CS-3510: Design and Analysis of Algorithms

NP Completeness I

Instructor: Shahrokh Shahi

College of Computing Georgia Institute of Technology Summer 2022



CS-3510: Design and Analysis of Algorithms | Summer 2022

NP-Completeness

- So far, we have seen a lot of good news!
 - Problems can be solved quickly/efficiently (i.e., linear time, or at least a time that is some small polynomial function of the input size)
- NP-completeness is a form of bad news!
 - There exist many important problems that cannot be solved quickly.
- NP-complete problems really come up all the time!



NP-Completeness

- Why should we care?
 - Knowing that they are hard lets you stop beating your head against a wall trying to solve them!
 - **Restrict** the problem: find special restrictions/variants to the problem for which there is a polynomial time algorithm.
 - Use a heuristic: come up with a method for solving a reasonable fraction of the common cases.
 - Solve approximately: come up with a method that finds solutions provably close to the optimal.
 - Use an exponential time solution: if you really have to solve the problem exactly and stop worrying about finding a better solution.



Optimization vs. Decision Problems

Decision problems

- Given <u>an input and a question regarding a problem</u>, determine if the answer is <u>yes or no</u>
- Optimization problems
 - Find a solution with the "best" value

• Optimization problems can be cast as decision problems that are easier to study



Optimization vs. Decision Problems

- Casting optimization problems as decision problems
- Ex.
 - <u>Problem</u>: shortest path problem in unweighted graphs
 - <u>Optimization problem</u>: Find a path between u and v that uses the <u>fewest</u> edges
 - <u>Decision problem</u>: Does a path exist from u to v consisting of <u>at most k</u>edges?



- Class P consists of [decision] problems that are solvable in polynomial time
- Recall from the first lecture:
 - [slide #36] <u>Polynomial time</u> \rightarrow Running time is $O(n^k)$ for some constant k > 0.
 - Examples
 - Linear search O(n)
 - Dynamic programming solutions $(O(n), O(n^2), O(n^3), ...)$
 - Sorting (O(n²), O(nlogn))
 - Divide-and-conquer solutions
 - Graph algorithms O(n+m), O(mlogn), ...
 - Non-polynomial time $\rightarrow O(2^n), O(a^n), O(n!), O(n^n), ...$



- Class P consists of [decision] problems that are solvable in polynomial time
- Recall from the first lecture:
 - [slide #36] <u>Polynomial time</u> \rightarrow Running time is $O(n^k)$ for some constant k > 0.
 - Examples
 - Linear search O(n)
 - Dynamic programming solutions $(O(n), O(n^2), O(n^3), ...)$
 - Sorting (O(n²), O(nlogn))
 - Divide-and-conquer solutions
 - Graph algorithms O(n+m), O(mlogn), ...

Problems in P are Considered/called tractable

Problem not in P are intractable

• Non-polynomial time $\rightarrow O(2^n), O(a^n), O(n!), O(n^n), ...$



- Class P consists of [decision] problems that are solvable in polynomial time
 - Problems in P are Considered/called tractable
 - Problem not in P are intractable
- Note this does not mean that non-polynomial algorithms are always worst than polynomial algorithms!
 - $O(n^{100...0000})$ technically tractable (polynomial time), but practically impossible.
 - $O(n^{\log \log \log n})$ technically intractable, but practiclly easy to solve.
- Recall the "asymptotic" meaning of running time.

• First of all: **NP does NOT stand for not-P!**

NP = <u>N</u>ondeterministic <u>P</u>olynomial

• NP is the class of problems for which a <u>candidate solution</u> can be <u>verified</u> in <u>polynomial time</u>.

• P is a subset of NP ($P \subseteq NP$)



• First of all: **NP does NOT stand for not-P!**

NP = <u>Nondeterministic</u> <u>P</u>olynomial

• NP is the class of problems for which a <u>candidate solution</u> can be <u>verified</u> in <u>polynomial time</u>.

