Structural Bioinformatics 2004-05 Semester B
Assignment 3.

Due to March 29

1. Given two equal size point sets $A = \{a_i\}_1^n$ and $B = \{b_i\}_1^n$ consider three similarity measures (in 3D):

   (a) Bottleneck (L_∞): there exists a permutation of B, B', and there exists a transformation T, such that:
   $\underset{i}{max}|a_i - T(b'_i)| \leq \epsilon$.

   (b) Hausdorff: $\exists T$, such that $H(A, T(B)) \leq \epsilon$, where
   $H(A, B) = max(h(A, B), h(B, A))$,
   $h(A, B) = \underset{a \in A}{max}\ \underset{b \in B}{min}||a - b||$

   (c) Distance difference: there exists a permutation of B, B', such that:
   $\forall(i, j)\ ||a_i - a_j| - |b'_i - b'_j|| \leq \epsilon$.

   Show relations between these measures, i.e. whether measure (m) implies measure (m'), for the same pair (A, B) and the same ϵ.

2. A nonnegative function g(x,y) describing the "distance" between neighboring points for a given set is a metric if it satisfies the following properties:

   (a) $g(x, x) = 0$.

   (b) if $g(x, y) = 0$ then $x = y$.

   (c) $g(x, y) = g(y, x)$.

   (d) Triangle inequality: $g(x, y) \leq g(x, z) + g(z, y)$.

   Show that the following distance measures are metrics (enough to show the triangle inequality):
   A=$\{a_i\}_1^n$, B=$\{b_i\}_1^n$,

   (a) $RMSD(A, B) = \sqrt{\frac{\sum_{i=1}^n |a_i - b_i|^2}{n}}$.

   (b) $RMSD_{opt}(A, B) = \min_T \sqrt{\frac{\sum_{i=1}^n |a_i - T(b_i)|^2}{n}}$
   (hint: use the fact that $RMSD$ is a metric).

3. We want to devise a fast method to support the following operations.

   (a) Create a matching set S and compute $f(S)$. Given two sets of 3-D points $\{u_i\}_1^n$ and $\{v_i\}_1^n$ a matching set S is defined as a correspondence set $S = \{(u_i, v_i)\}_1^n$. A scoring $f(S)$ is defined as:

$$f(S) = \min_T \sqrt{\frac{\sum_{i=1}^n |Tu_i - v_i|^2}{n}}$$

$$(Tu_i = Ru_i + a)$$

   (this operation is exactly the same as explained in the class)

   (b) Given two matching sets S_1 and S_2 create a joined matching set $S_3 = S_1 \cup S_2$ (assume that $S_1 \cap S_2 = \emptyset$) and compute $f(S_3)$.

   Explain how to support the second operation so its time complexity is only $O(1)$. After each operation you are allowed to store an additional information.

4. (*) Extend the algorithm, given in the class, to solve the optimal LCP for rotations in R^3. Give the time complexity.