Continuous Approximation Models with Temporal Constraints and Objectives

September 2023

A Research Report from the Pacific Southwest Region University Transportation Center

John Gunnar Carlsson, University of Southern California

Ying Peng, University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR-22-19	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Continuous Approximation Models with Temporal Constraints and Objectives		**5. Report Date** 09/30/23 **6. Performing Organization Code** N/A
7. Author(s) John Gunnar Carlsson, 0000-0001-5346-8529 Ying Peng		**8. Performing Organization Report No.** TBD
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (09/01/22-09/30/23) **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes https://doi.org/10.25554/qyys-za80		

16. Abstract

The purpose of this project is to discover new continuous approximation models for modern logistical problems in which time plays a significant role, with a specific focus on last-mile delivery. Famous examples of such problems include the vehicle routing problem with time windows (VRPTW) and the cumulative travelling salesperson problem (CTSP). The continuous approximation paradigm is a quantitative method for solving logistics problems in which one uses a small set of parameters to model a complex system, which results in simple algebraic equations that are easier to manage than (for example) large-scale optimization models. As a further benefit, one often obtains insights from these simpler formulations that help to determine what affects the outcome most significantly.

Although continuous approximation models have been used for over 60 years in logistics systems analysis, there has been very little research conducted on their use to problems with temporal features such as those described above. Based on our experience in this research area, this is likely because the addition of a time dimension complicates the problem in a way that is not readily accessible relative to classical models, which emphasize spatial aspects of modelling. However, our recent advances indicate that one can likely apply modern mathematical machinery to tackle these higher-dimensional problems. This project will combine tools from geospatial optimization, computational geometry, and geometric probability theory to formulate new models that will enable practitioners and policy-makers to solve these temporally-constrained problems, and most importantly, to identify what features are most impactful in their real-world application.

17. Key Words Vehicle routing; time constraints; last mile delivery	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	73	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Pacific Southwest Region UTC
University Transportation Center

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Disclosure

John Gunnar Carlsson conducted this research titled, "Continuous Approximation Models with Temporal Constraints and Objectives" at the Epstein Department of Industrial and Systems Engineering at the Viterbi School of Engineering at the University of Southern California. The research took place from August 2021 to August 2022 and was funded by a grant from METRANS in the amount of $99,998. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

The authors gratefully acknowledge the support of METRANS for funding this research.

Abstract

The purpose of this project is to discover new continuous approximation models for modern logistical problems in which time plays a significant role, with a specific focus on last-mile delivery. Famous examples of such problems include the vehicle routing problem with time windows (VRPTW) and the cumulative travelling salesperson problem (CTSP). The continuous approximation paradigm is a quantitative method for solving logistics problems in which one uses a small set of parameters to model a complex system, which results in simple algebraic equations that are easier to manage than (for example) large-scale optimization models. As a further benefit, one often obtains insights from these simpler formulations that help to determine what affects the outcome most significantly.

Although continuous approximation models have been used for over 60 years in logistics systems analysis, there has been very little research conducted on their use to problems with temporal features such as those described above. Based on our experience in this research area, this is likely because the addition of a time dimension complicates the problem in a way that is not readily accessible relative to classical models, which emphasize spatial aspects of modelling. However, our recent advances indicate that one can likely apply modern mathematical machinery to tackle these higher-dimensional problems.

Continuous Approximation Models with Temporal Constraints and Objectives

Executive Summary

The purpose of this project is to design simple and concise mathematical models for predicting trade-offs that arise in logistical problems with time constraints and objectives, such as the vehicle routing problem with time windows (VRPTW) and the cumulative travelling salesperson problem (CTSP). Examples of these trade-offs include the relationships between time to completion of service, average or worst-case customer satisfaction, vehicle miles travelled (VMT), or greenhouse gas (GHG) emissions. Traditionally, these problems have been solved in a discrete setting, involving fixed sets of (for example) demand points, time periods, and service facility locations; one then solves them with an integer mathematical programming solver such as CPLEX or Gurobi. A drawback of this approach is that the problems are almost always NP-hard, and hence solving large-scale instances would require enormous computational efforts which likely increase exponentially with the problem instance size. A further drawback is that such models are often extremely complex, which hinders understanding of salient problem features and managerial insights.

