Introduction to Algorithms and Data Structures

Tutorial 9

your tutor

University of Edinburgh

10th-14th March, 2025

Q1: SAT $\leq_P$ 3-SAT

Q1(a) Show that SAT and 3-SAT are in NP

- "solution/certificate" for a SAT instance $\phi = C_1 \wedge \ldots C_m$ will be an assignment to the logical variables $\{x_1, \ldots, x_n\}$
- Assignment can be written as binary string $b_1 \ldots b_n$ of length n (so polynomial in the size of the input formula ϕ).
- To verify $b_1 \ldots b_n$ against ϕ, verify each C_j in turn
  - For each literal $\ell = x_i$ in C_j, check whether $b_i = 1$?
    For each literal $\ell = \bar{x}_i$ in C_j, check whether $b_i = 0$?
    As long as one of these tests passes, C_j is satisfied by $b_1 \ldots b_n$.
  - It takes $O(|C_j|)$ "lookups" to check C_j.
- To check *all* C_j are satisfied takes $O\left(\sum_{j=1}^{m} |C_j|\right)$ lookups in total, which is polynomial-time, in the size of our input instance ϕ.

Hence SAT is in NP.
Same argument works for 3-SAT.

Q1: $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$

Q1(b) Show that $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$.

Instance $\Phi = C_1 \wedge \ldots \wedge C_m$ of $\textsc{Sat}$, over boolean variables $\{x_1, \ldots, x_n\}$.

Assume that no C_j includes any *complementary pair* of literals $x_i, \bar{x}_i$ (these would be trivially satisfied - just delete the clause).

Will convert Φ to an equivalent $3\text{-}\textsc{Sat}$ formula "clause by clause".
Take into account the size of each C_j.
4 cases:

- $|C_j| = 1$
- $|C_j| = 2$
- $|C_j| = 3$ ... *just leave these exactly as they are!*
- $|C_j| > 4$

Q1: $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$ cont'd.

$|C_j| = 1$: Suppose $C_j = (\ell_{j,1})$ for that specific literal $\ell_{j,1}$. Create *two* dummy variables $y_{j,1}, y_{j,2}$, then we will replace C_j by the following four clauses:

$$C_{j,1} = (\ell_{j,1} \vee y_{j,1} \vee y_{j,2}) \qquad C_{j,2} = (\ell_{j,1} \vee \overline{y_{j,1}} \vee y_{j,2})$$
$$C_{j,3} = (\ell_{j,1} \vee y_{j,1} \vee \overline{y_{j,2}}) \qquad C_{j,4} = (\ell_{j,1} \vee \overline{y_{j,1}} \vee \overline{y_{j,2}})$$

Claim: The "$\wedge$" of these 4 clauses is equivalent to $C_j = (\ell_{j,1})$.

Why? Note that *no matter what* values get assigned to $y_{j,1}, y_{j,2}$, there will be one pairing (from the options $\{y_{j,1}, \overline{y_{j,1}}\} \times \{y_{j,2}, \overline{y_{j,2}}\}$) that has *both of* the $y_{j,\cdot}$ literals fail. So this clause will force $\ell_{j,1}$ to be true, equivalent to C_j.

Q1: $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$ cont'd.

$|C_j| = 2$: Suppose $C_j = (\ell_{j,1} \vee \ell_{j,2})$ for its two specific literals. Use *one* dummy variable y_j to replace C_j by the following two clauses:

$$C_{j,1} = (\ell_{j,1} \vee \ell_{j,2} \vee y_j), C_{j,2} = (\ell_{j,1} \vee \ell_{j,2} \vee \bar{y_j}).$$

For similar reasons to the $|C_j| = 1$ case,

$$(\ell_{j,1} \vee \ell_{j,2} \vee y_j) \wedge (\ell_{j,1} \vee \ell_{j,2} \vee \bar{y_j})$$

is true in exactly the same circs as the original C_j.

Q1: $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$ cont'd.

