On properties of cell matrices

Gašper Jaklič*, Jolanda Modic†

March 10, 2010

Abstract

In this paper properties of cell matrices are studied. A determinant of such a matrix is given in a closed form. In the proof a general method for determining a determinant of a symbolic matrix with polynomial entries, based on multivariate polynomial Lagrange interpolation, is outlined. It is shown that a cell matrix of size $n > 1$ has exactly one positive eigenvalue. Using this result it is proven that cell matrices are (Circum-)Euclidean Distance Matrices ((C)EDM), and their generalization, $k$-cell matrices, are CEDM under certain natural restrictions. A characterization of $k$-cell matrices is outlined.

Keywords: Cell matrix, Star graph, Determinant, Eigenvalues, Euclidean distance matrix, Circum-Euclidean distance matrix.

1 Introduction

A matrix $M \in \mathbb{R}^{n \times n}$ is an Euclidean Distance Matrix (EDM), if there exist points $x_1, x_2, \ldots, x_n \in \mathbb{R}^r$ ($r \leq n$), such that $m_{ij} = \|x_i - x_j\|^2$ for all $i, j = 1, 2, \ldots, n$ ([1, 2]). These matrices were introduced by Schoenberg in [2, 3] and have received a considerable attention. They are used in applications in geodesy, economics, genetics, psychology, biochemistry, engineering, etc., where frequently a question arises, what facts can be deduced given only distance information. Some examples can be found in [4]: kissing-number of sphere packing, trilateration in wireless sensor or cellular telephone network, molecular conformation, convex polyhedron construction, etc.. In

*FMF and IMFM, University of Ljubljana and PINT, University of Primorska, Jadranska 21, 1000 Ljubljana, Slovenia, gasper.jaklic@fmf.uni-lj.si (Corresponding author)
†FMF, University of Ljubljana, Jadranska 21, 1000 Ljubljana, Slovenia, jolanda.modic@gmail.com
bioinformatics, distance matrices are used to represent protein structures in a coordinate-independent manner, and in DNA/RNA sequential alignment, determination of the conformation of biological molecules from information given by nuclear magnetic resonance data, etc..

An EDM matrix $M$ is circum-Euclidean (CEDM) (also spherical) if the points which generate it lie on the surface of some hypersphere ([5]). Circum-Euclidean distance matrices are important because every EDM is a limit of CEDMs.

Let $a = (a_i), i = 1, 2, \ldots, n,$ be given numbers and $a > 0,$ where the inequality is considered componentwise. A cell matrix $D \in \mathbb{R}^{n \times n},$ associated with $a,$ is defined as

$$d_{ij} := \begin{cases} a_i + a_j, & i \neq j \\ 0, & i = j \end{cases}.$$  

A cell matrix is a particular distance matrix of a star graph (also called a claw graph), i.e., a graph with outer vertices (leaves), connected only to one inner vertex ([6]), where only Euclidean distances between leaves are considered, measured through the inner vertex of the graph. Their name is derived from the approximation theory, where a triangulation (or, more generally, a simplicial partition) with only one inner vertex is called a cell (also star). Such matrices are used in graph theory and in biochemistry [7, 6, 8, 9]. Later on the assumption on positivity of $a$ will be loosened with $a_i \geq 0,$ and thus the inner vertex can be included in the distance information.

If we take $k$ star graphs and connect their inner vertices, we obtain so-called $k$-star graph. The matrix, whose elements are distances between leaves of a $k$-star graph, is a $k$-cell matrix. More precisely, let $C \in \mathbb{R}^{n \times n}$ be a $k$-cell matrix, where $k \leq n/2.$ Let $G$ be the associated connected $k$-star graph, consisting of star graphs $S_1, S_2, \ldots, S_k$ and let $d((u, v))$ be the distance of the edge $(u, v)$ in $G.$ Let $v_i$ be a leaf of the graph $G$ and $u_\ell$ the attached inner vertex. Let us denote $a_i := d(v_i, u_\ell)$ and let $h_{\ell, m} := d(u_\ell, u_m)$ be the distance between inner vertices $u_\ell$ and $u_m$ of star graphs $S_\ell$ and $S_m,$ respectively. Then

$$c_{ij} := \begin{cases} 0, & i = j \\ a_i + a_j, & i \neq j, v_i, v_j \text{ belong to the same star graph} \\ a_i + a_j + h_{\ell, m}, & i \neq j, v_i, v_j \text{ belong to distinct star graphs } S_\ell, S_m \end{cases}.$$ 

In this paper we study properties of cell matrices. Firstly, in Section 2, we establish determinants of principal submatrices of a cell matrix. In the proof a general method for confirming a determinant formula of a symbolic matrix with polynomial entries, based on multivariate polynomial Lagrange interpolation, is outlined. Using this, in Section 3, we show that such a
matrix has only one positive eigenvalue, and establish that cell matrices
belong to a class of well-known Euclidean distance matrices. Furthermore,
they are circum-Euclidean. Also their generalization, \( k \)-cell matrices,
are CEDMs under some natural assumptions. In Section 4, a characterization of
\( k \)-cell matrices is presented. The paper is concluded by an example in the
last section.

