Online Set Cover: Randomized Rounding and Extensions

Thomas Kesselheim Last Update: April 29, 2020

So far, we have seen an algorithmic approach to solve the fractional version of Online Set Cover. That is, we are presented the LP

$$\begin{aligned}
\text{minimize } & \sum_{S \in \mathcal{S}} c_S x_S \\
\text{subject to } & \sum_{S \colon e \in S} x_S \geq 1 && \text{for all } e \in U \\
& x_S \geq 0 && \text{for all } S \in \mathcal{S}
\end{aligned}$$

one constraint at a time. Whenever we see a constraint, we have to increase the variables to ensure it is fulfilled.

So far, the variables could take fractional values. By design, these values never exceed 1. In this lecture, we will see how the fractional solutions can be turned into integral ones. We will do this in a randomized way by interpreting fractional values as probabilities.

1 Ski Rental

Our first result is again for the *ski-rental problem*. Recall that we derived a primal-dual algorithm for the fractional problem, which is given by the following LP.

$$\begin{aligned}
\text{minimize } & B x_{\text{buy}} + \sum_t x_{\text{rent},t} \\
\text{subject to } & x_{\text{buy}} + x_{\text{rent},t} \geq 1 && \text{for all } t \\
& x_{\text{buy}}, x_{\text{rent},t} \geq 0 && \text{for all } t
\end{aligned}$$

Our algorithm is deterministic and $\frac{\mathrm{e}}{\mathrm{e}-1}$-competitive. Our goal now is to turn this algorithm into a randomized algorithm for the integral problem. To this end, we will interpret the fractions as probabilities. For example, for $B = 3$, our algorithm would set $x_{\text{buy}} = \frac{4}{19}$ on the first day, $x_{\text{buy}} = \frac{10}{19}$ on the second day, and $x_{\text{buy}} = 1$ on the third day. This translates to a randomized algorithm that buys skis with probability $\frac{4}{19}$ on the first day, $\frac{10}{19}$ within the first two days, and 1 within the first three days.

The easiest way to formalize this is as follows. We draw θ uniformly at random from $[0, 1]$ beforehand, buy the skis as soon as $x_{\text{buy}}^{(t)} \geq \theta$, and rent them before. Interestingly, this approach works for any deterministic algorithm for the fractional problem.

Theorem 4.1. *Given an α-competitive algorithm for the fractional relaxation of Ski Rental, the above algorithm is a randomized α-competitive algorithm for Ski Rental.*

Proof. Let Z_{buy} be the random variable indicating the cost of buying skis within the first m steps. We do so with probability $x_{\text{buy}}^{(m)}$. The expected cost from buying in the first m steps is $\mathbf{E}\left[Z_{\text{buy}}\right] = B \cdot x_{\text{buy}}^{(m)}$.

Let $Z_{\text{rent},t}$ be the random variable indicating the cost of renting skis in the t-th step. We rent skis with probability $1 - x_{\text{buy}}^{(t)}$. Note that $x_{\text{buy}}^{(t)} + x_{\text{rent},t}^{(t)} \geq 1$ because this is the new constraint that has to be fulfilled. In other words, we rent skis with probability at most $x^{(t)}_{\text{rent},t}$. So, $\mathbf{E}\left[Z_{\text{rent},t}\right] \leq x^{(t)}_{\text{rent},t} \leq x^{(m)}_{\text{rent},t}$.

Overall, our expected cost is

$$\mathbf{E}\left[Z_{\text{buy}} + \sum_{t=1}^{m} Z_{\text{rent},t}\right] \leq B \cdot x^{(m)}_{\text{buy}} + \sum_{t=1}^{m} x^{(m)}_{\text{rent},t} \leq \alpha \left(B \cdot x^{*}_{\text{buy}} + \sum_{t=1}^{m} x^{*}_{\text{rent},t}\right) \leq \alpha \cdot \text{cost}(\text{OPT}(\sigma)) \enspace ,$$

where x^* is an optimal fractional offline solution. □

2 Rounding Fractional Solutions for Set Cover

Our algorithm for general Set Cover is a little more complicated but follows the same principle.

A priori, we choose for each $S \in \mathcal{S}$ a threshold θ_S uniformly at random from $[0, 1]$.

Upon arrival of an element e, update $(x^{(t)}_S)_{S \in \mathcal{S}}$ according to the fractional algorithm.

