CSOR W4246–Fall, 2022
Eleni Drinea

Homework 4 Theoretical (110 points)

Out: Wednesday, November 23, 2022
Due: 11:59pm, Tuesday, December 6, 2022 — HARD DEADLINE

Homework Instructions.

1. For all algorithms that you are asked to "give" or "design", you should

   - Describe your algorithm clearly in English.
   - Give pseudocode.
   - Argue correctness, even if you don't give an entirely formal proof.
   - Give the best upper bound that you can for the running time.

   You are also encouraged to analyze the space required by your algorithm but we will not remove marks if you don't, unless the problem explicitly asks you to analyze space complexity.

2. If you give a DP algorithm, you should follow the instructions in `hw2-theoretical`.

3. If you give a reduction, you should do so as we did in class, that is

   (a) Give the inputs to the two problems.

   (b) Describe in English the reduction transformation and argue that it requires polynomial time. (You do not need to give pseudocode.)

   (c) Prove carefully equivalence of the original and the reduced instances.

4. You should submit your assignment as a **pdf** file on Gradescope. Other file formats will not be graded, and will automatically receive a score of 0.

5. I recommend you type your solutions using LaTeX. For every assignment, you will earn 5 extra credit points if you type your solutions using LaTeX or other software that prints equations and algorithms neatly. If you do not type your solutions, make sure that your handwriting is very clear and that your scan is high-quality.

6. **You should not use any external resources for this homework.** Failure to follow this instruction might have a negative impact on your performance in the second exam and interviews. For the same reason, **you should avoid collaborating** with your classmates unless you have thought through the problems on your own for a long time and are unable to make any further progress. I also encourage you to work on all the recommended exercises.

7. You should write up the solutions **entirely on your own**. Collaboration is limited to discussion of ideas only and you should adhere to the department's academic honesty policy (see the course syllabus). Similarity between your solutions and solutions of your classmates or solutions posted online will result in receiving a 0 in this assignment and possibly further disciplinary actions. You should list your collaborators on your write-up.

Homework Problems

1. (20 points) In the `MAXSAT` problem, the input is a `SAT` formula ϕ with m clauses over n variables and the output is a truth assignment that maximizes the number of satisfied clauses, as well as that number.

   State the decision version of `MAXSAT` and show that `MAXSAT(D)` is $\mathcal{NP}$-complete.

2. (25 points) Suppose you had a polynomial-time algorithm A for `SAT`. That is, on input a formula ϕ in CNF with m clauses over n variables, A runs in polynomial time and answers **yes** if and only if ϕ is satisfiable, that is, there is a truth assignment to the n variables that satisfies ϕ.

   Design a polynomial-time algorithm B that takes as input a formula ϕ and returns a satisfying truth assignment for ϕ if one exists, or **no** otherwise.

   For example, on input $\phi = (y_1 \vee \neg y_2) \wedge (\neg y_1 \vee y_2) \wedge (\neg y_1 \vee \neg y_2 \vee y_3)$, B could return $y_1 = y_2 = y_3 = 1$. (Note that, if a formula is satisfiable, there may exist more than one satisfying truth assignments; B should return one of them.)

3. (30 points) Let `X` be the following problem.

   - **Input:** A binary $m \times n$ matrix A. This is the matrix of customer purchases maintained by a large store that has m customers and n products; entry $A_{ij} = 1$ if customer i has purchased product j, otherwise, $A_{ij} = 0$.
   - **Output:** A maximum subset of *orthogonal* customers; two customers are called *orthogonal* if they did not purchase any products in common.

   State the decision version of X and prove that it is $\mathcal{NP}$-complete.

4. (35 points) Let `Y` be the following problem.

   - **Input:** A set F of m fire stations that we can potentially operate, and a set H of n homes. For each fire station $i \in F$ and each home $j \in H$, there is a cost $a_{ij} > 0$ if home j is assigned to fire station i. Further, there is a cost $f_i > 0$ for operating fire station i. A home can only be assigned to an operating fire station.
   - **Output:** A subset $F' \subseteq F$ of fire stations to operate and an assignment of homes to fire stations in F' such that
     (a) every home in H gets assigned to some fire station in F'; and
     (b) the sum of the following two costs is minimized: (i) the total cost of assigning homes to operating fire stations in F'; and (ii) the total cost of operating the fire stations in F'.

   State the decision version of Y and prove that is it $\mathcal{NP}$-complete.

RECOMMENDED exercises: do NOT return, they will not be graded.)

1. *(Using reductions to prove $\mathcal{NP}$-completeness)*

   (a) A *clique* in an undirected graph $G = (V, E)$ is a subset S of vertices such that *all* possible edges between the vertices in S appear in E. Computing the maximum clique in a network (or the number of cliques of at least a certain size) is useful in analyzing social networks, where cliques corresponds to groups of people who all know each other.

   State the decision version of the above maximization problem and show that it is $\mathcal{NP}$-complete. *Hint: reduction from* `Independent Set`.

   (b) We say that G is a *subgraph* of H if, by deleting certain vertices and edges of H we obtain a graph that is, up to renaming of the vertices, identical to G.

   The following problem has applications, e.g., in pattern discovery in databases and in analyzing the structure of social networks.

   `Subgraph Isomorphism:` Given two undirected graphs G and H, determine whether G is a subgraph of H and if so, return the corresponding mapping of vertices in G to vertices in H.

   Show that `Subgraph Isomorphism` is $\mathcal{NP}$-complete.

   (c) Similarly, consider the following problem.

   `Dense Subgraph:` Given a graph G and two integers a and b, find a set of a vertices of G such that there are at least b edges between them.

   Show that `Dense Subgraph` is $\mathcal{NP}$-complete.

2. You are asked to assist in the following crisis event.

   Due to large scale flooding, there is a set of n injured people distributed across a region that need to be rushed to hospitals. There are k hospitals in the region, and each of the n people needs to be brought to a hospital that is within a half-hour's driving time of their current location (so different people will have different options for hospitals, depending on where they are right now). However you cannot overload any single hospital; instead, every hospital must receive at most $\lceil n/k \rceil$ people.

   Give an efficient algorithm for this problem.

3. Suppose you are given n cities and a set of non-negative distances d_{ij} between pairs of cities.

   (a) Give an $O(n^2 2^n)$ dynamic programming algorithm to solve this instance of **TSP**; that is, compute the cost of the optimal tour and output the actual optimal tour.

   (b) What are the space requirements of your algorithm?

   Hint: Let $V = \{1, \ldots, n\}$ be the set of cities. Consider progressively larger subsets of cities; for every subset S of cities including city 1 and at least one other city, compute the shortest path that starts at city 1, visits all cities in S and ends up in city j, for every $j \in S$.