

APPROXIMATION ALGORITHMS FOR FACILITY LOCATION PROBLEMS

DILPREET KAUR



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

December 2023

©Indian Institute of Technology Delhi - 2023

All rights reserved.

APPROXIMATION ALGORITHMS FOR FACILITY LOCATION PROBLEMS

by

DILPREET KAUR

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of **Doctor of Philosophy**

to the



Indian Institute of Technology Delhi

December 2023

To Daddy Ji.

Though you didn't get to see this
you're in every page

Acknowledgements

“There are unknown unknowns”

– Donald Henry Rumsfeld

“There is no way around hard work. Embrace it. You have to put in the hours because there’s always something which you can improve.”

– Roger Federer

I am thankful to my advisor, Naveen Garg, whose unwavering support, mentorship, and guidance have been pivotal throughout my academic journey. I admire his perfected art of skillfully simplifying intricate problems to unveil elegant solutions. His adept analytical thinking and innovative problem-solving approach have been transformative in shaping my perspectives. The insights I gained have not only enhanced my academic prowess but also equipped me with life skills extending far beyond academia. Not just research pursuits, but I also learned the joy of teaching. My gratitude and respect for him cannot be quantified or repaid for the empathy and kindness he has bestowed on me. During the COVID-19 pandemic in the face of my personal loss, he extended beyond academic guidance, expressing genuine concern for my well-being and that of my family. He motivated me to return to my dream in a time of doubt. I am forever thankful and highly indebted for his wisdom and support.

I extend my gratitude to the Ministry of Electronics and Information Technology(MeitY), Gov-

ernment of India, for awarding me with the Visvesvaraya Fellowship, which supported me financially during my full-time enrolment as a Ph.D. I would also like to thank Walmart Global Technology Services, India, for their support in pursuing my studies on a part-time basis during the last phase of my Ph.D. while being employed by the company.

My Ph.D. journey could not have been complete without the immense 24×7 support from my friends and colleagues - Ovia Seshadri, Sandeep Kumar, Arindam Bhattacharya, Shikha Goel, Vinayak Gupta, Happy Mittal, Ankit Anand, Diksha Moolchandani, Divya Gautam and Varun Kumar. A special shout-out to Dishant Goyal for numerous technical discussions and constructive suggestions during the development of this research work. I would also like to thank the sports fraternity on campus with special mentions to Amit Kothari, Swati Agrawal, Ashim Garg, and Medha Agarwal. Their lessons in teamwork, sportsmanship and the cherished memories we crafted during Inter IIT events have been truly invaluable. I would also like to extend my heartfelt appreciation to my friends from M.E., B.Tech., and even back from school days, who have remained connected and been reliable pillars of support.

Lastly, I would like to express my heartfelt gratitude to my parents Ranjeet Singh and Rani, and my brother Jaspreet, for their unwavering love, unwavering support, and the endless sacrifices they have made to see me succeed. Their unconditional love is my constant motivation to strive for excellence and make them proud.

Dilpreet Kaur

Abstract

This thesis presents a study of facility location problems with a focus on extending the uncapacitated facility location problem to different settings. The (metric) uncapacitated facility location problem is defined as follows: given a set C of clients and a set F of locations where a facility can be opened, all situated within a metric space and governed by a cost function $c : F \times C \rightarrow \mathbb{R}^+$, with each facility i incurring a facility opening cost of f_i and can serve an unlimited number of clients. The objective is to identify the placement of optimal locations for opening facilities that collectively minimize the total cost of opening these facilities and connecting clients to them.

First, we consider the problem of locating service and charging stations to facilitate efficient commuting. This scenario deviates notably from traditional facility location problems, as the clients are not stationary vertices; instead, they navigate through the graph using specified paths or routes. While the underlying cost function retains the metric property, it's important to note that the shortest paths from a client to its nearest facility (also known as a detour) may not strictly adhere to metric characteristics.

