Approximation using linear programming

Huy L. Nguyễn

We have seen several examples where the LP gives exact solution for discrete optimization problems. In general, there might be no integral optimal solution for an LP relaxation and we cannot obtain the optimal solution for the discrete problem. Nonetheless, we can use the fraction solution to construct an *approximation* solution that is close in quality compared with the optimal solution. We will explore a few examples of this approach.

1 Vertex cover

Consider a graph $G = (V, E)$ where each node u has a weight $w_u \geq 0$. The goal is to find a vertex cover, which is a subset of vertices that is adjacent to all edges in the graph. Furthermore, we would like to find the vertex cover of minimum total weight.

We can write this problem as an integer linear program as follows.

$$\begin{aligned} \min \sum_{u \in V} w_u x_u &: \\ x_u + x_v \geq 1 \ \forall (u, v) &\in E \\ x_u &\in \{0, 1\} \end{aligned}$$

To obtain an LP, we can relax the integral constraints:

$$\begin{aligned} \min \sum_{u \in V} w_u x_u &: \\ x_u + x_v \geq 1 \ \forall (u, v) &\in E \\ 0 \leq x_u &\leq 1 \end{aligned}$$

From the fractional solution, we can simply round all x_u up to 1 if $x_u \geq 1/2$ and down to 0 if $x_u < 1/2$. Let S be the set of vertices selected by the algorithm. We will show that this is a valid solution whose weight is at most twice that of the optimal solution.

Lemma 1.1. *S is a valid vertex cover.*

Proof. For each edge (u, v), we know $x_u + x_v \geq 1$ so either x_u or x_v is at least $1/2$. Therefore, either u or v is selected in S and the edge (u, v) is covered. □

Lemma 1.2. *The weight of S is at most twice that of the optimal solution.*

Proof. Because we only pick vertices with $x_u \geq 1/2$, the weight of S is at most 2 times $\sum_u x_u w_u$. Because the LP is a relaxation, the optimal integral solution is a valid solution for the LP. Thus, the LP value is at most the weight of the optimal integral solution. Thus, the weight of S is at most 2 times the weight of the optimal integral solution. □

2 MAX2SAT

A 2CNF formula consists of n Boolean variables $x_1, \ldots, x_n$ and m clauses of the form $y \vee z$, where each y, z is called a *literal*, which is either a variable or its negation. Given a formula, our goal is to set the variables so as to maximize the number of satisfied clauses.

We start with an integral formulation. We use variable z_j to indicate whether clause j is satisfied or not.

$$\max \sum_{j=1}^{m} z_j$$

$$y_{j1} + y_{j2} \geq z_j \ \forall j$$

$$z_j \leq 1 \ \forall j x_i \in \{0, 1\}$$

where y_{j1} is the shorthand for x_i if the first literal in clause j is variable i and for $1 - x_i$ if the literal is the negation of variable i.

We relax the integral formulation to obtain an LP by replacing $x_i \in \{0, 1\}$ with $0 \leq x_i \leq 1$.

Now to obtain an integral solution, independently for each variable i, we randomly set it to 1 with probability x_i and to 0 with probability $1 - x_i$.

Lemma 2.1. *The expected number of satisfied clause is at least* $\frac{3}{4}$ *times the optimal value.*

Proof. We will prove that the probability a given clause is satisfied is at least $3z_j/4$. The lemma then follows from linearity of expectation.

Suppose the clause is $x_a \vee x_b$. Notice that at the optimal solution, $z_j = \min(1, x_a + x_b)$ since the LP tries to maximize $\sum_j z_j$.

The probability that randomized rounding satisfies this clause is exactly

$$1 - (1 - x_a)(1 - x_b) = x_a + x_b - x_a x_b$$

Consider two cases. First, consider the case $x_a + x_b \leq 1$. We have $x_a x_b \leq (x_a + x_b)^2/4 \leq (x_a + x_b)/4$. Thus,

$$x_a + x_b - x_a x_b \geq \frac{3}{4}(x_a + x_b) \geq 3z_j/4$$

Next, consider the case $t = x_a + x_b \geq 1$. We have $x_a x_b \leq (x_a + x_b)^2/4 = t^2/4$. Thus,

$$x_a + x_b - x_a x_b \geq t - t^2/4 \geq 3/4 \ \forall t \in [1, 2]$$

Thus, in both cases, the probability that randomized rounding satisfies a clause is at least $3z_j/4$. $\square$