Approximation Algorithms

Lecture 5:
LP-based Approximation Algorithms for SetCover

Part I:
SetCover as an ILP

Alexander Wolff

Winter 2024/25

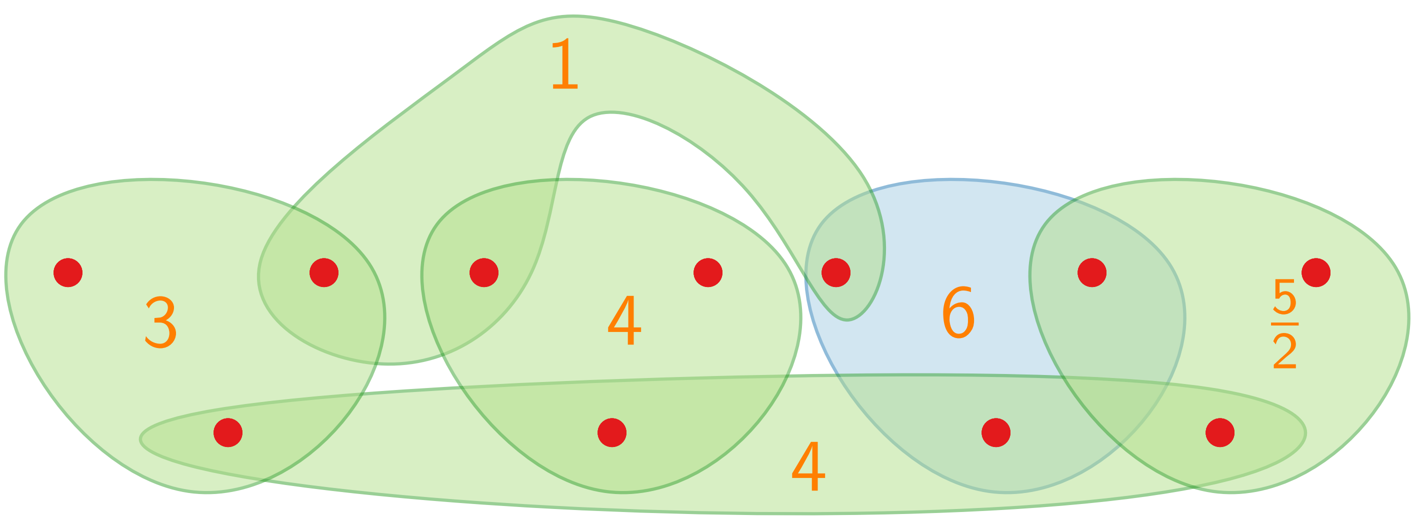
SetCover as an ILP

$$\begin{aligned} \textbf{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \textbf{subject to} \quad & \sum_{S \ni u} x_S \geq 1 && \forall u \in U \\ & x_S \in \{0,1\} && \forall S \in \mathcal{S} \end{aligned}$$

Ground set U

Family $\mathcal{S} \subseteq 2^U$ with $\bigcup \mathcal{S} = U$

Costs $c \colon \mathcal{S} \to \mathbb{Q}^+$



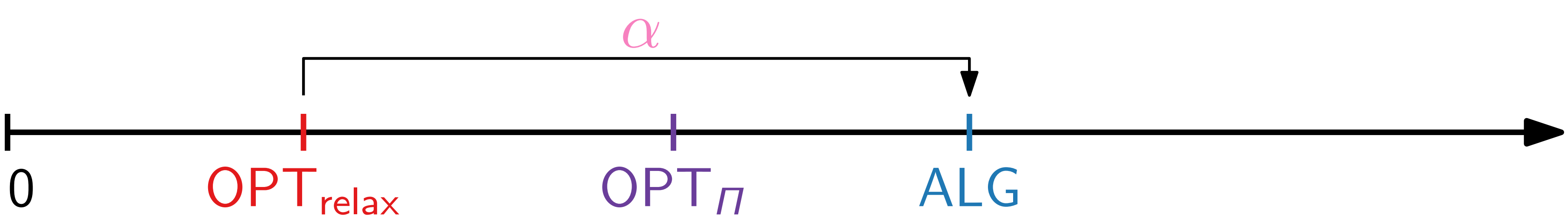
Find cover $\mathcal{S}' \subseteq \mathcal{S}$ of U with minimum cost.

Approximation Algorithms

Lecture 5:
LP-based Approximation Algorithms for SetCover

Part II:
LP-Rounding

Technique I) LP-Rounding



Consider a minimization problem Π in ILP form.

Compute a solution for the LP-relaxation.

Round to obtain an integer solution for Π.

