1 Brief Review of Previous Lecture

Last lecture, we introduced the $k$-means problem. We also gave an FPT $(1 + \varepsilon)$-approximation algorithm for the case when the optimal clusters were of size $n/k$, and introduced the $D^2$-sampling idea of [JKS12]. In this class, we prove correctness of the full $(1 + \varepsilon)$-approximation algorithm for arbitrary cluster sizes.

We review some important notation from the previous lecture. For a point set $P$, we write $cg(P)$ for $\frac{1}{|P|} \sum_{p \in P} p$. We define $\Delta(S,c) = \sum_{x \in S} \|x - c\|^2$.

We will make use of the following two facts.

\begin{align*}
  (*) & \quad \Delta(P,a) = \Delta(P,cg(P)) + |P| \cdot \|a - cg(P)\|^2, \\
  (**) & \quad \|a - b\|^2 \leq 2(\|a - c\|^2 + \|c - b\|^2)
\end{align*}

2 JKS

We first present the entire JKS algorithm in full, and then subsequently prove its correctness.

function JKS($P, A^{i-1}$)
  if $|A^{i-1}| = k$ then
    Record solution quality
  else
    For $p \in P$, let $d_p = \text{dist}(p, A^{i-1})$
    Sample $S_i$ of size $\Omega(k/\varepsilon^3)$ from $P$, where $p$ is sampled with probability proportional to $d_p^2$
    for $T \subseteq S_i$ of size $\Omega(1) / \varepsilon$ do
      JKS($P, A^{i-1} \cup \{cg(T)\}$)
    end for
  end if
end function

2.1 Stability

To ensure correctness of our algorithm, we need to ensure one more condition, called stability. Let $\text{OPT}_k$ denote the value of the optimal solution with $k$ clusters for point set $P$. We say that $P$ is stable if $\text{OPT}_k \geq (1 + \varepsilon) \text{OPT}_k$.

There are two good reasons that we are not missing out on much by considering only stable point sets. The first reason is: if our point set is not stable then intuitively the "true" number of clusters is not $k$, but smaller. If the true number of clusters were $k$, we should expect to gain a significant improvement moving from $k - 1$ to $k$ clusters.

To see the second reason, first note that there must be some maximal $r \leq k$ where stability does hold with $r$ clusters, or $\text{OPT}_{r-1} \geq (1 + \varepsilon) \text{OPT}_r$. Then we can find a $(1 + \varepsilon)$-approximation for $\text{OPT}_r$. Adding $k - r \leq k$ more points each increases the value of $\text{OPT}$ by at most a $1 + \varepsilon$ multiplicative
factor due to instability for point set sizes above \( r \), thus for any point set we can obtain a \((1 + \varepsilon)^k\) approximation. If we use \( \varepsilon/k \) instead of \( \varepsilon \), this yields a full \((1 + \varepsilon/k)^k \leq (1 + \varepsilon)\)-approximation algorithm.

Thus we are justified in considering only the case of stable point sets.

### 3 Correctness of JKS

In the analysis of the JKS algorithm, we will use \( O_i \) to denote the \( i \)-th cluster in an optimal \( k \)-means solution, and use \( c_i \) to denote the center of the \( i \)-th cluster. We define \( r_i = \frac{1}{|O_i|} \Delta(O_i, c_i) \), the average squared distance to the center of gravity of points in \( O_i \).

Let \( A^{i-1} = \{a_1, a_2, \ldots, a_{i-1}\} \) and assume inductively that \( \forall i, \|a_i - c_i\|^2 \leq \frac{\varepsilon}{20} \cdot r_i \). Throughout this section, we will show that with high probability, at least one of the recursive calls of JKS will maintain this property. If we reach the bottom of the recursion, \((*)\) applies to show that we obtain a \((1 + \varepsilon)\)-solution.

Given the IH, then the following two lemmas hold. Intuitively, Lemma 19.1 says that some some cluster \( O_j \) has a high enough probability of being hit by a \( D^2 \)-sample, and Lemma 19.2 says that given a chosen point is this cluster \( O_j \), each \( p \in O_j \) is chosen with probability not too far off from uniform.

**Lemma 19.1.** There exists \( j \geq i \) such that

\[
\frac{\sum_{p \in O_j} d_p^2}{\sum_{q \in P} d_q^2} \geq \frac{\varepsilon}{2k}
\]

**Lemma 19.2.** For the \( j \) guaranteed to exist by Lemma 19.1, for all \( p \in O_j \) we have

\[
\frac{d_p^2}{\sum_{q \in O_j} d_q^2} \geq \frac{1}{O_j} \cdot \frac{\varepsilon}{64}
\]

**Proof of Lemma 19.1.** Suppose no such \( j \) exists. Define \( w_j = \frac{\sum_{p \in O_j} d_p^2}{\sum_{q \in P} d_q^2} \) for ease of notation. Then our supposition says \( w_j \leq \frac{\varepsilon}{2k} \) for \( j > i \). This means

\[
\sum_{j \geq i} w_j \leq \sum_{j \geq i} \frac{\varepsilon}{2k} \leq \frac{\varepsilon}{2}.
\]