In the service station location problem, we are given the paths followed by m clients and wish to locate k service stations, from a set of feasible locations, such that the maximum detour that a client has to take is minimized. We give a solution that has a maximum detour $3\text{OPT} + L$ where L is the length of the longest client path when we are allowed to open facilities on a

subset of all vertices. If we are permitted to establish facilities at any vertex in the graph then the maximum detour is bounded by $20PT + L$.

Electric vehicles have a limited range and charging stations need to be located so that a client can drive from the source to the destination without running out of charge. We consider two variants of the problem. In the first, given source and destination vertices for each client, the task is to locate charging facilities at a subset of locations, ensuring each client can feasibly traverse from source to destination without running out of charge. In the second variant with clients specifying their intended paths, the objective is to locate the charging facilities such that each client can follow its path without any detours. In both instances, the objective remains the minimization of the number of charging stations while ensuring feasible paths for the clients. On general graphs, the problem with source-destination pairs is at least as hard as the node-weighted group-Steiner tree but if we allow vehicles in our solution to have a range of 4 times that of the optimum then we obtain an 8-approximation. For the problem with specified paths, we show that finding an $o(\log m)$ approximation is hard even when vehicles of our solution have a range that is a constant times larger than that of the optimum solution.

For all these variants of locating service and charging stations problem, when the underlying graph is a tree and the facility can be located at any vertex on the tree, we show that the problem can be solved in polynomial time.

We now shift our focus to the problem of hierarchical facility location (HFLP). This problem extends the uncapacitated facility location problem such that the clients also demand a service that can be fulfilled by a facility and the clients connect to a facility only if the demanded service is installed at an open facility. Each facility has its own facility opening cost. The cost of installing services is given by a rooted tree and the cost of installing any service(s) is the cost of the tree induced by the leaves corresponding to this service(s) and the root. The tree representing the service installation cost at every facility is identical. The connection costs

between the clients and the facilities form a metric. Thus given a set C of clients and a set F of locations, the objective is to identify a set $O \subseteq F$ of locations to open facilities, install services, and assign clients to these open facilities to minimize the sum of connection costs, service installation costs, and facility opening costs.

A local search algorithm by Svitkina and Tardos [46] gives a 4.237 approximation for this problem. Hierarchical facility location is a special case of the submodular facility location where the facility opening cost is a submodular function of the clients served by the facility. An $O(\log \log n)$ approximation for this problem was recently proposed by Abbasi *et al.* [1] and this is the best known. They also showed that the greedy algorithm cannot perform better than $\Omega(\log \log n)$ for submodular facility location.

We will assess the greedy algorithm for the hierarchical facility location problem and we show that the greedy algorithm is a constant approximation. In doing so we identify a suitable set of inequalities that capture the essence of the greedy algorithm. These inequalities then help us formulate a linear program, an optimum solution to which identifies the worst-case performance of the greedy algorithm. For instances of size up to 300, the greedy algorithm has an approximation ratio at most 4 and this is also our conjecture for the approximation guarantee of the greedy algorithm for arbitrary-sized instances. Factor-revealing LP to identify the approximation guarantee of the uncapacitated facility location problem was discussed by Jain *et al.* [30]. However, since the number of trees on a set of m clients/services could be exponentially large, this idea cannot be applied directly to the setting of HFLP. The inequalities we formulate can be interpreted as the greedy algorithm applied to instances where the tree is a path, and this lets us formulate a succinct linear program.

We believe that the tools and techniques we develop for the hierarchical facility location problem would find use in analyzing greedy algorithms for other related problems like the group Steiner tree and the directed Steiner tree. In particular, the possibility of capturing the key el-

ements of the greedy algorithm by a simplified structure (a skewed path) is something worth exploring for these other problems too.

