}=d_{54}+d_{4k}+d_{4i}$ two are equal and not smaller than the third one. Hence, by Theorem B, $\langle\{1,2,3,4\}\rangle$ or $\langle\{1,2,3,5\}\rangle$ is non-tree-realizable. By Theorem E, both must be.

This completes the proof of Theorem 3. ■

4. NON-TREE-REALIZABLE DISTANCE MATRICES OF ORDER n

In what follows for $i=4,\ldots,n$, $Q_i(D)$ denotes the number of principal, non-tree-realizable $i\times i$ submatrices of D. Bounds for $Q_i(D)$ will be derived and discussed.

THEOREM 4. *Let D be non-tree-realizable of order n. Then*

$$Q_i(D)\geqslant\binom{n-3}{i-3}\qquad \textit{for}\quad i=4,\ldots,n.$$

Proof. By Theorem B, the matrix D has a 4×4 non-tree-realizable submatrix, say $\langle\{1,2,3,4\}\rangle$.

Denote by $a_1,\ldots,a_{n-4}$ the indices of D other than $1,2,3,4$. Obviously, $\langle\{1,2,3,4,a_1\}\rangle,\ldots,\langle\{1,2,3,4,a_{n-4}\}\rangle$ are non-tree-realizable. By Theorem 3, each one contains at least two non-tree-realizable submatrices of order 4. Therefore, besides $\langle\{1,2,3,4\}\rangle$, there is, for each a_j, a non-tree-realizable matrix $\langle\{b_1,b_2,b_3,a_j\}\rangle$ with $\{b_1,b_2,b_3\}\subset\{1,2,3,4\}$. This shows that $Q_4(D)\geqslant n-3$, which is our statement for $i=4$.

For any i, $n\geqslant i>4$, we may obtain an $i\times i$ non-tree-realizable submatrix of D either by joining an $(i-4)$-subset of $\{a_1,\ldots,a_{n-4}\}$ to the 4-set $\{1,2,3,4\}$ or by joining an $(i-4)$-subset of $\{a_{j+1},\ldots,a_{n-4}\}$ to $\{b_1,b_2,b_3,a_j\}$. The matrices obtained are all distinct, and there are at least

$$\binom{n-4}{i-4}+\binom{n-4}{i-3}=\binom{n-3}{i-3},$$

which completes the proof of Theorem 4. ■

These lower bounds can be attained: consider the graph pictured in Figure 1, and replace one of the pendant edges, say $(x_4,4)$, by a tree with $n-3$ external vertices $4,a_1,\ldots,a_{n-4}$. For further reference we call such a graph a P-graph. Assign lengths to the new edges. For the distance matrix D

arising from a P-graph,

$$Q_i(D)=\binom{n-3}{i-3} \qquad \text{for} \quad i=4,\ldots,n.$$

Moreover, by Theorems G and 2, a P-graph is an optimal realization of its distance matrix.

Our next statement generalizes Theorem 3.

THEOREM 5. *A non-tree-realizable distance matrix of order j contains at least $j-3$ non-tree-realizable submatrices of order $j-1$.*

Proof. The statement is equivalent to saying that, if a matrix D of order j has 4 tree-realizable submatrices of order $j-1$, then D is tree-realizable. Now let $S=\{1,\ldots,j\}$ $D=\langle S\rangle$, and $\langle S-\{a\}\rangle, \langle S-\{b\}\rangle, \langle S-\{c\}\rangle, \langle S-\{d\}\rangle$ be tree-realizable. It follows that $\langle\{a,b,c,d\}\rangle$ is the only 4×4 principal submatrix of D which might be non-tree-realizable. This contradicts Theorem 4 and thus proves Theorem 5. ■

THEOREM 6. *For $j\geqslant 5$, $(n-j+1)Q_{j-1}\geqslant (j-3)Q_j$.*

Proof. Besides Theorem 5, use the fact that each non-tree-realizable submatrix of order $j-1$ is contained in $n-j+1$ submatrices of order j. ■

The following describes the optimal realizations of a class of matrices.