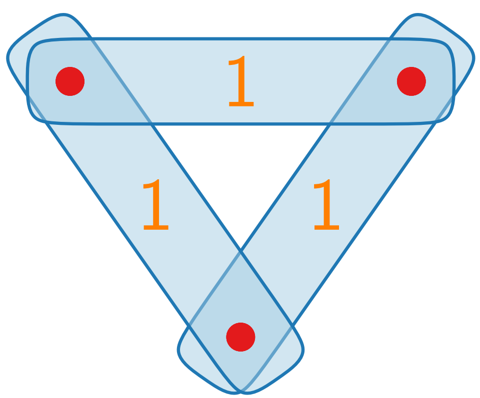
Difficulty: Ensure the **feasiblity** of the solution.

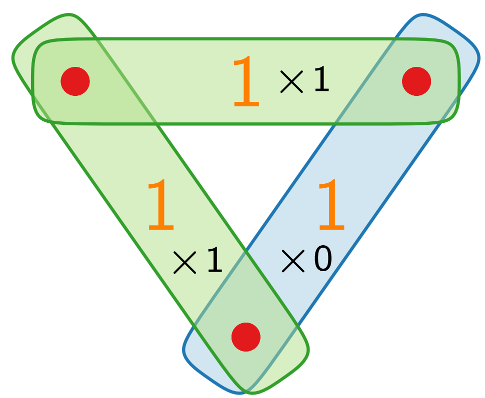
Approximation factor: $\quad \mathrm{ALG}/\mathrm{OPT}_{\Pi} \leq \mathrm{ALG}/\mathrm{OPT}_{\mathrm{relax}}$.

SetCover – LP-Relaxation

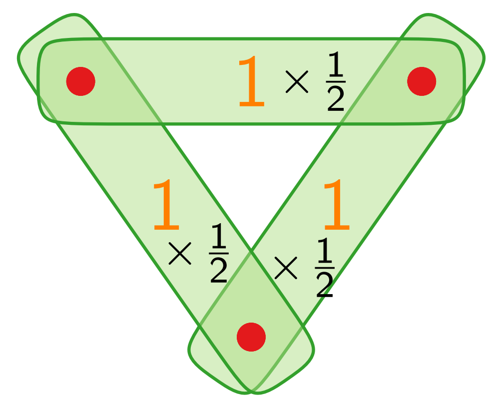
$$\begin{aligned}
\textbf{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\
\textbf{subject to} \quad & \sum_{S \ni u} x_S \geq 1 \quad \forall u \in U \\
& x_S \geq 0 \quad \forall S \in \mathcal{S}
\end{aligned}$$

Optimal?





integer: 2



fractional: $\frac{3}{2}$

LP-Rounding: Approach I

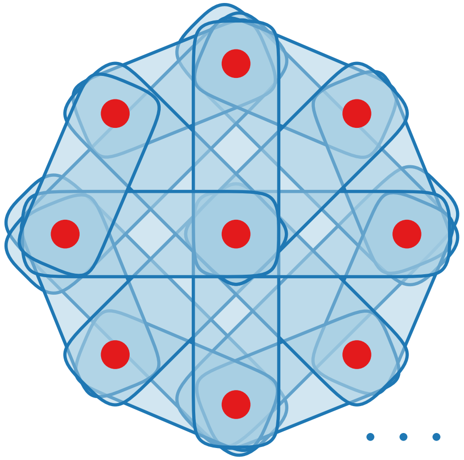
$$\begin{aligned} \textbf{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \textbf{subject to} \quad & \sum_{S \ni u} x_S \geq 1 \quad \forall u \in U \\ & x_S \geq 0 \quad \forall S \in \mathcal{S} \end{aligned}$$

LP-Rounding-One($U, \mathcal{S}, c$)

Compute optimal solution x for LP-relaxation.
Round each x_S with $x_S > 0$ to 1.

– Generates a feasible solution.
– Scaling factor arbitrarily large.

Use frequency f



LP-Rounding: Approach II

$$\begin{aligned} \textbf{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \textbf{subject to} \quad & \sum_{S \ni u} x_S \geq 1 \quad \forall u \in U \\ & x_S \geq 0 \quad \forall S \in \mathcal{S} \end{aligned}$$

LP-Rounding-Two($U, \mathcal{S}, c$)

Compute optimal solution x for LP-relaxation.
Round each x_S with $x_S \geq 1/f$ to 1; remaining to 0.

Let f be the frequency of (i.e., the number of sets containing) the most frequent element.