2 Determinant and spectrum of cell matrices

First let us determine determinants of principal submatrices of a cell matrix.

**Lemma 1.** Let \( D \in \mathbb{R}^{n \times n} \) be a cell matrix, associated with a vector \( \mathbf{a} > 0 \)
and let \( D^{(i)} := D(1 : i, 1 : i) \), \( i = 1, 2, \ldots, n \), be its principal submatrices. Then

\[
\det D^{(i)} = (-1)^{i-1}2^{i-2} \left( 4(i-1) + \sum_{j=1}^{i} \sum_{\ell=1}^{j-1} \frac{(a_j - a_\ell)^2}{a_j a_\ell} \right) \prod_{k=1}^{i} a_k.
\]

(3)

From Lemma 1 it can easily be seen that if one of the parameters \( a_i \) is
zero, the determinant formula simplifies. If at least two of the parameters \( a_i \)
are zero, a cell matrix is singular.

**Corollary 1.** If \( a_m = 0 \) for some \( m \in \{1, 2, \ldots, n\} \) and \( a_i > 0, \forall i \neq m \), then

\[
\det D = (-1)^{n-1}2^{n-2} \sum_{\ell=1}^{n} a_\ell^2 \prod_{k=1}^{n} a_k.
\]

If \( a_m = a_j = 0, j \neq m \), then \( \det D = 0 \).

**Proof.** Let \( \mathbb{P}_i(\mathbb{R}^d) \) denote the space of polynomials of total degree \( \leq i \) in
d \( d \) variables. Elements of the matrix \( D^{(i)} := D^{(i)}(a_1, a_2, \ldots, a_i) \) are linear
polynomials, therefore \( \det D^{(i)} \) is a polynomial in \( i \) variables \( a_1, a_2, \ldots, a_i \) of
total degree \( \leq i \),

\[
p_i := p_i(a_1, a_2, \ldots, a_i) := \det D^{(i)} \in \mathbb{P}_i(\mathbb{R}^d).
\]

Let us choose \( k = \dim \mathbb{P}_i(\mathbb{R}^d) = \binom{2i}{i} \) pairwise distinct points

\[
\mathbf{a}^{(j)} := \left( a_1^{(j)}, a_2^{(j)}, \ldots, a_i^{(j)} \right) \in \mathbb{Z}^i, \quad 1 \leq j \leq k,
\]
in such a way, that they do not lie on an algebraic hypersurface of degree \( \leq i \).
Thus the multivariate Lagrange polynomial interpolation problem is unisolvent [10]. Integer components are needed only to ensure exact computation.
later on. Now let us evaluate determinants of matrices $D^{(i)}$ at chosen points, $z_j = \det D^{(i)}(a_1^{(j)}, \ldots, a_i^{(j)})$ for $1 \leq j \leq k$. Let $q := q(a_1, a_2, \ldots, a_i) \in \mathbb{P}_i(\mathbb{R}^i)$ denote the polynomial on the right-hand side of (3). Now compute the values $w_j := q(a_1^{(j)}, a_2^{(j)}, \ldots, a_i^{(j)})$. If $w_j = z_j$ for all $j = 1, 2, \ldots, k$, the polynomials $p_i$ and $q$ have the same values at $k = \dim \mathbb{P}_i(\mathbb{R}^i)$ points and as there is precisely one Lagrange interpolation polynomial in $\mathbb{P}_i(\mathbb{R}^i)$ through prescribed data $(a^{(j)}, z_j)$, $p_i \equiv q$. This concludes the proof.

Note that in [11] a similar result has been proven, but is unfortunately inappropriate for our purposes. The presented approach can be efficiently applied in general for proving that a given polynomial expression is the determinant of a symbolic matrix with polynomial entries. Using this approach, the hard part is to somehow obtain the closed form of the determinant, later the proof is quite straightforward, since a powerful tool of approximation theory is applied. Further examples can be found in [12] and [13], where the dimension of a bivariate spline space was studied. An excellent overview of similar methods for determinant calculation is [14].

