Set Cover via Dual Fitting

Approximation Algorithms (V.V.Vazirani)

Chapter 13

Presentation: Evangelos Bampas

Set Cover

- Input:
  - $\mathcal{U}$: universe, $|\mathcal{U}| = n$,
  - $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$: a collection of subsets of $\mathcal{U}$,
  - $c : \mathcal{S} \to \mathbb{Q}^+$: a cost function.
- Feasible solution:
  - a subset $\mathcal{T}$ of $\mathcal{S}$ such that $\bigcup_{t \in \mathcal{T}} t = \mathcal{U}$.
- Goal:
  - minimize the cost of $\mathcal{T}$: $\sum_{t \in \mathcal{T}} c(t)$.

A Greedy algorithm for Set Cover

$C \leftarrow \emptyset$.

while $C \neq \mathcal{U}$ **do**

find the most cost-effective set in the current iteration, say S.

let $\alpha = \frac{c(S)}{|S-C|}$, i.e. the cost-effectiveness of S.

pick S, and for each $e \in S - C$ set $\text{price}(e) = \alpha$.

$C \leftarrow C \cup S$.

end while

output the picked sets.

Dual Fitting

- Solve the given IP (opt. sol. OPT) with some combinatorial algorithm (yielding sol. SOL).

- Find a solution of cost D for the dual program of the corresponding LP relaxation (opt. sol. $\mathrm{OPT_f}$), and a suitable quantity r such that:
  - $\mathrm{SOL} \leq \mathrm{D}$
  - $\frac{\mathrm{D}}{r} \leq \mathrm{OPT_f}$ *[fitting]*

- $\mathrm{SOL} \leq \mathrm{D} \leq r \cdot \mathrm{OPT_f} \leq r \cdot \mathrm{OPT}$, therefore the algorithm has an approximation factor of r.

Analysis of the Greedy algorithm for Set Cover

IP formulation of Set Cover:

- minimize $\sum_{S \in \mathcal{S}} c(S) \cdot x_S$
- subject to:
  - $\sum_{S : e \in S} x_S \geq 1$, $e \in \mathcal{U}$
  - $x_S \in \{0, 1\}$, $S \in \mathcal{S}$

LP relaxation:

- minimize $\sum_{S \in \mathcal{S}} c(S) \cdot x_S$
- subject to:
  - $\sum_{S : e \in S} x_S \geq 1$, $e \in \mathcal{U}$
  - $x_S \geq 0$, $S \in \mathcal{S}$

Dual program:

- maximize $\sum_{e \in \mathcal{U}} y_e$
- subject to:
  - $\sum_{e \in S} y_e \leq c(S)$, $S \in \mathcal{S}$
  - $y_e \geq 0$, $e \in \mathcal{U}$

Analysis of the Greedy algorithm for Set Cover

$$\text{SOL} = \sum_{e \in \mathcal{U}} \text{price}\,(e)$$

To obtain a solution for the dual, we set $y_e = \text{price}\,(e)$ for each $e \in \mathcal{U}$. Therefore, $\text{D} = \text{SOL}$. In general this is not a feasible solution.

Dividing it by $r = H_n$ yields a feasible solution.

Lemma 1 *The vector* $\mathbf{y}$ *with* $y_e = \frac{\text{price}(e)}{H_n}$ *for each* $e \in \mathcal{U}$ *is a feasible solution for the dual program.*

Therefore, $\text{SOL} = \text{D} \leq H_n \cdot \text{OPT}_\text{f} \leq H_n \cdot \text{OPT}$.

Analysis of the Greedy algorithm for Set Cover

Proof

Consider $S \in \mathcal{S}$ and its elements $e_1, e_2, \ldots, e_i, \ldots, e_k$ in the order in which they are covered by the algorithm.

Before the iteration in which e_i is covered, S contains at least $k - i + 1$ uncovered elements. Therefore, e_i can be covered with an average cost of at most $\frac{c(S)}{k-i+1}$.

$$\mathrm{price}\,(e_i) \leq \frac{c\,(S)}{k-i+1} \Rightarrow y_{e_i} \leq \frac{1}{H_n} \cdot \frac{c\,(S)}{k-i+1}$$

Summing over all elements in S, we get

$$\sum_{e \in S} y_e \leq \frac{c\,(S)}{H_n} \cdot \sum_{i=1}^{k} \frac{1}{k-i+1} = \frac{H_k}{H_n} \cdot c\,(S) \leq c\,(S)$$

Tight Example

Universe: $\mathcal{U} = \{e_1, e_2, \ldots, e_n\}$, where $n = 2^k - 1$.

