Randomised Algorithms

Lecture 10: Approximation Algorithms: Set-Cover and MAX-CNF

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2023

UNIVERSITY OF CAMBRIDGE

Outline

Weighted Set Cover

MAX-CNF

Appendix: An Approximation Algorithm of TSP (non-examin.)

The Weighted Set-Cover Problem

Set Cover Problem

- Given: set X and a family of subsets $\mathcal{F}$, and a cost function $c : \mathcal{F} \to \mathbb{R}^+$
- Goal: Find a minimum-cost subset $\mathcal{C} \subseteq \mathcal{F}$

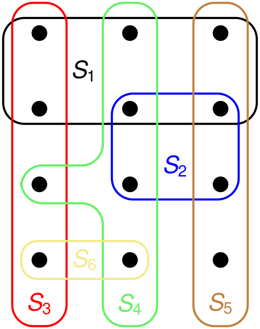
$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

The Weighted Set-Cover Problem

Set Cover Problem

- Given: set X and a family of subsets $\mathcal{F}$, and a cost function $c : \mathcal{F} \to \mathbb{R}^+$
- Goal: Find a minimum-cost subset $\mathcal{C} \subseteq \mathcal{F}$ s.t. $X = \bigcup_{S \in \mathcal{C}} S.$

Sum over the costs of all sets in $\mathcal{C}$



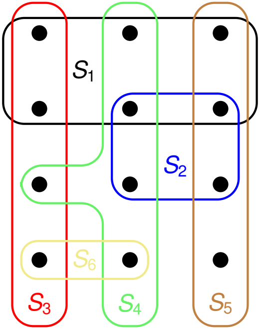
The Weighted Set-Cover Problem

Set Cover Problem

- Given: set X and a family of subsets $\mathcal{F}$, and a cost function $c : \mathcal{F} \to \mathbb{R}^+$
- Goal: Find a minimum-cost subset $\mathcal{C} \subseteq \mathcal{F}$

s.t. $$X = \bigcup_{S \in \mathcal{C}} S.$$

Sum over the costs of all sets in $\mathcal{C}$



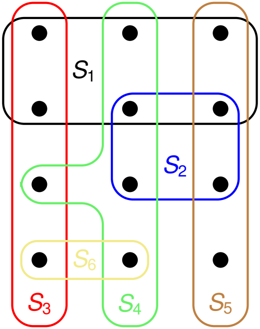
	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2

The Weighted Set-Cover Problem

Set Cover Problem

- Given: set X and a family of subsets $\mathcal{F}$, and a cost function $c : \mathcal{F} \to \mathbb{R}^+$
- Goal: Find a minimum-cost subset $\mathcal{C} \subseteq \mathcal{F}$ s.t. $X = \bigcup_{S \in \mathcal{C}} S.$

Sum over the costs of all sets in $\mathcal{C}$



	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2

Remarks:

- generalisation of the weighted Vertex-Cover problem
- models resource allocation problems

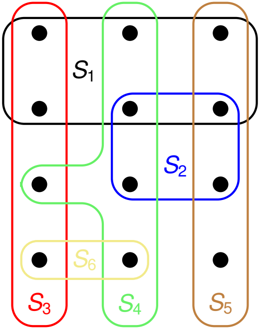
The Weighted Set-Cover Problem

Set Cover Problem

- Given: set X and a family of subsets $\mathcal{F}$, and a cost function $c : \mathcal{F} \to \mathbb{R}^+$
- Goal: Find a minimum-cost subset $\mathcal{C} \subseteq \mathcal{F}$

Sum over the costs of all sets in $\mathcal{C}$

s.t. $$X = \bigcup_{S \in \mathcal{C}} S.$$



	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2

Question: How can we reduce the Vertex-Cover problem to the Set-Cover problem?

Remarks:

- generalisation of the weighted Vertex-Cover problem
- models resource allocation problems

Setting up an Integer Program

Exercise: Try to formulate the integer program and linear program of the weighted SET-COVER problem (solution on next slide!)

Setting up an Integer Program

0-1 Integer Program

$$
\begin{array}{llcll}
\text{minimize} & \displaystyle\sum_{S \in \mathcal{F}} c(S) y(S) & & & \\
\text{subject to} & \displaystyle\sum_{S \in \mathcal{F} \colon x \in S} y(S) & \geq & 1 & \text{for each } x \in X \\
 & y(S) & \in & \{0, 1\} & \text{for each } S \in \mathcal{F}
\end{array}
$$

Setting up an Integer Program

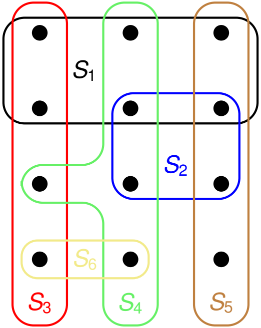
0-1 Integer Program

$$
\begin{array}{llcll}
\text{minimize} & \displaystyle\sum_{S \in \mathcal{F}} c(S) y(S) & & & \\
\text{subject to} & \displaystyle\sum_{S \in \mathcal{F}\colon x \in S} y(S) & \geq & 1 & \text{for each } x \in X \\
& y(S) & \in & \{0, 1\} & \text{for each } S \in \mathcal{F}
\end{array}
$$

Linear Program

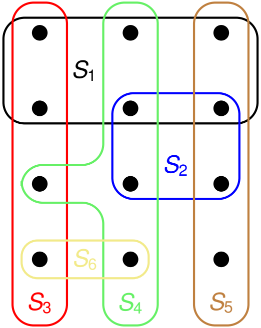
$$
\begin{array}{llcll}
\text{minimize} & \displaystyle\sum_{S \in \mathcal{F}} c(S) y(S) & & & \\
\text{subject to} & \displaystyle\sum_{S \in \mathcal{F}\colon x \in S} y(S) & \geq & 1 & \text{for each } x \in X \\
& y(S) & \in & [0, 1] & \text{for each } S \in \mathcal{F}
\end{array}
$$

Back to the Example



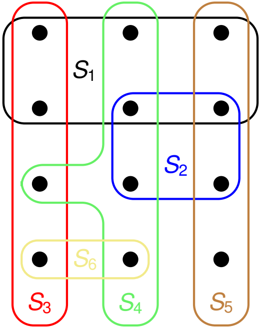
	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2

Back to the Example



	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	1/2	1/2	1/2	1/2	1	1/2

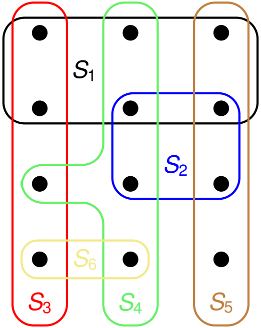
Back to the Example



	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Cost equals 8.5

Back to the Example

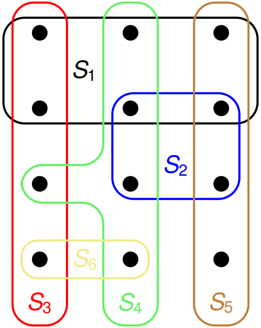


	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Cost equals 8.5

The strategy employed for Vertex-Cover would take all 6 sets!

Back to the Example



	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Cost equals 8.5

The strategy employed for Vertex-Cover would take all 6 sets!

Even worse: If all $\overline{y}$'s were below $1/2$, we would not even return a valid cover!

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Randomised Rounding

- Let $\mathcal{C} \subseteq \mathcal{F}$ be a random set with each set S being included independently with probability $\overline{y}(S)$.
- More precisely, if $\overline{y}$ denotes the optimal solution of the LP, then we compute an integral solution y by:

$$y(S) = \begin{cases} 1 & \text{with probability } \overline{y}(S) \\ 0 & \text{otherwise.} \end{cases} \qquad \text{for all } S \in \mathcal{F}.$$

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Randomised Rounding

- Let $\mathcal{C} \subseteq \mathcal{F}$ be a random set with each set S being included independently with probability $\overline{y}(S)$.
- More precisely, if $\overline{y}$ denotes the optimal solution of the LP, then we compute an integral solution y by:

$$y(S) = \begin{cases} 1 & \text{with probability } \overline{y}(S) \\ 0 & \text{otherwise.} \end{cases} \qquad \text{for all } S \in \mathcal{F}.$$

- Therefore, $\mathbf{E}\left[\, y(S) \,\right] = \overline{y}(S)$.

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Lemma

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Lemma

- The expected cost satisfies

$$\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$$

Randomised Rounding

	S_1	S_2	S_3	S_4	S_5	S_6
c :	2	3	3	5	1	2
$\overline{y}(.)$:	$1/2$	$1/2$	$1/2$	$1/2$	1	$1/2$

Idea: Interpret the $\overline{y}$-values as probabilities for picking the respective set.