$|C_j| = 2$: Suppose $C_j = (\ell_{j,1} \vee \ell_{j,2})$ for its two specific literals. Use *one* dummy variable y_j to replace C_j by the following two clauses:

$$C_{j,1} = (\ell_{j,1} \vee \ell_{j,2} \vee y_j), C_{j,2} = (\ell_{j,1} \vee \ell_{j,2} \vee \bar{y_j}).$$

For similar reasons to the $|C_j| = 1$ case,

$$(\ell_{j,1} \vee \ell_{j,2} \vee y_j) \wedge (\ell_{j,1} \vee \ell_{j,2} \vee \bar{y_j})$$

is true in exactly the same circs as the original C_j.

Q1: $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$ cont'd.

$|C_j| > 3$: We will add $|C_j| - 3$ new dummy variables $y_{j,2}, \ldots y_{j,|C_j|-2}$ (so 1 dummy variable if $|C_j|$ was 4, two dummy variables if $|C_j|$ was 5, ...)

Suppose $C_j = (\ell_{j,1} \vee \ell_{j,2} \vee \ldots \vee \ell_{j,|C_j|})$.

We then replace C_j with the following clauses $C_{j,1}, \ldots, C_{j,|C_j|-2}$ defined as follows:

$$C_{j,i} \;=\; \begin{cases} (\ell_{j,1} \vee \ell_{j,2} \vee y_{j,2}) & i = 1 \\ (\overline{y_{j,i}} \vee \ell_{j,i+1} \vee y_{j,i+1}) & i, 1 < i < |C_j| - 2 \\ (\overline{y_{j,|C_j|-2}} \vee \ell_{j,|C_j|-1} \vee \ell_{j,|C_j|}) & i = |C_j| - 2 \end{cases}$$

Claim: $C_{j,1} \wedge \ldots \wedge C_{j,|C_j|-2}$ is satisfiable $\Leftrightarrow$ C_j is satisfiable.

Q1: $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P$ 3-$\textsc{Sat}$ cont'd.

$$C_{j,i} \;=\; \begin{cases} (\ell_{j,1} \vee \ell_{j,2} \vee y_{j,2}) & i = 1 \\ (\overline{y_{j,i}} \vee \ell_{j,i+1} \vee y_{j,i+1}) & i, 1 < i < |C_j| - 2 \\ (\overline{y_{j,|C_j|-2}} \vee \ell_{j,|C_j|-1} \vee \ell_{j,|C_j|}) & i = |C_j| - 2 \end{cases}$$

Claim: $C_{j,1} \wedge \ldots \wedge C_{j,|C_j|-2}$ is satisfiable $\Leftrightarrow$ C_j is satisfiable.

$\Rightarrow$ direction: Suppose $C_{j,1} \wedge \ldots \wedge C_{j,|C_j|-2}$ is satisfiable.
Let $i*$ be the first such that $y_{j,i*} = 0$. If $i* = 2$, then $\ell_{j,1} \vee \ell_{j,2}$ must be true ... if $2 < i* \leq |C_j| - 2$, then $\ell_{j,i*}$ must be true ... if $y_{j,i} = 1$ for all i, then $\ell_{j,|C_j|-1} \vee \ell_{j,|C_j|}$ must be true.

$\Leftarrow$ direction: Choose any literal $\ell_{j,i*}$ of C_j made true by the satisfying assignment.
Set $y_{j,i} = 1$ for $i < i*$ and $y_{j,i} = 0$ for $i \geq i*$, this satisfies all $C_{j,i}$

Note: Need "fresh" dummy variables for each C_j with $|C_j|$.

Q1: $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$

Q1(b) $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$ cont'd.

We argued "equivalence" on a clause-by-clause basis.
By using "fresh" dummy variables for each C_j, this equivalence extends to the entire logical formula (no interdependence except among the original variables).

We have created an instance of $3\text{-}\textsc{Sat}$ of total size at most 12 times our original problem (if size is counted in "total number of literals").

Each of the conversions to 3-CNF are methodological, and can be done in time linear in the size of C_j.

Hence $\textsc{Sat} \leq_P 3\text{-}\textsc{Sat}$.