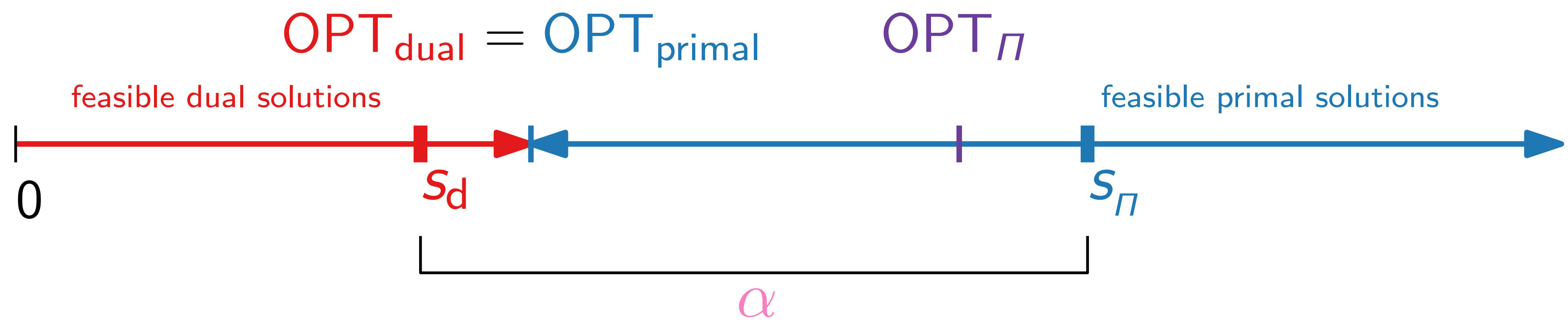
Theorem. LP-Rounding-Two is a factor-f approximation algorithm for SetCover.

Approximation Algorithms

Lecture 5:
LP-based Approximation Algorithms for SetCover

Part III:
The Primal-Dual Schema

Technique II) Primal–Dual Approach



Consider a minimization problem Π in ILP form.

- Start with (trivial) feasible dual solution and infeasible primal solution (e.g., all variables $= 0$).
- Compute dual solution s_{d} and integral primal solution s_{Π} for Π iteratively:
  Increase s_{d} according to CS and make s_{Π} "more feasible".

Approximation factor $\leq \mathrm{obj}(s_{\Pi})/\mathrm{obj}(s_{\mathrm{d}})$

Advantage: Don't need LP-"machinery"; possibly faster, more flexible.

SetCover – Dual LP

$$\begin{aligned} \textbf{minimize} \quad & \sum_{S \in \mathcal{S}} c_S x_S \\ \textbf{subject to} \quad & \sum_{S \ni u} x_S \geq 1 \quad \forall u \in U \\ & x_S \geq 0 \quad \forall S \in \mathcal{S} \end{aligned}$$

$$\begin{aligned} \textbf{maximize} \quad & \sum_{u \in U} y_u \\ \textbf{subject to} \quad & \sum_{u \in S} y_u \leq c_S \quad \forall S \in \mathcal{S} \\ & y_u \geq 0 \quad \forall u \in U \end{aligned}$$

Complementary Slackness

minimize	$c^\intercal x$		
subject to	Ax	$\geq$	b
	x	$\geq$	0

maximize	$b^\intercal y$		
subject to	$A^\intercal y$	$\leq$	c
	y	$\geq$	0

Theorem. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$ be valid solutions for the primal and dual program, respectively.
Then x and y are optimal $\Leftrightarrow$ following conditions are met:

Primal CS:
For each $j = 1, \ldots, n$: $\quad x_j = 0$ or $\sum_{i=1}^{m} a_{ij} y_i = c_j$

Dual CS:
For each $i = 1, \ldots, m$: $\quad y_i = 0$ or $\sum_{j=1}^{n} a_{ij} x_j = b_i$

Relaxing Complementary Slackness

minimize	$c^\intercal x$		
subject to	Ax	$\geq$	b
	x	$\geq$	0

maximize	$b^\intercal y$		
subject to	$A^\intercal y$	$\leq$	c
	y	$\geq$	0

~~**Primal CS**~~: Relaxed Primal CS

For each $j = 1, \ldots, n$: $x_j = 0$ or ~~$\sum_{i=1}^{m} a_{ij} y_i = c_j$~~

$$c_j/\alpha \leq \sum_{i=1}^{m} a_{ij} y_i \leq c_j$$

~~**Dual CS**~~: Relaxed Dual CS

For each $i = 1, \ldots, m$: $y_i = 0$ or ~~$\sum_{j=1}^{n} a_{ij} x_j = b_i$~~

$$b_i \leq \sum_{j=1}^{n} a_{ij} x_j \leq \beta \cdot b_i$$

$$\Leftrightarrow \cancel{\sum_{j=1}^{n} c_j x_j = \sum_{i=1}^{m} b_i y_i} \quad \Rightarrow \sum_{j=1}^{n} c_j x_j \leq \alpha\beta \sum_{i=1}^{m} b_i y_i \leq \alpha\beta \cdot \mathrm{OPT}_{\mathrm{LP}}$$

Primal–Dual Schema

Start with a feasible dual and infeasible primal solution (often trivial).

