Algorithms in Structural Bioinformatics

Assignment 5, I. Emiris

version 25/4. Deadline: Tue. 7/5

It is easier if you use an algebra package like Matlab (or its free equivalent, Scilab), Maple (available from our Lab), or Mathematica. This assignment concerns methods described in 3.dists.pdf.

Distances in the plane

Take border (Cayley-Menger) matrix B, with entries $\text{dist}_{ij}^2/2$:

$$B = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & a & b \\ 1 & a & 0 & c \\ 1 & b & c & 0 \end{bmatrix}$$

a) Let $c = 1$: is it always possible to set one distance to 1? why and how?
b) For $c = 1$, write Menger's inequality $(-1)^2 D(1,2,3) \geq 0$: What does it imply for a, b?
c) Let $a = 2, b = c = 1$ and construct the 2×2 Gram matrix G with $G_{ij} = \frac{1}{2}(d_{i0}^2 + d_{j0}^2 - d_{ij}^2)$, where you pick point p_0 having a special role. Then, compute the coordinates of the corresponding points by SVD applied to the planar case (3.dists.pdf describes the 3d case).

Cyclohexane

Consider 6 points in $\mathbb{R}^3$ and their Cayley-Menger (border) matrix B, where $u = 1.526, c = 2.285$ and x, y, z are unknown, of the form $\text{dist}_{ij}^2/2$. Suppose $x \in \{4.685396365, 11.2278561\}$, $y \in \{2.63120838,\ 3.81109533\}$ and $z \in \{3.8112039,\ 0.4330644\}$.

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

a) Using the triangle inequality, can you rule out any candidate value of x?

b) For which of the remaining combinations of x, y, z, the pointset embeds in $\mathbb{R}^3$? If you use the rank condition to decide embeddability, pay attention to *numerical* rank (computed by, e.g., Matlab), because it is higher than "true" rank, due to floating point calculations.

c) Pick one solution from (b), construct the 5×5 Gram matrix G where $x_{ij} = \frac{1}{2}(d_{i6}^2 + d_{j6}^2 - d_{ij}^2)$, $\forall i, j = 1, \ldots, 5$, and find the coordinates of the pointset in $\mathbb{R}^3$ (applying your code from 1c).