For these reasons, this project will use tools from geospatial optimization, computational geometry, and geometric probability theory to discover simple continuous approximation models that identify the key problem attributes that affect them most significantly. A continuous approximation model is characterized by its use of continuous representations of input data and decision variables as density functions over time and space, and the goal is to approximate the objective function into an expression that can be optimized by relatively simple analytical operations. Such an approximation enables transforming otherwise high-dimensional decision variables into a low-dimensional space, allowing the optimal solution to be obtained with mere calculus, even when significant operational complexities are present. The results from such models often bear closed-form analytical structures that help reveal managerial insights.

1 Introduction

The *traveling salesman problem* (TSP) is an NP-hard problem in computer science and combinatorial optimization. It is concerned with finding the optimal Hamiltonian cycle which has the minimum sum of edge weights. There are many problems in the literature that are based on the TSP, such as the traveling repairman problem (TRP), vehicle routing problem (VRP), and the traveling purchaser problem (TPP), which all have important applications in engineering and computer science problems such as optimal routing in communication networks, planning, logistics, manufacturing of microchips, transportation, and delivery systems.

However, as shown in recent publications, this kind of classical cost-minimizing problem may not properly reflect the need for fast service, or equity [58]. One specific example is the procurement of humanitarian aid in the context of natural disasters, such as tsunamis or earthquakes [58, 60, 21, 59]. When natural disasters strike, humanitarian supply chains (unlike commercial supply chains, which focus on quality and profitability) are focused on minimizing loss of life and suffering and should have higher priority and thus require the definition of more customer-centric or service-based objective functions. There are many performance measures that can be used to define such functions. Minimizing the average arrival time, or minimizing the sum of arrival times, are common. Apart from humanitarian aid after a natural disaster, there exist many other practical applications such as distribution, machine scheduling and power control and receiver optimization in wireless telecommunication systems [15].

The preceding examples suggest that a "customer-centric" problem should be considered, so as to better reflect priorities and ensure equity and fairness [21]. Instead of minimizing the total travel cost, which caused some commodities or customers to be served significantly later than others, this new problem's goal is to minimize the sum of customers' waiting times. This new problem is a variant of classical routing problems called the *Cumulative Routing Problem*. In this report, we are interested in the *Cumulative Travelling Salesman Problem* (CTSP) and the *Cumulative Capacitated Vehicle Routing Problem* (CCVRP), where CCVRP is a generalization of the CTSP.

Bianco, Mingozzi, and Ricciardelli [16] considered minimizing the sum of all distances traveled from the origin to all other cities in 1993. Later, Ngueveu, Prins, and Wolfler Calvo [58] took vehicle capacities into consideration and first introduced the *Cumulative Vehicle Routing Problem* (CVRP). Despite considerable research focus on this kind of problem and the generation of efficient heuristics and exact algorithms to attack it, they are still not flexible enough for real-world applications. One of the issues is that they cannot handle large-scale points distributed by different probability laws, which is the main characteristics of the real-world problems. For instance, the postal system relies on continuous approximations of tour length to partition the service territory [37]. Based on this motivation, this report studies the asymptotic analysis of this family of problems.

There is a long history of studies on asymptotic and probabilistic bounds over various graph structures. These works study many graph structures over Euclidean points in the context of the Beardwood-Halton-Hammersley (BHH) theorem and its extensions. This theorem is originally stated in [12] and later further developed by Steele in [72]. Instead of finding the optimal (shortest) tour length over random sample points embedded in Euclidean spaces, our goal of finding optimal (shortest) cumulative tour lengths adds an additional problem characteristic that compounds its difficulty.

This report is primarily concerned with the asymptotic behavior of both the *Cumulative Travelling Salesman Problem* (CTSP) and the *Cumulative Capacitated Vehicle Routing Problem* (CCVRP). We perform the analysis in the Euclidean plane, where points are uniformly and non-uniformly distributed respectively. We also describe the impact of vehicle capacities on the final cost. Furthermore, we extend our analysis to the case where multiple vehicles are involved, and we explore how coordinated service can affect the final solution. Finally, we demonstrate a practical application of our cumulative model by using the CTSP and CCVRP to predict the total waiting cost of a population, even when we only have access to a subset of the data. This application highlights the potential of our models to address real-world optimization problems, where incomplete information is often a significant challenge.