सार

यह थीसिस 'सुविधा स्थान' की समस्याओं का एक अध्ययन प्रस्तुत करती है, जिसमें 'सुविधा स्थान' के अक्षम विस्तार की समस्या का विभिन्न समायोजन पर ध्यान केंद्रित किया गया है। दी (मीट्रिक) अक्षम 'सुविधा स्थान' समस्या को इस प्रकार परिभाषित किया गया है: ग्राहकों का एक सेट C और स्थानों का एक सेट F जहां एक सुविधा खोली जा सकती है, सभी एक मीट्रिक स्थान के भीतर स्थित हैं और एक लागत फ़ंक्शन $c : F \times C \rightarrow \mathbb{R}^+$ द्वारा शासित होते हैं। प्रत्येक सुविधा i को खोलने की लागत f_i होगी और ये 'सुविधा स्थान' असीमित ग्राहकों को सेवा दे सकते हैं। इसका उद्देश्य सुविधाओं को खोलने के लिए इष्टतम स्थानों की पहचान करना है जो सामूहिक रूप से इन सुविधाओं को खोलने और ग्राहकों को उनसे जोड़ने की कुल लागत को कम करते हैं।

हम कुशल आवागमन की सुविधा के लिए सर्विस और चार्जिंग स्टेशन ढूंढने की समस्या पर विचार करते हैं। यह व्यवस्था रूढ़ि सुविधा स्थान की समस्याओं से उल्लेखनीय रूप से विचलित है, क्योंकि ग्राहक स्थिर शीर्ष नहीं हैं; इसके बजाय, वे निर्दिष्ट पथों या मार्गों का उपयोग करके ग्राफ़ के माध्यम से चलते हैं। जबकि अंतर्निहित लागत फ़ंक्शन मीट्रिक लक्षण को बरकरार रखता है, यह ध्यान रखना महत्वपूर्ण है कि ग्राहक से उसकी निकटतम सुविधा तक का सबसे छोटा रास्ता (जिसे चक्र के रूप में भी जाना जाता है) मीट्रिक विशेषताओं का सख्ती से पालन नहीं कर सकता है।

सर्विस स्टेशन स्थान समस्या में, हमें m ग्राहकों द्वारा अपनाए गए पथ दिए गए हैं और हम संभावित स्थानों के एक सेट से k सर्विस स्टेशनों का पता लगाना चाहते हैं, ताकि ग्राहक द्वारा लिया जाने वाला अधिकतम चक्र कम से कम हो। जब हमें सभी शीर्षों के सबसेट पर सुविधाएं खोलने की अनुमति होती है, हम एक समाधान देते हैं जिसमें अधिकतम चक्र $3OPT + L$ होता है जहां L सबसे लंबे क्लाइंट-पथ की लंबाई होती है। यदि हमें

ग्राफ़ में किसी भी शीर्ष पर सुविधाएं स्थापित करने की अनुमति है तो अधिकतम चक्र $2OPT + L$ से घिरा है।

इलेक्ट्रिक वाहनों की एक सीमित बैटरी होती है और चार्जिंग स्टेशन स्थापित करने की आवश्यकता होती है ताकि ग्राहक बिना चार्ज खत्म हुए स्रोत से गंतव्य तक वाहन द्वारा जा सके। हम समस्या के दो प्रकारों पर विचार करते हैं। पहले प्रकार में, प्रत्येक ग्राहक के लिए स्रोत और गंतव्य शीर्षों दिए गए हैं और कार्य स्थानों के सबसेट पर चार्जिंग सुविधाओं का पता लगाना है ताकि यह सुनिश्चित कर सकें कि प्रत्येक ग्राहक बिना चार्ज खत्म हुए स्रोत से गंतव्य तक यात्रा कर सकता है। दूसरे प्रकार में, ग्राहक अपने इच्छित पथ निर्दिष्ट करते हैं और उद्देश्य चार्जिंग सुविधाओं का पता लगाना है ताकि प्रत्येक ग्राहक बिना किसी रुकावट के अपने पथ का अनुसरण कर सके। दोनों उदाहरणों में, उद्देश्य ग्राहकों के लिए व्यवहार्य रास्ते सुनिश्चित करते हुए चार्जिंग स्टेशनों की संख्या को कम करना है। सामान्य ग्राफ़ पर, स्रोत-गंतव्य जोड़े के साथ समस्या कम से कम नोड-भारित समूह-स्टाइनर पेड़ जितनी कठिन है, लेकिन अगर हम अपने समाधान में वाहनों को बैटरी की 4 गुना सीमा की अनुमति देते हैं तो हमें 8-सन्निकटन प्राप्त होता है। निर्दिष्ट पथों वाली समस्या के लिए, हम दिखाते हैं कि $o(\log m)$ सन्निकटन खोजना तब भी कठिन होता है, जब हमारे समाधान के वाहनों की बैटरी की सीमा इष्टतम समाधान की तुलना में लगातार कई गुना अधिक होती है।

