TD 9 : Algorithmes d'approximation, en ligne et probabiliste

1 k-Centre (Metric k-Center)

We are given a metric space (X, d) and points $\mathbb{P} = \{p_1, \ldots, p_n\} \subset X$. We want to choose k points $Y = \{y_1, \ldots, y_k\} \subset \mathbb{P}$ so that

$$\max_{p \in \mathbb{P}} d(p, Y)$$

is minimized.

Problem 1.1 (Warm-up)**.** For any $Y \subset \mathbb{P}$, if R is the minimum value such that $\mathbb{P}$ is covered by the union of the circles of radius R centered at points in Y, then show that $R = \max_{p \in \mathbb{P}} d(p, Y)$.

Solution. This is easy to check. □

Problem 1.2. Give a 2-approximation algorithm for the metric k-center problem.

Solution. Consider the following algorithm.

> **Algorithm ALG :**
> Let y_1 be an arbitrary point in $\mathbb{P}$. For $i = 2$ to k, choose y_i to be the point that *maximizes* $d(y_i, \{y_1, \ldots, y_{i-1}\})$.

Let OPT be the radius given by the optimal solution. Let the circles of the optimal solution be $C_1, \ldots, C_k$ and their centers $c_1, \ldots, c_k$.

Consider the points chosen by ALG be $Y = \{y_1, \ldots, y_k\}$ in that order. We consider two cases.

1. Every y_ℓ belongs to a different C_{i_ℓ} : Then for any point $p \in \mathbb{P}$, there exists a y_ℓ such that p and y_ℓ belong to the same circle C_{i_ℓ}. Then
$$d(p, y_i) \leq d(p, c_{i_\ell}) + d(y_i, c_{i_\ell}) \leq 2 \cdot \mathrm{OPT}\,.$$
2. Suppose $y_1, \ldots, y_j$ where $j < k$ belong to different circles $C_{i_1}, \ldots, C_{i_j}$, but the next point y_{j+1} also belongs to some C_{i_ℓ} where $1 \leq \ell \leq j$.
   — Again by the triangle inequality, $d(y_{j+1}, y_\ell) \leq 2 \cdot \mathrm{OPT}$, and
   — for all points $p \in \mathbb{P} \setminus \{y_1, \ldots, y_{j+1}\}$, by design of ALG, we have $d(p, y_\ell) \leq d(y_{j+1}, y_j)$, and thus $d(p, y_\ell) \leq 2 \cdot \mathrm{OPT}$.
   Thus all eventual points lie within distance $2 \cdot \mathrm{OPT}$ of Y.

□

Problem 1.3. Show that unless $\mathbf{P} = \mathbf{NP}$, there is no $(2 - \varepsilon)$-approximation algorithm for any $\varepsilon > 0$ that runs in time polynomial in $|\mathbb{P}|$. *Hint :* Use the Dominating Set problem.

> **Dominating Set problem :**
> Given an undirected graph $G = (V, E)$, a subset of vertices $D \subseteq V$ is called a *dominating set* if for every vertex $u \in V \setminus D$, there is a vertex $v \in D$ such that $\{u, v\} \in E$.
> The problem of deciding for a given k whether there exists a dominating set of size at most k is known to be **NP**-complete.

Solution. Consider an instance $\langle G = (V, E), k\rangle$ of the Dominating Set problem. The gap reduction is as follows. Consider the metric space (V, d) and $\mathbb{P} = V$ such that for $u, v \in V$,

$$d(u, v) = \begin{cases} 1 & \text{if } \{u, v\} \in E \\ 2 & \text{otherwise} \end{cases}$$

For the sake of contradiction, suppose there is a $(2 - \varepsilon)$-approximation algorithm $\mathcal{A}$ for the k-center problem, so $\mathrm{OPT} \leq \mathcal{A} \leq (2 - \varepsilon) \cdot \mathrm{OPT}$. The following equivalences can be checked :

— G has dominating set D of size $\leq k$ *iff* there is a k-center (namely, one that contains D) with radius 1 *iff* $\text{OPT} = 1$ *iff* $\mathcal{A} \leq 2 - \varepsilon$.
— All dominating sets D in G have size $> k$ *iff* all k-centers have radius 2 *iff* $\text{OPT} = 2$ *iff* $\mathcal{A} \geq 2$.

Thus using $\mathcal{A}$, we can determine in polynomial time whether or not G has a dominating set of size at most k. □

2 Méthode probabiliste (The Probabilistic Method)

2.1 Basic argument

Problem 2.1 (Lower bounding Ramsey number $R(k,k)$)**.** If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then it is possible to color the edges of K_n with two colors such that it has no monochromatic K_k subgraph.

Solution. Consider a uniformly random coloring of the edges of K_n with the colors red and blue. For any set R of k vertices, let A_R be the event that the induced subgraph of K_n on R is monochromatic. Then $\mathbb{P}(A_R) = 2^{1-\binom{k}{2}}$. Thus

$$\mathbb{P}\left(\bigcup_{|R|=k} A_R\right) \leq \sum_{|R|=k} \mathbb{P}(A_R) \leq \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

Hence with a positive probability, no A_R occurs and there is no monochromatic K_k subgraph. □

2.2 Expectation argument

Problem 2.2. Suppose we have a probability space $\mathcal{S}$ (with discrete probability distribution) and a random variable X defined on $\mathcal{S}$ such that $\mathbb{E}[X] = \mu$. Then show that $\mathbb{P}(X \geq \mu) > 0$ and $\mathbb{P}(X \leq \mu) > 0$.

