4.3 Online Set Cover

The algorithm takes a fractional perspective: for each set S it maintains a variable $x_S \in [0,1]$ which denotes the fractional amount of S picked by the solution. We enforce that x_S starts at 0, and is *monotone* and only increases over time.

Suppose e is the element arriving at time t, and gives rise to the the covering constraint

$$\sum_{S:e\in S} x_S \geq 1.$$

Our algorithm uses the multiplicative update rule: simultaneously for each set S containing e, we raise its variable according to the rule

$$x'_S = \frac{x_S + \eta}{c_S} \tag{4.1}$$

until the covering $\sum_{S\ni e} x_S \geq 1$ is satisfied.

The *shifting* constant η needs to be strictly positive, else the update rule $x'_S \propto (x_S + 0)$ could never increase x_S from 0 to a non-zero value.

Theorem 4.1. *If r is the maximum number of sets containing any element, then the fracional algorithm has competitive ratio*

$$\alpha := (1 + r\eta)\ln(1 + 1/\eta)$$

against any optimal integer solution. Hence, setting $\eta = O(1/r)$ gives a competitive ratio of $2\ln(r+1)$.

For simplicity we assume that r is known. If r is unknown in advance, the algorithm can simply use the current value of r: as an exercise show that the competitive ratio only gets better.

Proof. Fix a benchmark integer solution $\mathcal{B}$, and consider the potential

$$\Phi := (1 + r\eta) \sum_{S\in\mathcal{B}} c_S \ln\left(\frac{1+\eta}{x_S + \eta}\right).$$

As each $x_S \in [0,1]$, each logarithmic term lies in $[0, \ln(1 + 1/\eta)]$. The term is large if the set S lies in the offline solution and x_S is small; this makes sense as the online algorithm may pay for this *mistake* later, and the potential provides the money in the bank for these future mistakes.

As always, we need to show

$$\Delta ON + \Delta\Phi \leq \alpha\Delta OFF. \tag{4.2}$$

First, the offline benchmark solution picks some set S containing e (unless e was already covered by its previous choices.) In that case, the potential increases by

$$(1 + r\eta)c_S \ln\left(\frac{1+\eta}{x_S+\eta}\right) \leq (1 + r\eta)c_S \ln\left(\frac{1+\eta}{\eta}\right) = \alpha c_S,$$

irrespective of x_S. Since the offline cost is c_S, (4.2) holds.

Next, the online algorithm moves. We use a continuous analysis: as the algorithm increase x_S for each S containing e according to the rule (4.1) as long as $\sum_{S:e\in S} x_S < 1$, the rate of increase of the online cost is

$$\sum_{S:e\in S} c_S x_S' = \sum_{S:e\in S} c_S \frac{x_S+\eta}{c_S} = \underbrace{\sum_{S:e\in S} x_S}_{<1} + \underbrace{\sum_{S:e\in S} \eta}_{\leq r\eta} \leq 1 + r\eta.$$

The rate of change of the potential satisfies

$$\begin{aligned}\Phi' &= (1+r\eta) \sum_{S\in OPT} c_S \left(-\frac{\partial}{\partial x_S} \log(x_S+\eta)\right) x_S' \\ &= (1+r\eta) \sum_{S\in OPT:e\in S} c_S \left(-\frac{1}{x_S+\eta}\right) \cdot \frac{x_S+\eta}{c_S} \\ &= -(1+r\eta) \sum_{S\in OPT:e\in S} -1 \leq -(1+r\eta),\end{aligned}$$

We use the chain rule:

$$\Phi' = \frac{d\Phi}{dz} = \sum_i \frac{\partial \Phi}{\partial x_S} \frac{dx_S}{dz}.$$

where the inequality follows as OPT must contain at least one set S that contains e. Thus (4.2) holds and the result follows. □

4.3.1 *Competitiveness Against a Fractional Optimum*

Theorem 4.1 compared the cost of the online algorithm to an integer benchmark. We now show that same result with respect to a fractional benchmark. Fix any fractional solution y that satisfies $\sum_{S:e\in S} y_S \geq 1$ for each element e. The potential is now

$$\Phi := (1+r\eta) \sum_S c_S y_S \ln\left(\frac{y_S+\eta}{x_S+\eta}\right).$$

If each $y_S \in \{0,1\}$, this is the same expression as (4.5) and the potential is always positive. But when y_S is fractional, the logarithmic term can now be positive or negative depending on whether $y_S > x_S$ or not. However, it is always in the range $[-\ln(1+1/\eta), \ln(1+1/\eta)]$.

When the benchmark solution increases y_S at rate 1, the potential Φ changes at rate

$$\begin{aligned}\frac{\partial \Phi}{\partial y_S} &= (1+r\eta)\, c_S \cdot \frac{\partial}{\partial y_S} y_S \ln\left(\frac{y_S+\eta}{x_S+\eta}\right) \\ &= (1+r\eta)\, c_S \left[\ln\left(\frac{y_S+\eta}{x_S+\eta}\right) + \frac{y_S}{(y_S+\eta)}\right] \leq \alpha'\, c_S,\end{aligned}$$

where $\alpha' = (1+r\eta)(1+\ln(1+1/\eta))$.