Lemma

- The expected cost satisfies

$$\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$$

- The probability that an element $x \in X$ is covered satisfies

$$\mathbf{P}\left[x \in \bigcup_{S \in \mathcal{C}} S \right] \geq 1 - \frac{1}{e}.$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}\left[\, x \in \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

$$\mathbf{E}[\, c(\mathcal{C}) \,]$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right]$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\,c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\,x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

$$\mathbf{E}[\,c(\mathcal{C})\,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$

$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\,S \in \mathcal{C}\,] \cdot c(S)$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$

$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,]$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C})\,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C}\,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S\,] = \prod_{S \in \mathcal{F}\colon\, x \in S} \mathbf{P}[\, S \notin \mathcal{C}\,]$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] = \prod_{S \in \mathcal{F}\colon\, x \in S} \mathbf{P}[\, S \notin \mathcal{C} \,] = \prod_{S \in \mathcal{F}\colon\, x \in S} (1 - \overline{y}(S))$$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C})\,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C}\,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S\,] = \prod_{S \in \mathcal{F}\colon x \in S} \mathbf{P}[\, S \notin \mathcal{C}\,] = \prod_{S \in \mathcal{F}\colon x \in S} (1 - \overline{y}(S))$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] = \prod_{S \in \mathcal{F}:\ x \in S} \mathbf{P}[\, S \notin \mathcal{C} \,] = \prod_{S \in \mathcal{F}:\ x \in S} (1 - \overline{y}(S))$$
$$\leq \prod_{S \in \mathcal{F}:\ x \in S} e^{-\overline{y}(S)}$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[c(\mathcal{C})] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[x \in \cup_{S \in \mathcal{C}} S] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[c(\mathcal{C})] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[S \in \mathcal{C}] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[x \notin \cup_{S \in \mathcal{C}} S] = \prod_{S \in \mathcal{F}:\, x \in S} \mathbf{P}[S \notin \mathcal{C}] = \prod_{S \in \mathcal{F}:\, x \in S} (1 - \overline{y}(S))$$
$$\leq \prod_{S \in \mathcal{F}:\, x \in S} e^{-\overline{y}(S)}$$
$$= e^{-\sum_{S \in \mathcal{F}:\, x \in S} \overline{y}(S)}$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C})\,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C}\,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S\,] = \prod_{S \in \mathcal{F}\colon\, x \in S} \mathbf{P}[\, S \notin \mathcal{C}\,] = \prod_{S \in \mathcal{F}\colon\, x \in S} (1 - \overline{y}(S))$$
$$\leq \prod_{S \in \mathcal{F}\colon\, x \in S} e^{-\overline{y}(S)}$$
$$= e^{-\sum_{S \in \mathcal{F}\colon\, x \in S} \overline{y}(S)}$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

$\overline{y}$ solves the LP!

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C}) \,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$

$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C} \,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] = \prod_{S \in \mathcal{F}:\, x \in S} \mathbf{P}[\, S \notin \mathcal{C} \,] = \prod_{S \in \mathcal{F}:\, x \in S} (1 - \overline{y}(S))$$

$$\leq \prod_{S \in \mathcal{F}:\, x \in S} e^{-\overline{y}(S)}$$

$$= e^{-\sum_{S \in \mathcal{F}:\, x \in S} \overline{y}(S)} \leq e^{-1}$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

$\overline{y}$ solves the LP!

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}[\, c(\mathcal{C})\,] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S)\right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S)\right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}[\, S \in \mathcal{C}\,] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered ✓

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S\,] = \prod_{S \in \mathcal{F}\colon x \in S} \mathbf{P}[\, S \notin \mathcal{C}\,] = \prod_{S \in \mathcal{F}\colon x \in S} (1 - \overline{y}(S))$$
$$\leq \prod_{S \in \mathcal{F}\colon x \in S} e^{-\overline{y}(S)}$$
$$= e^{-\sum_{S \in \mathcal{F}\colon x \in S} \overline{y}(S)} \leq e^{-1}$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

$\overline{y}$ solves the LP!

Proof of Lemma

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $\overline{y}(S)$.

- The expected cost satisfies $\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
- The probability that x is covered satisfies $\mathbf{P}\left[\, x \in \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - \frac{1}{e}$.

Proof:

- **Step 1**: The expected cost of the random set $\mathcal{C}$ ✓

$$\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \mathbf{E}\left[\sum_{S \in \mathcal{C}} c(S) \right] = \mathbf{E}\left[\sum_{S \in \mathcal{F}} \mathbf{1}_{S \in \mathcal{C}} \cdot c(S) \right]$$
$$= \sum_{S \in \mathcal{F}} \mathbf{P}\left[\, S \in \mathcal{C} \,\right] \cdot c(S) = \sum_{S \in \mathcal{F}} \overline{y}(S) \cdot c(S).$$

- **Step 2**: The probability for an element to be (not) covered ✓

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] = \prod_{S \in \mathcal{F}:\, x \in S} \mathbf{P}\left[\, S \notin \mathcal{C} \,\right] = \prod_{S \in \mathcal{F}:\, x \in S} (1 - \overline{y}(S))$$
$$\leq \prod_{S \in \mathcal{F}:\, x \in S} e^{-\overline{y}(S)}$$
$$= e^{-\sum_{S \in \mathcal{F}:\, x \in S} \overline{y}(S)} \leq e^{-1} \qquad \square$$

$1 + x \leq e^x$ for any $x \in \mathbb{R}$

$\overline{y}$ solves the LP!

The Final Step

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $y(S)$.

- The expected cost satisfies $\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot y(S)$.
- The probability that x is covered satisfies $\mathbf{P}\left[\, x \in \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - \frac{1}{e}$.

The Final Step

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $y(S)$.

- The expected cost satisfies $\mathbf{E}\left[\, c(\mathcal{C}) \,\right] = \sum_{S \in \mathcal{F}} c(S) \cdot y(S)$.
- The probability that x is covered satisfies $\mathbf{P}\left[\, x \in \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - \frac{1}{e}$.

Problem: Need to make sure that every element is covered!

The Final Step

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $y(S)$.

- The expected cost satisfies $\mathbf{E}\left[c(\mathcal{C}) \right] = \sum_{S \in \mathcal{F}} c(S) \cdot y(S)$.
- The probability that x is covered satisfies $\mathbf{P}\left[x \in \cup_{S \in \mathcal{C}} S \right] \geq 1 - \frac{1}{e}$.

Problem: Need to make sure that every element is covered!

Idea: Amplify this probability by taking the union of $\Omega(\log n)$ random sets $\mathcal{C}$.

The Final Step

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $y(S)$.

- The expected cost satisfies $\mathbf{E}[\, c(\mathcal{C}) \,] = \sum_{S \in \mathcal{F}} c(S) \cdot y(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\, x \in \cup_{S \in \mathcal{C}} S \,] \geq 1 - \frac{1}{e}$.

Problem: Need to make sure that every element is covered!

Idea: Amplify this probability by taking the union of $\Omega(\log n)$ random sets $\mathcal{C}$.

```
WEIGHTED SET COVER-LP(X, F, c)
compute ȳ, an optimal solution to the linear program
C = ∅
repeat 2 ln n times
    for each S ∈ F
        let C = C ∪ {S} with probability ȳ(S)
return C
```

The Final Step

Lemma

Let $\mathcal{C} \subseteq \mathcal{F}$ be a random subset with each set S being included independently with probability $y(S)$.

- The expected cost satisfies $\mathbf{E}[\,c(\mathcal{C})\,] = \sum_{S \in \mathcal{F}} c(S) \cdot y(S)$.
- The probability that x is covered satisfies $\mathbf{P}[\,x \in \cup_{S \in \mathcal{C}} S\,] \geq 1 - \frac{1}{e}$.

Problem: Need to make sure that every element is covered!

Idea: Amplify this probability by taking the union of $\Omega(\log n)$ random sets $\mathcal{C}$.

WEIGHTED SET COVER-LP$(X, \mathcal{F}, c)$

1: compute $\overline{y}$, an optimal solution to the linear program
2: $\mathcal{C} = \emptyset$
3: **repeat** $2 \ln n$ times
4:     **for** each $S \in \mathcal{F}$
5:         let $\mathcal{C} = \mathcal{C} \cup \{S\}$ with probability $\overline{y}(S)$
6: **return** $\mathcal{C}$

clearly runs in polynomial-time!

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2 \ln(n)$.

Proof:

Analysis of Weighted Set Cover-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n}$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S \in \mathcal{C}} S \,\right] =$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[x \notin \cup_{S \in \mathcal{C}} S\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[X = \cup_{S \in \mathcal{C}} S\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S \in \mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

Analysis of Weighted Set Cover-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S\in\mathcal{C}} S \,\right] \le \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S\in\mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x\in X}\{x \notin \cup_{S\in\mathcal{C}} S\}\right]$$

$\mathbf{P}[\,A \cup B\,] \le \mathbf{P}[\,A\,] + \mathbf{P}[\,B\,]$

$$\ge 1 - \sum_{x\in X} \mathbf{P}\left[\, x \notin \cup_{S\in\mathcal{C}} S \,\right]$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}[\, X = \cup_{S \in \mathcal{C}} S \,] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\} \right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \geq 1 - n \cdot \frac{1}{n^2}$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S \in \mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}[\, X = \cup_{S \in \mathcal{C}} S \,] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}[\, X = \cup_{S \in \mathcal{C}} S \,] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}[\, x \notin \cup_{S \in \mathcal{C}} S \,] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2 \ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2 \ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S \in \mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio
  - By previous lemma, the expected cost of one iteration is $\sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S\in\mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S\in\mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x\in X}\{x \notin \cup_{S\in\mathcal{C}} S\}\right]$$

$\mathbf{P}[\, A \cup B \,] \leq \mathbf{P}[\, A \,] + \mathbf{P}[\, B \,]$

$$\geq 1 - \sum_{x\in X} \mathbf{P}\left[\, x \notin \cup_{S\in\mathcal{C}} S \,\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio
  - By previous lemma, the expected cost of one iteration is $\sum_{S\in\mathcal{F}} c(S)\cdot\overline{y}(S)$.
  - Linearity $\Rightarrow \mathbf{E}\left[\, c(\mathcal{C}) \,\right] \leq 2\ln(n)\cdot\sum_{S\in\mathcal{F}} c(S)\cdot\overline{y}(S)$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[x \notin \cup_{S \in \mathcal{C}} S\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[X = \cup_{S \in \mathcal{C}} S\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[A \cup B] \leq \mathbf{P}[A] + \mathbf{P}[B]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}\left[x \notin \cup_{S \in \mathcal{C}} S\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio
  - By previous lemma, the expected cost of one iteration is $\sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
  - Linearity $\Rightarrow \mathbf{E}\left[c(\mathcal{C})\right] \leq 2\ln(n) \cdot \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S) \leq 2\ln(n) \cdot c(\mathcal{C}^*)$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[x \notin \cup_{S \in \mathcal{C}} S\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[X = \cup_{S \in \mathcal{C}} S\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[A \cup B] \leq \mathbf{P}[A] + \mathbf{P}[B]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}\left[x \notin \cup_{S \in \mathcal{C}} S\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio ✓
  - By previous lemma, the expected cost of one iteration is $\sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
  - Linearity $\Rightarrow \mathbf{E}\left[c(\mathcal{C})\right] \leq 2\ln(n) \cdot \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S) \leq 2\ln(n) \cdot c(\mathcal{C}^*)$