2 Literature Review

2.1 Research on TSP

The *traveling salesman problem* (TSP) is one of the most famous NP-hard combinatorial optimizations problems, which has captured the attention of mathematicians and computer scientists. There are many practical applications for this problem such as printed circuit board design [4], X-Ray Crystallography [17], Overhauling Gas Turbine Engines [61], the Order-Picking Problem in Warehouses [63], Computer Wiring [64], Vehicle Routing [11], Mask Plotting in PCB Production [33], control robots [49] and so on. This problem's statement is very simple: there are n cities on a map, a salesman wants to start from a home city, visit all the other cites exactly once, and come back to the home city, with the minimum amount of distance travelled.

One of the earliest studies on the TSP was conducted by Held and Karp [44]. They proposed an integer linear programming formulation for this problem and demonstrated its effectiveness through numerical experiments on various benchmark instances. They also proposed a dynamic programming algorithm that could optimally solve small instances of the problem. However, finding exact solution for large instances is infeasible due to its computational complexity. Therefore, researchers have turned to *heuristics* for TSP. These algorithms depend on an initial starting solution, and although they may not find the optimal solution, they can often find high-quality "sub-optimal" solutions much faster than exact approaches. One of the most widely used heuristic algorithms for the TSP is the Lin-Kernighan algorithm, developed by Lin and Kernighan [50]. Other approaches include genetic algorithms proposed by Grefenstette et al. [42], a hybrid heuristic algorithm combining genetic algorithms and local search techniques proposed by Bektaş and Laporte [13], simulated annealing introduced by Aarts, Korst, and Laarhoven [1], tabu search used by Fiechter [35], and an effective evolutionary algorithm proposed by Nagata [55].

In many applications, assuming a probability distribution on the sample points of the graph offers insights. There is a long history of studies on asymptotic bounds over various graph structures. One of the famous studies is the Beardwood-Halton-Hammersley (BHH) theorem and its extensions [12]. This analysis is to develop an effective approximation of the tour cost with a small computational price. Many graph structures over Eucilidean sample points have been studied in the context of BHH theorem.

2.2 Research on Cumulative TSP (CTSP)

The *Time-Dependent Traveling Salesman Problem* (TDTSP) is one of the most well-known variants of TSP. Given a graph $G = (V, A)$ where $V = \{1, 2, \ldots, n\}$ and $A = \{(i, j) : i, j = 1, \ldots, n, i \neq j\}$, this problem tries to find a minimum cost Hamiltonian circuit, where arc costs depend on its position in the tour [41].

One of the special case of TDTSP is called the *Cumulative Traveling Salesman Problem* (CTSP) also known as the traveling delivery problem [39], the *The Traveling Repairman Problem* (TRP) [54, 2, 9, 38], and the *the Minimum Latency Problem* (MLP) [7, 8, 18, 40, 70]. Its goal is to find a path, initiated at the depot and visiting every customer exactly once, such that the sum of the times required to reach every customer, along the path, is minimal.

Such problems are often used when a fairness criterion about visiting clients needs to be enforced. One of the application is delivering pizzas, which requires that the total time to visit all customers be minimal [36]. This problem can also be used in the area of computer networks [9]. Simchi-Levi and Berman [69] show how to use this problem to find the routing of automated guided vehicles through cells in a flexible manufacturing system. Arora and Karakostas[8] applied this problem in disk-head scheduling. However, even CTSP is a variant of TSP, it is an NP-hard problem and much harder to solve and approximate [18].

The CTSP problem is similar to TSP, and many different mathematical programming formulations have been proposed. Fischetti, Laporte, and Martello [36] used Integer Linear Programming to solve it. Méndez-Dıéaz, Zabala, and Lucena [54] exploited connections between TDP and the Linear Ordering Problem (LOP). Wu [76] used a dynamic programming algorithm to solve it, and Branch-and-Bound algorithm, which is used in TSP, can still be used to solve this problem [77]. However, due to the NP-hardness of this problem, all the approaches above cannot be used to solve medium- and large-scale instances.