"Improve" the feasibility of the primal solution...

...and simultaneously the objective value of the dual solution.

Do so until the relaxed CS conditions are met.

Maintain that the primal solution is integer-valued.

The feasibility of the primal solution and the relaxed CS conditions provide an approximation ratio.

Relaxed CS for SetCover

minimize $\sum_{S \in \mathcal{S}} c_S x_S$	**maximize** $\sum_{u \in U} y_u$
subject to $\sum_{S \ni u} x_S \geq 1 \quad \forall u \in U$	**subject to** $\sum_{u \in S} y_u \leq c_S \quad \forall S \in \mathcal{S}$
$x_S \geq 0 \quad \forall S \in \mathcal{S}$	$y_u \geq 0 \quad \forall u \in U$

critical set

(Unrelaxed) primal CS: $x_S \neq 0 \Rightarrow \sum_{u \in S} y_u = c_S$

only chooses critical sets

trivial for binary x

Relaxed dual CS: $y_u \neq 0 \Rightarrow 1 \leq \sum_{S \ni u} x_S \leq f \cdot 1$

Primal–Dual Schema for SetCover

PrimalDualSetCover($U, \mathcal{S}, c$)

$x \leftarrow 0, y \leftarrow 0$

repeat

Select an uncovered element u.

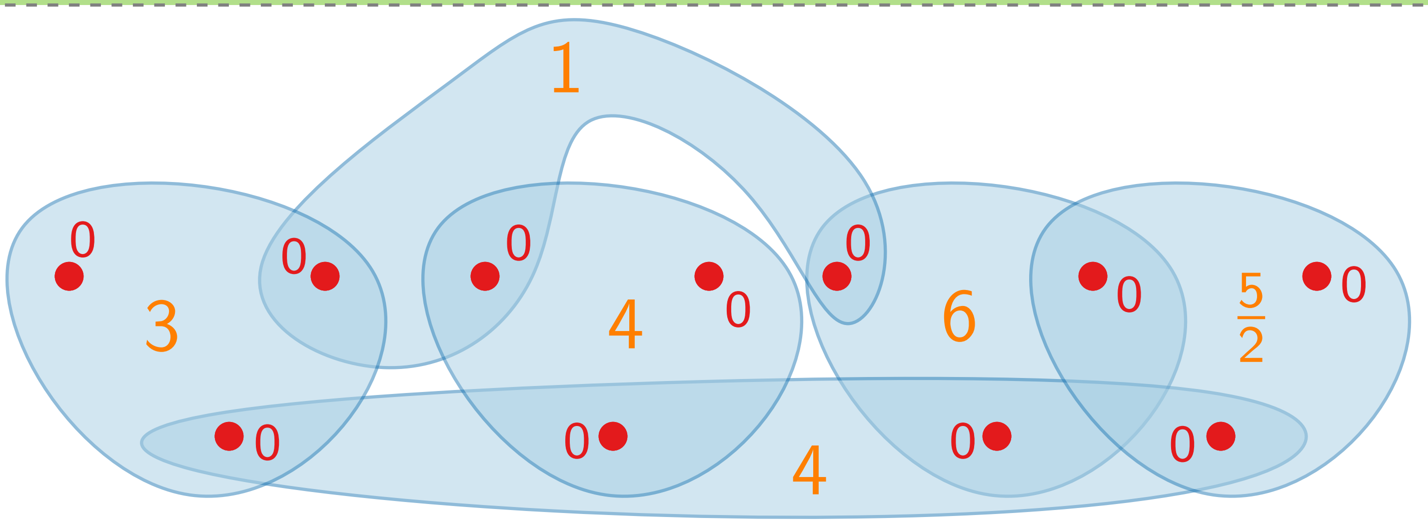
Increase y_u until a set S is critical ($\sum_{u' \in S} y_{u'} = c_S$).

Select all critical sets and update x.