Sets: $S_i = \{e_j : \mathbf{i} \cdot \mathbf{j} = 1\}$, $i = 1, \ldots, n$, each of cost 1.

Each set contains $2^{k-1} = \frac{n+1}{2}$ elements, and each element belongs to $\frac{n+1}{2}$ sets. A fractional cover is: $x_i = \frac{2}{n+1}$. Its cost is $\frac{2n}{n+1}$.

Any integral set cover must pick at least $k = \log(n+1)$ sets.

$$\frac{\mathrm{OPT}}{\mathrm{OPT_f}} \geq \frac{\log n}{2} \cdot \frac{n+1}{n} > \frac{H_n}{2}$$

Generalizations of Set Cover

- Set Multicover: Each element e has to be covered r_e times, possibly picking the same set $k \geq 1$ times at a cost of $k \cdot c(S)$.
- Multiset Multicover: Instead of sets, we consider multisets where the multiplicity of each element in any set does not exceed its coverage requirement: $M(S, e) \leq r_e$.
- Covering integer programs: minimize $\mathbf{c} \cdot \mathbf{x}$, subject to $\mathbf{A} \cdot \mathbf{x} \geq \mathbf{b}$. All entries in $\mathbf{A}, \mathbf{b}, \mathbf{c}$ are nonnegative and $\mathbf{x}$ is required to be nonnegative and integral.

A Greedy algorithm for constrained Set Multicover

consider a multiset $\mathcal{U}_m$ where each element e has multiplicity r_e.

multiset $C \leftarrow \emptyset$.

while $C \neq \mathcal{U}$ **do**

  compute the average cost per new element of each unpicked

    $S \in \mathcal{S}$ in this iteration: $\overline{c}\,(S) = \frac{c(S)}{|S \cap (\mathcal{U}_m - C)|}$.

  pick the most cost-effective set, say S.

  for each $e \in S \cap (\mathcal{U}_m - C)$ set price$\,(e, j_e) = \overline{c}\,(S)$.

  $C \leftarrow C \cup S$.

end while

output the picked sets.

Analysis of the Greedy algorithm for constrained Set Multicover

IP formulation:

- minimize $\sum_{S\in\mathcal{S}} c(S)\cdot x_S$
- subject to:
  - $\sum_{S:e\in S} x_S \geq r_e$, $e\in\mathcal{U}$
  - $x_S\in\{0,1\}$, $S\in\mathcal{S}$

LP relaxation:

- minimize $\sum_{S\in\mathcal{S}} c(S)\cdot x_S$
- subject to:
  - $\sum_{S:e\in S} x_S \geq r_e$, $e\in\mathcal{U}$
  - $-x_S \geq -1$, $S\in\mathcal{S}$
  - $x_S \geq 0$, $S\in\mathcal{S}$

Dual program:

- maximize $\sum_{e\in\mathcal{U}} r_e\cdot y_e - \sum_{S\in\mathcal{S}} z_S$
- subject to:
  - $\sum_{e\in S} y_e - z_S \leq c(S)$, $S\in\mathcal{S}$
  - $y_e \geq 0$, $e\in\mathcal{U}$
  - $z_S \geq 0$, $S\in\mathcal{S}$

Analysis of the Greedy algorithm for constrained Set Multicover

$$\text{SOL} = \sum_{e \in \mathcal{U}} \sum_{j=1}^{r_e} \text{price}\,(e, j)$$

Consider the objective function value D of the dual solution (α, β), where $\alpha_e = \text{price}\,(e, r_e)$ for each $e \in \mathcal{U}$ and

$$\beta_S = \sum_{e \text{ covered by } S} (\text{price}\,(e, r_e) - \text{price}\,(e, j_e))\,,$$

for each $S \in \mathcal{S}$ that was picked by the algorithm and $\beta_S = 0$ otherwise.

$\text{D} = \sum_{e \in \mathcal{U}} r_e \cdot \text{price}\,(e, r_e) -$
$\sum_{e \in \mathcal{U}} \left(r_e \cdot \text{price}\,(e, r_e) - \sum_{j=1}^{r_e} \text{price}\,(e, j) \right) = \text{SOL}$

Analysis of the Greedy algorithm for constrained Set Multicover

Lemma 2 *The pair* $(\mathbf{y}, \mathbf{z})$ *where* $y_e = \frac{\alpha_e}{H_n}$ *and* $z_S = \frac{\beta_S}{H_n}$ *is a feasible solution for the dual program.*

Therefore, $\frac{\mathrm{D}}{H_n} \leq \mathrm{OPT_f} \Rightarrow \mathrm{SOL} \leq H_n \cdot \mathrm{OPT}$ which implies an approximation factor of H_n for the previous algorithm.