सर्विस और चार्जिंग स्टेशन समस्या का पता लगाने के इन सभी प्रकारों के लिए, जब अंतर्निहित ग्राफ़ एक पेड़ है और सुविधा पेड़ पर किसी भी शीर्ष पर स्थित हो सकती है, तो हम दिखाते हैं कि समस्या को बहुपद समय में हल किया जा सकता है।

अब हम अपना ध्यान श्रेणीबद्ध सुविधा स्थान की समस्या पर केंद्रित करते हैं। यह समस्या असंबद्ध सुविधा स्थान की समस्या को इस तरह बढ़ाती है कि ग्राहक एक ऐसी सेवा की भी मांग करते हैं जिसे एक सुविधा द्वारा पूरा किया जा सकता है और ग्राहक किसी सुविधा से तभी जुड़ते हैं जब मांग की गई सेवा एक खुली सुविधा पर स्थापित की जाती है। प्रत्येक सुविधा की अपनी सुविधा खोलने की लागत होती है। सेवाओं को स्थापित करने की लागत एक जड़ वाले पेड़ द्वारा दी जाती है और किसी भी सेवा को स्थापित करने की लागत इस सेवा (सेवाओं) और जड़ से संबंधित पत्तियों द्वारा प्रेरित पेड़ की लागत होती है। प्रत्येक सुविधा पर सेवा स्थापना लागत का प्रतिनिधित्व करने वाला पेड़ समान है। ग्राहकों और सुविधाओं के बीच जोड़ने की लागत एक मीट्रिक बनाती है। इस प्रकार ग्राहकों का एक सेट C और स्थानों का एक सेट F दिया गया है, इसका उद्देश्य एक सेट $O \subseteq F$ की पहचान करना है ताकि जोड़ने की लागत, सेवा स्थापना लागत और सुविधा खोलने की लागत के

योग को कम किया जा सके।

स्वित्किना और टार्डोस [3] द्वारा एक स्थानीय खोज एल्गोरिदम इस समस्या के लिए 4.237 सन्निकटन देता है। पदानुक्रमित सुविधा स्थान सबमॉड्यूलर सुविधा स्थान का एक विशेष मामला है जहां सुविधा खोलने की लागत सुविधा द्वारा सेवा प्रदान किए गए ग्राहकों का एक सबमॉड्यूलर कार्य है। इस समस्या के लिए एक $O(\log \log n)$ सन्निकटन हाल ही में अब्बासी et al. [1] द्वारा प्रस्तावित किया गया था और यह अब तक का सबसे अच्छा ज्ञात परिणाम है। उन्होंने यह भी दिखाया कि ग्रीडी एल्गोरिदम सबमॉड्यूलर सुविधा स्थान के लिए $\Omega(\log \log n)$ से बेहतर प्रदर्शन नहीं कर सकता है।