Since cell matrices $D \in \mathbb{R}^{n \times n}$ are symmetric, their eigenvalues $\lambda_i$ are real. They have a zero diagonal, hence the sum of their eigenvalues is zero. The following theorem shows that they have exactly one positive eigenvalue, the rest of eigenvalues are negative. Thus cell matrices are nonsingular if $a > 0$.

**Theorem 1.** Let $D \in \mathbb{R}^{n \times n}$ be a cell matrix, associated with a vector $a > 0$ and let $D^{(i)} := D(1 : i, 1 : i)$, $i = 1, 2, \ldots, n$, be its principal submatrices. Let

$$
\lambda_i^{(i)} \leq \lambda_i^{(i-1)} \leq \cdots \leq \lambda_2^{(i)} \leq \lambda_1^{(i)}
$$

be the eigenvalues of the matrix $D^{(i)}$. Then $\lambda_1^{(i)} > 0$, $\lambda_2^{(i)} < 0$ for $i > 1$ and $\lambda_1^{(1)} = 0$.

**Proof.** Let $p_i(x) := \det(D^{(i)} - xI)$ denote the characteristic polynomial of the matrix $D^{(i)}$. Clearly, $\lambda_1^{(1)} = 0$. Lemma 1 yields the determinant $\det D^{(i)}$ in a closed form (3). Since trace $D^{(i)} = \sum_{j=1}^i \lambda_j^{(i)} = 0$ and det $D^{(i)} = \prod_{j=1}^i \lambda_j^{(i)} \neq 0$,

$$
\lambda_1^{(i)} > 0, \quad \lambda_i^{(i)} < 0, \quad i > 1,
$$

in particular $\lambda_2^{(2)} < 0$. Cauchy’s interlacing theorem [15] implies

$$
\lambda_3^{(i-1)} \leq \lambda_3^{(i)} \leq \lambda_2^{(i-1)} \leq \lambda_2^{(i)} \leq \lambda_1^{(i-1)} \leq \lambda_1^{(i)}.
$$

A quick calculation gives

$$
p_3(0) = 2(a_1 + a_2)(a_1 + a_3)(a_2 + a_3) > 0.
$$
Since \( p_3(0) > 0 \), \( \lambda_2^{(2)} = -(a_1 + a_2) < 0 \), \( \lambda_1^{(2)} = a_1 + a_2 > 0 \) and \( p_3(x) = (\lambda_1^{(3)} - x)(\lambda_2^{(2)} - x)(\lambda_3^{(3)} - x) \), therefore \( \lambda_2^{(3)} < 0 \).

Now let by inductive supposition \( \lambda_i^{(i-1)} < 0 \). Recall the interlacing of eigenvalues (4), and the facts \( \lambda_1^{(i)} > 0 \) and \( \lambda_j^{(i)} < 0 \), \( j \geq 3 \). Since \( p_i(x) = \prod_{j=1}^{i} (\lambda_j^{(i)} - x) \) and by (3), \( \text{sign}(p_i(0)) = (-1)^{i-1} \), thus \( \lambda_2^{(i)} < 0 \). This concludes the proof.

3 Euclidean distance matrices

It is interesting to consider a relation between cell matrices and well-known Euclidean distance matrices. By using Theorem 1 and a characterization of Euclidean distance matrices ([1]), we can prove the following claim.

**Theorem 2.** Cell matrices are Euclidean distance matrices. Furthermore, they are circum-Euclidean.

**Proof.** Let \( D \in \mathbb{R}^{n \times n} \) be a cell matrix, associated with a position vector \( a > 0 \) and let us define \( e := [1, 1, \ldots, 1]^T \in \mathbb{R}^n \). Since by Theorem 1 \( D \) has exactly one positive eigenvalue, it is by a characterization of Euclidean distance matrices [1, Thm. 2.2], enough to prove, that there exists \( w \in \mathbb{R}^n \), such that \( Dw = e \) and \( w^T e \geq 0 \).

Let us define \( z_\ell := \left( \sum_{j=1}^{n} \frac{1}{a_j} - (n - 2) \frac{1}{a_\ell} \right) \prod_{k=1}^{n} a_k \) for \( \ell = 1, 2, \ldots, n \), \( z := (z_\ell)_\ell \) and \( t := \left( \sum_{k=1}^{n} a_k \sum_{j=1, j \neq k}^{n} \frac{1}{a_j} - (n - 4)(n - 1) \right) \prod_{\ell=1}^{n} a_\ell. \)

Further, let \( w := 1/t \cdot z \). Then \( Dw = e \). A simple computation yields

\[
 w^T e = \frac{2}{t} \sum_{j=1}^{n} \frac{1}{a_j} \prod_{k=1}^{n} a_k > 0.
\]

