

Approximation Algorithm

In this week, we learn the approximation algorithms for NP-hard problem. We have two specific examples, one is travelling salesman problem and the other is the integer linear programming.

8.1 Travelling salesman problem

First, in order to learn about what the travelling salesman problem is. We review the Hamilton cycle problem.

1. Input: undirected graph G .
2. output: Yes, if there's a cycle that go through all vertices. No, otherwise.

Then, we talk about NP-hard problems. NP-hard (Non-deterministic Polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP". That is all the NP problems can be reduced to HP-hard problems.

The travelling salesman problem (TSP) asks the following question: Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city? It is an NP-hard problem in combinatorial optimization. We can reduce Hamilton cycle to the TSP problem. We can think every link in the Ham problem has a weight 1. Then this problem can be converted to TSP problem, that's find a cycle/tour with minimum weight.

We discussed two different approximation scenarios for TSP. One is TSP with Euclidean metric and the other is the general TSP. We conclude that TSP with Euclidean metric has 2OPT approximation and the TSP in general is hard to approximate within constant factor.

8.1.1 TSP with Euclidean metric

If the TSP has Euclidean metric, then it has 2OPT approximation the Euclidean metric. The Euclidean metric is:

1. $d(x, y) \geq 0$;
2. $d(x, y) = d(y, x)$;
3. $d(x, y) + d(y, z) \geq d(x, z)$.

We use the minimum spanning tree structure to get the approximation results. We think the sum weights of the minimum spanning tree as the approximation results.

1. Construct a minimum spanning tree MST for graph G. Then the sum weights of the MST is less than solution of TSP, as if the optimization solution of TSP is less than the MST, then the solution would be the MST. So we have $MST \leq OPT$.
2. Then we have shown $SOL \leq 2MST$. Let's consider that we can go each link as many times as we want. Then $2MST$ is a solution for the TSP problem. But we only need to go once of each link and if there is a triangle abc. Then if ab and ac are in the MST solution, then of course ab plus ac is greater than bc. So in this way if bc is in the TSP solution, then of course $SOL \leq 2MST$.
3. As we have shown in the above two statements, then we have $SOL \leq 2MST \leq 2OPT$

8.1.2 TSP in generate

To prove that is hard to approximate within constant factor, we need reduce Hamiltonian cycle to it.

Define $d(u, v) = \begin{cases} 1 & \text{if } (u,v) \in G. \\ kn + 1 & \text{else} \end{cases}$

If G has a Hamiltonian cycle, so OPT_{TSP} in G' is n. If G doesn't have a Hamiltonian cycle, $OPT_{TSP} \geq kn + 1 + n - 1 = (k + 1)n$. If A is k-OPT algorithm for TSP, create G' from G, then run A(G). Then we can decide Ham(G) using the result A(G') by

1. If G has a Hamiltonian cycle, $SOL_A \leq kn$;
2. If G doesn't have a Hamiltonian cycle, $(k+1)n \leq OPT \leq SOL_A$, $kn \leq (k+1)n \leq SOL_A$.

Then if A can be solved in polynomial time, then Ham can be solved in polynomial time, so NP will be equal to P. We don't know that P is equal to NP or not. So, right now, we can get the conclusion TSP is hard to approximate within a constant factor.

8.2 Integer Linear programming

Then we learn about that the Integer Linear programming is NP-hard problem. We reduce the Vertex Cover problem to it. We learn that LP is the lower bound for ILP.

The objective function of the VC is to minimize $\sum x_i$ with n vertices and m edges.

1. The ILP is: For each edge (i,j) in graph G, $x_i + x_j \geq 1$, $x_i \in \{0, 1\}$.
2. The LP is: For each edge (i,j) in graph G, $x_i + x_j \geq 1$, $0 \leq x_i \leq 1$.

Then we have $OPT_{LP} \leq OPT_{ILP}$, because the solution for ILP is a subset of LP. We use a rounding program to show the approximation for the ILP using LP's result. Suppose we have got the optimization solution for LP. Let $\bar{x}_i=0.5$ if $x_i \geq 0.5$ or $\bar{x}_i=0$ o.w. Let x_i^* the variables in the LP.

If $(i,j) \in E$, then $\bar{x}_i + \bar{x}_j \geq 1$. We can get $x_i^* + x_j^* \geq 1$, because at least one of x_i^* is greater than 0.5. Then we can get $\bar{x}_i \leq 2x_i^*$, $OPT_{ILP} \leq \sum \bar{x}_i \leq 2*OPT_{LP}$.