हम पदानुक्रमित सुविधा स्थान समस्या के लिए ग्रीडी एल्गोरिदम का आकलन करेंगे और हम दिखाएंगे कि ग्रीडी एल्गोरिदम एक निश्चित सन्निकटन है। ऐसा करने पर हम असमानताओं के एक उपयुक्त सेट की पहचान करते हैं जो ग्रीडी एल्गोरिदम के सार को पकड़ लेता है। ये असमानताएं हमें एक रैखिक कार्यक्रम तैयार करने में मदद करती हैं, एक इष्टतम समाधान जो ग्रीडी एल्गोरिदम के सबसे खराब प्रदर्शन की पहचान करता है। 300 तक के आकार के उदाहरणों के लिए, ग्रीडी एल्गोरिदम का सन्निकटन अनुपात अधिकतम 4 होता है और यह मनमाने आकार के उदाहरणों के लिए ग्रीडी एल्गोरिदम की सन्निकटन गारंटी के लिए हमारा अनुमान भी है। अक्षम सुविधा स्थान समस्या की अनुमानित गारंटी की पहचान करने के लिए कारक-खुलासा एलपी पर जैन et al. [2] द्वारा चर्चा की गई थी। हालाँकि, चूँकि m ग्राहकों/सेवाओं के सेट पर पेड़ों की संख्या तेजी से बड़ी हो सकती है, इसलिए इस विचार को सीधे पदानुक्रमित सुविधा स्थान समस्या पर लागू नहीं किया जा सकता है। जिन असमानताओं को हम तैयार करते हैं उनकी व्याख्या ग्रीडी एल्गोरिथ्म के रूप में की जा सकती है जो उन उदाहरणों पर लागू होती है जहां पेड़ एक पथ है, और यह हमें एक संक्षिप्त रैखिक कार्यक्रम तैयार करने की सुविधा देता है।

हमारा मानना है कि पदानुक्रमित सुविधा स्थान समस्या के लिए हम जो उपकरण और तकनीक विकसित करते हैं, उनका उपयोग समूह स्टीनर ट्री और निर्देशित स्टीनर ट्री जैसी अन्य संबंधित समस्याओं के लिए ग्रीडी एल्गोरिदम का विश्लेषण करने में किया जाएगा। विशेष रूप से, एक सरलीकृत संरचना (एक तिरछा पथ) द्वारा ग्रीडी एल्गोरिथ्म के प्रमुख तत्वों को पकड़ने की संभावना इन अन्य समस्याओं के लिए भी तलाशने लायक है।

Contents

Certificate	i
Acknowledgements	v
Abstract	vii
1 Introduction	1
1.1 Clustering Problem	3
1.2 Uncapacitated Facility Location Problem	4
1.3 k -Facility Location Problem	10
1.4 Capacitated Facility Location Problem	12
1.5 Organization of the Thesis	13
2 Locating Service Station	15
2.1 Overview	16
2.2 Related Work	18

2.3	Notations and Preliminaries	19
2.4	Linear program and the integrality gap	22
2.5	The graph is a Tree	25
2.6	General Graphs	28
2.6.1	F is a subset of V	28
2.6.2	$F = V$	31
3	Locating Charging Station	35
3.1	Overview	37
3.2	Related Work	39
3.3	Notations and Preliminaries	40
3.4	Locating Charging Stations for given source-destination pairs	42
3.4.1	The graph is a Tree	42
3.4.2	General Graphs	52
3.5	Locating Charging Stations with specified routes	59
3.5.1	The graph is a Tree	59
3.5.2	General Graphs	61
4	Hierarchical Facility Location	65
4.1	Overview	67

4.2	Related Work	69
4.3	Notations and Preliminaries	70
4.3.1	Linear Programming Formulation	72
4.4	The Greedy Algorithm	74
4.5	The Analysis of the Greedy Algorithm	77
4.5.1	The Worst Instance can be modeled as a Path	78
4.5.2	Analyzing Greedy on a Path: A special case	83
4.5.3	Analyzing Greedy on a Path: The general case	86
4.6	Conclusion	94
5	Experimental Results	95
5.1	Related Work	96
5.1.1	The Factor Revealing Linear Program	97
5.2	Hierarchical Facility Location Problem	99
5.2.1	The Factor-Revealing Linear Program	100
5.2.2	Results	102
5.3	Modified Greedy Approach for HFLP	103
5.3.1	Greedy Algorithm 2	104
5.3.2	The Factor-Revealing Linear Program	105

5.3.3 Results	109
6 Conclusion and Future Work	111
APPENDICES	113
A Factor-Revealing Linear Program	113
Bibliography	121
List of Publications	127
Biography	129

