TECHNISCHE UNIVERSITEIT DELFT
Faculteit Elektrotechniek, Wiskunde en Informatica

TU Delft

Test Exam IN4301 Advanced Algorithms

January 2010

- Use a separate sheet for each question.
- This is an closed book examination with 6 questions worth of 99 points in total.
- Your mark for this exam will be the number of points divided by 10.
- If you have at least a 5.0 for both this exam and the practical work, your final mark for this course is the average of both.
- Use of book, readers, notes, and slides is not allowed.
- Use of (graphical) calculators is not permitted.
- Specify your name, student number and degree program, and indicate the total number of submitted pages on the first page.
- Write clearly, use correct English, and avoid verbose explanations. Giving irrelevant information may lead to a reduction in your score.
- This exam covers Chapters 10 and 11 of Kleinberg, J. and Tardos, E. (2005). *Algorithm Design*, all information on the slides of the course, and the papers as described in the study guide.
- The total number of pages of this exam is 2 (excluding this front page).

1. Consider the problem of placing a minimum number of broadcasting stations in a communication network of cities in such a way that each city either has a broadcasting station or is directly connected to a city with a broadcasting station. This problem can be modeled as the problem of finding a dominating set of minimum size in a graph G.
   (a) (3 points) Give the definition of a dominating set S in a graph $G = (V, E)$.
   (b) (6 points) To solve this problem we could try to apply the technique of pruning the search tree. Give an example of how to construct such a search tree for this problem, and explain how to compute the run-time of such a suggestion.
   (c) (6 points) In case G is a tree, we can efficiently solve this problem with a dynamic programming approach. Describe the main steps of this approach, illustrated by an example (preferably for the dominating set problem).
   (d) (4 points) When G is not a tree, but looks a bit like a tree, a similar method based on dynamic programming can be used using a tree decomposition of G. Give a precise definition of a tree decomposition of a graph G.
   (e) (6 points) Often it is useful to apply kernelization to solve this problem. Give a precise definition of when a decision problem with input (I, k) is kernelizable, where I is the problem instance and k is a parameter. Also, give an example of a rule that can contribute to obtaining a kernel of the minimum dominating set problem.

2. In the KNAPSACK problem, we have a knapsack of volume V and a collection of n objects whose volumes are $v_1 \ldots v_n$ and whose costs are $c_1, ..., c_n$. Your task is to select items to pack in your knapsack so that the total cost of those items is maximized, subject to the constraint that the total volume of the selected items does not exceed V.
   (a) (4 points) It seems reasonable in selecting items to base the selection upon the ratio $\frac{c_i}{v_i}$ of cost to volume. Specify a greedy algorithm based on this principle
   (b) (5 points) Show by giving an example with 3 items that your greedy algorithm does not always provide an optimal solution to the Knapsack problem.
   (c) (5 points) Present the definition of an approximation ratio to measure the approximation quality of the greedy algorithm.
   (d) (2 points) Explain what the difference is between the approximation ratio r_A of an approximation algorithm A for P and the approximation threshold r of P.
   (e) (3 points) What is a *tight example* for an approximation algorithm with a given approximation ratio r_A?
   (f) (6 points) Discuss the gap introducing reduction technique presented in class for showing that a given problem cannot be approximated within a certain ratio r (<10 lines).

3. Given a primal Linear Programming problem P. Call D its dual.
   (a) (3 points) If P has n variables and m constraints (apart from the natural non-negativity constraints) how many variables has D and how many constraints?
   (b) (3 points) Suppose we add a new constraint to P. What happens with D?
   (c) (3 points) If P is a maximization problem what is the orientation of the optimization in D?
   (d) (3 points) If only the coefficients for the second variable change throughout the functions in P, what changes in D?

4. P is changed into a Boolean problem PB, that is, the variables now are assumed to be 0,1.
   (a) (3 points) If we relax PB again into a Linear Program P(rel), how are P and P(rel) related?
   (b) (3 points) If we solve PB with branch and bound, explain why the process terminates.
   (c) (3 points) If Q is the semi-definite relaxation of PB, what are the dimensions of the matrix variables in Q?
   (d) (3 points) If we solve Q, does the answer provide an upper bound or a lower bound for PB?

5. ***Estimation-of-Distribution Algorithms*** Consider the optimization problem "bitcounting" for four variables, i.e. $f(\vec{x}) = x_0 + x_1 + x_2 + x_3$. Alice wants to solve this problem using an EDA and the univariate factorization. Bob wants to solve this problem using an EDA and the factorization $P(\vec{X}) = P(X_0, X_1)P(X_2, X_3)$. Given is the following initial population:

0	0	1	0
1	0	0	0
0	1	1	0
1	1	0	0
0	0	1	1

   (a) (3 points) What is the definition of the univariate factorization for l random variables?
   (b) (5 points) Give maximum-likelihood estimates for the distributions of Alice and Bob for the initial population.
   (c) (3 points) What is the definition of likelihood?
   (d) (6 points) Bob says that his EDA is superior to that of Alice because the likelihood of his distribution is always at least that of Alice's distribution because Alice's distribution doesn't consider combinations of variables. Prove Bob wrong by giving a population of 5 solutions from which it is impossible for Bob's EDA to find the optimal solution to the bitcounting problem, but Alice's EDA does not have zero probability.

6. ***Multi-Objective Optimization***
   (a) (4 points) In multi-objective optimization, what is a performance indicator and give an example of a typical performance indicator.
   (b) (4 points) What is the definition of Pareto dominance for two solutions x and y?

End of test