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

Proof:

- **Step 1**: The probability that $\mathcal{C}$ is a cover ✓
  - By previous Lemma, an element $x \in X$ is covered in one of the $2\ln n$ iterations with probability at least $1 - \frac{1}{e}$, so that

$$\mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \leq \left(\frac{1}{e}\right)^{2\ln n} = \frac{1}{n^2}.$$

  - This implies for the event that all elements are covered:

$$\mathbf{P}\left[\, X = \cup_{S \in \mathcal{C}} S \,\right] = 1 - \mathbf{P}\left[\bigcup_{x \in X} \{x \notin \cup_{S \in \mathcal{C}} S\}\right]$$

$\mathbf{P}[\,A \cup B\,] \leq \mathbf{P}[\,A\,] + \mathbf{P}[\,B\,]$

$$\geq 1 - \sum_{x \in X} \mathbf{P}\left[\, x \notin \cup_{S \in \mathcal{C}} S \,\right] \geq 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}.$$

- **Step 2**: The expected approximation ratio ✓
  - By previous lemma, the expected cost of one iteration is $\sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S)$.
  - Linearity $\Rightarrow \mathbf{E}\left[\, c(\mathcal{C}) \,\right] \leq 2\ln(n) \cdot \sum_{S \in \mathcal{F}} c(S) \cdot \overline{y}(S) \leq 2\ln(n) \cdot c(\mathcal{C}^*)$ □

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}[\, c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*) \,] \geq 1/2$.

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}\left[\, c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*) \,\right] \geq 1/2$.

Hence with probability at least $1 - \frac{1}{n} - \frac{1}{2} > \frac{1}{3}$, solution is valid and within a factor of $4\ln(n)$ of the optimum.

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}[\,c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*)\,] \geq 1/2$.

Hence with probability at least $1 - \frac{1}{n} - \frac{1}{2} > \frac{1}{3}$, solution is valid and within a factor of $4\ln(n)$ of the optimum.

probability could be further increased by repeating

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}[\, c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*) \,] \geq 1/2$.

Hence with probability at least $1 - \frac{1}{n} - \frac{1}{2} > \frac{1}{3}$, solution is valid and within a factor of $4\ln(n)$ of the optimum.

probability could be further increased by repeating

Typical Approach for Designing Approximation Algorithms based on LPs

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}[\, c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*) \,] \geq 1/2$.

Hence with probability at least $1 - \frac{1}{n} - \frac{1}{2} > \frac{1}{3}$, solution is valid and within a factor of $4\ln(n)$ of the optimum.

probability could be further increased by repeating

Typical Approach for Designing Approximation Algorithms based on LPs

Analysis of WEIGHTED SET COVER-LP

Theorem

- With probability at least $1 - \frac{1}{n}$, the returned set $\mathcal{C}$ is a valid cover of X.
- The expected approximation ratio is $2\ln(n)$.

By Markov's inequality, $\mathbf{P}[\, c(\mathcal{C}) \leq 4\ln(n) \cdot c(\mathcal{C}^*) \,] \geq 1/2$.

Hence with probability at least $1 - \frac{1}{n} - \frac{1}{2} > \frac{1}{3}$, solution is valid and within a factor of $4\ln(n)$ of the optimum.

probability could be further increased by repeating

Typical Approach for Designing Approximation Algorithms based on LPs

[Exercise Question (9/10).10] gives a different perspective on the amplification procedure through non-linear randomised rounding.

Outline

Weighted Set Cover

MAX-CNF

Appendix: An Approximation Algorithm of TSP (non-examin.)

MAX-CNF

Recall:

MAX-3-CNF Satisfiability

- Given: 3-CNF formula, e.g.: $(x_1 \vee x_3 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

MAX-CNF Satisfiability (MAX-SAT)

MAX-CNF

Recall:

MAX-3-CNF Satisfiability

- Given: 3-CNF formula, e.g.: $(x_1 \vee x_3 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

MAX-CNF Satisfiability (MAX-SAT)

- Given: CNF formula, e.g.: $(x_1 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee x_4 \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

MAX-CNF

Recall:

MAX-3-CNF Satisfiability

- Given: 3-CNF formula, e.g.: $(x_1 \vee x_3 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

MAX-CNF Satisfiability (MAX-SAT)

- Given: CNF formula, e.g.: $(x_1 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee x_4 \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

Why study this generalised problem?

MAX-CNF

Recall:

MAX-3-CNF Satisfiability

- Given: 3-CNF formula, e.g.: $(x_1 \vee x_3 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

MAX-CNF Satisfiability (MAX-SAT)

- Given: CNF formula, e.g.: $(x_1 \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee x_4 \vee \overline{x_5}) \wedge \cdots$
- Goal: Find an assignment of the variables that satisfies as many clauses as possible.

Why study this generalised problem?

- Allowing arbitrary clause lengths makes the problem more interesting (we will see that simply guessing is not the best!)
- a nice concluding example where we can practice previously learned approaches

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Recall: This was the successful approach to solve MAX-3-CNF!

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Recall: This was the successful approach to solve MAX-3-CNF!

Analysis

For any clause i which has length ℓ,

$$\mathbf{P}[\,\text{clause } i \text{ is satisfied}\,] = 1 - 2^{-\ell} := \alpha_\ell.$$

In particular, the guessing algorithm is a randomised 2-approximation.

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Recall: This was the successful approach to solve MAX-3-CNF!

Analysis

For any clause i which has length ℓ,

$$\mathbf{P}[\,\text{clause } i \text{ is satisfied}\,] = 1 - 2^{-\ell} := \alpha_\ell.$$

In particular, the guessing algorithm is a randomised 2-approximation.

Proof:

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Recall: This was the successful approach to solve MAX-3-CNF!

Analysis

For any clause *i* which has length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] = 1 - 2^{-\ell} := \alpha_\ell.$$

In particular, the guessing algorithm is a randomised 2-approximation.

Proof:

- First statement as in the proof of Theorem 35.6. For clause *i* not to be satisfied, all ℓ occurring variables must be set to a specific value.

Approach 1: Guessing the Assignment

Assign each variable true or false uniformly and independently at random.

Recall: This was the successful approach to solve MAX-3-CNF!

Analysis

For any clause i which has length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] = 1 - 2^{-\ell} := \alpha_\ell.$$

In particular, the guessing algorithm is a randomised 2-approximation.

Proof:

- First statement as in the proof of Theorem 35.6. For clause i not to be satisfied, all ℓ occurring variables must be set to a specific value.
- As before, let $Y := \sum_{i=1}^{m} Y_i$ be the number of satisfied clauses. Then,

$$\mathbf{E}[Y] = \mathbf{E}\left[\sum_{i=1}^{m} Y_i\right] = \sum_{i=1}^{m} \mathbf{E}[Y_i] \geq \sum_{i=1}^{m} \frac{1}{2} = \frac{1}{2} \cdot m. \qquad \square$$

Approach 2: Guessing with a “Hunch” (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

Approach 2: Guessing with a “Hunch” (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

The same as randomised rounding!

Approach 2: Guessing with a "Hunch" (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

The same as randomised rounding!

0-1 Integer Program

$$\begin{array}{llll} \text{maximize} & \sum_{i=1}^{m} z_i & & \\ \text{subject to} & \sum_{j \in C_i^+} y_j + \sum_{j \in C_i^-} (1 - y_j) \geq z_i & & \text{for each } i = 1, 2, \ldots, m \\ & z_i \in \{0, 1\} & & \text{for each } i = 1, 2, \ldots, m \\ & y_j \in \{0, 1\} & & \text{for each } j = 1, 2, \ldots, n \end{array}$$

Approach 2: Guessing with a "Hunch" (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

The same as randomised rounding!

0-1 Integer Program

$$
\begin{array}{llcll}
\text{maximize} & \displaystyle\sum_{i=1}^{m} z_i & & & \\
\text{subject to} & \displaystyle\sum_{j \in C_i^+} y_j + \sum_{j \in C_i^-} (1 - y_j) & \geq & z_i & \text{for each } i = 1, 2, \ldots, m \\
& z_i & \in & \{0, 1\} & \text{for each } i = 1, 2, \ldots, m \\
& y_j & \in & \{0, 1\} & \text{for each } j = 1, 2, \ldots, n
\end{array}
$$

C_i^+ is the index set of the un-negated variables of clause i.

Approach 2: Guessing with a "Hunch" (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

The same as randomised rounding!

0-1 Integer Program

$$\text{maximize} \quad \sum_{i=1}^{m} z_i$$

$$\text{subject to} \quad \sum_{j \in C_i^+} y_j + \sum_{j \in C_i^-} (1 - y_j) \geq z_i \qquad \text{for each } i = 1, 2, \ldots, m$$

$$z_i \in \{0, 1\} \qquad \text{for each } i = 1, 2, \ldots, m$$

$$y_j \in \{0, 1\} \qquad \text{for each } j = 1, 2, \ldots, n$$

These auxiliary variables are used to reflect whether a clause is satisfied or not

C_i^+ is the index set of the unnegated variables of clause i.

Approach 2: Guessing with a "Hunch" (Randomised Rounding)

First solve a linear program and use fractional values for a **biased** coin flip.

The same as randomised rounding!

0-1 Integer Program

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{m} z_i \\ \text{subject to} \quad & \sum_{j \in C_i^+} y_j + \sum_{j \in C_i^-} (1 - y_j) \geq z_i && \text{for each } i = 1, 2, \ldots, m \\ & z_i \in \{0, 1\} && \text{for each } i = 1, 2, \ldots, m \\ & y_j \in \{0, 1\} && \text{for each } j = 1, 2, \ldots, n \end{aligned}$$

These auxiliary variables are used to reflect whether a clause is satisfied or not

C_i^+ is the index set of the un-negated variables of clause i.