Next, the online algorithm increases variables x_S for sets S that contain the element e at rate (4.1): this causes the online cost to increase at rate $1+r\eta$ as in the integer case, and the potential changes at rate

$$\begin{aligned}\Phi' &= (1+r\eta)\sum_{S\ni e} c_S y_S \cdot \frac{\partial}{\partial x_S}\ln\left(\frac{y_S+\eta}{x_S+\eta}\right)\cdot x'_S \\ &= (1+r\eta)\sum_{S\ni e} c_S y_S \cdot \frac{-1}{x_S+\eta}\cdot\frac{x_S+\eta}{c_S} \\ &= (1+r\eta)\sum_{S\ni e}(-y_S) \leq -(1+r\eta).\end{aligned}$$

Above, we used that the solution y is feasible and hence $\sum_{S\ni e} y_S \geq 1$. As always, this ensures that

$$ON + \Phi_T - \Phi_0 \leq \alpha' \, OFF.$$

The starting potential $\Phi_0 = 0$, but the final potential can be negative. Nevertheless, one can verify that

$$-\Phi_T \leq \alpha' OFF$$

So we have $ON \leq 2\alpha' \, OFF$, which proves the following result:

Theorem 4.2. *If r is the maximum number of sets containing any element, then the fracional algorithm has competitive ratio*

$$\alpha' := 2(1+r\eta)\cdot(1+\ln(1+1/\eta))$$

against any optimal fractional solution. Hence, setting $\eta = O(1/r)$ gives a competitive ratio of $O(\ln(r+1))$.

4.4 *The Shifted KL Divergence*

Let us abstract out the argument in the previous sections. For values $p, q \in [0,1]^m$, we define the *shifted KL divergence* function:

$$KL_\eta(p\|q) := \sum_i \left[p_i \ln\left(\frac{p_i+\eta}{q_i+\eta}\right)\right].$$

Given weights $c_i \geq 0$ along with the shift η, a weighted version is the following:

$$KL_{c,\eta}(p\|q) := \sum_i c_i \left[p_i \ln\left(\frac{p_i+\eta}{q_i+\eta}\right)\right].$$

The potential function Φ we chose in the previous section was precisely $C := (1+r\eta)$ times the weighted KL divergence between the optimal solution and the algorithm's solution.

Here are some facts for the unweighted case (the weighted version has analogous facts):

1. $KL_{c,\eta} \geq 0$ if $p_i \geq q_i$ for all i where $p_i > 0$.

2. The partial derivative with respect to the first argument is:

$$\frac{\partial KL_\eta}{\partial p_i} = \ln\left(\frac{p_i+\eta}{q_i+\eta}\right) + \frac{p_i}{p_i+\eta} \le \ln\left(\frac{p_i+\eta}{q_i+\eta}\right) + 1. \tag{4.3}$$

Hence, when OPT increases p_i at rate p_i' then Φ increases at rate $\Phi' \le C\, p_i'(1+\log\frac{1+\eta}{\eta})$, meaning that

$$\Phi' \le OPT' \cdot C \cdot \left(1 + \log\left(\frac{1+\eta}{\eta}\right)\right).$$

In other words, the potential is Lipschitz with respect to its first argument (which in this case is the optimal solution): the parameter η affects the Lipschitz-ness, and larger values make the Lipschitz constant better.

3. The partial derivative of the potential with respect to the second argument is

$$\frac{\partial KL_\eta}{\partial q_i} = -\frac{p_i}{q_i+\eta}. \tag{4.4}$$

In set cover, ALG picks some subset A indices corresponding to the sets that cover the yet-uncovered element, and changes q_i for each $i \in A$ at some rate q_i'. The rate of potential change is $\Phi' = -C\sum_{i\in A}\frac{p_i}{q_i+\eta}\cdot q_i'$. So setting the rate of change of q_i for $i \in A$ to be

$$q_i' = q_i + \eta$$

ensures that $\Phi' = -C\sum_{i\in A} p_i$. We want to show that $ALG' + \Phi' \le 0$. Observe that $ALG = \sum_i q_i$, and hence

$$ALG' + \Phi' = \sum_{i\in A}[q_i' - Cp_i] = \sum_{i\in A}[q_i + \eta - Cp_i].$$

Since the optimal solution must cover the element, $\sum_{i\in A} p_i \ge 1$. Since the online solution was not yet feasible, $\sum_{i\in A} q_i < 1$. So we want that $1+\eta|A| - C \le 0$, or $C \ge 1+\eta|A|$. Finally, defining r to be the maximum possible size of A completes the argument.