Chapter 1

An $O(log^*n)$ Approximation Algorithm for the Asymmetric p-Center Problem

Sundar Vishwanathan
Department of Computer Science
Indian Institute of Technology
Bombay, India 400076.
sundar@cse.iitb.ernet.in

Abstract

The input to the asymmetric p-center problem consists of an integer p and an $n \times n$ distance matrix D defined on a vertex set V of size n, where d_{ij} gives the distance from i to j. The distances are assumed to obey the triangle inequality. For a subset $S \subseteq V$ the radius of S is the minimum distance R such that every point in V is at a distance at most R from some point in S. The p-center problem consists of picking a set $S \subseteq V$ of size p to minimize the radius. This problem is known to be NP-complete.

For the symmetric case, when $d_{ij} = d_{ji}$, approximation algorithms that deliver a solution to within 2 of the optimal are known. David Shmoys, in his article [12], mentions that nothing was known about the asymmetric case. Rina Panigrahy [11] recently gave a simple $O(\log n)$ approximation algorithm. We improve this substantially: our algorithm achieves a factor of $O(\log^* n)$.

1 Introduction

The p-center problem is a canonical problem of the "facility location" type. Imagine that you are given a map of a city, along with the time it takes to reach point x from point y, for all important pairs of points x, y in the city. You have to decide where to place p facilities, say p fire-stations, so that any important point is reachable quickly from at least one of these fire stations. We will assume that the fire-stations have to be located at one of the important points in the city. Because of the distributions in traffic density, or perhaps because of one-way streets, it is very likely that the time it takes to go from x to y is very different from the time it takes to travel from y to x. This is precisely the asymmetric p-center problem. If, however, the time taken to travel from x to y is the same as the time taken to travel from y to x, then it becomes an instance of the symmetric p-center problem or simply the p-center problem.

More formally, the input to the (asymmetric) p-center problem consists of an $n \times n$ distance matrix D, and a positive integer p. An entry d_{ij} in the distance matrix defines the distance from point i to point j (In the example, the time taken to reach destination j from point i.). We will assume that the distances obey the triangle inequality, that is, $d_{ij} + d_{jk} \geq d_{ik}$, for all i, j, k. For a set of points S, the radius of S is the minimum distance R such that every point is within a distance R of some point in S. As output to the p-center problem we require a set of p points, called the *centers*, with minimum radius. The decision version of this problem has an additional parameter, a positive real number R. The question one asks is whether one can find a set C of p points, the centers, such that every other point is at a distance at most R from some point in C.

This problem is known to be NP-complete [5]. In fact, it remains NP-complete even when the distance matrix D is symmetric and when the distances are restricted to be either 1 or 2. It is then natural to ask how good a solution can one find in polynomial time. Before we discuss the substantial amount of work that has already been done, we introduce the notion of an *Approximation Algorithm* and related terminology. See [5] for more details.

An approximation algorithm for a problem, loosely speaking, is an algorithm that runs in polynomial time and produces an "approximate solution" to the problem. Let Π be a minimization problem. Let $\mathcal{A}$ be an algorithm for Π. For an instance I of Π, let $\mathcal{A}(I)$ denote the value of the output of $\mathcal{A}$ on input I and let $OPT(I)$ denote value of the optimum solution. We define the absolute performance ratio for $\mathcal{A}$ to be $\inf\{r \geq 1 : \frac{A(I)}{OPT(I)} \leq r$ for all instances $I\}$. For maximization problems we define the absolute performance ratio to be $\sup\{r \leq 1 : \frac{A(I)}{OPT(I)} \geq r$ for all instances $I\}$. The *asymptotic performance ratio* is given by $\inf\{r \geq 1 :$ for some $N \in Z^+, \frac{A(I)}{OPT(I)} \leq r$ for all I satisfying $OPT(I) \geq N\}$. For the purposes of this paper, this distinction between the absolute and the asymptotic performance ratios is not necessary and and hence we will just refer to this quantity as the approximation ratio. We will say that an algorithm is an α-approximation algorithm if it always delivers a solution to within a factor α of the optimum.