Note that \( \sum_{i=1}^k w_i = 1 \) since the \( O_j \) are disjoint and union to \( P \). Then \( \sum_{j \leq i-1} w_j \geq 1 - \frac{\varepsilon}{2} \), and
multiplying both sides by \( \frac{\sum_{q \in P} d_q^2}{1 - \varepsilon/2} \) we find

\[
\sum_{q \in P} d_q^2 \leq \frac{1}{1 - \varepsilon/2} \sum_{j \leq i - 1} \sum_{p \in O_j} d_p^2 \\
\leq \frac{1}{1 - \varepsilon/2} \sum_{j \leq i - 1} \Delta(O_j, a_j) \\
\leq \frac{1}{1 - \varepsilon/2} \sum_{j \leq i - 1} \Delta(O_j, c_j) + |O_j||a_j - c_j|^2 \\
\leq \frac{1}{1 - \varepsilon/2} \left( Q + \sum_{j \leq i - 1} |O_j||a_j - c_j|^2 \right) \\
\leq \frac{1}{1 - \varepsilon/2} \left( Q + \sum_{j \leq i - 1} |O_j|\frac{\varepsilon r_i}{20} \right) \\
\leq \frac{1 + \varepsilon}{1 - \varepsilon/2} \cdot Q \\
\leq (1 + \varepsilon)Q \tag{by (\star)}
\]

However this means \( \text{OPT}_{k-1} \leq (\text{cost of using } A^{i-1} \text{ as centers}) \leq (1 + \varepsilon)Q \leq (1 + \varepsilon) \text{OPT}_k \), which contradicts the stability of \( P \). Thus our assumption no such \( j \) exists must have been false.

We now prove a lemma that will be helpful in establishing Lemma 19.2. Intuitively, the lemma says that the stability condition ensures the centers of optimal clusters cannot be too close together.

**Lemma 19.3.** For \( i \neq j \), we have

\[
\|c_i - c_j\|^2 \geq \varepsilon(r_i + r_j)
\]

**Proof.** Suppose to the contrary that there are \( i, j \) with \( \|c_i - c_j\|^2 < \varepsilon(r_i + r_j) \) Say WLOG that \( |O_i| \geq |O_j| \). Then we analyze the change in cost by removing \( c_j \) and putting all points in \( C_j \) into \( O_i \)'s cluster

\[
\Delta(O_i \cup O_j, c_i) \leq |O_i|r_i + |O_j|r_j + |O_j||c_i - c_j|^2 \\
< |O_i|r_i + |O_j|r_j + |O_j| \cdot \varepsilon(r_i + r_j) \\
\leq (1 + \varepsilon)(|O_i|r_i + |O_j|r_j)
\]

Thus removing \( c_j \) the cost goes up by less than a \( 1 + \varepsilon \) multiplicative factor. This contradicts the stability of \( P \), so no such \( i \) and \( j \) can exist. \( \Box \)

We are now ready to present the proof of Lemma 19.2.

**Proof of Lemma 19.2.** Suppose \( p \in O_j \), and let \( a_\ell \) for \( \ell \leq i - 1 \) be the closest point in \( A^{i-1} \) to \( p \). We will lower-bound the numerator and upper-bound the denominator of \( d_p^2/\sum_{q \in O_j} d_q^2 \).
To bound the numerator, see
\[ d_p^2 = \|p - a_\ell\|^2 \]
\[ \geq \frac{1}{2} \|p - c_\ell\|^2 - \|a_\ell - c_\ell\|^2 \quad \text{(By (**))} \]
\[ \geq \frac{1}{8} \|c_j - c_\ell\|^2 - \|a_\ell - c_\ell\|^2 \quad (p \text{ is closer to } c_j \text{ than } c_\ell) \]
\[ \geq \frac{1}{8} \|c_j - c_\ell\|^2 - \frac{\varepsilon}{20} \cdot r_\ell \quad \text{(By IH)} \]
\[ \geq \frac{1}{16} \|c_j - c_\ell\|^2 \quad \text{(By Lemma 19.3, } \|c_j - c_\ell\|^2 \geq \varepsilon r_\ell) \]

To bound the denominator, see
\[ \sum_{q \in O_j} d_q^2 \leq \sum_{q \in O_j} \|q - a_\ell\|^2 \]
\[ = \sum_{q \in O_j} \|q - c_j\|^2 + \|c_j - a_\ell\|^2 \quad \text{(By (*))} \]
\[ = |O_j| r_j + |O_j| \|c_j - a_\ell\|^2 \]
\[ \leq |O_j| r_j + 2|O_j| (\|c_j - c_\ell\|^2 + \|c_\ell - a_\ell\|^2) \quad \text{(By (**))} \]
\[ \leq |O_j| (r_j + 2\|c_j - c_\ell\|^2 + r_\ell) \quad \text{(By IH)} \]
\[ \leq |O_j| (2 + 1/\varepsilon) \|c_j - c_\ell\|^2 \quad \text{(By Lemma 19.3)} \]