- In the corresponding LP each $\in \{0, 1\}$ is replaced by $\in [0, 1]$
- Let $(\overline{y}, \overline{z})$ be the optimal solution of the LP
- Obtain an integer solution y through randomised rounding of $\overline{y}$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}[\text{clause } i \text{ is satisfied}] =$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] = 1 - \prod_{j=1}^{\ell} \mathbf{P}\left[\; y_j \text{ is false} \;\right]$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}[\text{clause } i \text{ is satisfied}] = 1 - \prod_{j=1}^{\ell} \mathbf{P}[\ y_j \text{ is false }\ ] = 1 - \prod_{j=1}^{\ell} (1 - \overline{y}_j)$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}[\text{clause } i \text{ is satisfied}] = 1 - \prod_{j=1}^{\ell} \mathbf{P}[\ y_j \text{ is false}\ ] = 1 - \prod_{j=1}^{\ell} (1 - \overline{y}_j)$$

Arithmetic vs. geometric mean:

$$\frac{a_1 + \ldots + a_k}{k} \geq \sqrt[k]{a_1 \times \ldots \times a_k}.$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] = 1 - \prod_{j=1}^{\ell} \mathbf{P}\left[\; y_j \text{ is false }\;\right] = 1 - \prod_{j=1}^{\ell} \left(1 - \overline{y}_j\right)$$

$$\geq 1 - \left(\frac{\sum_{j=1}^{\ell}(1 - \overline{y}_j)}{\ell}\right)^{\ell}$$

Arithmetic vs. geometric mean:

$$\frac{a_1 + \ldots + a_k}{k} \geq \sqrt[k]{a_1 \times \ldots \times a_k}.$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\,\text{clause } i \text{ is satisfied}\,] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}[\,\text{clause } i \text{ is satisfied}\,] = 1 - \prod_{j=1}^{\ell} \mathbf{P}[\; y_j \text{ is false }\;] = 1 - \prod_{j=1}^{\ell} (1 - \overline{y}_j)$$

$$\geq 1 - \left(\frac{\sum_{j=1}^{\ell}(1 - \overline{y}_j)}{\ell}\right)^{\ell}$$

$$= 1 - \left(1 - \frac{\sum_{j=1}^{\ell} \overline{y}_j}{\ell}\right)^{\ell}$$

Arithmetic vs. geometric mean:

$$\frac{a_1 + \ldots + a_k}{k} \geq \sqrt[k]{a_1 \times \ldots \times a_k}.$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (1/2):

- Assume w.l.o.g. all literals in clause i appear non-negated (otherwise replace every occurrence of x_j by $\overline{x_j}$ in the whole formula)
- Further, by relabelling assume $C_i = (x_1 \vee \cdots \vee x_\ell)$

$$\Rightarrow \quad \mathbf{P}[\text{clause } i \text{ is satisfied}] = 1 - \prod_{j=1}^{\ell} \mathbf{P}[\ y_j \text{ is false }\ ] = 1 - \prod_{j=1}^{\ell} (1 - \overline{y}_j)$$

Arithmetic vs. geometric mean:

$$\frac{a_1 + \ldots + a_k}{k} \geq \sqrt[k]{a_1 \times \ldots \times a_k}.$$

$$\geq 1 - \left(\frac{\sum_{j=1}^{\ell}(1 - \overline{y}_j)}{\ell}\right)^{\ell}$$

$$= 1 - \left(1 - \frac{\sum_{j=1}^{\ell} \overline{y}_j}{\ell}\right)^{\ell} \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}.$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\text{clause } i \text{ is satisfied}\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}\left[\text{clause } i \text{ is satisfied}\right] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - \left(1 - \frac{z}{\ell}\right)^{\ell}$.

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - \left(1 - \frac{z}{\ell}\right)^{\ell}$. This is a concave function with $g(0) = 0$ and $g(1) = 1 - \left(1 - \frac{1}{\ell}\right)^{\ell} =: \beta_\ell$.

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

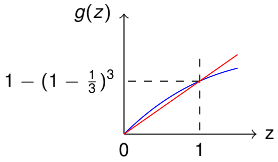
$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - \left(1 - \frac{z}{\ell}\right)^{\ell}$. This is a concave function with $g(0) = 0$ and $g(1) = 1 - \left(1 - \frac{1}{\ell}\right)^{\ell} =: \beta_\ell$.



Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

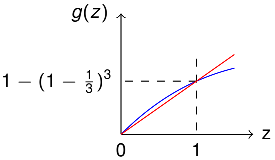
Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - \left(1 - \frac{z}{\ell}\right)^{\ell}$. This is a concave function with $g(0) = 0$ and $g(1) = 1 - \left(1 - \frac{1}{\ell}\right)^{\ell} =: \beta_\ell$.

$$\Rightarrow \quad g(z) \geq \beta_\ell \cdot z \qquad \text{for any } z \in [0,1]$$



Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

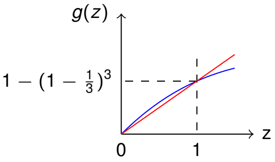
Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - (1 - \frac{z}{\ell})^{\ell}$. This is a concave function with $g(0) = 0$ and $g(1) = 1 - \left(1 - \frac{1}{\ell}\right)^{\ell} =: \beta_\ell$.

$$\Rightarrow \quad g(z) \geq \beta_\ell \cdot z \qquad \text{for any } z \in [0, 1]$$



- Therefore, $\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \beta_\ell \cdot \overline{z}_i$.

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

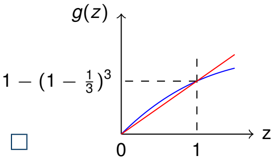
Proof of Lemma (2/2):

- So far we have shown:

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq 1 - \left(1 - \frac{\overline{z}_i}{\ell}\right)^{\ell}$$

- For any $\ell \geq 1$, define $g(z) := 1 - (1 - \frac{z}{\ell})^{\ell}$. This is a concave function with $g(0) = 0$ and $g(1) = 1 - \left(1 - \frac{1}{\ell}\right)^{\ell} =: \beta_\ell$.

$$\Rightarrow \quad g(z) \geq \beta_\ell \cdot z \qquad \text{for any } z \in [0,1]$$



- Therefore, $\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \beta_\ell \cdot \overline{z}_i$. □

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1-1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\text{clause } i \text{ is satisfied}\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1-1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\text{clause } i \text{ is satisfied}\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1-1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

- For any clause $i = 1, 2, \ldots, m$, let ℓ_i be the corresponding length.

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1 - 1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

- For any clause $i = 1, 2, \ldots, m$, let ℓ_i be the corresponding length.
- Then the expected number of satisfied clauses is:

$$\mathbf{E}[Y] = \sum_{i=1}^{m} \mathbf{E}[Y_i] \geq$$

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\,\text{clause } i \text{ is satisfied}\,] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1-1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

- For any clause $i = 1, 2, \ldots, m$, let ℓ_i be the corresponding length.
- Then the expected number of satisfied clauses is:

$$\mathbf{E}[\,Y\,] = \sum_{i=1}^{m} \mathbf{E}[\,Y_i\,] \geq \sum_{i=1}^{m} \left(1 - \left(1 - \frac{1}{\ell_i}\right)^{\ell_i}\right) \cdot \overline{z}_i$$

By Lemma

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}\left[\,\text{clause } i \text{ is satisfied}\,\right] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1 - 1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

- For any clause $i = 1, 2, \ldots, m$, let ℓ_i be the corresponding length.
- Then the expected number of satisfied clauses is:

$$\mathbf{E}\left[\,Y\,\right] = \sum_{i=1}^{m} \mathbf{E}\left[\,Y_i\,\right] \geq \sum_{i=1}^{m} \left(1 - \left(1 - \frac{1}{\ell_i}\right)^{\ell_i}\right) \cdot \overline{z}_i \geq \sum_{i=1}^{m} \left(1 - \frac{1}{e}\right) \cdot \overline{z}_i$$

By Lemma

Since $(1 - 1/x)^x \leq 1/e$

Analysis of Randomised Rounding

Lemma

For any clause i of length ℓ,

$$\mathbf{P}[\text{clause } i \text{ is satisfied}] \geq \left(1 - \left(1 - \frac{1}{\ell}\right)^{\ell}\right) \cdot \overline{z}_i.$$

Theorem

Randomised Rounding yields a $1/(1-1/e) \approx 1.5820$ randomised approximation algorithm for MAX-CNF.

Proof of Theorem:

- For any clause $i = 1, 2, \ldots, m$, let ℓ_i be the corresponding length.
- Then the expected number of satisfied clauses is:

$$\mathbf{E}[Y] = \sum_{i=1}^{m} \mathbf{E}[Y_i] \geq \sum_{i=1}^{m} \left(1 - \left(1 - \frac{1}{\ell_i}\right)^{\ell_i}\right) \cdot \overline{z}_i \geq \sum_{i=1}^{m} \left(1 - \frac{1}{e}\right) \cdot \overline{z}_i \geq \left(1 - \frac{1}{e}\right) \cdot \mathrm{OPT}$$

By Lemma

Since $(1 - 1/x)^x \leq 1/e$

LP solution at least as good as optimum

Approach 3: Hybrid Algorithm

Summary

- Approach 1 (Guessing) achieves better guarantee on longer clauses
- Approach 2 (Rounding) achieves better guarantee on shorter clauses

Approach 3: Hybrid Algorithm

Summary

- Approach 1 (Guessing) achieves better guarantee on longer clauses
- Approach 2 (Rounding) achieves better guarantee on shorter clauses

Idea: Consider a hybrid algorithm which interpolates between the two approaches

Approach 3: Hybrid Algorithm

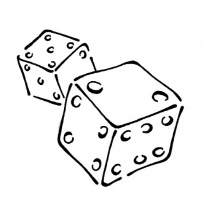
Summary

- Approach 1 (Guessing) achieves better guarantee on longer clauses
- Approach 2 (Rounding) achieves better guarantee on shorter clauses

Idea: Consider a hybrid algorithm which interpolates between the two approaches

HYBRID-MAX-CNF(φ, n, m)

Let $b \in \{0, 1\}$ be the flip of a fair coin
If $b = 0$ **then** perform random guessing
If $b = 1$ **then** perform randomised rounding
return the computed solution



Approach 3: Hybrid Algorithm

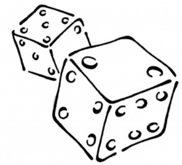
Summary

- Approach 1 (Guessing) achieves better guarantee on longer clauses
- Approach 2 (Rounding) achieves better guarantee on shorter clauses

Idea: Consider a hybrid algorithm which interpolates between the two approaches

HYBRID-MAX-CNF(φ, n, m)

Let $b \in \{0, 1\}$ be the flip of a fair coin
If $b = 0$ **then** perform random guessing
If $b = 1$ **then** perform randomised rounding
return the computed solution



Algorithm sets each variable x_i to TRUE with prob. $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \overline{y}_i$.
Note, however, that variables are **not** independently assigned!