THEOREM 7. *Optimal realizations of distance matrices D such that*

$$Q_i(D)=\binom{n-3}{i-3} \qquad \textit{for} \quad i=4,\ldots,n$$

are P-graphs.

Proof. First we claim that, when $Q_4(D)=n-3$, all non-tree-realizable submatrices of order 4 share a set of 3 indices. The claim is obvious for $n=5$. Suppose $n\geqslant 6$. As in the proof of Theorem 4, let $\langle\{1,2,3,4\}\rangle$ be non-tree-realizable, and for each $a_j\in\{5,\ldots,n\}$ let $\langle\{b_1,b_2,b_3,a_j\}\rangle$ be non-tree-realizable for some set $\{b_1,b_2,b_3\}\subset\{1,2,3,4\}$. These are $n-4+1$ non-tree-realizable submatrices. We show that they have the same set $\{b_1,b_2,b_3\}$. In fact, if say $\langle\{b_1,b_2,b_3,a_1\}\rangle$ and $\langle\{b_1',b_2',b_3',a_2\}\rangle$ are non-tree-realizable, then so are $\langle\{b_1, b_2, b_3, a_1, a_2\}\rangle$ and $\langle\{b_1', b_2', b_3', a_1, a_2\}\rangle$, and each one contains two 4×4 non-tree-realizable submatrices. Among these, either one contains a_1 and a_2 in its index set (and it is distinct from the remaining $n-3$), or $\langle\{b_1, b_2, b_3, a_2\}\rangle$ and $\langle\{b_1', b_2', b_3', a_1\}\rangle$ are also non-tree-realizable (and distinct from the remaining $n-3$). In both cases we would have $Q_4(D)>n-3$, which proves our claim.

Without loss of generality, suppose that $\langle\{1,2,3,j\}\rangle$ for $j=4,\dots,n$ are the non-tree-realizable submatrices of D and that D is compactified for indices $1,2,3$. Let $1,3$ (and $2,4$) be the opposite indices in $\langle\{1,2,3,4\}\rangle$. We claim that

$$d_{13}=d_{12}+d_{23}. \tag{24}$$

To prove the claim, suppose $d_{13}<d_{12}+d_{23}$. For some pair a,b, we must have $d_{ab}=d_{a2}+d_{2b}$. Since $1,3$ are opposite in $\langle\{1,2,3,4\}\rangle$, the pair a,b cannot be $4,3$ or $4,1$. We may therefore choose $j\in\{5,\dots,n\}\cap\{a,b\}$ and distinguish two cases:

Case 1a. $\{a,b\}=\{j,1\}$ *or* $\{a,b\}=\{j,3\}$. The arguments being similar for $\{j,1\}$ and $\{j,3\}$, we give the details for the case when we have

$$d_{j1}=d_{j2}+d_{21}. \tag{25}$$

By hypothesis, $\langle\{1,2,4,j\}\rangle$ and $\langle\{2,3,4,j\}\rangle$ are tree-realizable. Set $A_1=d_{12}+d_{4j}$, $A_2=d_{14}+d_{2j}$, $A_3=d_{1j}+d_{24}$, $C_1=d_{23}+d_{4j}$, $C_2=d_{24}+d_{3j}$, $C_3=d_{2j}+d_{34}$. By (25), $A_3\geqslant A_1$ and $A_3\geqslant A_2$. Since $2,4$ are opposite in $\langle\{1,2,3,4\}\rangle$, we cannot have $d_{14}=d_{21}+d_{24}$; hence $A_3\neq A_2$. Thus we have $A_3=A_1$, which means $d_{4j}=d_{j2}+d_{24}$. It follows that $C_1=d_{23}+d_{j2}+d_{24}$ and, consequently, $C_1\geqslant C_2$ and $C_1\geqslant C_3$. Now $C_1=C_2$ if and only if $d_{3j}=d_{23}+d_{j2}$; we claim that this equality does not hold. In fact, by (25), $d_{j1}+d_{23}=d_{j2}+d_{21}+d_{23}>d_{j2}+d_{13}$. By (25) and the fact that $\langle\{1,2,3,j\}\rangle$ is non-tree-realizable, we see that $j,1$ and $2,3$ are opposite, and this proves our claim. Therefore $C_1=C_3$, that is, $d_{34}=d_{23}+d_{24}$. Since $2,4$ are opposite in $\langle\{1,2,3,4\}\rangle$, this is a contradiction.