Mark all elements in these sets as covered.

until all elements are covered.

return x

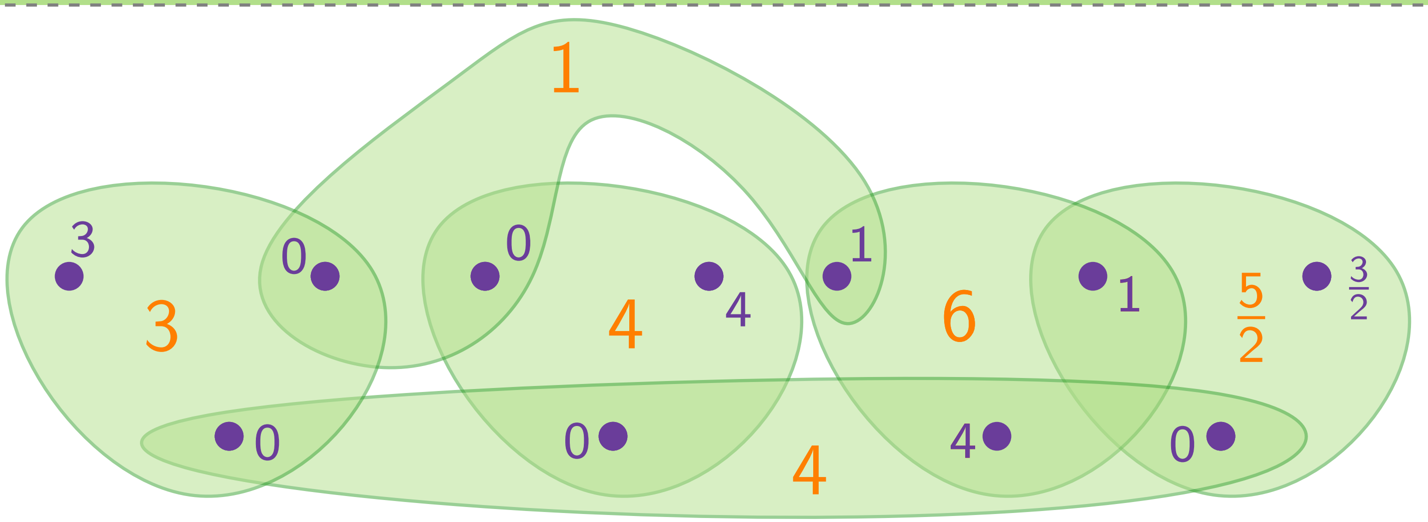


Primal–Dual Schema for SetCover

PrimalDualSetCover($U, \mathcal{S}, c$)

$x \leftarrow 0$, $y \leftarrow 0$
repeat
- Select an uncovered element u.
- Increase y_u until a set S is critical ($\sum_{u' \in S} y_{u'} = c_S$).
- Select all critical sets and update x.
- Mark all elements in these sets as covered.

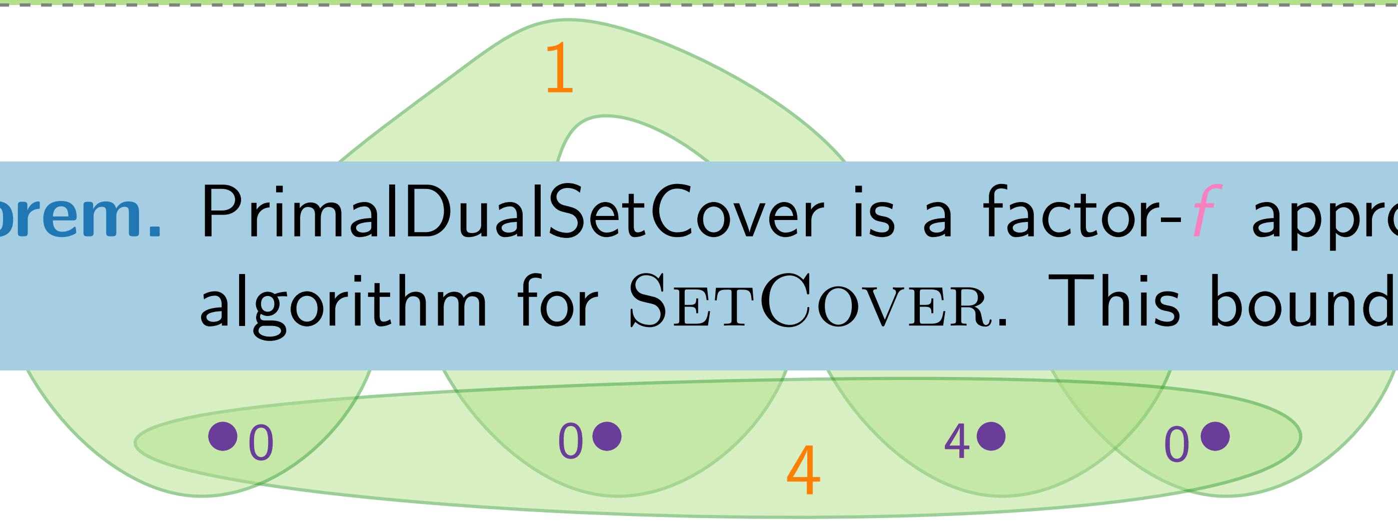
until all elements are covered.
return x