Due to the NP-hardness of CTSP, researchers have turned to heuristic algorithms to obtain more tractable time complexity guarantees. Some promising approaches have been proposed. Dewilde et al.[31] described a tabu search algorithm with multiple neighborhoods that can quickly generate high-quality solutions. Beraldi et al.[14] formulated the problem via a non-linear model and used a beam search heuristic to solve it heuristically. Bruni, Beraldi, and Khodaparasti[20] presented a new heuristic based on General Variable Neighborhood Search that combines multiple neighborhoods in an effective way, outperforming previous heuristics in experiments. Salehipour et al.[68] developed a new approach called GRASP+VND/VNS, a multi-start method that consists of a greedy randomized construction phase and a variable neighborhood descent or search improvement phase in each iteration.

2.3 Research on Cumulative CVRP (CCVRP)

The *Vehicle Routing Problem* (VRP), a classic combinatorial optimization problem introduced by Dantzig and Ramser[30], has practical applications in resource allocation for electric power distribution [78], pharmaceutical distribution [47], and waste collection [62].

By considering additional constraints on route constructions, different Vehicle Routing Problems (VRPs) have been formulated. One of the practical and central variants is the Capacitated Vehicle Routing Problem (CVRP), which aims to design vehicle routes from a depot to a set of geographically-scattered customers, with capacity limit constraints. This problem's objective is to minimize the total travel cost. In such case, some customers may be served later than others, which may (in some contexts) violate equity and fairness [32]. To better reflect priorities and to ensure equity and fairness, it has been argued that the waiting time of a service system from the customer's point of view should be considered [21]. Based on this need, a new variant of the classical CVRP, the cumulative capacitated vehicle routing problem (CCVRP), has arisen.

The CCVRP was first introduced by Ngueveu, Prins, and Wolfler Calvo [58]. In contrast to the traditional VRP, whose objective function is cost-based, CCVRP's objective goal can be viewed as service-based. Its aims to minimize the sum of arrival times at all intermediate customers. Campbell, Vandenbussche, and Hermann [21] states that the CCVRP's and traditional VRP's optimal solutions are significantly different.

Ngueveu, Prins, and Wolfler Calvo [58] presented a mathematical model based on a cumulative vehicle routing problem with time windows [45] and developed a memetic algorithm (MA) as a solution method [58]. It is assessed using instance whose size range from 50 to 199 nodes. Rivera, Murat Afsar, and Prins [67] used mixed integer linear programs, a flow-based model and a set partitioning model for small instances with 20 sites. Their exact algorithm outperforms a commercial MIP solver on small instances and can solve cases with 40 sites to optimality. Mattos Ribeiro and Laporte [53] proposed an Adaptive Large Neighborhood Search (ALNS) for this problem. Chen, Dong, and Niu [25] provided a comparison with MA and ALNS. They showed that ALNS outperformed MA in terms of computational time and quality of the solution. Later, Ke and Feng [46] developed a two-phase metaheuristic and its result is better than MA and ALNS. In 2014, Lysgaard and Wøhlk [52] investigated the first exact algorithm for the CCVRP based on Branch-and-cut-and-price procedure (BCP). This algorithm is capable of solving instances with up to 150 customers in reasonable computational time.

When one has multiple depots for this cumulative routing problem, Rivera, Murat Afsar, and Prins [67] transformed it into a resource-constrained shortest path problem where each node corresponds to one trip and the sites to visit become resources. This transferred problem can be solved via an adaptation of Bellman–Ford algorithm to a directed acyclic graph with resource constraints and a cumulative objective function. Rivera, Afsar, and Prins [66] compared a mixed integer linear program (MILP), a dominance rule, and a hybrid metaheuristic: a multi-start iterated local search (MS-ILS) calling a variable neighborhood descent with $\mathcal{O}(1)$ move evaluations. On three sets of instances, MS-ILS obtains good solutions.

2.4 Research on continuous approximation models

There exist numerous discrete models for routing problems which have also been developed to address issues at the operational level in both deterministic and stochastic environments [10, 27, 48]. However, there exist many drawbacks. First, they generally have a relatively complex formulation structure that may hinder one's understanding of problem properties and managerial insights [5]. Often, the routing problems belong to the class of NP-hard problems, and hence solving large-scale instances would require enormous computational efforts, which likely increase exponentially with the problem instance size. Hence, it is not practical to solve large-scale logistics problems to optimality. All of this drawbacks are compounded further when we need to make decisions in stochastic, time-varying, competitive and coupled environment.