Analysis of the Greedy algorithm for constrained Set Multicover

Proof

Consider $S \in \mathcal{S}$ and its elements $e_1, e_2, \ldots, e_i, \ldots, e_k$ in the order in which their requirements are covered by the algorithm.

Case 1: S is not picked

Just before the last copy of e_i is covered, S contains at least $k - i + 1$ alive elements. So, $\text{price}\,(e_i, r_{e_i}) \leq \frac{c(S)}{k-i+1}$. Moreover, since $z_S = 0$ we get:

$$\sum_{e \in S} y_e - z_S \leq \frac{1}{H_n} \cdot \sum_{i=1}^{k} \frac{c\,(S)}{k-i+1} \leq c\,(S)$$

Analysis of the Greedy algorithm for constrained Set Multicover

Case 2: S is picked

Assume that just before S is picked, k' of its elements are completely covered.

$$\sum_{e\in S} y_e - z_S = \frac{1}{H_n}\cdot\left(\sum_{i=1}^{k}\operatorname{price}(e_i, r_{e_i}) - \sum_{i=k'+1}^{k}\left(\operatorname{price}(e_i, r_{e_i}) - \operatorname{price}(e_i, j_i)\right)\right)$$

$$= \frac{1}{H_n}\cdot\left(\sum_{i=1}^{k'}\operatorname{price}(e_i, r_{e_i}) + \sum_{i=k'+1}^{k}\operatorname{price}(e_i, j_i)\right) \leq c(S)$$

Rounding Applied to Set Cover

Approximation Algorithms (V.V.Vazirani)

Chapter 14

Presentation: Evangelos Bampas

Simple Rounding

We convert an (optimal) fractional solution for Set Cover into an integral solution by picking only those sets S for which $x_S \geq \frac{1}{f}$.

Since in the optimal solution we have $0 \leq x_S \leq 1$ for all S, it follows that $\mathrm{SOL} \leq f \cdot \mathrm{OPT_f}$.

Moreover, the sets picked in this manner form a valid set cover. For any $e \in \mathcal{U}$, we have

$$\sum_{S:e\in S} x_S \geq 1 \Rightarrow f \cdot \max_{S:e\in S} x_S \geq 1 \Rightarrow \max_{S:e\in S} x_S \geq \frac{1}{f}$$

Therefore, e is covered by at least one set in the integral set cover.

Tight Example

Let $V_1, \ldots, V_k$ be k disjoint sets with n elements each.

The instance has universe $\mathcal{U} = V_1 \cup \ldots \cup V_k$ and all n^k possible sets which contain exactly one element from each V_i. The cost of each set is 1.

$$f = n^{k-1}$$

$$\mathrm{OPT_f} = \frac{1}{n^{k-1}}$$

$$\mathrm{SOL} = n^k$$

$$\mathrm{OPT} = n \Rightarrow \mathrm{SOL} = f \cdot \mathrm{OPT}$$

Randomized Rounding algorithm for Set Cover

solve the LP relaxation and treat the values of the fractional solution $\mathbf{x} = \left(x_1, \ldots, x_{|\mathcal{S}|}\right)$ as a probability vector.

for each set $S \in \mathcal{S}$, include S in the integral set cover C with probability x_S.

repeat the above step t times and take the union C' of all created covers.

- The solution returned by this algorithm need not be feasible. We obtain bounds on its expected cost and on the probability that it is not feasible.

Analysis of the Randomized Rounding algorithm

The expected cost of the cover C returned by a single iteration is:

$$\mathbf{E}\left[c\left(C\right)\right] = \sum_{S \in \mathcal{S}} \mathbf{Pr}\left[S \in C\right] \cdot c\left(S\right) = \sum_{S \in \mathcal{S}} x_S \cdot c\left(S\right) = \mathrm{OPT_f}$$

Now, consider an element e and let P_e be the probability that e is not covered by C. Suppose that e belongs to k of the sets in $\mathcal{S}$ and let $p_1, \ldots, p_k$ be the probabilities with which each of them was included in C. We wish to obtain an upper bound on P_e. Since e was fractionally covered in the optimal solution of the LP relaxation, it must be that $p_1 + \ldots + p_k = d \geq 1$.