Q3 (a): derandomization for 3-SAT

Want specific assignment to satisfy $Y \geq \frac{7}{8}9$, ie at least 8, of the clauses in the following Φ.

$$\begin{aligned}\Phi \;=\; & (x_1 \vee x_2 \vee x_3) \wedge (\bar{x_1} \vee \bar{x_2} \vee \bar{x_3}) \wedge (\bar{x_1} \vee x_2 \vee x_3) \wedge \\ & (x_1 \vee \bar{x_2} \vee \bar{x_3}) \wedge (x_1 \vee \bar{x_2} \vee x_4) \wedge (x_2 \vee x_3 \vee \bar{x_4}) \wedge \\ & (\bar{x_1} \vee \bar{x_3} \vee \bar{x_4}) \wedge (\bar{x_2} \vee \bar{x_3} \vee x_4) \wedge (\bar{x_1} \vee x_3 \vee x_4).\end{aligned}$$

variable x_1: Two options: $x_1 \leftarrow 0$ and $x_1 \leftarrow 1$.

To compute $Exp_0 = \mathrm{E}[Y \mid x_1 \leftarrow 0]$, the expected number of satisfied clauses *conditional on x_1 being 0*, we notice that Φ has

- 4 clauses containing the negative literal $\bar{x_1}$
- 3 clauses containing the positive literal x_1
- 2 clauses not involving this variable at all.

Q3 (a): derandomization for 3-SAT

$$
\begin{aligned}
\Phi \quad = \quad & (x_1 \vee x_2 \vee x_3) \wedge (\bar{x_1} \vee \bar{x_2} \vee \bar{x_3}) \wedge (\bar{x_1} \vee x_2 \vee x_3) \wedge \\
& (x_1 \vee \bar{x_2} \vee \bar{x_3}) \wedge (x_1 \vee \bar{x_2} \vee x_4) \wedge (x_2 \vee x_3 \vee \bar{x_4}) \wedge \\
& (\bar{x_1} \vee \bar{x_3} \vee \bar{x_4}) \wedge (\bar{x_2} \vee \bar{x_3} \vee x_4) \wedge (\bar{x_1} \vee x_3 \vee x_4).
\end{aligned}
$$

variable x_1: Two options: $x_1 \leftarrow 0$ and $x_1 \leftarrow 1$.
If we set $x_1 \leftarrow 0$, we satisfy all clauses with $\bar{x_1}$, we delete the x_1 literal from the clauses containing it (probability drops to $\frac{3}{4}$) ... the two other clauses still have probability $\frac{7}{8}$. This gives

$$\mathrm{E}[Y \mid x_1 \leftarrow 0] \;=\; 4 + 3 \cdot \tfrac{3}{4} + 2 \cdot \tfrac{7}{8} \;=\; 8.$$

To compute $Exp_1 = \mathrm{E}[Y \mid x_1 \leftarrow 1]$, the circs for positive literals (3) and negative literals (4) are reversed, hence the value Exp_1 can be computed as

$$\mathrm{E}[Y \mid x_1 \leftarrow 1] \;=\; 3 + 4 \cdot \tfrac{3}{4} + 2 \cdot \tfrac{7}{8} \;=\; 7.75.$$

Verdict: $x_1 \leftarrow 0$

Q3 (a): derandomization for 3-SAT

$$
\begin{aligned}
\Phi' \quad = \quad & (\cancel{x_1 \vee}\ x_2 \vee x_3) \wedge (\bar{x_1} \vee \bar{x_2} \vee \bar{x_3}) \wedge (\bar{x_1} \vee x_2 \vee x_3) \wedge \\
& (\cancel{x_1 \vee}\ \bar{x_2} \vee \bar{x_3}) \wedge (\cancel{x_1 \vee}\ \bar{x_2} \vee x_4) \wedge (x_2 \vee x_3 \vee \bar{x_4}) \wedge \\
& (\bar{x_1} \vee \bar{x_3} \vee \bar{x_4}) \wedge (\bar{x_2} \vee \bar{x_3} \vee x_4) \wedge (\bar{x_1} \vee x_3 \vee x_4)
\end{aligned}
$$

4 clauses already satisfied. Only 5 “active” clauses remain.