The *continuous approximation* paradigm was proposed as a means of partially addressing the preceding challenges. Continous approximation approaches were first proposed by Newell[56] and Newell [57]. It features continuous representations of input data and decision variables as density functions over time and space, and the key idea is to approximate the objective into a functional (e.g. and integral) of localized functions that can be optimized by relatively simple analytical operations. Each localized function approximates the cost structure of a local neighborhood with nearly homogeneous settings. Such homogeneous approximations enable mapping otherwise high-dimensional decision variables into a low-dimensional space, which allows the optimal design for this neighborhood to be obtained with simple calculus, even when spatial stochastic, temporal dynamics and other operational complexities are present.

One of the famous applications of continuous approximations is in routing problem, which determines the most economic routes for vehicles to deliver or pickup commodities or people across a continuous space. In routing problems, one of the most famous such theorems is the aforementioned BHH theorem[12]. It states that when the number of customer points approaches infinity on a compact area, the optimal tour length can be approximated by a simple analytic expression.

$$TSP(n) \sim k_{TSP}\sqrt{An}$$

where k_{TSP} is a constant, whose value is determined by the distance metric (typically Euclidean).

It is known that in practice, the BHH theorem result underestimates the tour length when the area is an elongated shape [29]. In order to address this issue, Daganzo [29] introduced a strip strategy method, which can efficiently compute the optimal tour length in different shapes. These two methods [12, 29] gave rise to several extensions. Later, Webb [74], Christofides and Eilon [26] [34] introduced the approximation to length of capacitated vehicle routing problem, which shows that the length is related to three terms: capacity, number of customers, and the average distance between the customers and the depot and area. Daganzo [28] shows when the depot is not necessarily located in the area that contains the customers, the tour length admits the following approximation:

$$\text{CVRP(n)} \sim \frac{2\bar{r}n}{c} + k_{CVRP}\sqrt{nA}$$

Where $\bar{r}$ represent the average distance between the customers, c corresponds to the capacity limit per vehicle, and A is the area.

Results of this kind inspired many subsequent studies. For example, Rifki et al. [65] focused on an asymptotic approximation of the traveling salesman problem with uniform non-overlapping time windows. Instead of visiting all points, Aldous and Krikun [3] formalized the idea of minimum average edge-length in a path linking some infinite subset of points of a Poisson process.

Based on the literature review, we conclude that both CTSP and CCVRP have been extensively studied, but have not yet been analyzed from the continuous approximation perspective. In this paper, we try to connect the bridge between both and fill the gap.

3 Problem Statement

3.1 Problem Assumption

We begin with the following assumptions about our four problems of interest, the Cumulative TSP (CTSP), the Cumulative Capacitated VRP (CCVRP) and their generalizations, which are the multiple Vehicle Cumulative TSP (m-CTSP) and the multiple vehicle Cumulative Capacitated VRP (m-CCVRP) respectively:

- The travel speed is a constant, which is proportional to the distance between points. Therefore, “distance” and “time” are essentially interchangeable.
- Regarding the Cumulative Capacitated Vehicle Routing Problem (CCVRP), we restrict our analysis to the scenario where a single vehicle is available, and it must periodically return to the depot as a result of its capacity constraints. It should be noted that the generalized version of CTSP and CCVRP will not be subject to this assumption.

3.2 Notation

The notational conventions for this report are summarized in table 3.1. We will also make use of some standard conventions in asymptotic analysis:

- We say that $f(x) \in \mathcal{O}(g(x))$ if there exists a constant c and a value x_0 such that $f(x) \leq c \cdot g(x)$ for all $x \geq x_0$.
- We say that $f(x) \in \Omega(g(x))$ if there exists a constant c and a value x_0 such that $f(x) \geq c \cdot g(x)$ for all $x \geq x_0$, and
- We say that $f(x) \sim g(x)$ if $\lim\limits_{x \to \infty} f(x)/g(x) = 1$.

and we also make the following definition:

Definition 3.2.1. Let $\phi(x) \equiv \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ be a step density function with compact support $\mathcal{R}$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area($\mathcal{R}$) = 1). Define $\Psi : \mathcal{R} \to \mathbb{R}^2$ be the union of all Ψ_i, where $\{\Psi_i : \boxdot_i \to \boxdot_i' | \Psi_i(y) = \sqrt{a_i} y + \xi_i\}$. ξ_i is chosen to make:

- Each $\boxdot_i'$ disjoint;
- Area$(\boxdot_i') = \frac{1}{s}$ for all i;
- Points in $\boxdot_i'$ are uniform distribution.