List of Figures

1.1	An example graph for metric k -supplier	3
1.2	An example graph for metric UFLP	5
1.3	Integer Linear Program for the metric Uncapacitated Facility Location Problem	6
1.4	Dual Program of LP-relaxation of Figure 1.3	7
1.5	A different Integer Programming formulation for the metric UFLP	8
1.6	Dual program of the Linear relaxation of ILP in Figure 1.5 for the metric UFLP	9
1.7	Integer Linear Program for the metric k -Facility Location Problem	11
1.8	Integer Linear Program for the metric Capacitated Facility Location Problem .	12
2.1	An example graph for $S_{\forall SP}()$	17
2.2	Metric property may not hold on client-paths in a graph	20
2.3	Integer Linear Program for $S_{\forall SP}(G, l, \mathcal{P}, k)$	22
2.4	A feasible solution of linear relaxation on complete graph	24
2.5	Reducing a graph H into a corresponding tree T	25

2.6	Proof of Lemma 2.	27
2.7	The problem $\text{SVSP}(G, l, \mathcal{P}, k)$ for $F \subset V$ has cost at most $30\text{PT} + L$	29
2.8	Reducing an instance of k -supplier to the problem $\text{SVSP}(G, l, \mathcal{P}, k)$	30
2.9	Reducing an instance of k -center to the problem $\text{SVSP}(G, l, \mathcal{P}, k)$	33
3.1	Reducing a graph H into a corresponding tree T with edges of length B	43
3.2	Subtree rooted at r_j to demonstrate claim 3	44
3.3	Subtree rooted at r_j to demonstrate claim 4	45
3.4	Three cases highlighting the position of z_1 in the tree T'	50
3.5	Analysis when r_j is in T' or not	51
3.6	Reducing an instance of the set cover problem to an instance of the $\text{ChSD}()$ with $F = V$	53
3.7	Reducing an instance of the GST problem to an instance of the FST problem (from Figure 1 in [7]).	55
3.8	Reducing an instance of the FST problem to an instance of the GST problem (from Figure 2 in [7]).	55
3.9	Short-circuit around terminal vertices which are internal nodes in T	57
3.10	Illustrating the reduction from set cover to $\text{ChSP}(G, l, \mathcal{P}, \alpha B)$	63
4.1	Illustration of an offline retail store that provides pre-sales and after-sales support.	67
4.2	A facility i and its services in tree T	70

4.3	A graph G constructed for HFLP	73
4.4	A directed graph G with facilities and active terminals	75
4.5	Apply Decomposability to determine a mapping $\phi : Y \rightarrow X$	80
4.6	The Worst instance can be modeled as a Path	83
4.7	Visual representation of the used notations	87
5.1	Dual program of the Linear relaxation of ILP in Figure 1.5 for metric UFLP . .	96
5.2	Factor-Revealing Linear Program for metric Uncapacitated Facility Location Problem (from [30])	98
5.3	A facility and its services on a skewed tree	100
5.4	Linear Program for metric hierarchical facility location problem	101
5.5	Factor-Revealing LP for metric Hierarchical facility location problem	102
5.6	Linear Program for metric hierarchical facility location problem when clients are allowed to switch	106
5.7	A facility and its services on a skewed tree to allow clients to switch	107
5.8	Factor-Revealing LP for metric Hierarchical Facility Location Problem when clients are allowed to switch	108

List of Tables

1.1	Notations for metric Uncapacitated Facility Location Problem	5
1.2	Approximation Bounds for the metric Uncapacitated Facility Location Problems	10
2.1	Summary of results for Locating Service Station Problem	17
2.2	Notations for Locating Service Station Problem	20
3.1	Summary of results for Locating Charging Station Problem	38
4.1	Notations for hierarchical facility location problem (HFLP)	71
5.1	Solution of the factor-revealing LP if clients cannot switch.	103
5.2	Solution of the factor-revealing LP if already connected clients can switch. . . .	109