

Structural Bioinformatics 2003-04 Semester B

Assignment 3.

Mar 23

1. Given two equal size point sets $A = \{a_i\}_1^n$ and $B = \{b_i\}_1^n$ consider three similarity measures:
 - (a) Bottleneck (L_∞): there exists a permutation of B , B' , and there exists a transformation T , such that:

$$\max_i |a_i - T(b'_i)| \leq \epsilon.$$
 - (b) Hausdorff: $\exists T$, such that $H(A, T(B)) \leq \epsilon$, where

$$H(A, B) = \max(h(A, B), h(B, A)),$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$
 - (c) Distance difference: there exists a permutation of B , B' , such that:

$$\forall (i, j) \quad ||a_i - a_j| - |b'_i - b'_j|| \leq \epsilon.$$

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) $RMSD(A, B)$.
 - (b) $RMSD_{opt}(A, B)$ (hint: use the fact that $RMSD$ is a metric).
3. Devise an algorithm that answers in $O(n \log(n))$ time whether two point sets in R^2 of size n are congruent (i.e. exact congruence under rigid (translation + rotation) transformation). Assume that points are in general position, i.e. without degeneracies, e.g. no 3 points lie on the same line and so on.

4. 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 $S_3 = S_1 \cup S_2$ 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.

5. Given two ordered point sets $(u_1, \dots, u_n)$ and $(v_1, \dots, v_m)$ and ϵ give an algorithm that detects a largest alignment which preserves the sequence order, i.e. the goal is to detect a maximal cardinality set $S = \{(u_{i_k}, v_{j_k})\}_{k=1}^t$, such that $\forall k \|u_{i_k} - v_{j_k}\| \leq \epsilon$ and matching indices are in ascending order $i_{k+1} > i_k$ and $j_{k+1} > j_k$.
(!!! no optimization on transformations is needed, the points are fixed in 3D).
6. The same as in (2) but alignment doesn't preserve the sequence order (sequence order independent alignment).
7. Given two sets of 3-D points $\{u_i\}_1^n$ and $\{v_i\}_1^n$ assume that the minimal distance between points in each set is strictly more than $\epsilon * 2$. Also assume that the diameter (maximal distance between two points) is bounded by D . Give a linear time algorithm (linear in the number of points and $(D/\epsilon)^3$) that detects the sequence order independent alignment of maximal size. The distance between the matched points should be less than ϵ . (no optimization on transformations is needed, the points are fixed in 3D).