Primal–Dual Schema for SetCover

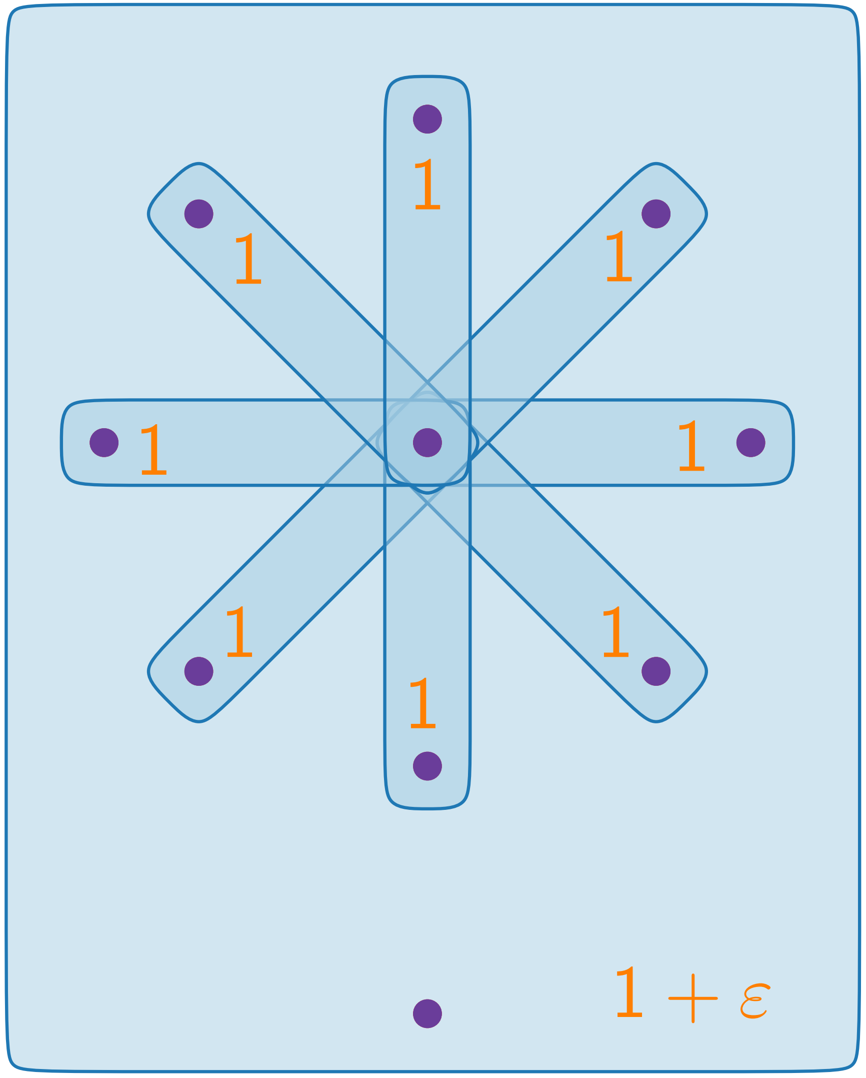
PrimalDualSetCover($U, \mathcal{S}, c$)

$x \leftarrow 0$, $y \leftarrow 0$
repeat
  Select an uncovered element u.
  Increase y_u until a set S is critical ($\sum_{u' \in S} y_{u'} = c_S$).
  Select all critical sets and update x.
  Mark all elements in these sets as covered.
until all elements are covered.
return x