• P is a subset of NP ($P \subseteq NP$)



- <u>Nondeterministic</u> algorithms entail a two-stage procedure:
 - 1. Nondeterministic "guessing" stage
 - Generate randomly an arbitrary candidate solution (≡ "certificate")

2. Deterministic "verifying" stage

• Take the certificate and the instance to the problem and returns <u>YES</u> if the certificate represents a solution (verifying in polynomial time)



- <u>Nondeterministic</u> algorithms entail a two-stage procedure: Note in NP algorithms the verification step is polynomial
 - 1. Nondeterministic "guessing" stage
 - Generate randomly an arbitrary candidate solution (≡ "certificate")
 - 2. Deterministic "verifying" stage
 - Take the certificate and the instance to the problem and returns <u>YES</u> if the certificate represents a solution (verifying in polynomial time)



- <u>Nondeterministic</u> algorithms entail a two-stage procedure: But what does it mean "verifying" a candidate solution?
 - 1. Nondeterministic "guessing" stage
 - Generate randomly an arbitrary candidate solution (≡ "certificate")
 - 2. Deterministic "verifying" stage
 - Take the certificate and the instance to the problem and returns <u>YES</u> if the certificate represents a solution (verifying in polynomial time)



• Difference between <u>solving a problem</u> and <u>verifying a</u> <u>candidate solution</u>:

- <u>Solving a problem</u>: is there a path in graph G from vertex u to vertex v with at most k edges?
- <u>Verifying a candidate solution</u>: is $v_0, v_1, ..., v_m$ a path in graph G from vertex u to vertex v with at most k edges?



Class "NP" solving vs verifying

- Difference between <u>solving a problem</u> and <u>verifying a</u> <u>candidate solution</u>:
- Example:
 - A Hamiltonian cycle in an undirected graph is a cycle that visits every vertex exactly once.



- <u>Solving a problem</u>: is there a Hamiltonian cycle in graph G?
- Verifying a candidate solution: is $v_0, ..., v_m$ a Hamiltonian cycle on graph G?

Class "NP" solving vs verifying

• Example:

- A Hamiltonian cycle in an undirected graph is a cycle that visits every vertex exactly once.
- <u>Solving a problem</u>: is there a Hamiltonian cycle in graph G?
- Verifying a candidate solution: is v₀,..., v_m a Hamiltonian cycle on graph G?
- <u>Certificate</u>: A list of n nodes.
- <u>Certifier</u>: Check that the list contains each node in V exactly once, and that there is an edge between each pair of adjacent nodes in the permutation.
- <u>Conclusion</u>: HAM-CYCLE is in NP.





Class "NP" solving vs verifying

• Intuitively, solving a problem from scratch seems much harder (and more time consuming) in comparison to just verifying whether a candidate solution can solve the problem or not.

• Note if there are many candidate solutions to check, then even if each individual one is quick to check, overall, it can take a long time.



• Is P = NP?

- Mentioned earlier that any problem in P is also in NP. So, P is a subset of NP ($P \subseteq NP$)
- But the big (and open) question is whether NP \subseteq P, and so P=NP.



• Is P = NP?

- Mentioned earlier that any problem in P is also in NP. So, P is a subset of NP ($P \subseteq NP$)
- But the big (and open) question is whether $NP \subseteq P$, and so P=NP.

What does it mean?



• Is P = NP?

- Mentioned earlier that any problem in P is also in NP. So, P is a subset of NP ($P \subseteq NP$)
- But the big (and open) question is whether $NP \subseteq P$, and so P=NP.
 - It means if it is always easy to check a candidate solution, should it also be easy to find a solution?
 - Answer? Most computer scientists believe that this is false, but we do not have a proof



• Is P = NP?

- Answer? Most computer scientists believe that this is false, but we do not have a proof
- Therefore, there are two possibilities/beliefs:



NP-Complete (NPC)

- NP-complete problems are a class of "hardest" problems in NP.
- If you can solve an NP-complete problem, then you can solve all NP problems (show later).
- Hence, if any NP-complete problem can be solved in polynomial time, then all problems in NP can be, and thus P = NP.
- Precise/formal definition coming later...



Possible Worlds

• Therefore, there are two possibilities:





- Reduction from A to B is showing that we can solve A using the algorithm that solves B
- We say that problem A is easier than problem B, and
- We write $A \leq B$



- Reduction from A to B is showing that we can solve A using the algorithm that solves B
- We say that problem A is easier than problem B, and
- We write $A \leq B$



- $A \leq B$: Reduction from A to B is showing that we can solve A using the algorithm that solves B
- If we have an oracle for solving B, then we can solve A by making polynomial number of computations and polynomial number of calls to the oracle for B
- We can transform the inputs of A to inputs of B



- Before discussing further regarding reduction in NPC, note that we can also do reductions on polynomial time (poly-time) algorithms.
- Examples:
 - Transforming a given problem to a graph, and solving the problem using graph algorithms (for example, SCC)
 - Solving all-pairs shortest path problem using multiple (polynomial number of) calls to Dijkstra's algorithm



Polynomial Reductions

- Given two problems, A and B, we say that A is polynomially reducible to B, and write it as $A \leq_p B$ if:
 - 1. There exists a function f that converts the input of A to inputs of B in polynomial time

2.
$$A(i) = YES \iff B(f(i)) = YES$$



Proving Polynomial Time

- 1. Use a polynomial time reduction algorithm to transform A into B.
- 2. Run a known polynomial time algorithm for B.
- 3. Use the answer for B as the answer for A.