Case 1b. $\{a,b\}=\{j,k\}$ *with* $k\in\{4,\dots,n\}-\{j\}$. In this case we have

$$d_{jk}=d_{j2}+d_{2k}. \tag{26}$$

We have also

$$d_{j1}<d_{j2}+d_{21}, \tag{27}$$

$$d_{j3}<d_{j2}+d_{23}; \tag{28}$$

otherwise we would have case 1a. Since $\langle\{1,2,j,k\}\rangle$ is tree-realizable, setting $G_1 = d_{12} + d_{jk} = d_{12} + d_{j2} + d_{2k}$, $G_2 = d_{1j} + d_{2k}$, $G_3 = d_{1k} + d_{2j}$, we have, by (27), $G_2 < G_1$. It follows that $G_1 = G_3$ and therefore $d_{1k} = d_{12} + d_{2k}$. Since $\langle\{2,3,j,k\}\rangle$ is tree-realizable, setting $H_1 = d_{23} + d_{jk} \equiv d_{23} + d_{j2} + d_{2k}$, $H_2 = d_{2j} + d_{3k}$, $H_3 = d_{2k} + d_{3j}$ we have, by (28), $H_3 < H_1$. It follows that $H_1 = H_2$ and thus $d_{3k} = d_{23} + d_{2k}$. By hypothesis, $\langle\{1,2,3,k\}\rangle$ is non-tree-realizable. Setting $F_1 = d_{12} + d_{3k} \equiv d_{12} + d_{23} + d_{2k}$, $F_2 = d_{13} + d_{2k}$, $F_3 = d_{1k} + d_{23} \equiv d_{12} + d_{2k} + d_{23}$, we have $F_1 = F_3 > F_2$, a contradiction.

Our claim that (24) holds is thus proved.

Now we claim that, for any j, $4 \leqslant j \leqslant n$,

$$d_{21} + d_{1j} = d_{2j} = d_{23} + d_{3j}. \tag{29}$$

We give a detailed proof for the first equality (the second is similar). Choose a fixed j and suppose that

$$d_{2j} < d_{21} + d_{1j}. \tag{30}$$

By the hypothesis of compactification, for some pair a, b, $d_{ab} = d_{a1} + d_{1b}$. By (24), (30), and the fact that 1,3 are opposite in $\langle\{1,2,3,j\}\rangle$, the pair a, b is neither 3,2 nor j,2 nor j,3. It is therefore k,3 or k,2 or k, j with $k \in \{4,\dots,n\} - \{j\}$.

Case 2a. Suppose $d_{k3} = d_{k1} + d_{13}$. The matrix $\langle\{1,2,3,k\}\rangle$ is non-tree-realizable. Setting $A_1 = d_{12} + d_{3k} \equiv d_{12} + d_{k1} + d_{12} + d_{23}$, $A_2 = d_{13} + d_{2k} \equiv d_{12} + d_{23} + d_{2k}$, $A_3 = d_{1k} + d_{23}$, we get $A_1 > A_3$, $A_2 \geqslant A_3$; hence $A_1 \neq A_2$, and thus we have