Table 3.1: Notational Conventions

$\mathcal{R}$	An area that customers showing up having Area($\mathcal{R}) = A$
n	Number of clients in $\mathcal{R}$
x_0	Depot
X_i	i'th customer where $X_i \overset{i.i.d}{\sim} \mathcal{P}$, $i = \{1, \ldots, n\}$
c	Each vehicle's capacity
d_{ij}	Euclidean distance between customer i and customer j
r_i	Euclidean distance between customer i and depot
$\bar{r}$	Average euclidean distance between customers and depot over all n customers
$\boxdot$	A rectangle in $\mathbb{R}^2$
$\mathbb{1}\{E\}$	Indicator function of event E
ϕ	A step density function defined on $\mathcal{R}$
f	An absolutely continuous probability distribution defined on $\mathcal{R}$
TSP($X_1, \ldots, X_n$)	Length of shortest TSP tour that visits all n points
VRP($X_1, \ldots, X_n; c$)	Length of shortest VRP tour that visits all n points with vehicle capacity c
$L(X_1, \ldots, X_n; pn)$	length of the shortest tour that visits pn out of all n points, where $p \in (0, 1)$
$CumL(X_1, \ldots, X_n)$	Length of the shortest cumulative TSP tour that visits all n points
$CumL(X_1, \ldots, X_n; c)$	Length of the shortest cumulative VRP tour that visits all n points with vehicle capacity c
$CumL(m; X_1, \ldots, X_n)$	Length of the shortest cumulative TSP tour that visits all n points using m different vehicles
$CumL(m; X_1, \ldots, X_n; c)$	Length of the shortest cumulative VRP tour that visits all n points with vehicle capacity c using m different vehicles.

4 Preliminaries

This section consists of preliminary results that we will make of in our subsequent analysis; the first component consists of both well-known results (such as the Borel-Cantelli lemma or Stirling's approximation), or routine technical exercises such as application of the union bound or Cavalieri's principle. The second component deals with results that are particularly relevant to the capacitated VRP.

4.1 General preliminary results

Theorem 1 is the famous *Beardwood-Halton-Hammersley (BHH) [12]* theorem.

Theorem 1 (BHH theorem)**.** *There is a constant β_d such that, for almost any sequence of independent variables $\{X_i\}$ sampled from an absolutely continuous density f on $\mathbb{R}^d$ with compact support, we have*

$$\lim_{n\to\infty} \frac{\mathit{TSP}(X_1,\ldots,X_n)}{n^{(d-1)/d}} = \beta_d \int_{\mathbb{R}^d} f(\mathbf{x})^{(d-1)/d} d\mathbf{x}$$

with probability one.

For d = 2, although the exact β_2 is unknown, numerical computations suggest $\beta_2 \approx 0.714$ [6].

Lemma 2 (Borel-Cantelli)**.** *Let $\{E_n\}$ be a sequence of events in a sample space. Then if $\sum_{n=1}^{\infty} \mathit{Pr}(E_n) < \infty$, we have*

$$\mathit{Pr}(E_n \text{ occurs infinitely often}) = 0$$

which is equivalent to

$$\mathit{Pr}(\limsup_{n\to\infty} E_n) = \mathit{Pr}(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} E_m) = 0$$

Lemma 3 (Approximation with a step function)**.** *Let f be a probability density function with compact support $\mathcal{R} \subset \mathbb{R}^2$ whose level sets have Lebesgue measure zero. Define*

$$P(x) := \mathit{Pr}(f(X) \le f(x)) = \iint_{x': f(x') \le f(x)} f(x')dx'$$

For any $\epsilon > 0$, there exists a step density function $\phi(x) := \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ and corresponding

$$\Pi(x) := \mathit{Pr}(\phi(X) \le \phi(x)) = \iint_{x': \phi(x') \le \phi(x)} \phi(x')dx'$$

such that the following conditions hold:

1. $\int_{\mathcal{R}} |\phi(x) - f(x)|dx \leq \epsilon$,
2. $|\mathbb{1}(P(X) \geq p) - \mathbb{1}(\Pi(X)) \geq p)| \leq \epsilon \quad \forall x \in \mathcal{R}$,
3. *All of the components of ϕ have the same mass, i.e. a_iArea$(\boxdot_i) = 1/s$.*

