

# Algorithms for Biological Sequence Analysis (Midterm #2)

Instructor: Kun–Mao Chao

TA: Yi-Ching Chen

December 18, 2007

**Problem 1 (30%):** In this problem, you are given a number sequence  $A = \langle 8, 3, 5, 7, 9, 6, 8, 5, 9, 6 \rangle$ .

- (a) (10%): Draw a left-skew decomposition for every prefix of  $A$ .
- (b) (8%): Use this example sequence to explain why the left-skew decomposition can be used to find the maximum-average segment ending at each index of  $A$ .
- (c) (7%): Use this example sequence to explain why the left-skew decompositions for all prefixes can be done in linear time.
- (a) (5%): Draw a right-skew decomposition for every suffix of  $A$ .

**Problem 2 (15%):** Given a number sequence  $A = \langle a_1, a_2, \dots, a_n \rangle$ , design a linear-time algorithm for finding the nearest smaller element for every element of the sequence.

**Problem 3 (10%):** Give an algorithm that solves the RMQ (Range Minima Query) problem in  $O(n \log n)$ -preprocessing time and  $O(1)$ -query time, where  $n$  is the length of the number sequence. (Hint: For each index, compute the minima of the segments starting from it which are of lengths  $1, 2, 2^2, 2^3, \dots$ )

**Problem 4 (15%):** Let  $X$  be an algorithm for solving the RMSQ (Range Maximum-Sum Segment Query) problem in  $O(g(n))$ -preprocessing time and  $O(1)$ -query time, where  $n$  is the length of the number sequence. Show that  $X$  can be used to solve the RMQ (Range Minima Query) problem for a number sequence of length  $n$  in  $O(g(n) + n)$ -preprocessing time and  $O(1)$ -query time.

**Problem 5 (30%):** The goal of the International HapMap Project is to develop a haplotype map of the human genome, the HapMap, which will describe the common patterns of human DNA sequence variation. (<http://www.hapmap.org/abouthapmap.html>)

- (a) (10%): Define the haplotype inference problem?
- (b) (5%): Give an Integer Quadratic Programming formulation for haplotype inference.
- (d) (5%): What are tag SNPs?
- (e) (5%): Use an example to show how to reduce the tag SNP problem to the minimum test set problem introduced in class.
- (e) (5%): Use an example to describe the idea of LD bins.