Implications of Polynomial-Time Reductions

- **<u>Purpose</u>**. Classify problems according to relative difficulty.
- **Design algorithms**. If $X \leq_p Y$ and Y can be solved in polynomial-time, then X can also be solved in polynomial time.
- Establish intractability. If $X \leq_p Y$ and X cannot be solved in polynomial-time, then Y cannot be solved in polynomial time.
- Establish equivalence. If $X \leq_p Y$ and $Y \leq_p X$, we use notation $X \equiv_p Y$.
- <u>Transitivity</u>. If $X \leq_p Y$ and $Y \leq_p Z$, then $X \leq_p Z$.

Reductions Strategies

- Given two problems, A and B, we say that A is polynomially reducible to B, and write it as $A \leq_p B$ if:
 - 1. There exists a function f that converts the input of A to inputs of B in polynomial time
 - 2. $A(i) = YES \iff B(f(i)) = YES$
- Reductions Strategies
 - Reduction by simple equivalence.
 - Reduction from <u>a special case to a general case</u>.
 - Reduction by encoding with gadgets.



• We want to show the problem VERTEX-COVER is polynomially reducible to the SET-COVER problem, i.e.,

$\texttt{VERTEX-COVER} \leq_{P} \texttt{SET-COVER}$

- **VERTEX-COVER** problem?
- **SET-COVER** problem?
- Reduction process?



- Vertex Cover
 - MINIMUM VERTEX COVER: Given a graph G = (V, E), find the smallest subset of vertices S ⊆ V, such that for each edge at least one of its endpoints is in S?
 - VERTEX COVER: Given a graph G = (V, E) and an integer k, is there a subset of vertices S ⊆ V such that |S| ≤ k, and for each edge, at least one of its endpoints is in S?
 - Ex. Is there a vertex cover of size ≤ 4 ? Yes.
 - Ex. Is there a vertex cover of size ≤ 3 ? No.



- Vertex Cover
 - SET COVER: Given a set U of elements, a collection S_1, S_2, \ldots, S_m of subsets of U, and an integer k, does there exist a collection of k of these sets whose union is equal to U?

• Ex. U =
$$\{1, 2, 3, 4, 5, 6, 7\}$$
 k=2

- $S_1 = \{3,7\}$
- $S_2 = \{3, 4, 5, 6\}$
- $S_3 = \{1\}$
- $S_4 = \{2, 4\}$
- $S_5 = \{5\}$
- $S_6 = \{1, 2, 6, 7\}$



- We want to show the problem VERTEX-COVER is polynomially reducible to the SET-COVER problem.
- Theorem. VERTEX-COVER \leq_P SET-COVER
- **Proof**. Given a VERTEX-COVER instance G = (V, E) and k, we construct a SET-COVER instance (U, S, k) that has a set cover of size k iff <u>G</u> has a vertex cover of size k.
- Construction.
 - Universe U = E.
 - Include one subset for each node $v \in V$: $S_v = \{e \in E : e \text{ incident to } v\}$.
 - The transformation takes linear time on the size of the VC instance.

• Construction.

- Universe U = E.
- Include one subset for each node $v \in V$: $S_v = \{e \in E : e \text{ incident to } v\}$.
- The transformation takes linear time on the size of the VC instance.



$$U = \{ 1, 2, 3, 4, 5, 6, 7 \}$$

$$S_a = \{ 3, 7 \} \qquad S_b = \{ 2, 4 \}$$

$$S_c = \{ 3, 4, 5, 6 \} \qquad S_d = \{ 5 \}$$

$$S_e = \{ 1 \} \qquad S_f = \{ 1, 2, 6, 7 \}$$

set cover instance (k = 2)



• Lemma. G = (V, E) contains a vertex cover of size k iff (U, S, k) contains a set cover of size k.

That is, $VC(i) = yes \iff SC(f(i)) = yes$

• **Proof**. (⇒)

- Let $X \subseteq V$ be a vertex cover of size k in G.
- Then, $Y = \{S_v: v \in X\}$ is a set cover of size k.