Hochbaum and Shmoys [6] give a 2-approximation algorithm for the p-center problem when D is symmetric. Dyer and Frieze [4] describe a different 2-approximation algorithm for the same problem. Hsu and Nemhauser [7] prove that unless P = NP, this is the best possible. This lower bound however does not hold for specific metrics, like for example the L_2 norm. Slightly worse lower bounds are however known. The exact complexity in these cases remains an interesting open problem.

We refer the reader to the article by Shmoys [12] for an excellent account of the current status of this problem (and many others!). His concluding remarks on the p-center problem are relevant to this work and we reproduce them below.

"A natural generalization of the problem is to relax the restriction that the distance matrix be symmetric. This turns out to be a non-trivial generalization and essentially nothing is known about the performance guarantee for this extension."

The asymmetric case does turn out to be harder in other problems too. The most celebrated such example is the travelling salesman problem. While for the symmetric case a 3/2-approximation algorithm is known [5], for the asymmetric case the best known algorithm only achieves a factor of $O(\log n)$.

Rina Panigrahy [11] observed that a recursive application of the 'greedy set cover heuristic' yields an $O(\log n)$ approximation algorithm for the asymmetric p-center problem.

In this paper we substantially improve this bound. We achieve a bound of $O(\log^* n)$.

2 Preliminaries

Our notation is standard. Consider n points with pairwise distance matrix $D =< d_{ij} >$. Occasionally, for convenience, we will also use $d(i,j)$ instead of d_{ij}. For a point v, $\Gamma^+{}_d(v)$ will denote the set $\{x : d_{vx} \leq d\}$. and $\Gamma^-{}_d(v)$ will denote the set $\{x : d_{xv} \leq d\}$. We will say that a center v covers a point x within R if $d_{vx} \leq R$. We will also say that v R–covers x. Extending this notation we will say that a set of centers C covers a set A within R if for every $a \in A$ there is a $c \in C$ that covers a within R.

The **Set Cover** problem is the following:

Instance: Set S, $\mathcal{F}$, a collection of subsets of S, and a positive integer k.

Question: Does there exist a $\mathcal{F}' \subseteq \mathcal{F}$, $|\mathcal{F}'| \leq k$, such that every element of S occurs at least once in some set in $\mathcal{F}'$?

The performance of the greedy heuristic for the Set Cover problem has been analyzed by many people [3, 9, 8]. The following theorem is well known.

THEOREM 2.1. *Supposing that the optimal solution to the set cover problem above had a value p. Then the greedy algorithm that chooses a set which covers the maximum number of elements at each stage outputs a cover of size at most $p(1 + \ln(\frac{n}{p}))$, where $|S| = n$.*

We sketch a simple proof, told to the author by Jaikumar Radhakrishnan. It is easy to see that when one picks the set that covers the most number of elements, the number of as yet uncovered elements drops by a factor of $(1 - \frac{1}{p})$. Hence, after having picked $p\ln(\frac{n}{p})$ sets the number of uncovered elements is at most p. The result follows. For another clever proof that gives better lower order terms see [11].

3 The Algorithm

We will assume that we know the optimum radius R. This is not a serious problem since there are just $O(n^2)$ possible values for R and we can run our algorithm for each of these values and choose the best solution.

During the course of the algorithm we would have picked some vertices as centers. These would cover some vertices. The set A of vertices that have not yet been covered will be called *active*.

Rina Panigrahy [11] proved that a recursive application of the greedy set cover heuristic yields the following theorem:

THEOREM 3.1. *There is a polynomial time algorithm that finds a solution to the asymmetric p-center problem to within a factor of $O(\log p + \log^* n)$ of the optimal.*

The following theorem can be inferred from Panigrahy's proof.

THEOREM 3.2. *Given $A \subseteq V, p, R$, so that V has an R–cover of size p, one can find in polynomial time a set of $2p$ vertices that cover A within a radius of $O(R\log^* |A|)$.*

Our algorithm will have two phases. For phase 2 we will need the following stronger theorem, for which we modify Panigrahy's algorithm slightly and strengthen the analysis.

THEOREM 3.3. *Suppose you are given $A \subseteq V, p, R$, so that A has an R–cover of size p. Also assume that there is a set C_1 of vertices that R-cover $V \setminus A$. Then one can find in polynomial time, a set of $2p$ vertices C_2, that together with C_1 cover A (and hence V) to within a radius of $O(R\log^* |A|)$.*

Here is the algorithm to prove theorem 3.3.