The inequality follows from the observation

\[
 t = \frac{\det D}{(-1)^{n-1}2^{n-2}}
\]

and thus from (3) clearly \( t > 0 \). This confirms that \( D \) is an Euclidean distance matrix.
Now let \( s = (s_i)_i \), with

\[
    s_i := \frac{1}{2} \left( 1 - \frac{n - 2}{a_i \sum_{k=1}^{n} \frac{1}{a_k}} \right), \quad i = 1, 2, \ldots, n - 1,
\]

and

\[
    s_n := 1 - \sum_{i=1}^{n-1} s_i, \quad \beta := \frac{1}{2} \left( \sum_{k=1}^{n} a_k - \frac{(n - 2)^2}{\sum_{k=1}^{n} \frac{1}{a_k}} \right).
\]

By [5, Thm. 3.4], an EDM matrix \( D \in \mathbb{R}^{n \times n} \) is CEDM iff there exist \( s \in \mathbb{R}^n \) and \( \beta \in \mathbb{R} \), such that \( Ds = \beta e \) and \( s^T e = 1 \). Since the constructed \( s \) and \( \beta \) satisfy these relations, the matrix \( D \) is CEDM, and the proof is completed.

In the case when \( \ell \geq 1 \) of the parameters \( a_i \) are zero, the corresponding cell matrix \( D \) is still a CEDM. But now \( D \) has \( \ell - 1 \) zero eigenvalues, and the proof is more complicated.

**Theorem 3.** Let a cell matrix \( D \in \mathbb{R}^{n \times n} \) be associated with parameters \( a \), and let there be \( 1 \leq \ell \leq n - 1 \) zero parameters, i.e., \( a_{i_1} = a_{i_2} = \cdots = a_{i_\ell} = 0 \). Then \( D \) is a circum-Euclidean distance matrix.

**Proof.** First, let us prove, that the matrix \( D \) has only one positive eigenvalue. If \( \ell = 1 \), i.e., \( a_{i_1} = 0 \), \( a_j \neq 0 \), \( j \neq i_1 \), the proof is analogous to the one of Theorem 1, where we use the determinant formula, given by Corollary 1, instead of (3).

Now, let \( \ell > 1 \). The matrix \( D \) is singular (see Corollary 1), and since trace \( D = 0 \) and \( D \neq 0 \), it has at least one positive and one negative eigenvalue. Clearly the rows \( i_1, i_2, \ldots, i_\ell \) of the matrix \( D \) are equal. Thus there are at least \( \ell - 1 \) zero eigenvalues. Let \( \mathcal{I} = \{1, 2, \ldots, n\} \setminus \{i_1, i_2, \ldots, i_\ell\} \) be an increasing sequence of indices. Now let \( P \in \mathbb{R}^{n \times n} \) be a permutation matrix, associated with the index sequence \( \{i_1, \mathcal{I}, i_2, \ldots, i_\ell\} \). The matrix \( D' = PDPT^T \) is similar to \( D \), since \( P^T = P^{-1} \).

Now let \( M \in \mathbb{R}^{n \times n} \) be an identity matrix, and additionally set \( m_{i_k, i_1} = -1 \) for \( k = 2, 3, \ldots, \ell \). The matrix \( M \) is clearly nonsingular, and the transformation \( D'' := MD'M^T \) sets the dependent rows and columns of the matrix \( D' \) to zero. The principal submatrix \( F \) of size \( n - \ell + 1 \) of the matrix \( D'' \) is a cell matrix, associated with parameters \( a_j \), \( j \neq i_2, i_3, \ldots, i_\ell \) and \( a_{i_1} = 0 \). By Corollary 1, the matrix \( F \) is of full rank, \( n - \ell + 1 \), and from the first part of the proof (the case \( \ell = 1 \)), the matrix \( F \) has one positive and \( n - \ell \) negative eigenvalues. But the spectrum of \( D'' \) consists of the eigenvalues of \( F \) and \( \ell - 1 \) zeros. By Sylvester’s law of inertia the number of positive (also zero,
and negative) eigenvalues of the matrices $D'$ and $D$ are the same. Thus the matrix $D$ has only one positive eigenvalue, $\ell - 1$ zero eigenvalues, and $n - \ell$ negative eigenvalues.

We have proven that the matrix $D$ has precisely one positive eigenvalue. Now we have to find a specific vector $w$, such that $Dw = e$ and $w^T e \geq 0$.

Let us define $A := \sum_{j \neq i_1, i_2, \ldots, i_\ell} a_j$ and construct a vector $w \in \mathbb{R}^n$ as follows

$$
\begin{align*}
w_{i_1} & := -\frac{n - 2 - \ell}{A}, \\
w_{i_k} & := 0, \quad k = 2, 3, \ldots, \ell, \\
w_j & := \frac{1}{A}, \quad j \neq i_1, i_2, \ldots, i_\ell.
\end{align*}
$$

A simple computation shows that $Dw = e$ and $w^T e = 2/A > 0$. By [1, Thm. 2.2], $D$ is an EDM matrix.