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised $4/3$-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised $4/3$-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.

Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$,

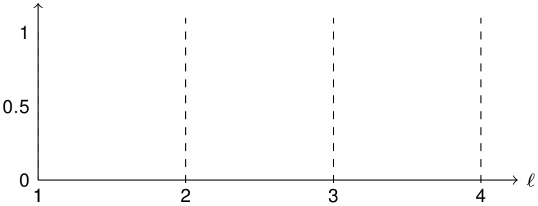
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell + \beta_\ell}{2} \geq 3/4$ (see figure)



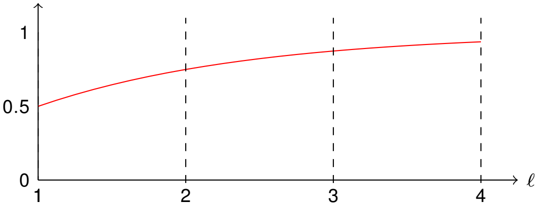
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell + \beta_\ell}{2} \geq 3/4$ (see figure)



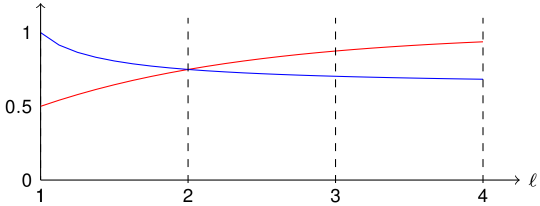
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised $4/3$-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell + \beta_\ell}{2} \geq 3/4$ (see figure)



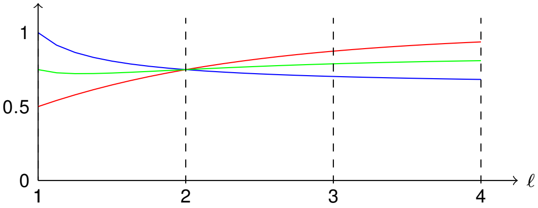
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised 4/3-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell+\beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell+\beta_\ell}{2} \geq 3/4$ (see figure)



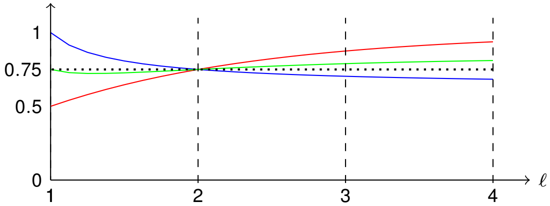
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised $4/3$-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell + \beta_\ell}{2} \geq 3/4$ (see figure)



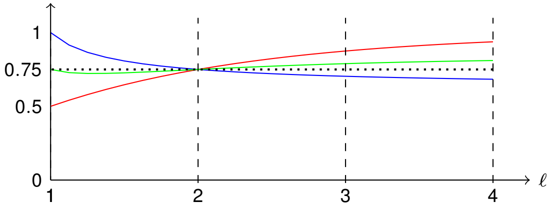
Analysis of Hybrid Algorithm

Theorem

HYBRID-MAX-CNF(φ, n, m) is a randomised $4/3$-approx. algorithm.

Proof:

- It suffices to prove that clause i is satisfied with probability at least $3/4 \cdot \overline{z}_i$
- For any clause i of length ℓ:
  - Algorithm 1 satisfies it with probability $1 - 2^{-\ell} = \alpha_\ell \geq \alpha_\ell \cdot \overline{z}_i$.
  - Algorithm 2 satisfies it with probability $\beta_\ell \cdot \overline{z}_i$.
  - HYBRID-MAX-CNF(φ, n, m) satisfies it with probability $\frac{1}{2} \cdot \alpha_\ell \cdot \overline{z}_i + \frac{1}{2} \cdot \beta_\ell \cdot \overline{z}_i$.
- Note $\frac{\alpha_\ell + \beta_\ell}{2} = 3/4$ for $\ell \in \{1, 2\}$, and for $\ell \geq 3$, $\frac{\alpha_\ell + \beta_\ell}{2} \geq 3/4$ (see figure)
- $\Rightarrow$ HYBRID-MAX-CNF(φ, n, m) satisfies it with prob. at least $3/4 \cdot \overline{z}_i$ □



MAX-CNF Conclusion

Summary

- Since $\alpha_2 = \beta_2 = 3/4$, we cannot achieve a better approximation ratio than $4/3$ by combining Algorithm 1 & 2 in a different way
- The $4/3$-approximation algorithm can be easily derandomised
  - Idea: use the conditional expectation trick for both Algorithm 1 & 2 and output the better solution
- The $4/3$-approximation algorithm applies unchanged to a weighted version of MAX-CNF, where each clause has a non-negative weight
- Even MAX-2-CNF (every clause has length 2) is NP-hard!

Outline

Weighted Set Cover

MAX-CNF

Appendix: An Approximation Algorithm of TSP (non-examin.)

Metric TSP (TSP Problem with the Triangle Inequality)

Idea: First compute an MST, and then create a tour based on the tree.

Metric TSP (TSP Problem with the Triangle Inequality)

Idea: First compute an MST, and then create a tour based on the tree.

APPROX-TSP-TOUR(G, c)
select a vertex $r \in G.V$ to be a "root" vertex
compute a minimum spanning tree $T_{\min}$ for G from root r
    using MST-PRIM(G, c, r)
let H be a list of vertices, ordered according to when they are first visited
    in a preorder walk of $T_{\min}$
return the hamiltonian cycle H

Metric TSP (TSP Problem with the Triangle Inequality)

Idea: First compute an MST, and then create a tour based on the tree.

APPROX-TSP-TOUR(G, c)
select a vertex $r \in G.V$ to be a “root” vertex
compute a minimum spanning tree T_{min} for G from root r
    using MST-PRIM(G, c, r)
let H be a list of vertices, ordered according to when they are first visited
    in a preorder walk of T_{min}
return the hamiltonian cycle H

Runtime is dominated by MST-PRIM, which is $\Theta(V^2)$.

Metric TSP (TSP Problem with the Triangle Inequality)

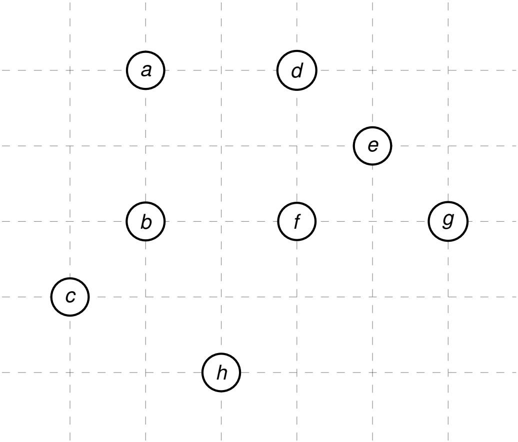
Idea: First compute an MST, and then create a tour based on the tree.

```
APPROX-TSP-TOUR(G, c)
select a vertex r ∈ G.V to be a "root" vertex
compute a minimum spanning tree T_min for G from root r
        using MST-PRIM(G, c, r)
let H be a list of vertices, ordered according to when they are first visited
        in a preorder walk of T_min
return the hamiltonian cycle H
```

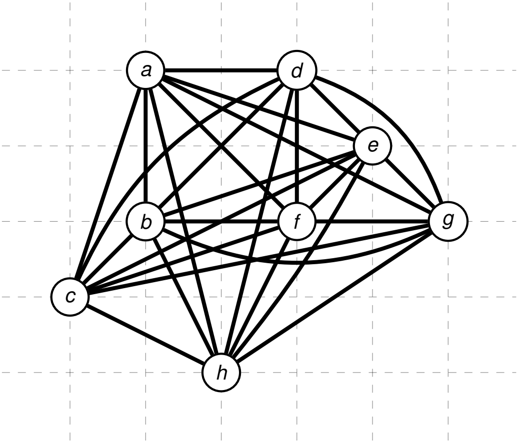
Runtime is dominated by MST-PRIM, which is $\Theta(V^2)$.

Remember: In the Metric-TSP problem, G is a complete graph.