$$d_{2k} < d_{k1} + d_{12}. \tag{31}$$

Now $\langle\{1,3,j,k\}\rangle$ is tree-realizable. Setting $Z_1 = d_{13} + d_{jk} \equiv d_{12} + d_{23} + d_{jk}$, $Z_2 = d_{1j} + d_{3k} \equiv d_{1j} + d_{k1} + d_{12} + d_{23}$, $Z_3 = d_{1k} + d_{3j}$, we get $Z_3 < Z_2$ because $d_{3j} < d_{31} + d_{1j}$. Hence $Z_1 = Z_2$ and therefore $d_{jk} = d_{k1} + d_{1j}$. Further, $\langle\{1,2,j,k\}\rangle$ is tree-realizable. Setting $W_1 = d_{12} + d_{jk} \equiv d_{12} + d_{1k} + d_{1j}$, $W_2 = d_{1j} + d_{2k}$, $W_3 = d_{1k} + d_{2j}$, we get $W_3 < W_1$; hence $W_2 = W_1$; this means $d_{2k} = d_{21} + d_{1k}$, which contradicts (31).

Case 2b. Suppose $d_{k2} = d_{k1} + d_{12}$. The matrix $\langle\{1,2,3,k\}\rangle$ is non-tree-realizable. Setting $A_1 = d_{12} + d_{3k}$, $A_2 = d_{13} + d_{2k} \equiv d_{13} + d_{12} + d_{1k}$, $A_3 = d_{1k} + d_{23}$, we get $A_1 \leqslant A_2$ and, by (24), $A_3 < A_2$. We thus have $A_1 < A_2$, which means

$$d_{3k} < d_{13} + d_{1k}. \tag{32}$$

The matrix $\langle\{1,2,j,k\}\rangle$ is tree-realizable. Setting $Z_1=d_{12}+d_{jk}$, $Z_2=d_{1j}+d_{2k}\equiv d_{1j}+d_{12}+d_{1k}$, $Z_3=d_{1k}+d_{2j}$, we get, by (30), $Z_3<Z_2$; hence $Z_1=Z_2$, which means $d_{jk}=d_{1j}+d_{1k}$. Finally, $\langle\{1,3,j,k\}\rangle$ is tree-realizable. Setting $W_1=d_{13}+d_{jk}\equiv d_{13}+d_{1j}+d_{1k}$, $W_2=d_{1j}+d_{3k}$, $W_3=d_{1k}+d_{3j}$, we get, by (32), $W_2<W_1$; hence $W_3=W_1$, which means that $d_{3j}=d_{31}+d_{1j}$. This is impossible because 1,3 are opposite in $\langle\{1,2,3,j\}\rangle$.

Case 2c. Suppose $d_{kj}=d_{k1}+d_{1j}$, and j is neither 3 (case 2a) nor 2 (case 2b). The matrix $\langle\{1,2,j,k\}\rangle$ is tree-realizable. Setting $Z_1=d_{12}+d_{jk}\equiv d_{12}+d_{j1}+d_{1k}$, $Z_2=d_{1j}+d_{2k}$, $Z_3=d_{1k}+d_{2j}$, we get $Z_2\leqslant Z_1$ and, by (30), $Z_3<Z_1$. Hence $Z_2=Z_1$, which means $d_{2k}=d_{21}+d_{1k}$, which is case 2b.

Our claim that (29) holds is thus proved.

Now let G' be a tree realization of $\langle\{1,3,4,\ldots,n\}\rangle$. Let G be obtained from G' by adding vertex 2 and edges (2,1) and (2,3) with weights d_{21} and d_{23}, respectively. The graph G is a P-graph and realizes $\langle\{1,2,3,\ldots,n\}\rangle$. As we have seen, P-graphs are optimal realizations. By Theorems G and 2, they are unique, which completes the proof of Theorem 7. ∎

5. A FASTER ALGORITHM FOR TESTING TREE-REALIZABILITY

Theorems B and E provide an algorithm to check whether an $n\times n$ distance matrix is tree-realizable. Such an algorithm requires apparently, in the worst case, $6\binom{n}{4}$ operations, since testing each principal submatrix of order 4 requires 3 sums and 3 comparisons. It is therefore polynomial of degree 4 in n.