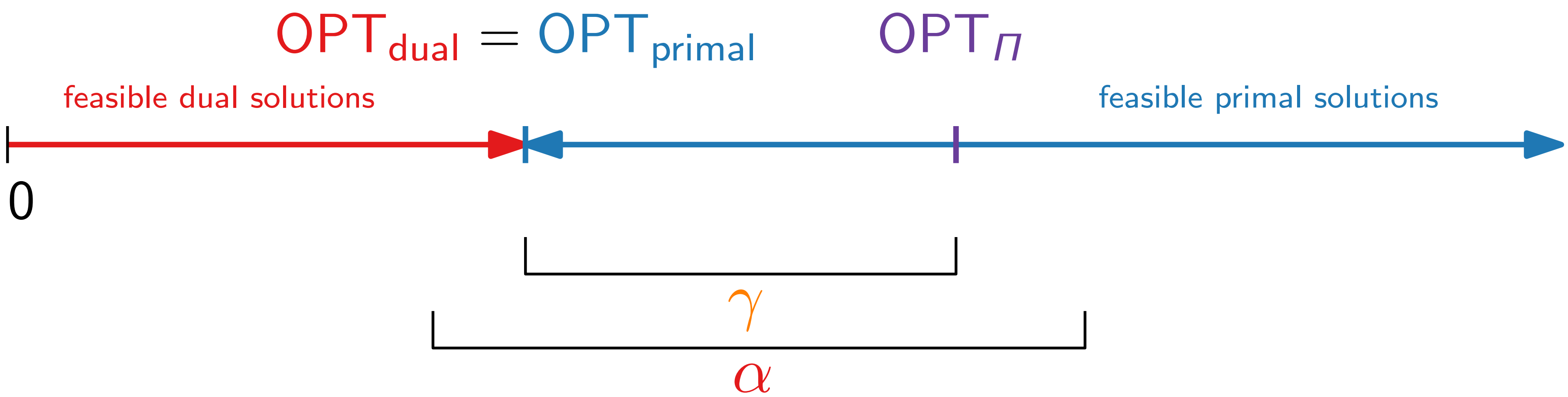


Theorem. PrimalDualSetCover is a factor-f approximation algorithm for SetCover. This bound is tight.

Tight Example



Integrality Gap



Consider a minimization problem Π in ILP form.

Dual methods (without outside help) are limited by the *integrality gap* of the LP-relaxation:

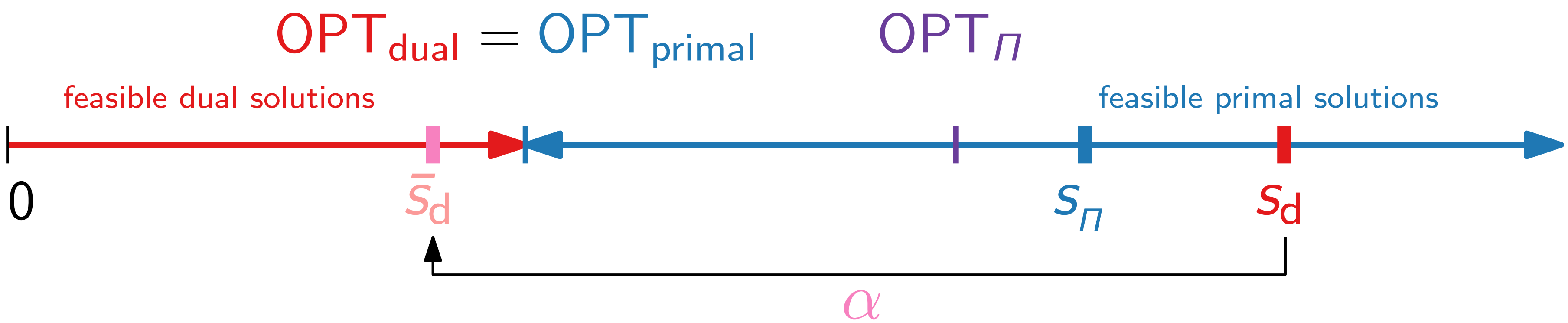
$$\alpha \geq \gamma = \sup_{I} \frac{\mathrm{OPT}_{\Pi}(I)}{\mathrm{OPT}_{\mathrm{primal}}(I)}$$

Approximation Algorithms

Lecture 5:
LP-based Approximation Algorithms for SetCover

Part IV:
Dual Fitting

Technique III) Dual Fitting



Consider a minimization problem Π in ILP form.

Combinatorial algorithm (e.g., greedy) computes feasible primal solution s_{Π} and infeasible dual solution s_{d} that completely "pays" for s_{Π}, i.e., $\mathrm{obj}(s_{\Pi}) \leq \mathrm{obj}(s_{\mathrm{d}})$.

Scale the dual variables $\rightsquigarrow$ feasible dual solution $\bar{s}_{\mathrm{d}}$.

$\Rightarrow$ $\mathrm{obj}(s_{\Pi})/\alpha \leq \mathrm{obj}(s_{\mathrm{d}})/\alpha = \mathrm{obj}(\bar{s}_{\mathrm{d}}) \leq \mathrm{OPT}_{\mathrm{dual}} \leq \mathrm{OPT}_{\Pi}$

$\Rightarrow$ Scaling factor α is approximation factor :-)

Dual Fitting for SetCover

Combinatorial (greedy) algorithm (see Lecture #2):

```
GreedySetCover(universe U, 𝒮 ⊆ 2^U, costs c: 𝒮 → ℚ_{≥0})
  C ← ∅
  𝒮' ← ∅
  while C ≠ U do
      S ← set from 𝒮 that minimizes c(S)/|S \ C|
      foreach u ∈ S \ C do
          price(u) ← c(S)/|S \ C|
      C ← C ∪ S
      𝒮' ← 𝒮' ∪ {S}
  return 𝒮'                                   // Cover of U
```

Reminder: $\sum_{u \in U} \text{price}(u)$ completely pays for $\mathcal{S}'$.