If $s := A/2 \cdot w$ and $\beta := A/2$, then $Ds = \beta e$ and $s^T e = 1$. By [5, Thm. 3.4], the matrix $D$ is CEDM. \hfill \square

Remark 1. Note that by setting $a_i = 0$ one can consider the inner vertex of a star graph as one of its leaves. Thus a cell matrix with an additional parameter $a_i$ equal to zero describes the whole distance information of a star graph.

Now let us consider $k$-cell matrices, $k > 1$. Let $C \in \mathbb{R}^{n \times n}$ be a $k$-cell matrix, associated to star graphs $S_1, S_2, \ldots, S_k$, $k \leq n/2$. Since $C$ is defined as (2), it can easily be seen, that it can be rewritten as $C = D + R$, where

$$
R := \begin{bmatrix}
0 & h_{12}E & h_{13}E & \ldots & h_{1k}E \\
h_{12}E & 0 & h_{23}E & \ldots & h_{2k}E \\
\vdots & \vdots & \ddots & \ldots & \vdots \\
h_{1,k-1}E & h_{2,k-1}E & \ldots & h_{k-1,k}E \\
h_{1k}E & h_{2k}E & h_{3k}E & \ldots & 0
\end{bmatrix}. \tag{5}
$$

Here $h_{ij} \geq 0$ is the distance between inner vertices of star graphs $S_i$ and $S_j$, $E$ is a matrix of ones, and $D$ is a cell matrix (1), associated with a vector $(a_1, a_2, \ldots, a_n)$. Let $H := (h_{ij}) \in \mathbb{R}^{k \times k}$, $h_{ii} = 0$, $H = H^T$, denote a matrix of distances between inner vertices.

Also $k$-cell matrices are Euclidean distance matrices, if some relations on the distances $h_{ij}$ between inner vertices of the underlying star-graphs are satisfied.
Theorem 4. A $k$-cell matrix $C$ is an Euclidean distance matrix if the inner-distance matrix $H$ is an Euclidean distance matrix. Furthermore, if at most one parameter $a_{i_k}$ is zero, $C$ is also CEDM. If $a_{i_1} = a_{i_2} = \cdots = a_{i_\ell} = 0$, $\ell > 1$, the matrix $C$ is CEDM if the matrix $H$ is CEDM.

Remark 2. In order to verify that the matrix $H$ is EDM, one can define a matrix $F := (f_{ij}) \in \mathbb{R}^{(k-1) \times (k-1)}$, $f_{ij} = h_{ik} + h_{jk} - h_{ij}$, and check that it is positive semidefinite. For example, if $k = 3$, the following inequality should be satisfied
\[ 2(h_{12}h_{23} + h_{13}h_{23} + h_{12}h_{13}) \geq h_{12}^2 + h_{13}^2 + h_{23}^2. \]  

Proof. The proof will consist of two parts. First it will be shown that the matrix $C$ is EDM under given assumptions, and in the second part, the CEDM property will be established.

The case $k = 1$ is covered by Theorem 2 and Theorem 3. Now assume $k > 1$, and consider the matrix $C = D + R$, where $R$ is defined by (5).

First, let us prove, that the matrix $C$ is an EDM. By Theorem 2, $D$ is EDM, thus $x^TDx \leq 0$ for all $x$, such that $x^Te = 0$ (cf. [2, 3]).

Now take an arbitrary $x$, such that $x^Te = 0$, and consider $x^TCx$. By EDM characterization [2, 3], it is enough to prove that the matrix $C$ is negative semidefinite, i.e., $x^TCx \leq 0$ for all $x$, such that $x^Te = 0$. Since $x^TDx \leq 0$, let us consider $x^TRx$. To simplify the notation, let us introduce the vector sum, $s(y) := y^Te = \sum y_j$. Now let us write the vector $x$ in a block form, $x = [x_1^T, x_2^T, \ldots, x_k^T]^T$, where the dimensions of vectors $x_j$ match the dimensions of blocks in $R$. Note that $Ey = s(y)e$. Thus a simple computation yields a simplified quadratic form
\[ x^TRx = 2 \sum_{i < j} h_{ij} s(x_i)s(x_j). \]  

Since $s(x) = \sum_{i=1}^k s(x_i) = 0$, the expression (7) is by Schoenberg’s theorem [1] equivalent to the fact that the matrix $H$ is an EDM. By using a substitution $s(x_k) = -\sum_{i=1}^{k-1} s(x_i)$ in (7), a new quadratic form is obtained with coefficients $-1/2f_{ij}$, where $f_{ij} = h_{ik} + h_{jk} - h_{ij}$, and $h_{ii} = 0$. To show that (7) is negative semidefinite for particular vectors $x$, such that $x^Te = 0$, is equivalent to show that the matrix $F \in \mathbb{R}^{(k-1) \times (k-1)}$ is positive semidefinite, and Remark 2 follows. Thus $x^TRx \leq 0$ and the first part of the proof is completed.