Algorithm RecursiveCover (A, p, R)

{We assume that there are p vertices that can cover A to within a radius of R. }

$A_0 \leftarrow A$.
$i \leftarrow 0$.
As long as $|A_i| > 2p$ repeat the following 3 steps:

1. Construct the following instance of Set Cover, $< S, \mathcal{F} >$. Set $S = A_i$. There is one set $X \in \mathcal{F}$ for each point $x \in V$. X consists of all points in A_i, R-covered by x.

2. We now run the greedy Set-Cover heuristic to get a set A'_{i+1} of points that R-cover A_i. $A_{i+1} \leftarrow A'_{i+1} \cap A$.

3. $i \leftarrow i + 1$.

Output A_i.
End RecursiveCover

We sketch the analysis. Details will be given in the full paper. We argue by induction on i that $A_i \cup \mathcal{C}_1$ covers A to within a radius of $2iR$. Notice that A'_{i+1} covers A_i to within R. Let A_{i+1} R-cover H_i and let $B_{i+1} = A'_{i+1} \setminus A_{i+1}$ R-cover D_i, where $D_i = A_i \setminus H_i$. We note that $\mathcal{C}_1$ covers B_{i+1} to within R and hence covers D_i to within $2R$. Hence $A_{i+1} \cup \mathcal{C}_1$ $2R$-cover A_i and by induction $(2iR + 2R)$-cover A.

Also, $|A_{i+1}|$ is approximately $p \log(|A_i|/p)$ and hence, by induction, approximately $p \log^{(i)}(|A|/p)$. (We use the fact that $1 + \ln x \leq \log_a x$ for small enough a and large enough x.) This ensures that the size falls to $2p$ in $O(\log^*(|A|))$ iterations. We can continue using this procedure, but that is inefficient. It takes a further $\log p$ iterations to decrease the size to p.

It would seem that once we get down to $2p$ vertices, perhaps, we can work with these efficiently and get a cover. But this does not seem possible. We will use RecursiveCover; but as a *last step* of our algorithm.

To describe our algorithm, we need the following notion. A vertex u will be called a **Center Capturing Vertex (CCV)** if $\Gamma^-{}_R(v) \subseteq \Gamma^+{}_R(v)$. It is easy to see that if a vertex v is a **CCV** then either it is a center or there is a center in $\Gamma^+{}_R(v)$.

We begin with an informal description of the algorithm.

There are two phases to the algorithm. The find or halve phase and the augment phase. In the first phase we repeatedly look for a **CCV** u, and include it in our cover as long as we can find one. We also remove from the active set every vertex covered within a radius of $2R$ from u.

We enter the second phase if we are left with some vertices to cover, none of which are **CCV**s. In other words, we are left with a non-empty set of active vertices A, none of which is a **CCV**. The combinatorial crux of this work is in proving that in such a case, roughly, if p vertices R-cover the vertices in A, then there exist $p/2$ vertices that $5R$-cover them. Hence, in a way, we have set ourselves up for an application of RecursiveCover.

Here are the formal details.

Algorithm Approximate-p-Center
Input is the distance matrix, the number of centers to pick p, and the optimum radius R.
The Find or Halve phase
$A \leftarrow V$
Repeat the following steps as long as there is a **CCV**, $v \in A$ and $p > 0$.

1. $\mathcal{C}_1 \leftarrow \mathcal{C}_1 \cup \{v\}$
{ we include v as a center }

2. $A \leftarrow A \setminus \Gamma^+{}_{2R}(v)$
{ we remove from the active set, vertices covered to within a radius of $2R$ by v }

3. $p \leftarrow p - 1$
{ note that the remaining vertices can be covered with one less center }

End the Find or Halve phase
The Augment phase
Set $A' \leftarrow A \setminus \Gamma^+{}_{5R}(\mathcal{C}_1)$.
If $|A'| \neq \emptyset$ then run *RecursiveCover(A',p/2,5R)* to get centers $\mathcal{C}_2$ of size p that together with $\mathcal{C}_1$ covers A within a radius of $O(R \log^* |A'|)$.
End the Augment phase
Output $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$.
End Algorithm

4 Analysis

This section is devoted to the proof of the following theorem.