Variable x_2:
$x_2 \leftarrow 0$: we have $\bar{x_2}$ in 3 active clauses (satisfied by $x_2 \leftarrow 0$), x_2 in one length-2 clause (length-1 after $x_2 \leftarrow 0$), one length-3 clause (becomes length 1). Hence

$$
\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0] \;=\; \mathbf{4} + 3 + 1 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{3}{4} \;=\; 8.25
$$

(the $\mathbf{4}$ is from the previously satisfied clauses).
$x_2 \leftarrow 1$: For $\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 1]$, just observe

$$
\mathrm{E}[Y \mid x_1 \leftarrow 0] \;=\; \frac{\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0] + \mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 1]}{2}.
$$

By $\mathrm{E}[Y \mid x_1 \leftarrow 0] = 8$ and $\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0] = 8.25$, we know $\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 1] < \mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0]$.
Verdict: $x_2 \leftarrow 0$

Q3 (a): derandomization for 3-SAT

$$\begin{aligned}\Phi' \quad = \quad & (\cancel{x_1 \vee x_2 \vee}\ x_3) \wedge (\bar{x_1} \vee \bar{x_2} \vee \bar{x_3}) \wedge (\bar{x_1} \vee x_2 \vee x_3) \wedge \\ & (\cancel{x_1 \vee}\ \bar{x_2} \vee \bar{x_3}) \wedge (\cancel{x_1 \vee}\ \bar{x_2} \vee x_4) \wedge (\cancel{x_2 \vee}\ x_3 \vee \bar{x_4}) \wedge \\ & (\bar{x_1} \vee \bar{x_3} \vee \bar{x_4}) \wedge (\bar{x_2} \vee \bar{x_3} \vee x_4) \wedge (\bar{x_1} \vee x_3 \vee x_4)\end{aligned}$$

7 clauses satisfied, just 2 active clauses.

Variable x_3: Both remaining clauses have x_3 as a positive literal. Hence:

$$\begin{aligned}\mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0, x_3 \leftarrow 1] \quad &= \quad \mathbf{7} + 2 \ = \ 9 \\ \mathrm{E}[Y \mid x_1 \leftarrow 0, x_2 \leftarrow 0, x_3 \leftarrow 0] \quad &= \quad \mathbf{7} + 0 + 1 \cdot \tfrac{1}{2} \ = 7.5\end{aligned}$$

Verdict: $x_3 \leftarrow 1$

Choose either value for x_4, doesn't matter which.
Overall assignment is $x_1 \leftarrow 0, x_2 \leftarrow 0, x_3 \leftarrow 1, x_4 \in \{0, 1\}$.

Q3 (b): derandomization for general SAT

Suppose we want to do the same process for general CNF?

Obs 1: $\mathrm{E}[Y] = \frac{7}{8}m$ no longer fits.

- Can use *linearity of expectation* for $\mathrm{E}[Y]$, but clauses have variable length
  - Let m_k be the number of clauses of length k in Φ (each $k = 1, \ldots, n$)
  - Then $\mathrm{E}[Y]$ is

$$\mathrm{E}[Y] \; = \; \sum_{k=1}^{n} m_k (1 - \tfrac{1}{2^k})$$

Obs 2: Can do a derandomization to achieve at least $\mathrm{E}[Y]$ satisfied clauses.

- While calculating the $\mathrm{E}[Y \mid x_1 = b_1, \ldots, x_j = b_j]$ values, we will be working with a Φ' formula with clauses of varying sizes.
- Probability is $(1 - \frac{1}{2^k})$ for clauses with $k > 3$ active literals.
- Calculations are still feasible.

Won't necessarily satisfy $\geq \frac{7}{8}m$ clauses, because $\mathrm{E}[Y]$ might have been smaller (especially if Φ has a lot of 1-literal and/or 2-literal clauses).

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

proof: By definition, the set $\mathcal{I}$ is an Independent set if (and only if) for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

proof: By definition, the set $\mathcal{I}$ is an Independent set if (and only if) for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.