Now assume that at most one $a_{i_k}$ is zero. Since the matrix $C$ is EDM, there exists a vector $w$, such that $Cw = e$ and $w^Te \geq 0$. First, let us consider the case $w^Te = 0$. Clearly, $w \neq 0$. Since
\[ Cw = Dw + Rw = e, \]
\( w^TDw + w^TRw = 0 \), and the fact that \( D \) and \( R \) are EDMs yields \( w^TDw = w^TRw = 0 \).

Since \( D \) is a CEDM, there exists \( \beta \), such that a matrix \( B := \beta ee^T - D \) is positive semidefinite (see [5, Cor. 3.1]). Thus \( Bw = -Dw \) and \( w^TBw = 0 \), hence \( Bw = 0 \), since \( B \) is positive semidefinite. This yields \( Dw = 0 \), a contradiction, since \( \ker D \) is trivial by Corollary 1 and Theorem 1.

Therefore, \( 1/\kappa := w^Te > 0 \), and let \( s = \kappa w \). Consequently, \( Cs = \kappa e \) and \( s^Te = 1 \), and by CEDM characterization, the matrix \( C \) is CEDM.

We are left with the case \( a_{i_1} = a_{i_2} = \cdots = a_{i_\ell} = 0 \), \( \ell > 1 \). Here the CEDM vector \( s \) can be given in a closed form. Define \( s := w + z \), with

\[
\begin{align*}
w_{i_1} &:= -\frac{1}{2}(n - 2 - \ell), & z_{i_1} &:= -y_2 - y_3 - \cdots - y_\ell, \\
w_{i_j} &:= 0, & z_{i_j} &:= y_j, \quad j = 2, 3, \ldots, \ell, \quad (8) \\
w_j &:= \frac{1}{2}, & z_j &:= 0, \quad j \neq i_1, i_2, \ldots, i_\ell,
\end{align*}
\]

where \( y_2, y_3, \ldots, y_\ell \) are unknown parameters. Clearly \( s^Te = 1 \) by construction. Thus \( Cs = \kappa e + Dz + Rs \), since \( D \) is CEDM, and by using the proof of Theorem 3. Therefore it is enough to solve a system \( Dz + Rs = \gamma e \). But the choice of \( z \) in (8) guarantees \( Dz = 0 \), hence only the system \( Rs = \gamma e \) has to be studied. Recall the idea of the first part of the proof, and write the vectors \( w \) and \( z \) in a block form, i.e., \( w = [w_1^T, \ldots, w_k^T]^T \), and \( z = [z_1^T, \ldots, z_k^T]^T \).

By using this notation, it turns out that \( R_iw = \sum_{j=1}^k h_{ij}s(w_j)e \), where \( R_i \) denotes the \( i \)-th block row of the matrix \( R \). Similarly, \( R_iz = \sum_{j=1}^k h_{ij}s(z_j)e \).

Denote \( t_i := s(w_i) + s(z_i) \) and \( t := (t_i)_{i=1}^k \), which yields \( s(t) = 1 \). Therefore \( Rs = \gamma e \) if \( Ht = \gamma e \), where \( t^Te = 1 \). But this is precisely CEDM characterization for \( H \), and the proof of the theorem is completed.

As an example let us consider a \( k \)-cell matrix, \( k > 1 \), where the inner vertices of star graphs are considered as leaves (i.e., the parameters \( a_{i_1} = a_{i_2} = \cdots = a_{i_k} = 0 \)). In this case the CEDM vector \( s \) can be given in a closed form \( s = w + z \) (see (8)). Clearly \( s^Te = 1 \) by construction. If \( k = 2 \) this yields \( Cs = 1/2(A + h_{12})e \) with \( A := \sum_{j\neq i_1, i_2, \ldots, i_k} a_j \). If \( k = 3 \), we obtain a system for \( y_2 \) and \( y_3 \), which has a solution if there is a strict inequality in (6).

