AA-Entrega-5

Oriol Miró i Jan Samaranch

May 2023

Problema 18

> The 3-Set Cover problem is defined as follows: Given a family of sets $S_1, \dots, S_m \subseteq U$ such that, for $1 \le i \le m$, $|S_i| = 3$ and has profit $c(S_i)$, find a subset of these sets that minimizes the profit, while their union covers all elements of U. You can assume that U is the union of all the given sets.
>
> Provide a integer linear programming formulation for the problem. Use this to devise a primal-dual algorithm and analyze whether it provides a constant approximation for the problem.

ILP Formulation

First, we formulate the problem as an Integer Linear Programming problem, in the following way:

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^{m} c(S_i) \cdot x_i \\
\text{Subject to} \quad & \sum_{i=1}^{m} x_i \cdot a_{ij} \ge 1, \quad \forall j \in U \\
& x_i \in \{0, 1\} \quad \forall i \in \{1, 2, \dots, m\}
\end{aligned}
$$

Where x_i denotes if we picked set S_i, and $a_i j = 1$ if S_i contains the element j, 0 otherwise.

Dual Formulation

To obtain the Dual formulation, we first relax the ILP formulation to an LP by allowing x_i to take fractional values between 0 and 1. The relaxed ILP can be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{m} c(S_i) \cdot x_i \\ \text{Subject to} \quad & \sum_{i=1}^{m} x_i \cdot a_{ij} \geq 1, \quad \forall j \in U \\ & x_i \geq 0 \quad \forall i \in \{1, 2, \ldots, m\} \end{aligned}$$

Now we construct the Dual LP in the following way:

- The primal problem is a minimization problem, so the dual problem is a maximization problem.
- We introduce dual variables y_j for each constraint in the primal problem. Since there is one constraint for each element in U, we have a y_j for each $j \in U$. The objective of the dual problem is to maximize the sum of dual variables, $\sum_{j \in U} y_j$.
- For each primal variable x_i, we create a constraint in the dual problem. The coefficients of the dual constraints are determined by the coefficients of the primal constraints. In this case, since $a_{ij} = 1$ if S_i contains element j, the constraint for the dual problem is $\sum_{j \in S_i} y_j \leq c(S_i)$ for each set S_i in the family of sets.
- The dual variables are non-negative: $y_j \geq 0$ for all $j \in U$.

$$\begin{aligned} \text{Maximize} \quad & \sum_{j \in U} y_j \\ \text{Subject to} \quad & \sum_{j \in S_i} y_j \leq c(S_i), \quad \forall i \in 1, 2, \ldots, m \\ & y_j \geq 0 \quad \forall j \in U \end{aligned}$$

Primal-Dual algorithm

The primal and dual problems are connected through weak duality, meaning that the optimal solution of the dual problem provides an upper bound on the optimal solution of the primal problem.

The primal-dual algorithm iteratively updates the primal and dual variables based on their corresponding constraints. The algorithm consists of the following steps:

Initialization: Set $x_i = 0$ for all $i \in 1, 2, \ldots, m$ and $y_j = 0$ for all $j \in U$. All elements in U are initially uncovered, and no sets are in the solution.

Dual Update: While there are uncovered elements, pick an element j from U that is not yet covered. Increase the dual variable y_j until one of the constraints in the dual becomes tight. This means that for some set S_i, the sum of the dual variables corresponding to elements in S_i equals $c(S_i)$.

Primal Update: Set $x_i = 1$ for the set S_i that has a tight constraint. Add set S_i to the solution and remove all elements in S_i from the uncovered list.

Termination: Repeat steps 2 and 3 until all elements in U are covered.

The algorithm works by gradually increasing the dual variables in the dual update step, which in turn tightens the constraints in the primal problem. The primal update step adds sets to the solution and removes covered elements when a tight constraint is found. The algorithm terminates when all elements in U are covered, and the final solution is an approximation of the optimal solution to the 3-Set Cover problem.

Constant aproximation analysis

To analyze whether the primal-dual algorithm provides a constant approximation, we need to compare the solution obtained by the algorithm with the optimal solution for the 3-Set Cover problem. We can use the weak duality and integral properties to show that the algorithm provides a constant-factor approximation.

The weak duality theorem states that the optimal solution of the dual problem provides an upper bound for the optimal solution of the primal problem. Since the primal-dual algorithm terminates when all elements in U are covered, the sum of the dual variables at termination, $\sum_{j \in U} y_j$, is a lower bound for the optimal solution of the primal problem.

Now, let's consider the cost of the solution produced by the primal-dual algorithm, denoted by C_P. For each element $j \in U$, the algorithm adds at most one set S_i to the solution. Since each set S_i contains exactly three elements, the algorithm can add at most $\frac{1}{3}$ of the cost of $c(S_i)$ for each element j. Therefore, the cost of the solution produced by the algorithm can be written as:

$$C_P \leq \frac{1}{3} \sum_{i=1}^{m} c(S_i) x_i \leq \frac{1}{3} \sum_{j \in U} \sum_{i=1}^{m} a_{ij} c(S_i) x_i,$$

where $a_{ij} = 1$ if S_i contains element j, and 0 otherwise. Since the dual variable y_j is increased until $\sum_{j \in S_i} y_j = c(S_i)$ for some set S_i, we have:

$$\sum_{j \in U} y_j \leq \sum_{j \in U} \sum_{i=1}^{m} a_{ij} c(S_i) x_i.$$

Combining the inequalities, we get:

$$C_P \leq \frac{1}{3} \sum_{j \in U} y_j.$$

Since $\sum_{j \in U} y_j$ is a lower bound for the optimal solution of the primal problem (denoted by C^*), we have:

$$C_P \leq \frac{1}{3} C^*.$$

This inequality shows that the primal-dual algorithm provides a 3-approximation for the 3-Set Cover problem. Therefore, the algorithm indeed provides a constant-factor approximation for the problem.