Putting these two together, we get
\[ \frac{d_p^2}{\sum_{q \in O_j} d_q^2} \geq \frac{\|c_j - c_\ell\|^2/16}{|O_j| \|c_j - c_\ell\|^2 (2 + 1/\varepsilon)} = \frac{1}{|O_j|} \cdot \frac{1}{32 + 2/\varepsilon} \geq \frac{1}{|O_j|} \cdot \frac{\varepsilon}{64} \]

We will reference one more result, presented without proof here as it was proved last lecture as Theorem 18.4

**Lemma 19.4.** Let \( S \) be an i.i.d uniform sample of size \( M \) from a point set \( P \). Then
\[ \Delta(P, cg(S)) \leq (1 + \frac{1}{\delta M}) \Delta(P, cg(P)) \]
with probability at least \( 1 - \delta \).
We now tie everything together and sketch the proof that with high probability, there exists $T \subseteq S^i$ such that $\|c_j - cg(T)\| \leq \frac{\epsilon}{20} \cdot r_i$, establishing the inductive step. Let $Y_k$ denote a random variable over $P$ which is the $k$-th point picked for $S^i$, so $S^i = \bigcup_k Y_k$. Let $j$ be the index guaranteed to exist by Lemma 19.1. We define a new random variable $X_k$ dependent on $Y_k$ over $O_j \cup \{\emptyset\}$, where a value of $\emptyset$ denotes that $X_k$ is “unset”. If $Y_k \notin O_j$, then $X_k = \emptyset$. Otherwise if $Y_k = p$, set $X_k = p$ or $X_k = \emptyset$ with probability specifically chosen so that $P[X_k = p] = \frac{\epsilon}{2k} \cdot \frac{\epsilon}{|Oj|^{1/2k}} = \frac{\epsilon^2}{|Oj|^{1/2k}}$, which must be possible by Lemmas 19.1 and 19.2 as $p$ has at least this probability of being chosen as $Y_k$. Then with probability $\frac{\epsilon^2}{|Oj|^{1/2k}}$ we have $X_k$ chosen uniformly at random from $O_j$, and otherwise unset. Let $S_i = \bigcup_k X_k$. Note that $S_i \subseteq S^i$, so if a set $T \subset S^i$ has the desired property, then the same subset of $S^i$ does as well. To finish the sketch, we make the constant factor in the size $\Omega(k/\epsilon^3)$ of $S^i$ large enough such that with high probability $|S_i|$ is large enough to apply Lemma 19.4 and show $\Delta(O_j, cg(S^i)) \leq (1 + \frac{\epsilon}{20})\Delta(O_j, c_j)$.

4 TSP Parameterized by $d$

In this section we give an overview of $(1 + \epsilon)$-approximation algorithms for the travelling salesman problem, possibly parameterized by dimension $d$. For the case where points in the TSP problem lie in $\mathbb{R}^2$ with the Euclidean metric, Arora and Mitchell [Aro98] [Mit99] achieve a $(1+\epsilon)$-approximation in $n^{O(1/\epsilon^2)}$ time. For general points in $\mathbb{R}^d$, Arora [Aro98] gives an $n^{(1/\epsilon)\Theta(d)}$ algorithm. It has been shown that in general a doubly exponential dependence on $d$ is required. Rao and Smith [RS98] give an algorithm with runtime $2^{(1/\epsilon)\Theta(d)}n + \log n$. The algorithm of Arora and Mitchell and of Arora generalize to arbitrary metric spaces, while the algorithm of Rao and Smith is not known to. Of course, for this assertion to make sense we must define the dimension of an arbitrary metric space, which is due to Nagata [Nag64] and Assouad [Ass79].

4.1 Dimension of a Metric Space

Let $M$ be a matrix space with metric $d$. For $S \subseteq M$, define the diameter of $S$ as $\text{diam}(S) = \max_{x,y \in S} d(x,y)$. Let $\ell(S)$ be the minimum $\ell$ such that there exist $C_1, C_2, \ldots, C_\ell \subseteq M$ such that $S \subseteq \bigcup_{i \in [\ell]} C_i$ and $\text{diam}(C_i) \leq \frac{1}{2} \text{diam}(S)$. The dimension of $M$ is defined as

$$\text{dim}(M) = \max_{S \subseteq M} \log d(S).$$

To obtain some intuition for this definition, think of $M = \mathbb{R}^d$ and let $S$ be the $d$-hypercube with side length 2. Then $2^d$ length $d$-hypercubes of side length 1 are needed to cover all of $M$, and this is the maximum over all such sets in $\mathbb{R}^d$.

References