New: LP-based Analysis

Observation. For each $u \in U$, $\text{price}(u)$ is a dual variable

But this dual solution is in general not feasible.

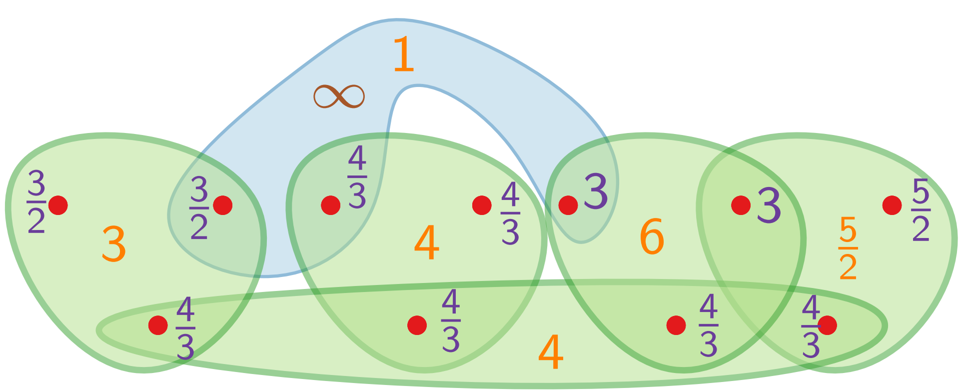
Homework exercise: Construct instance where some S are "overpacked" by factor $\approx \mathcal{H}_{|S|}$.

Dual-fitting trick:

Scale dual variables such that no set is overpacked.

Take $\bar{y}_u = \text{price}(u)/\mathcal{H}_k$. ($k =$ cardinality of largest set in $\mathcal{S}$.)

The greedy algorithm uses *these* dual variables as lower bound for OPT.



$$\begin{aligned} \textbf{maximize} \quad & \sum_{u \in U} y_u \\ \textbf{subject to} \quad & \sum_{u \in S} y_u \le c_S \quad \forall S \in \mathcal{S} \\ & y_u \ge 0 \quad \forall u \in U \end{aligned}$$

Proof. To prove: No set is overpacked by $\bar{y}$.

Let $S \in \mathcal{S}$ and $\ell = |S| \le k$.

Let $u_1, \dots, u_\ell$ be the elements of S –
in the order in which they are covered by greedy.

Consider the iteration in which u_i is covered.

Before that, $\ge \ell - i + 1$ elem. of S are uncovered.

So $\text{price}(u_i) \le c(S)/(\ell - i + 1)$.

$$\Rightarrow \bar{y}_{u_i} \le \frac{c(S)}{\mathcal{H}_k} \cdot \frac{1}{\ell - i + 1} \Rightarrow \sum_{i=1}^{\ell} \bar{y}_{u_i} \le \frac{c(S)}{\mathcal{H}_k} \cdot \overbrace{\left(\frac{1}{\ell} + \dots + \frac{1}{1}\right)}^{= \mathcal{H}_\ell \le \mathcal{H}_k}$$

$$\le c(S) \qquad \square$$

Lemma.
The vector $\bar{y} = (\bar{y}_u)_{u \in U}$ is a feasible solution for the dual LP.

maximize $\sum_{u \in U} y_u$

subject to $\sum_{u \in S} y_u \le c_S \quad \forall S \in \mathcal{S}$

$y_u \ge 0 \quad \forall u \in U$

Result for Dual Fitting

Theorem. GreedySetCover is a factor-$\mathcal{H}_k$ approximation algorithm for $\textsc{SetCover}$, where $k = \max_{S \in \mathcal{S}} |S|$.

Proof.

$$\text{ALG} = c(\mathcal{S}') \leq \sum_{u \in U} \text{price}(u) = \mathcal{H}_k \cdot \sum_{u \in U} \bar{y}_u \leq$$

$$\leq \mathcal{H}_k \cdot \text{OPT}_{\text{relax}}$$

$$\leq \mathcal{H}_k \cdot \text{OPT} \qquad \square$$

Strengthened bound with respect to $\text{OPT}_{\text{relax}} \leq \text{OPT}$.

Dual solution allows a *per-instance* estimation $c(\mathcal{S}')/\text{OPT}_{\text{relax}}$

of the quality of the greedy solution

…which may be stronger than the worst-case bound $\mathcal{H}_k$:

$\text{ALG}/\text{OPT} \leq \text{ALG}/\text{OPT}_{\text{relax}} \leq \sum_{u \in U} \text{price}(u)/\text{OPT}_{\text{relax}} \leq \mathcal{H}_k$.