Solution. We have

$$\mu = \mathbb{E}[X] = \sum_x x\mathbb{P}(X = x),$$

where the summation ranges over all values in the range of X. If $\mathbb{P}(X \geq \mu) = 0$, then

$$\mu = \sum_x \mathbb{P}(X = x) = \sum_{x<\mu} x\mathbb{P}(X = x) < \sum_{x<\mu} \mu\mathbb{P}(X = x) = \mu,$$

which is absurd. The same argument can be used to show that $\mathbb{P}(X \leq \mu) = 0$. □

Problem 2.3 (Existence of a large cut)**.** Given an undirected graph G with n vertices and m edges, show that there is a partition of V into two disjoint sets A and B such that at least $m/2$ edges connect a vertex in A to a vertex in B, i.e., there is a cut with value at least $m/2$.

Solution. Construct sets A and B by randomly and independently assigning each vertex to one of the two sets. Let $e_1, \ldots, e_m$ be the edges of G. For $i = 1, \ldots, m$, define X_i such that

$$X_i = \begin{cases} 1 & \text{if } e_i \text{ connects } A \text{ to } B, \\ 0 & \text{otherwise.} \end{cases}$$

The probability that edge e_i connects a vertex in A to a vertex in B is $1/2$, and thus $\mathbb{E}[X_i] = \frac{1}{2}$.

Let $C(A,B)$ be a random variable denoting the value of the cut corresponding to the sets A and B. Then

$$\mathbb{E}[C(A,B)] = \mathbb{E}\left[\sum_{i=1}^m X_i\right] = \sum_{i=1}^m \mathbb{E}[X_i] = \frac{m}{2}.$$

Since the expectation of $C(A,B)$ is $m/2$, there exists a partition A and B with at least $m/2$ edges connecting A to B. □

Problem 2.4 (Las-Vegas algorithm)**.** A sample involves assigning every vertex in V to one of A or B independently and uniformly at random. What is the expected number of samples before one obtains a cut with value at least $m/2$?

Solution. Let

$$p = \mathbb{P}\left(C(A,B) \geq \frac{m}{2}\right).$$

Then

$$\begin{aligned}\frac{m}{2} &= \mathbb{E}[C(A,B)] \\ &= \sum_{i \leq m/2-1} i \cdot \mathbb{P}(C(A,B)=i) + \sum_{i=m/2}^{m} i \cdot \mathbb{P}(C(A,B)=i) \\ &\leq (1-p)\left(\frac{m}{2}-1\right) + pm,\end{aligned}$$

implying that

$$p \geq \frac{1}{m/2+1}.$$

The number of samples required before obtaining a cut of size $m/2$ is given by a random variable N with a geometric distribution with success probability p. Thus $\mathbb{E}[N] \leq m/2+1$. □

2.3 Derandomization

Problem 2.5 (Derandomization)**.** The goal of this problem is to use Problem 2.3 to explicitly construct a cut of size $m/2$.

Let X be the random variable counting the number of edges in the cut (A,B). Suppose the vertices are ordered arbitrarily as $v_1,\dots,v_n$. We will assign them one-by-one to either A or B. Give a deterministic algorithm using the conditional expectations $\mathbb{E}[X \mid \text{choices for } v_1,\dots v_{i-1}]$ in order to construct a cut of size at least $m/2$ in polynomial time.

Solution. Given a placement of vertices $v_1,\dots,v_i$ ($i \geq 0$), we show how to place the next vertex v_{i+1} such that

$$\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_i] \leq \mathbb{E}[X \mid \text{choices of } v_1,\dots,v_i,v_{i+1}].$$

The base case is true as the placement of v_1 does not matter due to symmetry :

$$\mathbb{E}[X \mid \text{choice of } v_1] = \mathbb{E}[X].$$

Suppose we have placed $v_1,\dots,v_k$. Let Y_{i+1} be the random variable representing the set where v_{i+1} is placed. By the law of total expectations

$$\begin{aligned}\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_{k+1}] = &\frac{1}{2}\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=A] \\ &+ \frac{1}{2}\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=B]\end{aligned}$$

from which it follows that

$$\begin{aligned}\max\left(\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=A], \mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=B]\right) \\ \geq \mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k]\end{aligned}$$

We only need to compute the larger value among $\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=A]$ and $\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=B]$ and place v_{k+1} in the set which yields the larger expectation. The value of $\mathbb{E}[X \mid \text{choices of } v_1,\dots,v_k, Y_{k+1}=A]$ is the sum of the number of edges crossing the sub-cut given by the $k+1$ vertices in the conditioning and half of the remaining edges (use the same argument as in Problem 2.3). This can be easily computed. □

Remark. If you are a fan of this method, see the book of the same name by Alon and Spencer, and for a compilation of problems, see “Unexpected Uses of Probability" by Ravi Boppana.

2.4 Because why not

Problem 2.6 (Improving Problem 2.3). Show that there exists a partition in Problem 2.3 such that at least $mn/(2n-1)$ edges cross the partition.