As a by-product of Theorem 4, an improvement of this algorithm may be obtained. As seen in the proof of Theorem 4, for each index $1\leqslant j\leqslant n$ of a non-tree-realizable matrix D of order n there is a non-tree-realizable principal submatrix of order 4 which contains j. We therefore need only to test, in the worst case, $\binom{n-1}{3}$ principal submatrices of order 4, which corresponds to a polynomial of degree 3 in n.

6. PRESCRIBING TREE-REALIZABILITY OF A GIVEN NUMBER OF SUBMATRICES: SOME EXTREMAL CASES

By the definitions, for any distance matrix D of order n and for any k such that $4\leqslant k\leqslant n$, we have

$$Q_k(D)\leqslant\binom{n}{k}.$$

If D has a tree-realizable submatrix D' of order $n-m$, then all submatrices of D' are tree-realizable. An immediate consequence is the following result.

THEOREM 8. *If for some k such that $4 \leqslant k \leqslant n-m \leqslant n$,*

$$Q_k(D) > \binom{n}{k} - \binom{n-m}{k},$$

then no submatrix of order $n-m$ of D is tree-realizable.

Even if these bounds seem weak, they have some value: Firstly, as we have shown, $Q_4(D)$ can be found in polynomial time; secondly, the problem of finding a tree-realizable submatrix of maximum order of a non-tree-realizable distance matrix is NP-hard. This follows from work by Yannakakis [12], who proved that if Π is a property of graphs which is hereditary (this means, if G has Π, then all subgraphs of G have Π), interesting (graphs of arbitrarily large order may have Π) and nontrivial (some graphs do have Π, some do not), then, given a graph G, the problem of finding a maximum (induced) subgraph with Π is NP-hard. Our problem falls into this category. To each distance matrix D of order n we associate a complete graph G on n vertices whose edge weights are the entries of the matrix. G has property Π if a tree T can be found, a subset V of the vertex set of T can be specified, and an assignment of weights to the edges of T can be made so that the distances between the vertices in V along the paths of T are the weights assigned to the edges of G. Obviously, Π is hereditary, interesting, and nontrivial; to find the maximum order of a submatrix of D which is tree-realizable is equivalent to finding the maximum subgraph of G which has Π.

The bounds on $Q_k(D)$ which we present in this paper can be attained. Before giving examples, we recall the elementary fact that, in the euclidean plane, the sum of the lengths of the diagonals of a convex quadrilateral is always greater than the sum of the lengths of any two opposite sides. This means, by Theorem E, that the 4×4 matrix of the euclidean distances among the vertices of a plane convex quadrilateral is non-tree-realizable. It follows that, for the $n \times n$ matrix D whose entries are the euclidean distances between pairs of points of a set of n points lying on a circle, we have

$$Q_k(D) = \binom{n}{k} \qquad \text{for} \quad k \geqslant 4.$$

Consider now a circle and a chord. Let $n-1$ points $\beta_2, \dots, \beta_n$ lie on the chord, and i points $\alpha_1, \dots, \alpha_i$ lie on one of the arcs of the circle determined by the chord. The matrix D of order $q \equiv i+n-1$ whose entries are the euclidean distances between pairs of points of this set has

$$Q_j(D) = \binom{q}{j} - \binom{q-i}{j}$$

for $4 \leqslant j \leqslant n-1$. In fact, the only submatrices of D which are tree-realizable are those whose entries are the distances between points lying on the chord.

Matrices whose $Q_j(D)$ attain our bounds and which have rational entries can also be given. The interested reader may verify the following results, using Theorem E and straightforward albeit lengthy calculations.