### 4 Cell matrix characterization

A natural question arises: Is it possible to characterize cell matrices? The answer is affirmative.
Theorem 5. A matrix \( D \in \mathbb{R}^{n \times n} \) is a cell matrix, iff \( d_{ij} \geq 0, d_{ii} = 0, i, j = 1, 2, \ldots, n, \ D = D^T, \) and the following relations are fulfilled:

\[
d_{i,n-2} + d_{n-1,n} = d_{i,n-1} + d_{n-2,n}, \quad i = 1, 2, \ldots, n - 3,
\]
\[
d_{ij} + d_{n-2,n} + d_{n-1,n} = d_{i,n} + d_{j,n} + d_{n-2,n-1}, \quad \forall i < j, \ i, j \in \{1, 2, \ldots, n - 3\},
\]

and

\[
(n - 1)b_i - s(d) \geq 0, \quad i = 1, 2, \ldots, n,
\]

where

\[
b_i := \sum_{k=1}^{n} d_{ik}, \quad s(d) := \sum_{i,j=1 \atop i < j}^{n} d_{ij}.
\]

Proof. The matrix \( D \) is a cell matrix, if one can find \( a_i \geq 0, i = 1, 2, \ldots, n, \) such that \( a_i + a_j = d_{ij}, i < j. \) This is an overdetermined system of linear equations \( Aa = d \) with a very nice structure, if the lexicographic ordering of indices of \( d_{ij} \) is considered \( (d_{12}, d_{13}, \ldots, d_{1n}, d_{23}, \ldots, d_{2n}, \ldots, d_{n-1,n}). \) The matrix \( A \) is of full rank, thus the solution \( a = (a_i)_{i=1}^{n} \) can be obtained by solving the normal equations. A quick look at the normal equations yields \( A^TA = (n - 2)I + E, \) and it can be shown that

\[
(A^TA)^{-1} = \frac{1}{2(n - 2)(n - 1)} \cdot ((2n - 2)I - E),
\]

thus the solution

\[
a = (A^T A)^{-1} A^T d,
\]

can be easily obtained, and \( \|Aa - d\|_2 = 0 \) if the relations (9) are satisfied.

In order for \( a_i \) to be nonnegative, one can study the inequality \( a_i = e^T_i a \geq 0. \) By (12) and (11), it is enough to consider the relation \( e^T_i ((2n - 2)I - E)A^T d \geq 0. \) A simplification of the expression, and \( b_i = (A^T d)_i \) yield the conditions (10).

The converse of the theorem can be straightforwardly proven by substituting \( d_{ij} = a_i + a_j \) in (9) and verifying the relations (9) and (10) using \( b_i = (n - 2)a_i + \sum_{k=1}^{n} a_k \) and \( s(d) = (n - 1) \sum_{k=1}^{n} a_k. \)

Remark 3. The cases \( n = 1, 2, 3 \) are special. If \( n = 1, \) the matrix \( D = [0] \) is a cell matrix. If \( n = 2, \) \( d_{12} = a_1 + a_2 \geq 0 \) for suitable nonnegative pairs of \( a_1 \) and \( a_2. \) In the case \( n = 3, \) the system \( Aa = d \) is square, and only the relations (10) have to be satisfied. They simplify into the triangle inequality for \( d_{12}, d_{13} \) and \( d_{23}. \)
A similar characterization can be given for $k$-cell matrices.

**Theorem 6.** A matrix $C \in \mathbb{R}^{n \times n}$ is a $k$-cell matrix iff there exist $i_1, i_2, \ldots, i_{k-1}$, such that the matrices $C_{11} = C(1 : i_1, 1 : i_1)$, $C_{22} = C(i_1 + 1 : i_2, i_1 + 1 : i_2)$, \ldots $C_{kk} = C(i_{k-1} + 1 : n, i_{k-1} + 1 : n)$ are cell matrices of dimension at least 2 (satisfying conditions of Theorem 5), and the corresponding matrix $R = C - D$ is of the form (5) with $h_{ij} \geq 0$.

**Proof.** The matrices $C_{jj}$, $j = 1, 2, \ldots, k$, should satisfy assumptions of Theorem 5. By following its proof, the corresponding generating parameters $a_{i_{j-1} + 1}$, $a_{i_{j-1} + 2}$, \ldots, $a_{i_j}$ are computed. Using the obtained parameters $a_1, a_2$, \ldots, $a_n$, the cell matrix $D = (a_i + a_j)_{i \neq j}$ can be constructed. In order for $C$ to be a $k$-cell matrix, the matrix $R = C - D$ should be of the form (5). The converse of the claim is obvious. \hfill \Box

Cell matrices could be considered in relation with line distance matrices, introduced in [16]. For a given sequence $t_1 < t_2 < \cdots < t_n$, a **line distance matrix** $L$ is defined as a $n \times n$ matrix with elements $\ell_{ij} = |t_i - t_j|$. In [16] it was proven that such a matrix is EDM, and its spectral properties were applied for studying the DNA sequence alignment problem. Let us briefly demonstrate, that the proof of [16, Thm. 2] can be simplified, and furthermore, it can be shown that such a matrix is CEDM.