Run of APPROX-TSP-TOUR

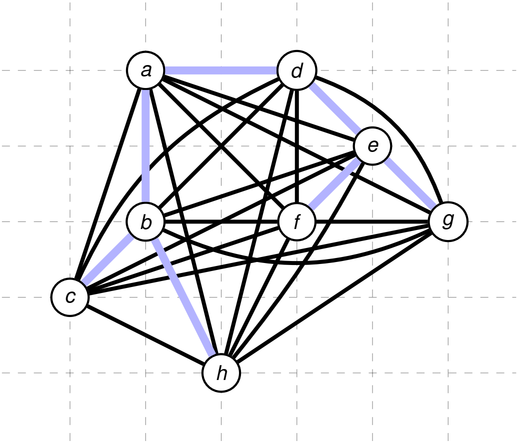


Run of APPROX-TSP-TOUR



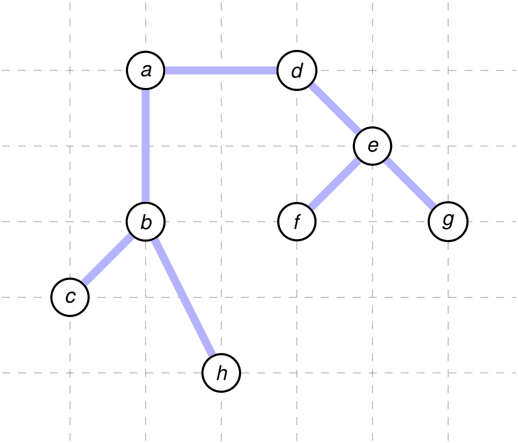
1. Compute MST $T_{\min}$

Run of APPROX-TSP-TOUR



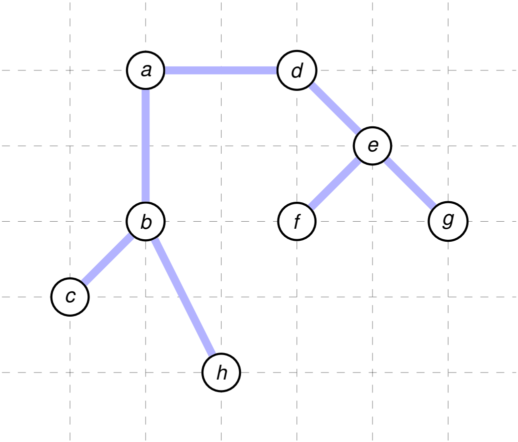
1. Compute MST $T_{\min}$

Run of APPROX-TSP-TOUR



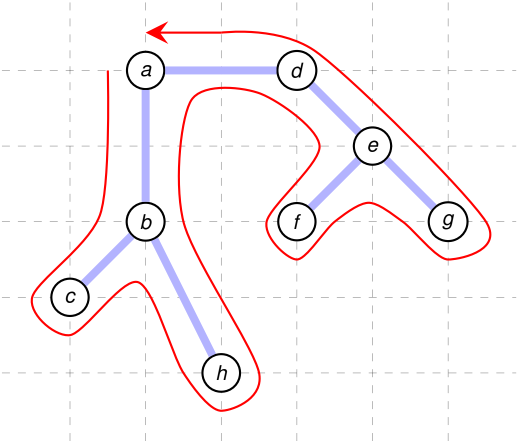
1. Compute MST T_{min} ✓

Run of APPROX-TSP-TOUR



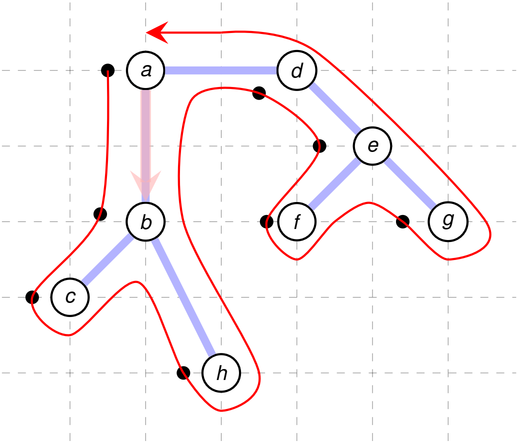
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$

Run of APPROX-TSP-TOUR



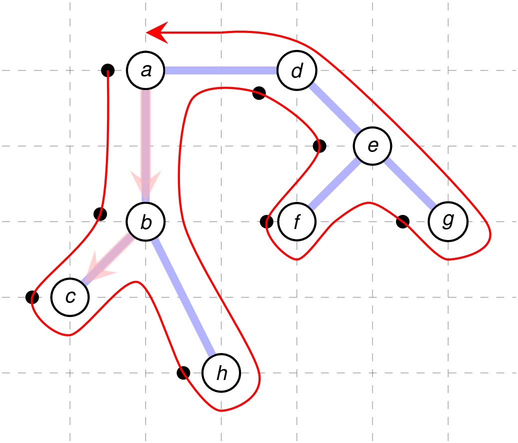
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓

Run of APPROX-TSP-TOUR



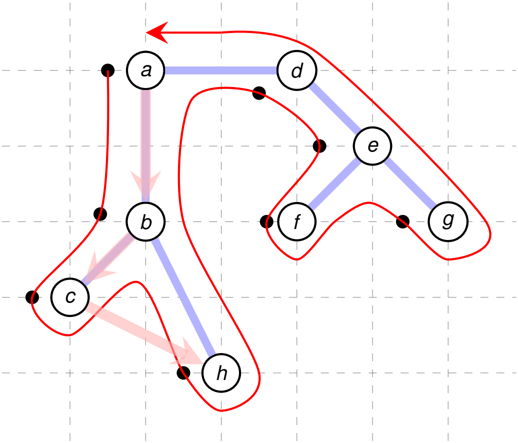
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



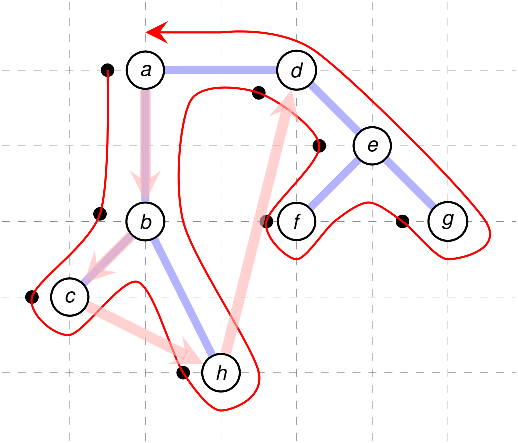
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



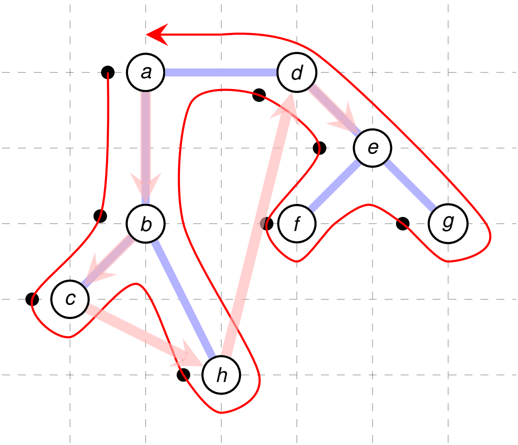
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



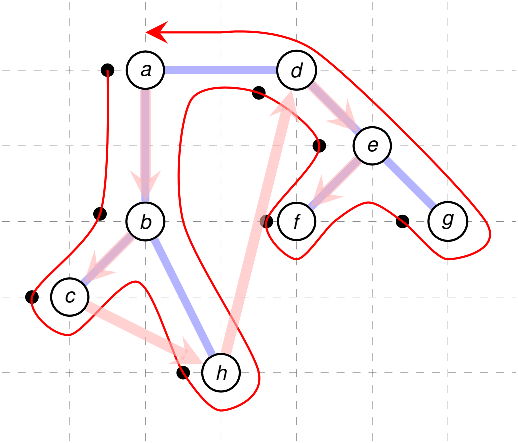
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of Approx-Tsp-Tour



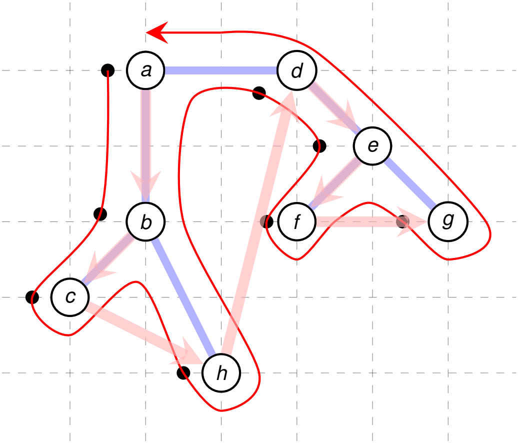
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



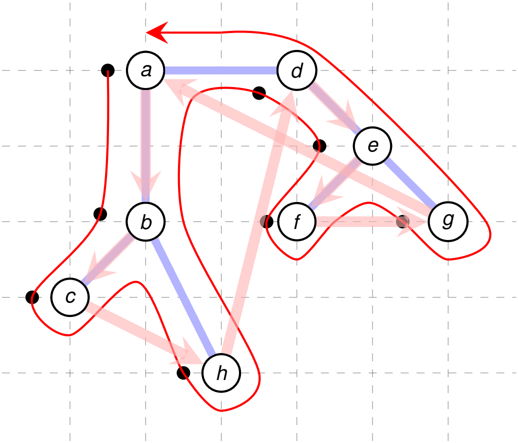
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



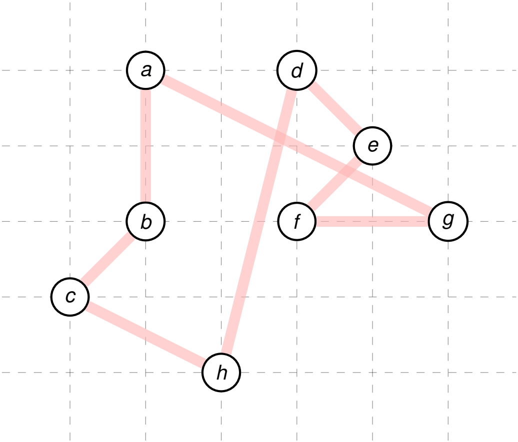
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



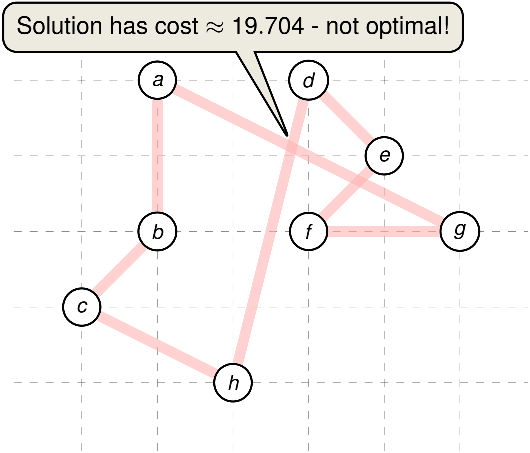
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk

Run of APPROX-TSP-TOUR



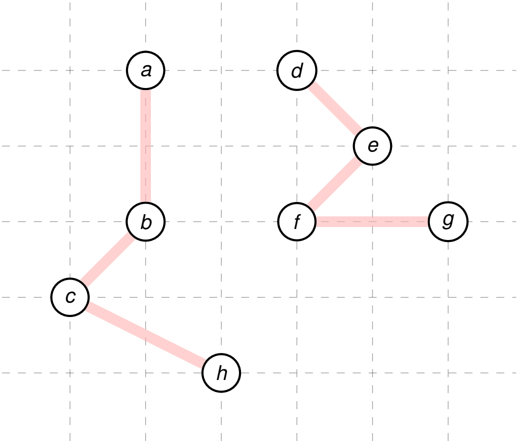
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk ✓

Run of APPROX-TSP-TOUR



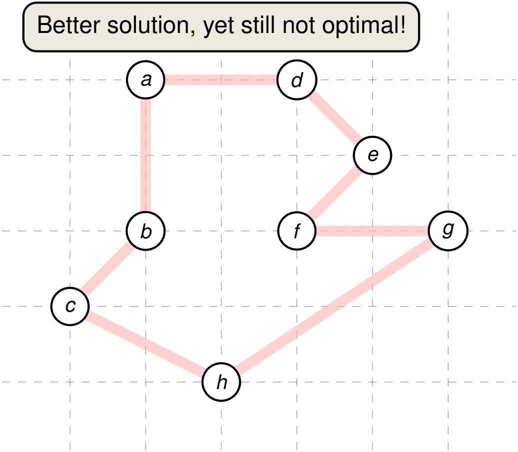
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk ✓

Run of APPROX-TSP-TOUR



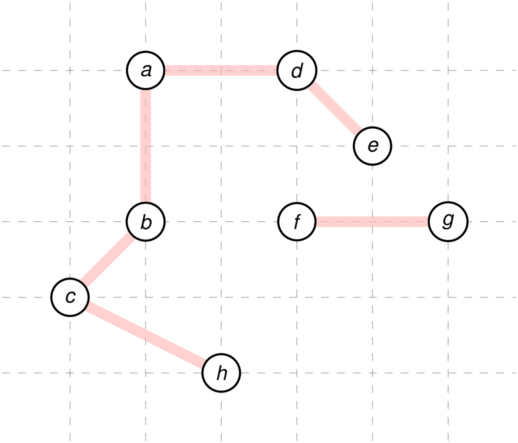
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk ✓

Run of APPROX-TSP-TOUR



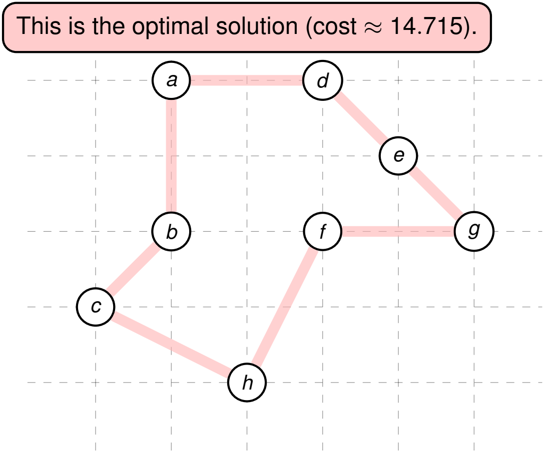
1. Compute MST T_{min} ✓
2. Perform preorder walk on MST T_{min} ✓
3. Return list of vertices according to the preorder tree walk ✓

Run of APPROX-TSP-TOUR



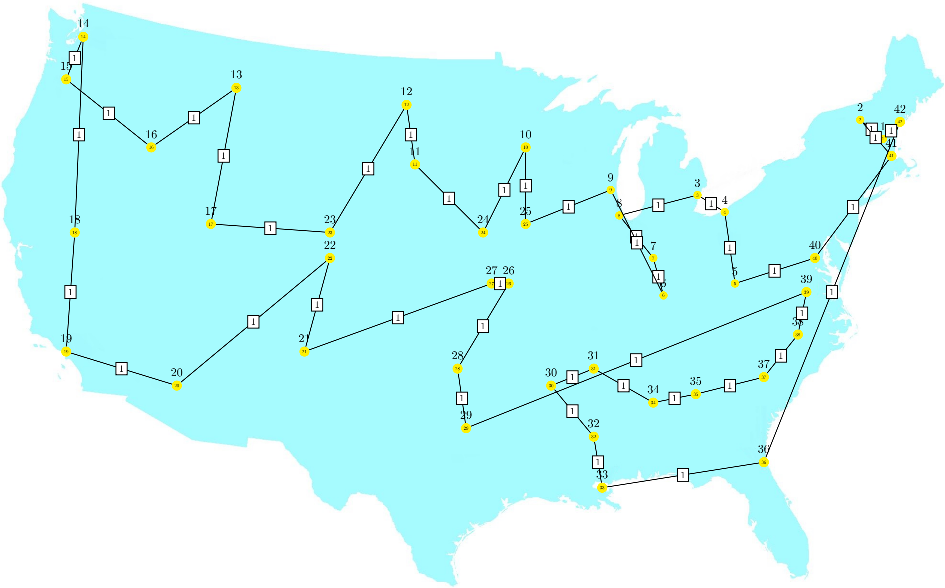
1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk ✓

Run of APPROX-TSP-TOUR

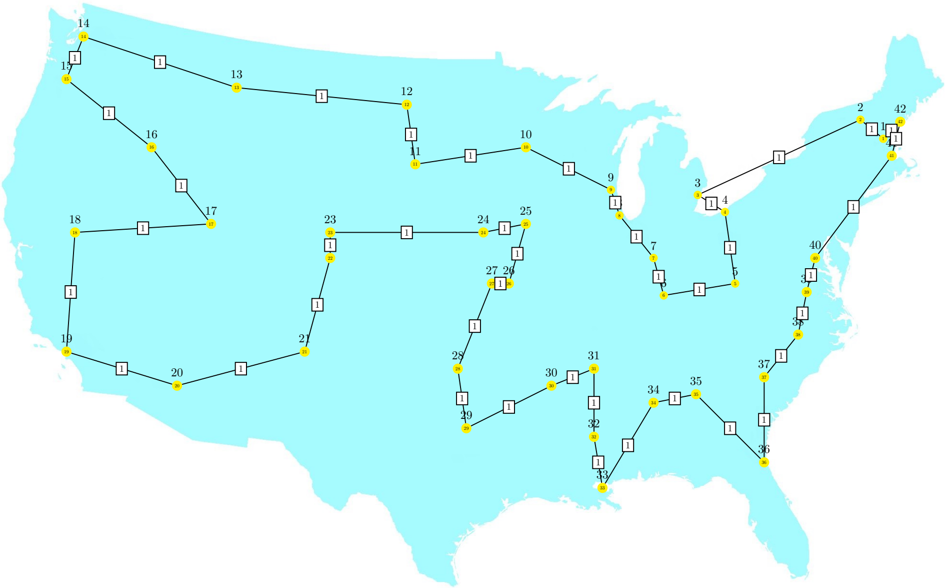


1. Compute MST $T_{\min}$ ✓
2. Perform preorder walk on MST $T_{\min}$ ✓
3. Return list of vertices according to the preorder tree walk ✓

Approximate Solution: Objective 921



Optimal Solution: Objective 699



Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

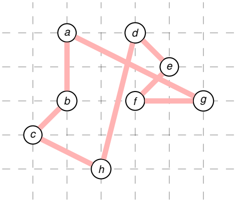
Proof:

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:



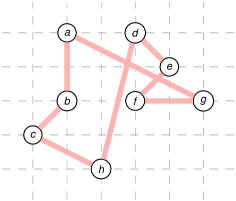
solution H of APPROX-TSP

Proof of the Approximation Ratio

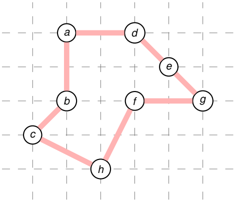
Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:



solution H of APPROX-TSP



optimal solution H^*

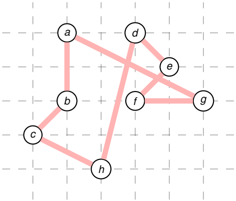
Proof of the Approximation Ratio

Theorem 35.2

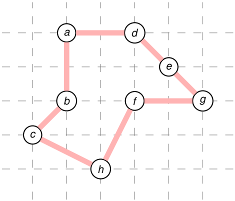
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge



solution H of APPROX-TSP



optimal solution H^*

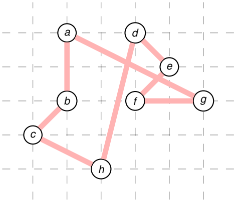
Proof of the Approximation Ratio

Theorem 35.2

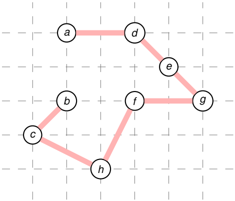
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge



solution H of APPROX-TSP



spanning tree T as a subset of H^*

Proof of the Approximation Ratio

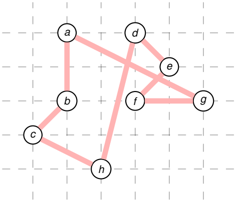
Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

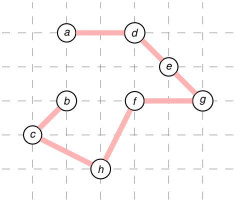
Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and



solution H of APPROX-TSP



spanning tree T as a subset of H^*

Proof of the Approximation Ratio

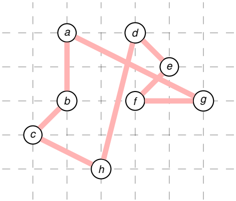
Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