- Lemma. $VC(i) = yes \iff SC(f(i)) = yes$
- **Proof**. (\Rightarrow) VC(i) = yes \Rightarrow SC(f(i)) = yes
 - VC(i) is a yes instance \Rightarrow it has a solution; let V' \subseteq V be such a solution $|V'| \le k$, every edge has at least one end point in V'
 - Consider V' = $\{i_1, i_2, i_3, ..., i_l\}, l \le k$, and therefore, A = $\{S_{i_1}, S_{i_2}, ..., S_{i_l}\}$.
 - For the sake of contradiction assume A is not a solution to SC(f(i)):
 - The number of sets in A is $1 \le k$. Thus, it must be the case that $S_{i_1} \cup S_{i_2} \cup \ldots \cup S_{i_l} \neq U$
 - This means there is at least an edge $e \in U$ that is not in $S_{i_1} \cup S_{i_2} \cup \ldots \cup S_{i_l}$.
 - This *e* is also corresponds to an edge in VC(i), *e* = (*u*, *v*), so S_u and S_v are not in A, i.e.,
 S_u, S_v ∉ A ⇒ u, v ∉ V' (by construction of A)
 - This means e = (u, v) would not have been covered by V'
 - So, V' is not solution to VC, which is a contradiction.

- Lemma. $VC(i) = yes \iff SC(f(i)) = yes$
- **Proof.** (\Leftarrow) (VC(i) = yes \Leftarrow SC(f(i)) = yes) <u>Or</u> (SC(f(i)) = yes \Rightarrow VC(i) = yes)
 - Let $Y \subseteq S$ be a set cover of size k in (U, S, k)
 - Then, $X = \{v: S_v \in Y\}$ is a vertex cover of size k in G.





- Lemma. $VC(i) = yes \iff SC(f(i)) = yes$
- **Proof.** (\Leftarrow) (VC(i) = yes \Leftarrow SC(f(i)) = yes) <u>Or</u> (SC(f(i)) = yes \Rightarrow VC(i) = yes)
 - SC(f(i)) is a yes instance \Rightarrow it has a solution; let A = { $S_{i_1}, S_{i_2}, ..., S_{i_l}$ } be such a solution.
 - \Rightarrow $l \le k$ and $S_{i_1} \cup S_{i_2} \cup \ldots \cup S_{i_l} = U$ (by definition of SC)
 - Consider the vertex set V' = $\{i_1, i_2, i_3, ..., i_l\}$
 - For the sake of contradiction assume V' is not a solution to VC(i):
 - The number of vertices in V' is $l \le k$. Thus, it must be breaking the edge covering requirement of VC.
 - Therefore, there must be at least an edge $e \in E$ such that $u \notin V'$, $v \notin V'$
 - This implies S_u and S_v were not included in solution A.
 - By construction of f(i), $e = (u, v) \in U$, and S_u , S_v were the only sets containing e.
 - Thus, e ∉ S_{i1} ∪ S_{i2} ∪ ... ∪ S_{il}, i.e., e is not covered by the solution set A. So, A is not a solution (Contradiction)
 - \Rightarrow V' is a solution to VC(i)

NP-Completeness (Formal Definition)

• A problem Y is **NP-hard** if $X \leq_p Y$ for all $X \in \mathbf{NP}$

- A problem is NP-hard <u>if and only if</u> a polynomial-time algorithm for it implies a polynomial-time algorithm for every problem in NP
- NP-hard problems are at least as hard as any NP problem
- A problem Y is **NP-complete** if:
 - *1.* $Y \in \mathbf{NP}$
 - 2. Y is NP-hard



https://en.wikipedia.org/wiki/P_versus_NP_problem

- Establishing NP-completeness \rightarrow using "reduction"
- Once we establish the first "natural" NP-complete problem, others fall like dominoes!
- Recipe to establish <u>NP-completeness of problem Y.</u>
 - Step1. Show that Y is in NP. $(Y \in NP)$
 - Step 2. Choose <u>an</u> NP-complete problem *X*.
 - Step 3. Prove that $X \leq_p Y$ (poly-time reduction).



- Establishing NP-completeness \rightarrow using "reduction"
- Once we establish the first "natural" NP-complete problem, others fall like dominoes!
- Recipe to establish <u>NP-completeness of problem Y.</u>
 - Step1. Show that Y is in NP. $(Y \in NP)$
 - Step 2. Choose <u>an</u> NP-complete problem *X*.
 - Step 3. Prove that $X \leq_p Y$ (poly-time reduction).

Why does it work?