THEOREM 4.1. *Algorithm Approximate-p-Center outputs a set of size p, that covers V to within a radius of $O(R\log^*|V|)$.*

Consider each execution of the **Find or Halve** phase. Suppose v were a **CCV**. Then either v is a center, or there is some center in $\Gamma^+{}_R(v)$. To see this consider the center c that covers v. Since v is a **CCV**, $c \in \Gamma^+{}_R(v)$. So at the end of the find or halve phase it is clear that every vertex not in A is $2R$-covered by vertices in $\mathcal{C}_1$. We prove, next, that if we reach the augment phase then there exist $p/2$ vertices that together with $\mathcal{C}_1$, $5R$-cover A'. (Note that the value of p now is that specified at the end of the find or halve phase. Hence, for instance if this value of p is 0 then all vertices must be covered at the end of the first phase.)

We begin with a simple combinatorial lemma.

LEMMA 4.1. *Let $D =< U, F >$ be a digraph. Then there is a subset $W \subseteq U, |W| \leq \frac{|U|}{2}$ such that every vertex with indegree at least 1 is reachable in at most two steps from some vertex in W.*

Proof. Induction on $|U|$. The basis, $n = 2$, is trivial. Also note that if the arc set F were empty we are done. For the inductive step, pick any vertex u with non-zero out-degree. Remove u and its neighbours. Since u has non-zero outdegree we reduce the size of the graph by at least 2. Hence by induction, there is a set of size at most $\frac{n-2}{2}$, which works for the remaining graph. Now, the only vertices in the original graph not covered by this set could be u, some of its neighbours, or perhaps vertices at distance 2 from u. This last case could occur if the indegree of some vertex at distance 2 from u fell to zero while removing neighbours of u. Hence adding u to this set gives us a set of the required size for the original graph. ∎

The next lemma is crucial to the working of the algorithm. We note that as long as we keep finding **CCV**s, we keep removing centers. The next lemma explains why the augment phase works.

LEMMA 4.2. *Consider a subset of the vertex set, A, that has an R-cover consisting of p vertices. Assume that we have already picked a set C_1 of vertices to $2R$-cover the vertices in $V \setminus A$. Suppose there are no* **CCV**s *in A. Then there are $p/2$ vertices C_2 that together with C_1, $5R$-cover A.*

Proof. Let $x_1, \ldots, x_p$ be the centers that form an R-covering of A. We partition these centers to be of three types:

1. centers in $V \setminus A$
2. centers $x \in A$ such that $\Gamma^-{}_{2R}(x) \cap (V \setminus A) \neq \emptyset$
3. centers $x \in A$ such that $\Gamma^-{}_{2R}(x) \cap (V \setminus A) = \emptyset$.

We note that the centers outside A are already $2R$-covered by the set C_1 of vertices that we have picked. Hence the vertices that these centers cover, are covered by vertices in C_1 within a radius of $3R$.

Consider any vertex covered by some center z of type (2). Since z is of type (2) there is a $u \in V \setminus A$ such that $d_{uz} \leq 2R$. But u is $2R$-covered by C_1. Hence C_1 $5R$-covers the points covered by z.

Summarizing, C_1 $5R$-covers all points except points covered by centers of type 3. Hence, for the proof of the lemma we need to prove the existence of $p/2$ points that $5R$-cover the points R-covered by centers of type 3.

Without loss of generality, let $x_1, \ldots, x_q$ be the centers of type 3, and $x_{q+1}, \ldots, x_s$ be the centers of type 2. Since the vertices $x_1, \ldots, x_q$ are not **CCV**s, there are vertices $y_1, \ldots, y_q \in V$ (perhaps not all distinct) such that $d_{y_i x_i} \leq R$ while $d_{x_i y_i} > R$. We note that $y_i \in A$, for otherwise, x_i would be a type 2 center. Consider some index $j \leq q$. Since x_j is of type 3, y_j is necessarily covered by some center u in A. Suppose to the contrary that some center $z \in V \setminus A$ covers y_j. Then $d(z, x_j) \leq d(z, y_j) + d(y_j, x_j) \leq 2R$ which contradicts the fact that x_j is a type 3 center. Moreover, this center u is distinct from x_j since $d_{x_j y_j} > R$.