This is the case if and only if for every $(u, v) \in E$, at least one of u, v is *not* in $\mathcal{I}$.

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

proof: By definition, the set $\mathcal{I}$ is an Independent set if (and only if) for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.

This is the case if and only if for every $(u, v) \in E$, at least one of u, v is *not* in $\mathcal{I}$.

This is the case if and only if if for every $(u, v) \in E$, either $u \in V \setminus \mathcal{I}$ or $v \in V \setminus \mathcal{I}$.

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

proof: By definition, the set $\mathcal{I}$ is an Independent set if (and only if) for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.

This is the case if and only if for every $(u, v) \in E$, at least one of u, v is *not* in $\mathcal{I}$.

This is the case if and only if if for every $(u, v) \in E$, either $u \in V \setminus \mathcal{I}$ or $v \in V \setminus \mathcal{I}$.

This is the case (by definition) if and only if $V \setminus \mathcal{I}$ is a Vertex Cover for G.

Q4 (a): Vertex Cover and Independent Set

- $\mathcal{I}$ is an *Independent Set* of G if for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.
- $\mathcal{K}$ is a *Vertex Cover* of G if for every $e = (u, v) \in E$, either $u \in \mathcal{K}$ or $v \in \mathcal{K}$.

proof: By definition, the set $\mathcal{I}$ is an Independent set if (and only if) for every $u \in \mathcal{I}, v \in \mathcal{I} \setminus \{u\}$, that $(u, v) \notin E$.

This is the case if and only if for every $(u, v) \in E$, at least one of u, v is *not* in $\mathcal{I}$.

This is the case if and only if if for every $(u, v) \in E$, either $u \in V \setminus \mathcal{I}$ or $v \in V \setminus \mathcal{I}$.

This is the case (by definition) if and only if $V \setminus \mathcal{I}$ is a Vertex Cover for G.

Q4 (a): Vertex Cover and Independent Set

Implications for the two decision problems:

$$\begin{gathered} G \textit{ has an Independent Set } \mathcal{I} \textit{ of size } |\mathcal{I}| \geq k \\ \Leftrightarrow G \textit{ has a Vertex Cover } V \setminus \mathcal{I} \textit{ such that } |\mathcal{I}| \geq k \\ \Leftrightarrow G \textit{ has a Vertex Cover } \mathcal{K} \textit{ such that } |\mathcal{K}| \leq (n-k). \end{gathered}$$

- Very straightforward "reduction" from Independent Set to Vertex Cover, *and vice versa*.
- Really simple reduction, graph doesn't change, just flip between k and $n-k$ for size parameter.

Therefore, Independent Set is NP-complete ⇔ Vertex Cover is NP-complete.

Rare to have "reductions" which work in both directions.

Q4 (b): Vertex Cover and Independent Set

Implications for the approximation problems:

Hypothesis: We have an *approximation algorithm* for Vertex Cover with approximation ration of α, for $\alpha > 1$.
So the algorithm will return ℓ satisfying $\ell \leq \alpha \cdot OPT_{VC}(G)$, where $OPT_{VC}(G)$ is the optimum/minimum size of a VC for G.

What does $n - \ell$ mean in relation to the maximum Independent Set for G? (remember maximum has size $n - OPT_{VC}(G)$). We **know** that

$$n - \ell \;\geq\; n - \alpha \cdot OPT_{VC}(G) \;=\; (n - OPT_{VC}(G)) - (\alpha - 1)OPT_{VC}(G)$$

We **would like** $n - \ell \geq \frac{1}{\beta}(n - OPT_{VC}(G))$ for some $\beta > 1$

But imagine α of the Vertex Cover is $\alpha = 2$. Then taking $n - \ell$ for Independent Set gives the bound

$$n - \ell \;\geq\; (n - OPT_{VC}(G)) - OPT_{VC}(G)$$

However, there may be graphs where $OPT_{VC}(G)$ is $n/2$ or even greater.
So $n - \ell$ may be arbitrarily close to 0.
Problem is the $-$ in the conversion between the two problems: subtraction does not preserve approximation.