(a) Pick all sets S, for which $x^{(t)}_S \geq \frac{1}{2 \ln t} \theta_S$.

(b) If e is still uncovered, choose one set S from the probability distribution defined by $(x^{(t)}_S)_{S: e \in S}$ and pick it. Note that this is possible since $\sum_{S: e \in S} x^{(t)}_S \geq 1$.

Theorem 4.2. *Using an α-competitive algorithm for the fractional problem, the algorithm for the integral problem is $O(\alpha \cdot \log m)$-competitive.*

Lemma 4.3. *The probability that part (b) is executed in the t-th step is at most $\frac{1}{t^2}$.*

Proof. Let $X^{(t)}_S = 1$ if set S has been picked until round t because of case (a), that is, $x^{(t)}_S \geq \frac{1}{2 \ln t} \theta_S$, otherwise set $X^{(t)}_S = 0$.

By design, if $(2 \ln t) x^{(t)}_S \leq 1$,

$$\mathbf{Pr}\left[X^{(t)}_S = 1\right] = \mathbf{Pr}\left[x^{(t)}_S \geq \frac{1}{2 \ln t} \theta_S\right] = \mathbf{Pr}\left[\theta_S \leq (2 \ln t) x^{(t)}_S\right] = (2 \ln t) x^{(t)}_S \enspace .$$

As $1 - q \leq \exp(-q)$ for $q \in [0, 1]$, this implies

$$\mathbf{Pr}\left[X^{(t)}_S = 0\right] = 1 - (2 \ln t) x^{(t)}_S \leq \exp(-(2 \ln t) x^{(t)}_S) \enspace .$$

If $(2 \ln t) x^{(t)}_S > 1$, then $\mathbf{Pr}\left[X^{(t)}_S = 0\right] = 0$, so this bound holds as well.

Note that these choices are independent, so

$$\mathbf{Pr}\left[\bigwedge_{S: e \in S} X^{(t)}_S = 0\right] = \prod_{S: e \in S} \mathbf{Pr}\left[X^{(t)}_S = 0\right] \leq \prod_{S: e \in S} \exp(-(2 \ln t) x^{(t)}_S) = \exp\left(-(2 \ln t) \sum_{S: e \in S} x^{(t)}_S\right) \leq \frac{1}{t^2} \enspace .$$

□

Lemma 4.4. *The expected cost due to set S within the first m rounds is at most $(\sum_{t=1}^{m} \frac{1}{t^2} + 2 \ln m) c_S x^{(m)}_S$.*

Proof. Let $X_{t,S} = 1$ if set S is chosen in part (b) statement of step t. Note that we have $\mathbf{E}[X_{t,S}] \leq \frac{1}{t^2} \cdot x_S^{(t)} \leq \frac{1}{t^2} \cdot x_S^{(m)}$.

The expected cost due to set S within the first m rounds is

$$\mathbf{E}\left[\sum_{t=1}^{m} c_S X_{t,S} + c_S X_S^{(m)}\right] = \sum_{t=1}^{m} c_S \mathbf{E}[X_{t,S}] + c_S \mathbf{E}\left[X_S^{(m)}\right] \leq c_S \sum_{t=1}^{m} \frac{1}{t^2} x_S^{(m)} + c_S(2\ln m) x_S^{(m)} \ .$$

□

Proof of Theorem 4.2. We consider the outcome after m rounds. Let x^* be an optimal offline fractional solution to the LP relaxation. As the fractional algorithm is α-competitive, we have $\sum_{S \in \mathcal{S}} c_S x_S^{(m)} \leq \alpha \cdot \sum_{S \in \mathcal{S}} c_S x_S^*$. Furthermore, by Lemma 4.4, $\mathbf{E}[\mathrm{cost}(\mathrm{ALG}(\sigma))] \leq \sum_{S \in \mathcal{S}} \left(\sum_{t=1}^{m} \frac{1}{t^2} + 2\ln m\right) c_S x_S^{(m)}$.

Note that $\sum_{t=1}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6}$. So overall, $\mathbf{E}[\mathrm{cost}(\mathrm{ALG}(\sigma))] \leq \alpha \cdot \left(\frac{\pi^2}{6} + 2\ln m\right) \sum_{S \in \mathcal{S}} c_S x_S^* = \alpha \cdot \left(\frac{\pi^2}{6} + 2\ln m\right) \mathrm{cost}(\mathrm{OPT}(\sigma))$. □

3 Extensions

The primal-dual framework for Online Set Cover can be extended in various ways, some of which we will highlight here.