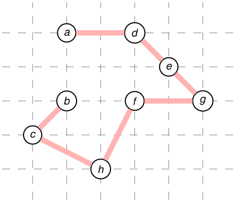
Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$



solution H of APPROX-TSP



spanning tree T as a subset of H^*

Proof of the Approximation Ratio

Theorem 35.2

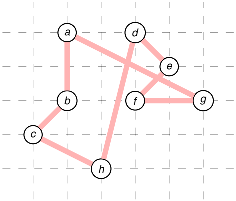
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

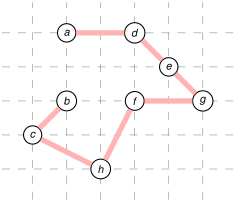
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

exploiting that all edge costs are non-negative!



solution H of APPROX-TSP



spanning tree T as a subset of H^*

Proof of the Approximation Ratio

Theorem 35.2

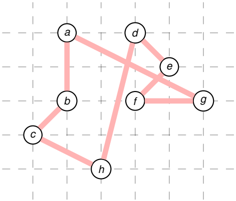
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

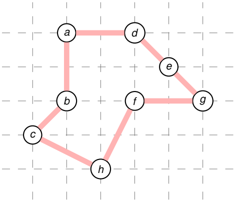
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)



solution H of APPROX-TSP



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

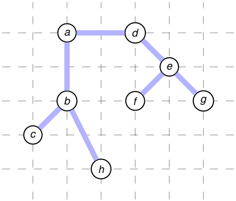
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

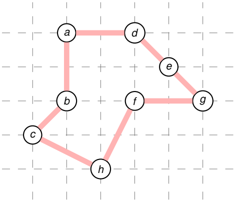
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)



minimum spanning tree $T_{\min}$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

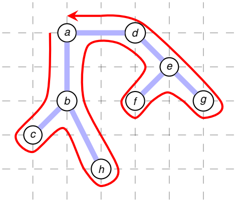
APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

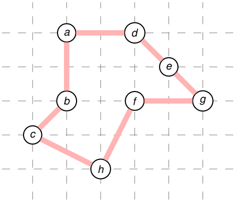
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

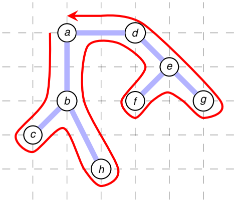
Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

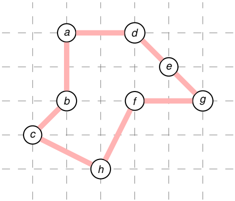
$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

$\Rightarrow$ Full walk traverses every edge exactly twice, so



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

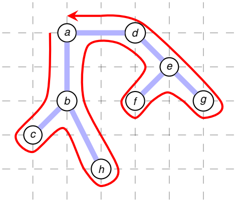
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

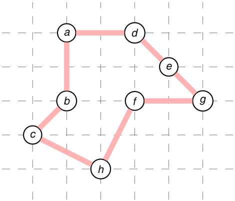
- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min})$$



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

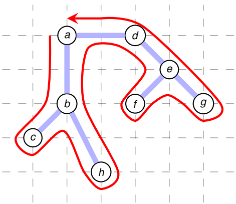
- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

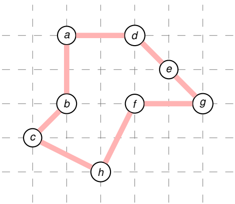
- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

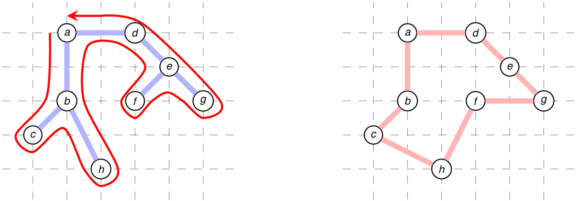
$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

- Deleting duplicate vertices from W yields a tour H



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$ optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

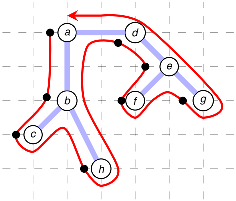
$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

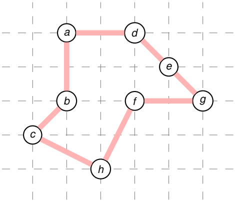
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

- Deleting duplicate vertices from W yields a tour H



Walk $W = (a, b, c, b, h, b, a, d, e, f, e, g, e, d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

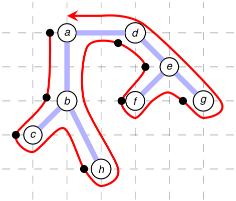
$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

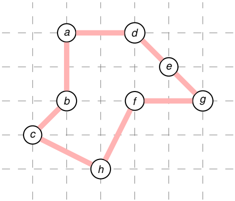
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

- Deleting duplicate vertices from W yields a tour H



Walk $W = (a, b, c, \not b, h, \not b, \not a, d, e, f, \not e, g, \not e, \not d, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

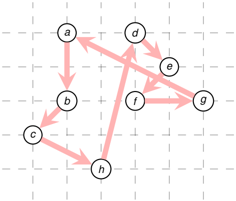
$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

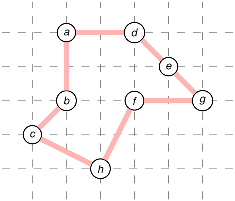
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

- Deleting duplicate vertices from W yields a tour H



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

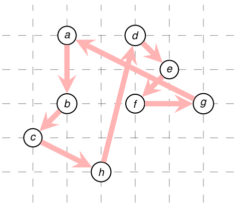
- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

$\Rightarrow$ Full walk traverses every edge exactly twice, so

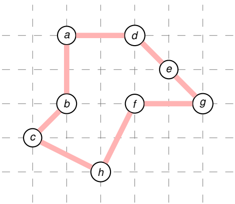
$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

exploiting triangle inequality!

- Deleting duplicate vertices from W yields a tour H with smaller cost:



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

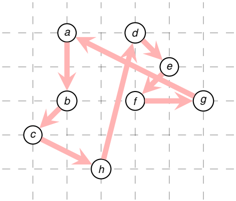
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

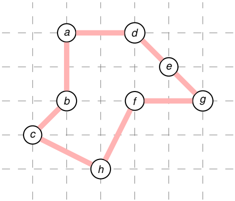
exploiting triangle inequality!

- Deleting duplicate vertices from W yields a tour H with smaller cost:

$$c(H) \leq c(W)$$



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

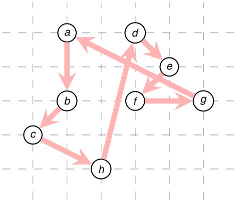
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

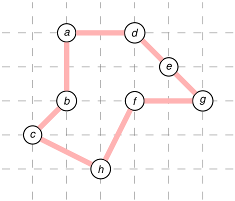
exploiting triangle inequality!

- Deleting duplicate vertices from W yields a tour H with smaller cost:

$$c(H) \leq c(W) \leq 2c(H^*)$$



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

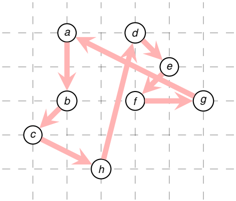
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

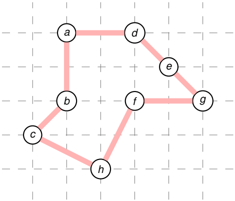
exploiting triangle inequality!

- Deleting duplicate vertices from W yields a tour H with smaller cost:

$$c(H) \leq c(W) \leq 2c(H^*)$$



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Proof of the Approximation Ratio

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Proof:

- Consider the optimal tour H^* and remove an arbitrary edge

$\Rightarrow$ yields a spanning tree T and $c(T_{\min}) \leq c(T) \leq c(H^*)$

- Let W be the full walk of the minimum spanning tree $T_{\min}$ (including repeated visits)

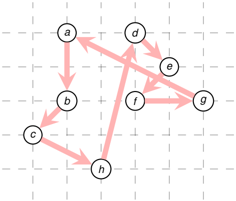
$\Rightarrow$ Full walk traverses every edge exactly twice, so

$$c(W) = 2c(T_{\min}) \leq 2c(T) \leq 2c(H^*)$$

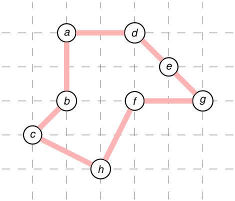
exploiting triangle inequality!

- Deleting duplicate vertices from W yields a tour H with smaller cost:

$$c(H) \leq c(W) \leq 2c(H^*) \qquad \square$$



Tour $H = (a, b, c, h, d, e, f, g, a)$



optimal solution H^*

Christofides Algorithm

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Christofides Algorithm

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Can we get a better approximation ratio?

Christofides Algorithm

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Can we get a better approximation ratio?

CHRISTOFIDES(G, c)
select a vertex $r \in G.V$ to be a "root" vertex
compute a minimum spanning tree $T_{\min}$ for G from root r
    using MST-PRIM(G, c, r)
compute a perfect matching $M_{\min}$ with minimum weight in the complete graph
    over the odd-degree vertices in $T_{\min}$
let H be a list of vertices, ordered according to when they are first visited
    in a Eulearian circuit of $T_{\min} \cup M_{\min}$
return the hamiltonian cycle H

Christofides Algorithm

Theorem 35.2

APPROX-TSP-TOUR is a polynomial-time 2-approximation for the traveling-salesman problem with the triangle inequality.

Can we get a better approximation ratio?

CHRISTOFIDES(G, c)
select a vertex $r \in G.V$ to be a "root" vertex
compute a minimum spanning tree $T_{\min}$ for G from root r
    using MST-PRIM(G, c, r)
compute a perfect matching $M_{\min}$ with minimum weight in the complete graph
    over the odd-degree vertices in $T_{\min}$
let H be a list of vertices, ordered according to when they are first visited
    in a Eulearian circuit of $T_{\min} \cup M_{\min}$
return the hamiltonian cycle H

Theorem (Christofides'76)

There is a polynomial-time $\frac{3}{2}$-approximation algorithm for the travelling salesman problem with the triangle inequality.