NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

School of Science

Information Technologies in Medicine and Biology

Direction: ***Bioinformatics***

Algorithms in Structural Bioinformatics

Postgraduate Student: *Begetis Nikolaos* Professor: *Emiris Ioannis*

Deadline Date: *07/05/2013*

Assignment 5

1. Distances in the plane

$$\textbf{\textit{Matrix B}} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \boldsymbol{a} & \boldsymbol{b} \\ \mathbf{1} & \boldsymbol{a} & \mathbf{0} & \boldsymbol{c} \\ \mathbf{1} & \boldsymbol{b} & \boldsymbol{c} & \mathbf{0} \end{bmatrix}$$

1(a).

We are assigned to take the border (Cayley-Menger) matrix B, with entries $dist_{ij}^2/2$ and let c=1. The question we have to answer is if it is possible to set one distance to 1 and explain how we do it.

As it was referred in class lecture and as we can see from the definitions of Caley and Menger the rank of one matrix depends on its distance, and $\forall k \geq d+3, \forall k\ x\ k\ the\ minor\ matrices = 0\ and\ vice-verca\ \forall k \leq d+2, \forall k\ x\ k\ the$

$minor\ matrices \neq 0.$

So,

1(b).

As a second task we are assigned to write Menger´s inequality$(-1)^3 D(1,2,3) \geq 0$, for c=1. Doing so, we should show what does it imply for a and b.

So, we construct the 3x3 M distance matrix which should be:

M=$\begin{bmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{bmatrix}$, with $i_1 = [0, a, b], i_2 = [a, 0, c]\ and\ i_3 = [b, c, 0]$.

Moreover, because of the definition of the distance matrix M, the formula $(-1)^k D(i_1,..,i_k) \geq, \forall k \geq 2$ should be valid. So, setting c=1, we find the determinant:

$$det \begin{vmatrix} 0 & a & b \\ a & 0 & 1 \\ b & 1 & 0 \end{vmatrix} = -a(-b) + ba = ab + ba = 2ab.$$

At follow, due to the Menger's inequality we have:

$$(-1)^3 \cdot 2ab \geq 0 = -2ab \geq 0 = ab \leq 0 \leftrightarrow sign(a) \neq sign(b).$$

So, doing the above calculations we can imply that the signs of a and b are opposite one another.

1(c).

As the final task by letting a=2, and b=c=1, we should construct the 2x2 Gram matrix G with $G_{ij} = \frac{1}{2}(d_{i0}^2 + d_{j0}^2 - d_{ij}^2)$, where by picking the point p_0, which has a special role, we should compute the coordinates of the corresponding points by SVD applied to the planar case.

At this time, the distance matrix M should be:

M=$\begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, with $i_1 = [0,2,1], i_2 = [2,0,1]$ and $i_3 = [1,1,0]$.

Taking into account the formula $dist_{ij}^2/2$ and $G_{ij} = \frac{1}{2}(d_{i0}^2 + d_{j0}^2 - d_{ij}^2)$, we are going to construct the Gram Matrix G. We have got:

$$G_{ij} = \frac{1}{2}(d_{i0}^2 + d_{j0}^2 - d_{ij}^2)$$

So,

$$G_{11} = \frac{1}{2}(d_{10}^2 + d_{10}^2 - d_{11}^2),\ with\ M_{21} = \frac{1}{2}d_{10}^2 = 2, and\ M_{22} = \frac{1}{2}d_{11}^2 = 0$$

and as a result,

$G_{11} = \frac{1}{2}(2 + 2 - 0) = 4$, and so on we continue.

After that we find the Singular Value Decomposition (SVD) and finally, using the formula $P := \sqrt{\Sigma}\, V^T$ (4.dists.pdf/p.14) we conclude to the following results:

$$U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

In the deliverable we provide the Matlab implementation which concluded in the above results.

2. Cyclohexane

$$\textbf{\textit{Matrix B}} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & u & c & x & c & u \\ 1 & u & 0 & u & c & y & c \\ 1 & c & u & 0 & u & c & z \\ 1 & x & c & u & 0 & u & c \\ 1 & c & y & c & u & 0 & u \\ 1 & u & c & z & c & u & 0 \end{bmatrix}$$

<u>2(a).</u>

We consider 6 points in R^3 and their Cayley-Menger (border) matrix B as shown above, where $u = 1.526,\ c = 2.285\ and\ x, y, z$ are unknown, of the form $dist_{ij}^2 = 2$. We suppose $x \in \{4.685396365, 11.2278561\}, y \in \{2.63120838, 3.81109533\}\ and\ z \in \{3.8112039, 0.4330644\}$

We create the distance matrix M as:

$$\mathbf{M} = \begin{bmatrix} 0 & u & c & x & c & u \\ u & 0 & u & c & y & c \\ c & u & 0 & u & c & z \\ x & c & u & 0 & u & c \\ c & y & c & u & 0 & u \\ u & c & z & c & u & 0 \end{bmatrix}$$

and for M, using the triangle inequality $\left|d_{ij} - d_{jk}\right| \le d_{ik} \le d_{ij} + d_{jk}$ and considering the formula $M_{ij} = dist_{ij}^2/2 \rightarrow d_{ij} = \sqrt{2M_{ij}}$ we can assume that for every matrix element there should be:

$$\left|\sqrt{2M_{ij}} - \sqrt{2M_{ik}}\right| \le \sqrt{2M_{ik}} \le \sqrt{2M_{ij}} + \sqrt{2M_{ik}}$$

So, consider the points p1, p3, p4 (cyclohexane carbon points), then for $\mathrm{i} = 1,\ \mathrm{j} = 2,\ \mathrm{k} = 4$ we have

$$\left|\sqrt{2M_{12}} - \sqrt{2M_{24}}\right| \le \sqrt{2M_{14}} \le \sqrt{2M_{12}} + \sqrt{2M_{24}} \rightarrow \left|\sqrt{2u} - \sqrt{2c}\right| \le \sqrt{2x} \le \sqrt{2u} + \sqrt{2c} = 0.0763 \ \le\ x\ \le\ 7.5460.$$

So the values of x are ranging in the intelval $x \in \{7.5460, 11.2278561\}$ and all the other candidates are ruled out.

2(b) + 2(c).

Due to Cayley's definition[4.dists.pdf/p.11] $rank(B) = d + 2$. In our case $rank(B) = 5$, because d=3.
We made an implementation for that, and set a threshold of 0.001.
The results for $x = 4.685396365, y = 2.63120838, z = 0.4330644$ are shown below:

For the Gram matrix G computation we used the implementation of task 1c by appending it in the previous implementation of the task 2(b). The results are shown below:

$$U = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.655 & 0.023 & 0 & 0 & 0 \\ 0 & 0.658 & 0 & 0.173 & 0 \\ 0.574 & 0 & 0 & 0.643 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 11.419 & 0 & 0 & 0 & 0 \\ 0 & 4.684 & 0 & 0 & 0 \\ 0 & 0 & 1.350 & 0 & 0 \\ 0.0000 & 0 & 0 & 0 & 0 \\ 0 & 0.000 & 0.000 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.655 & 0.023 & 0 & 0 & 0 \\ 0 & 0.658 & 0 & 0.173 & 0 \\ 0.574 & 0 & 0 & 0.643 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 1.419 & 0 & 0 & 0 \\ 0 & 0.050 & 0.764 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.201 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In the deliverable we provide the Matlab implementation which concluded in the above results.