Analysis of the Randomized Rounding algorithm

$$P_e = \prod_{i=1}^{k} (1 - p_i)$$

We shall find the values of p_i for which $L = \log P_e = \sum_{i=1}^{k} \log(1 - p_i)$ is maximized. Substituting $d - p_1 - \ldots - p_{k-1}$ for p_k in L, we have that for each $i \neq k$:

$$\frac{\partial L}{\partial p_i} = -\frac{1}{1 - p_i} + \frac{1}{1 - d + p_1 + \ldots + p_{k-1}} = -\frac{1}{1 - p_i} + \frac{1}{1 - p_k}$$

Therefore at the maximum,
$p_1 = \ldots = p_k = p \Rightarrow k \cdot p = d \geq 1 \Rightarrow p \geq \frac{1}{k}$. So,
$P_e = (1 - p)^k \leq \left(1 - \frac{1}{k}\right)^k \leq \frac{1}{e}$.

Analysis of the Randomized Rounding algorithm

The probability that e is not covered in C' is at most $\left(\frac{1}{e}\right)^t$. So, the probability that C' is not feasible is at most $n \cdot \left(\frac{1}{e}\right)^t$. The expected cost of C' is $\mathbf{E}\left[c\left(C'\right)\right] \leq t \cdot \mathbf{E}\left[c\left(C\right)\right] = t \cdot \mathrm{OPT}_{\mathrm{f}}$.

Pick t such that $n \cdot \left(\frac{1}{e}\right)^t \leq \frac{1}{4} \Rightarrow t = O\left(\log n\right)$. Then, the probability that C' is not valid is at most $\frac{1}{4}$ and, by Markov's inequality:

$$\mathbf{Pr}\left[c\left(C'\right) \geq 4t \cdot \mathrm{OPT}_{\mathrm{f}}\right] \leq \frac{\mathbf{E}\left[c\left(C'\right)\right]}{4t \cdot \mathrm{OPT}_{\mathrm{f}}} \leq \frac{1}{4}$$

Therefore, the probability that the algorithm outputs a valid set cover with cost no more than $O\left(\log n\right) \cdot \mathrm{OPT}$ is at least $\frac{1}{2}$.

Half-integrality of Vertex Cover

IP formulation:

- minimize $\sum_{v \in V} c(v) \cdot x_v$
- subject to:
  - $x_u + x_v \geq 1$, $(u, v) \in E$
  - $x_v \in \{0, 1\}$, $v \in V$

LP relaxation:

- minimize $\sum_{v \in V} c(v) \cdot x_v$
- subject to:
  - $x_u + x_v \geq 1$, $(u, v) \in E$
  - $x_v \geq 0$, $v \in V$

Extreme point solution: a feasible solution that cannot be expressed as convex combination of two other feasible solutions.

Half-integral solution: a feasible solution in which each variable is 0, 1, or $\frac{1}{2}$.

Half-integrality of Vertex Cover

Lemma 3 *Let* $\mathbf{x}$ *be a feasible, non-half-integral solution for the LP relaxation. Then* $\mathbf{x}$ *is not an extreme point solution.*

Proof

Let V_- be the set of vertices to which $\mathbf{x}$ assigns a non-half-integral value less than $\frac{1}{2}$ and V_+ be the set of vertices to which $\mathbf{x}$ assigns a non-half-integral value greater than $\frac{1}{2}$. For $\varepsilon > 0$ define the following vectors:

$$\mathbf{y} : y_v = \begin{cases} x_v + \varepsilon & , v \in V_+ \\ x_v - \varepsilon & , v \in V_- \\ x_v & , \text{otherwise} \end{cases} , \mathbf{z} : z_v = \begin{cases} x_v - \varepsilon & , v \in V_+ \\ x_v + \varepsilon & , v \in V_- \\ x_v & , \text{otherwise} \end{cases}$$

$\mathbf{y}, \mathbf{z}$ are feasible solutions, and $\mathbf{x} = \frac{1}{2} \cdot (\mathbf{y} + \mathbf{z})$.

Half-integrality of Vertex Cover

Since any extreme point solution of the LP relaxation is half-integral, we immediately have a factor 2 approximation algorithm for Vertex Cover.

The algorithm finds an extreme point solution $\mathbf{x}$ of the LP relaxation and picks those vertices v for which $x_v = \frac{1}{2}$ or $x_v = 1$.