- Recipe to establish <u>NP-completeness of problem Y.</u>
 - Step1. Show that Y is in NP. $(Y \in NP)$
 - Step 2. Choose <u>an</u> NP-complete problem *X*.
 - Step 3. Prove that $X \leq_p Y$ (poly-time reduction).
- Justification: If X is an NP-complete problem, and Y is a problem in NP with the property that $X \leq_P Y$ then Y is NP-complete.
- Proof.
 - Let W be any problem in NP. Then, $W \leq_P X \leq_P Y$
 - By transitivity, $W \leq_P Y$.
 - Hence, Y is NP-complete.

by definition of NPC

- Recipe to establish <u>NP-completeness of problem Y.</u>
 - Step1. Show that Y is in NP. $(Y \in NP)$
 - Describe how a potential solution will be represented
 - Describe a procedure to check whether the potential solution is a correct solution to the problem instance, and argue that this procedure takes polynomial time
 - Step 2. Choose <u>an</u> NP-complete problem X.
 - Step 3. Prove that $X \leq_p Y$ (X is **poly-time reducible** to Y).
 - Describe a procedure f that converts the inputs i of X to inputs of Y in polynomial time
 - Show that the reduction is correct by showing that $X(i) = YES \iff Y(f(i)) = YES$ Note this is an "if and only if" condition, so proofs are needed for both directions.



• Important note about step 2 and 3:

- To establish NP-completeness of problem Y, we show that some other NP-complete problem X is polynomially reducable to this algorithm.
- Note the reduction is from algorithm X to Y (X \leq_P Y), not the reverse direction!



Revisit "Is P = NP?"

- <u>Theorem</u>. Suppose Y is an NP-complete problem. Y is solvable in poly-time if and only if P = NP.
- <u>Proof (\Leftarrow)</u> If P=NP then Y is in P. Hence, Y can be solved in poly-time.
- <u>Proof (\Rightarrow)</u> Suppose Y can be solved in poly-time.
 - Let X be any problem in NP. Then, we know that $X \leq_P Y$ by definition of NP-complete and Y being NP-complete problem. Then we can solve X in poly-time by solving Y in poly-time. This implies any problem X in NP is also in P, i.e., NP \subseteq P.
 - We already know $P \subseteq NP$. Thus, P=NP.



 $P \neq NP$

https://en.wikipedia.org/wiki/P_versus_NP_problem

P = NP

Examples of NPC problems

Shortest simple path

- Given a graph G = (V, E) find a **shortest** path from a source to all other vertices
- Polynomial solution: Bellman-Ford O(VE) (complexity class P)

Longest simple path

- Given a graph G = (V, E) find a **longest** path from a source to all other vertices
- NP-complete



Examples of NPC problems

• Euler tour

- G = (V, E) a connected, directed graph find a cycle that traverses **each edge** of G exactly once (may visit a vertex multiple times)
- Polynomial solution O(E)

• Hamiltonian cycle

- G = (V, E) a connected, directed graph find a cycle that visits **each vertex** of G exactly once
- NP-complete



The First NPC Problem

- The satisfiability (SAT) problem was the first problem shown to be NP-complete (Cook–Levin theorem)
- Satisfiability problem: given a logical expression Φ , find an assignment of True/False values to binary variables x_i that causes Φ to evaluate to T.
- Ex. $\Phi = x_1 \vee \neg x_2 \wedge x_3 \vee \neg x_4$



Quick Review

- Boolean variables: take on values T (or 1) or F (or 0)
- Literal: variable or negation of a variable, e.g., x_2 , $\neg x_2$

• Notation: $\neg x_2 = \overline{x_2} = \operatorname{not} x_2$

<i>x</i> ₁	<i>x</i> ₂	$\begin{array}{c} x_1 \wedge x_2 \\ (AND) \end{array}$	$(OR) x_1 \forall x_2 \\ (OR)$
Т	Т	Т	Т
Т	F	F	Т
0	Т	F	Т
0	F	F	F



The First NPC Problem

- Satisfiability problem: given a logical expression Φ , find an assignment of True/False values to binary variables x_i that causes Φ to evaluate to T.
- <u>SAT is in NP</u>: given a value assignment, check the Boolean logic of Φ evaluates to True (linear time)
- The satisfiability (SAT) problem was the first problem shown to be NP-complete (Cook–Levin theorem)









References

- The lecture slides are heavily based on the <u>suggested textbooks</u> and the corresponding published lecture notes:
 - <u>Slides by Umit Catalyurek</u>, Georgia Institute of Technology. (Based on slides by <u>Bistra Dilkina</u>, <u>Anne Benoit</u>, <u>Jennifer Welch</u>, <u>George Bebis</u>, and <u>Kevin Wayne</u>)
 - CLRS: Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. Introduction to Algorithms, Third Edition, MIT Press, 2009.
 - KT: Kleinberg, J., & Tardos, E. Algorithm design. Pearson/Addison-Wesley, 2006.