We claim that at most $\frac{s}{2}$ points from $x_1, \ldots, x_s$ suffice to $5R$-cover the vertices R-covered by $x_1, \ldots, x_q$. In proof, consider the following auxiliary digraph B on s vertices, say $z_1, \ldots, z_s$. There is an arc from z_i to z_j if and only if y_j is covered by center x_i. Lemma 4.1 applies and it outputs a subset $\{z_{i_1}, \ldots, z_{i_m}\}$ of size at most $\frac{s}{2}$. We now verify that $x_{i_1}, \ldots, x_{i_m}$ form a $5R$-cover of the vertices R-covered by $x_1, \ldots, x_q$. To see this consider any vertex u R-covered by a center $x_j, j \leq q$. In the auxiliary digraph B, z_j will have indegree at least one since there is some $x_i, i \neq j$, that covers y_j. Hence there is some z_{i_h} such that z_j is reachable in at most two steps from it. We claim that u is $5R$-covered by x_{i_h}. So, consider a path $z_{i_h} \to z_p \to z_j$ in B. (The case when the path is of smaller length is analogously handled.) This means $d(x_{i_h}, y_p) \leq R$ and $d(x_p, y_j) \leq R$. We also know that $d(y_p, x_p) \leq R$ and $d(y_j, x_j) \leq R$. Hence $d(x_{i_h}, x_j) \leq d(x_{i_h}, y_p) + d(y_p . x_p) + d(x_p, y_j) + d(y_j, x_j) \leq 4R$. ∎

The proof of theorem 4.1 can now be inferred from the preceding discussion and theorem 3.3.

5 Conclusion

The interesting question here is if there exists an algorithm with constant approximation ratio for the asymmetric p-center problem. Perhaps a more careful look at the relaxation of the combinatorial structure that occurs when one imposes asymmetry will lead to such an algorithm.

However for the theorist it would be more interesting if this were not the case! This would then furnish the first example of a natural problem whose approximability is a very slowly growing function of the input size. The Feige-Lovasz 2-Prover-1-Round proof systems were used ingeneously by Lund and Yannakakis [10] to prove the hardness of set-cover. Can these proof systems be used again in the present context?

Acknowlegements. Thanks to Rina Panigrahy for discussions on this problem. His insight, to use the more refined estimate for the set cover problem, spurred this work. Thanks to Ajit Diwan and Howard Karloff for comments on an earlier version of this paper. Also, but for Howard sending me a preprint of Shmoys' article, I would never have worked on this problem. Last, I thank the University of Chicago for letting me use their computational facilities, which has been very helpful.

References

[1] A. V. Aho, J. E. Hopcroft and J. D. Ullman, *The Design and Analysis of Computer Algorithms*, Addison–Wesley, Reading, MA, 1974.

[2] B. Bollobas, *Combinatorics*, Cambridge University Press, 1986.

[3] V. Chvatal, *A greedy heuristic for the set covering problem*, Mathematics of Operations Research, 4 (1979), pp. 233–235.

[4] M. Dyer and A. Frieze, *A simple heuristic for the p-center problem*, Oper. Res. Lett., 3 (1991), pp. 285–288.

[5] M. R. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Fransisco, 1979.

[6] D. S. Hochbaum and D. B. Shmoys, *A best possible approximation algorithm for the k-center problem*, Math. Oper. Res. 10 (1985), pp. 180–184.

[7] W. L. Hsu and G. L. Nemhauser, *Easy and hard bottleneck location problems*, Discrete Appl. Math. , 1 (1979), pp. 209–216.

[8] D. S. Johnson, *Approximation Algorithms for Combinatorial Problems*, JCSS, 9 (1978), pp. 256–278.

[9] L. Lovasz, *On the ratio of optimal integral and fractional covers*, Discrete Mathematics, 13 (1975), pp. 383–390.

[10] C. Lund and M. Yannakakis, *On the hardness of approximating minimization problems*, Proceedings of ACM STOC, pages 286–293, 1993.

[11] Rina Panigrahy, *An $O(\log n)$ approximation algorithm for the asymmetric p-center problem*, Submitted.

[12] D. B. Shmoys, *Computing Near-Optimal Solutions to Combinatorial Optimization Problems*, Manuscript.

[13] R. E. Tarjan, *Data Structures and Network Algorithms*, SIAM, Pittsburgh, PA, 1983.