THEOREM 9. *Let n and p be integers, $p > 2n^3$, and D be a symmetric matrix of order n whose entries are, for $y > x$, $d_{xy} = y - x + 1/(y - x + p)$. Then D is a distance matrix and*

$$Q_j(D) = \binom{n}{j} \qquad \textit{for} \quad j = 4, \ldots, n.$$

THEOREM 10. *Let p, i, k, n be positive integers, $i \geqslant 2$, $n \geqslant 5$, $k > 2(4i + 2n)^3$, and $p = 2k + 3i + 2n > 2(k + i + n)$. Let $D = \langle \{\alpha_1, \ldots, \alpha_i, \beta_2, \ldots, \beta_n\} \rangle$ be a symmetric matrix of order $q \equiv n + i - 1$ whose entries are defined as follows:*

$$\begin{array}{ll}
d_{xx} = 0 & \textit{for} \quad x = \alpha_1, \ldots, \alpha_i, \beta_2, \ldots, \beta_n, \\
d_{\alpha_m \alpha_h} = m - h + 1/(m - h + p) & \textit{for all } m \textit{ and } h, \\
d_{\beta_x \beta_y} = 2(y - x)i & \textit{for all } x \textit{ and } y, \\
d_{\alpha_r \beta_2} = i(n-2) + 1/(p + r) & \textit{for all } r, \\
d_{\alpha_r \beta_x} = in + x - 4 + 1/(r + k + x - 4) & \textit{for} \quad 3 \leqslant x \leqslant \lceil (n+1)/2 \rceil, \\
d_{\alpha_r \beta_x} = in + n - x - 2 + 1/(r + k + x - 4) & \textit{for} \quad \lceil (n+1)/2 \rceil < x \leqslant n - 1, \\
d_{\alpha_r \beta_n} = i(n-2) - 1/(p + r) & \textit{for all } r.
\end{array}$$

Then D is a distance matrix, and for $4 \leqslant j \leqslant n + i - 1 - i \equiv n - 1$,

$$Q_j(D) = \binom{q}{j} - \binom{q-i}{j}.$$

Thanks are due to Professor Wilfried Imrich for suggestions leading to a substantially shorter proof of Theorem 2 and for contributing the geometrically constructed examples which we present before our Theorems 9 and 10 in Section 6. Thanks are also due to Professor Andreas Dress for pointing out to us the work done by Professor Manfred Eigen and others on genetic information and for sending us his unpublished manuscript "A characterization of tree like metric spaces or how to construct an evolutionary tree."

REFERENCES

1 F. T. Boesch, Properties of the distance matrix of a tree, *Quart. Appl. Math.* 26:607–609 (1968–69).
2 P. Buneman, A note on the metric properties of trees, *J. Combin. Theory Ser. B* 17:48–50 (1974).
3 A. K. Dewdney, Diagonal tree codes, *Inform. and Control* 40:234–239 (1979).
4 M. Eigen, W. Gardiner, P. Schuster, and R. Winkler-Oswatitsch, The origin of genetic information, *Scientific American*, Apr. 1981, pp. 88–118.
5 S. L. Hakimi and S. S. Yau, Distance matrix of a graph and its realizability, *Quart. Appl. Math.* 22:305–317 (1964–65).
6 W. Imrich, On metric properties of tree-like spaces, in *Contributions to Graph Theory and its Applications* (Internat. Colloq., Oberhof, 1977), Technische Hochschule Ilmenau, Ilmenau, 1977, pp. 129–156.
7 J. Lawrence, C. R. Johnson, and E. Howe, The structure of distances in networks, *Networks*, to appear.
8 A. N. Patrinos and S. L. Hakimi, The distance matrix of a graph and its tree realization, *Quart. Appl. Math.* 30:255–269 (1972–73).
9 J. M. S. Simões-Pereira, Some results on the tree realization of a distance matrix, in *Théorie des Graphes — Journées Internationales d'Étude*, (Dunod, Ed.), Rome, 1966, pp. 383–388.
10 J. M. S. Simões-Pereira, A note on the tree realizability of a distance matrix, *J. Combin. Theory* 6:303–310 (1969).
11 E. A. Smolenskii, Ob odnom sposobe lineĭnoĭ zapisi grafov, *Ž. Vyčisl. Mat. i Mat. Fiz.* 2:371–372 (1962).
12 M. Yannakakis, The effect of a connectivity requirement on the complexity of maximum subgraph problems, *J. Assoc. Comput. Mach.* 26:618–630 (1979).

Received 16 January 1981; revised 10 July 1981