**Theorem 7.** A line distance matrix $L \in \mathbb{R}^{n \times n}$, defined by a sequence $t_1 < t_2 < \cdots < t_n$, is CEDM.

**Proof.** Let us consider a matrix $D := (d_{ij})$ with $d_{ij} = (t_i - t_j)^2$. Clearly, the matrix $D$ is EDM, defined by points $t_i$ on the real line. But by [4, 3], the matrix with elements $\sqrt{d_{ij}} = \ell_{ij}$ is EDM.

Now let $s := [1/2, 0, \ldots, 0, 1/2]^T$ and $\beta := 1/2(t_n - t_1)$. Then it can easily be verified that $Ls = \beta e$, and $s^T e = 1$, where $e = [1, 1, \ldots, 1]^T$. By a CEDM characterization [5, Thm. 3.4], the matrix $L$ is CEDM. \hfill \Box

Some more interesting properties of EDMs and CEDMs can be found in [17, 18].

For a given EDM $D \in \mathbb{R}^{n \times n}$ it is possible to obtain a set of points $x_i$, such that $d_{ij} = \|x_i - x_j\|_2^2$ (see [1]). First one needs to construct a Gower’s centered matrix

$$G := -\frac{1}{2}(I - es^T)D(I - se^T),$$

where $s$ is a vector such that $s^T e = 1$ and $e$ is a vector of ones. Usually, one takes $s = 1/n \cdot e$. Since $G = X^T X$ is positive semidefinite, $X =
\[ \text{diag}(\sqrt{\sigma_i}) U^T, \] where \( G = U\Sigma U^T \) is the singular value decomposition of \( G \) and \( \Sigma = \text{diag}(\sigma_i) \).

In order to obtain generating points \( x_i \) for a CEDM, one has to use \( s \), which satisfies the relations \( Ds = \beta e \), \( s^T e = 1 \). Thus the obtained points lie on a hypersphere with the center \( 0 \) and the radius \( \sqrt{\beta/2} \).

The points \( x_i \) are obtained as columns of \( X \). The embedding dimension of \( D \) thus equals the rank of the matrix \( G \). Of course the points \( x_i \) are not unique, since an arbitrary translation, rotation or a mirror map preserves their distance matrix. This can be used to get a dual representation of vertices of a generalized star graph, as will be demonstrated in the following section.

### 5 Example

As an example, consider traveling by an airplane from a small airport to another one. Assume that there is no direct connection, so one has to fly first from the beginning airport to a larger airport (hub), maybe fly to another hub, and from there to the final destination. If there are several hubs, and there are no direct connections between smaller airports, this can be modeled as a \( k \)-cell matrix. Distances \( a_i \) are the distances between the small airport \( i \) and the attached hub. Hubs can be represented by parameters \( a_i \) equal to zero.

Let us take 3 hubs, Frankfurt (FRA), Atlanta (ATL) and Hong Kong (HKG) airports. Let Strasbourg (SXB) and Vienna (VIE) be connected to FRA, Miami (MIA) to ATL, and Taipei (TPE) to HKG (see Fig. 1, left). Then the associated 3-cell matrix is

\[
D = \begin{bmatrix}
0 & 802 & 178 & 8561 & 7604 & 10156 & 9349 \\
802 & 0 & 624 & 9007 & 8050 & 10602 & 9795 \\
178 & 624 & 0 & 8383 & 7426 & 9978 & 9171 \\
8561 & 9007 & 8383 & 0 & 957 & 15281 & 14474 \\
7604 & 8050 & 7426 & 957 & 0 & 14324 & 13517 \\
10156 & 10602 & 9978 & 15281 & 14324 & 0 & 807 \\
9349 & 9795 & 9171 & 14474 & 13517 & 807 & 0
\end{bmatrix}.
\] (13)

In (13), the real distance information in km is used, obtained by Great Circle Mapper, [http://gc.kls2.com/](http://gc.kls2.com/).

It can easily be seen that the matrix \( D \) has one positive and 6 negative eigenvalues. The matrix is CEDM, and by Gower’s construction, described in the previous section, a dual representation of airports can be obtained.
Figure 1: The graph of considered airports and its dual representation.

(Fig. 1, right). The best rank 3 approximation of the matrix $X$ is used, easily obtained from the singular value decomposition.

For some applications of cell matrices in chemistry, see for example [11, 6].

References