3.1 Arbitrary Covering Constraints

It is pretty straightforward to extend the approach to arbitrary covering constraints. That is, we have a covering LP of the form

$$\begin{aligned} \text{minimize } & \sum_j c_j x_j & \\ \text{subject to } & \sum_j a_{i,j} x_j \geq 1 & \text{for all } i \\ & x_j \geq 0 & \text{for all } j \end{aligned}$$

where all c_j and $a_{i,j}$ are non-negative reals.

3.2 Packing LPs

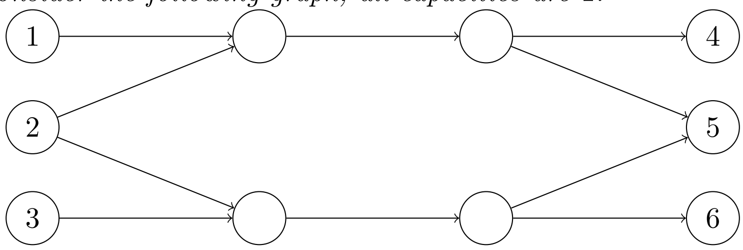
We can also consider packing problem, which are given by the following kind of linear program.

$$\begin{aligned} \text{maximize } & \sum_i y_i & \\ \text{subject to } & \sum_i a_{i,j} y_i \leq c_j & \text{for all } j \\ & y_i \geq 0 & \text{for all } i \end{aligned}$$

Now, we know all constraints in advance. In each step a variable appears and we have to decide on its value. Note that this is exactly what we already did with the dual in the covering problem. As it turns out, slight adaptations of our Set Cover algorithm give us an $O(\log n)$-competitive algorithm for the fractional problem with n packing constraints if $a_{i,j} \in \{0, 1\}$. If each variable appears in at most f constraints, this can be improved to $O(\log f)$ because this corresponds to the frequency of the set system of the Set Cover instance.

One particular application is the following. We are given a directed graph $G = (V, E)$, in which each edge $e \in E$ has a capacity c_e. Now, we are presented pairs of a source node $s_k \in V$ and a sink node $t_k \in V$ one after the other. For each of these pairs (s_k, t_k), we have to find a path connecting s_k and t_k or we can reject it. The number of paths using edge e may not exceed c_e. The goal is to connect as many source-sink pairs as possible.

Example 4.5. *Consider the following graph, all capacities are* 1.



In advance, we do not know which pairs of vertices should be connected. If the first request is to connect vertices 2 *and* 5*, we may either choose the top or the bottom path or reject the request. Suppose we choose the top path. If then the second request is to connect vertices* 1 *and* 4*, we have to reject the request because there is no more capacity left. So, we served only one request whereas the offline optimum can serve both.*

The fractional relaxation of the problem looks as follows. We have variables $f_{k,P}$ denoting the *flow* routed along path P for every path connecting s_k and t_k.

$$\begin{aligned}
\text{maximize} \quad & \sum_k \sum_P f_{k,P} \\
\text{subject to} \quad & \sum_{k,P: e \in P} f_{k,P} \leq c_e && \text{for all } e \in E && \text{(capacity constraints of edges)} \\
& \sum_P f_{k,P} \leq 1 && \text{for all } k && \text{(only one unit of flow between any } s_k \text{ and } t_k\text{)} \\
& f_{k,P} \geq 0 && \text{for all } k, P
\end{aligned}$$

To design an algorithm for the fractional problem, when we see a pair (s_k, t_k) we add all variables $f_{k,P}$ for paths connecting s_k to t_k to the LP one after the other and run an algorithm for packing LPs.

Note that in this case the number of constraints n corresponds to the number of edges plus the number of rounds. However, this LP is sparse in the following sense: Let d be the maximum length of a path. Then each variable appears in at most $d + 1$ constraints. Therefore, there is an $O(\log d)$-competitive algorithm for the fractional problem.

For more details and many other applications, see the survey by Buchbinder and Naor.

References

- N. Buchbinder, J. Naor: The Design of Competitive Online Algorithms via a Primal-Dual Approach. Foundations and Trends in Theoretical Computer Science 3(2-3): 93-263 (2009)