Set Cover via the Primal-Dual Schema

Approximation Algorithms
(V.V.Vazirani)

Chapter 15

Presentation: Evangelos Bampas

Complementary Slackness Conditions

Primal

- minimize $\sum_{j=1}^{n} c_j \cdot x_j$
- subject to:
  - $\sum_{j=1}^{n} a_{ij} \cdot x_j \geq b_i$, $i = 1, \ldots, m$
  - $x_j \geq 0$, $j = 1, \ldots, n$

Dual

- maximize $\sum_{i=1}^{m} b_i \cdot y_i$
- subject to:
  - $\sum_{i=1}^{m} a_{ij} \cdot y_i \leq c_j$, $j = 1, \ldots, n$
  - $y_i \geq 0$, $i = 1, \ldots, m$

Primal Complementary Slackness Conditions

For each $1 \leq j \leq n$: either $x_j = 0$ or $\sum_{i=1}^{m} a_{ij} \cdot y_i = c_j$

Dual Complementary Slackness Conditions

For each $1 \leq i \leq m$: either $y_i = 0$ or $\sum_{j=1}^{n} a_{ij} \cdot x_j = b_i$

If $\mathbf{x}$ and $\mathbf{y}$ are primal and dual feasible, respectively, and they satisfy the complementary slackness conditions then they are both optimal.

Relaxed Complementary Slackness Conditions

Relaxed Primal Complementary Slackness Conditions

For $\alpha \geq 1$, for each $1 \leq j \leq n$: either $x_j = 0$ or $\frac{c_j}{\alpha} \leq \sum_{i=1}^{m} a_{ij} \cdot y_i \leq c_j$

Relaxed Dual Complementary Slackness Conditions

For $\beta \geq 1$, for each $1 \leq i \leq m$: either $y_i = 0$ or $b_i \leq \sum_{j=1}^{n} a_{ij} \cdot x_j \leq \beta \cdot b_i$

If $\mathbf{x}$ and $\mathbf{y}$ are primal and dual feasible, respectively, and they satisfy the relaxed complementary slackness conditions then:

$$\sum_{j=1}^{n} c_j \cdot x_j \leq \alpha \cdot \beta \cdot \sum_{i=1}^{m} b_i \cdot y_i$$

Overview of the Primal-Dual schema

- We construct an approximation algorithm for a given IP by ensuring one set of conditions (for example, $\alpha := 1$) and suitably relaxing the other.
- The algorithm starts with a primal infeasible solution and a dual feasible solution (usually $\mathbf{x} = \mathbf{0}$, $\mathbf{y} = \mathbf{0}$).
- It iteratively improves the feasibility of the primal solution and the optimality of the dual solution.
- In the end, we obtain $\mathbf{x}'$ and $\mathbf{y}'$ which satisfy the relaxed complementary slackness conditions, with a suitable choice of β. We take care to always modify $\mathbf{x}$ integrally, so that $\mathbf{x}'$ is also a feasible solution of the original IP.
- $\mathrm{SOL} \leq \beta \cdot \mathrm{OPT}$.

Application to Set Cover

Primal

- minimize $\sum_{S \in \mathcal{S}} c(S) \cdot x_S$
- subject to:
  - $\sum_{S : e \in S} x_S \geq 1$, $e \in \mathcal{U}$
  - $x_S \geq 0$, $S \in \mathcal{S}$

Dual

- maximize $\sum_{e \in \mathcal{U}} y_e$
- subject to:
  - $\sum_{e \in S} y_e \leq c(S)$, $S \in \mathcal{S}$
  - $y_e \geq 0$, $e \in \mathcal{U}$

We work with $\alpha = 1$, $\beta = f$.

Primal conditions

For each $S \in \mathcal{S}$: $x_S = 0$ or $\sum_{e \in S} y_e = c(S)$

Dual conditions

For each $e \in \mathcal{U}$: $y_e = 0$ or $1 \leq \sum_{S : e \in S} x_S \leq f$

An f-approximation algorithm for Set Cover

$\mathbf{x} \leftarrow \mathbf{0}$.

$\mathbf{y} \leftarrow \mathbf{0}$.

while there is an uncovered element e **do**

  raise y_e until some set becomes tight.

  pick all tight sets and update $\mathbf{x}$.

  consider all elements occurring in these sets covered.

end while

output the set cover $\mathbf{x}$.

“Another” f-approximation algorithm for Set Cover

for each $S \in \mathcal{S}$, set $t(S) := c(S)$.

$C \leftarrow \emptyset$.

while $C \neq \mathcal{U}$ **do**

  choose $e \in \mathcal{U} - C$.

  let $m := \min_{S: e \in S} t(S)$.

  for each S containing e, set $t(S) := t(S) - m$.

  pick all sets S with $t(S) = 0$.

  $C \leftarrow C \cup \bigcup_{S: t(S)=0} S$.

end while

output all picked sets.

Tight Example