Proof. Even though the level sets having measure zero is not necessary, in order to keep notation consistent, we keep it (this requirement is violated when f is a uniform distribution, which is our base case for all of the instances in this paper). For a large integer q, define contour sets $\mathcal{S}_i = \{x : (i-1)/q \leq P(x) \leq i/q\}$. For each $\mathcal{S}_i$, we can approximate the restriction of f to $\mathcal{S}_i$ (i.e., $f(x)\mathbb{1}(x \in \mathcal{S}_i)$) to arbitrary precision ϵ' by a step function $\psi_i(x) = \sum_j a_{ij}\mathbb{1}\{x \in \boxdot_{ij}\}$ (This is the classical result of measure theory, see e.g., Theorem 2.4 (ii) of [73], i.e., step functions are dense in $L^1(\mathbb{R}^d)$). Based on the claim shown in Appendix, for all i and j, we can assume without loss of generality that

- $\boxdot_{ij} \subset \mathcal{S}_i$ for all i and j (i.e., the support of ψ_i is contained in $\mathcal{S}_i$),
- $a_{ij} < a_{(i+1)j'}$ for all i, j and j',
- all a_{ij} and Area$(\boxdot_{ij})$ are rational, and
- $\int_{\mathcal{S}_i} \psi_i = \int_{\mathcal{S}_i} f = 1/q$.

Given any $\epsilon > 0$, we set $q = \lceil 1/\epsilon' \rceil$ and $\epsilon' = \epsilon/q$ in the above construction. The function $\psi := \sum_{i=1}^{q} \psi_i$ is therefore a step density approximation of f whose aggregate error over $\mathcal{R}$ is at most ϵ, so condition 1 is satisfied. If we define $\Pi'(x) = \int_{x':\psi(x') \leq \psi(x)} \phi(x)dx$, then condition 2 is satisfied as well (using triangular inequality and add subtract Π').

For ease of notation, we now re-index all of the components of ψ (i.e., we disregard the fact that ψ decomposes into a sum of ψ_i's) so that we simply have $\psi(x) = \sum_j b_j \mathbb{1}(x \in \boxdot_j)$, where b_j and Area$(\boxdot_j)$ are rational. If we take δ to be the lowest common denominator over all b_jArea$(\boxdot_j)$, then we can write b_jArea$(\boxdot_j)$ = $z_j\delta$, with z_j a positive integer. To satisfy the last condition, all that remains is to decompose each $\boxdot_j$ into z_j pieces of equal area, and let ϕ denote the step function resulting thereof, which completes the proof. □

Lemma 4 (Super- and sub-additivity of the TSP). *Let $\mathcal{R} \subset \mathbb{R}^2$ be a compact Lebesgue measurable set, partitioned into pieces $\mathcal{P}_1, \ldots, \mathcal{P}_m$ whose common boundaries have finite length. There exists a constant C that depends only on the partition such that, for any set of points $\mathcal{X} = X_1, ..., X_n \subset \mathcal{R}$, we have:*

$$-C + \sum_{i=1}^{m} \text{TSP}(\mathcal{X} \cap \mathcal{P}_i) \leq \text{TSP}(\mathcal{X}) \leq C + \sum_{i=1}^{m} \text{TSP}(\mathcal{X} \cap \mathcal{P}_i)$$

Proof. The proof is shown in Lemma 2.3.1 of [71]. □

Lemma 5. *Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function and let $\mathfrak{B}_d(r) \subset \mathbb{R}^d$ be a ball of radius r centered about the origin. We have*

$$\int_{\mathfrak{B}_d(r)} f(\| \mathbf{x} \|)d\mathbf{x} = \int_0^r S_{d-1}(t)f(t)dt$$

Where $S_{d-1}(t)$ is the surface area of a $(d-1)$-sphere of radius t, which is given by

$$S_{d-1}(t) = \frac{2\pi^{d/2}}{\Gamma(d/2)} t^{d-1}$$

Lemma 6. *The volume of a d-dimensional ball of radius r is $\frac{\pi^{d/2} r^d}{\Gamma(d/2+1)}$*

Lemma 7 (Stirling's formula)**.** *The gamma function $\Gamma(x)$ satisfies $\log \Gamma(x+1) = x \log x - x + \frac{1}{2} \log x + \frac{1}{2} \log 2 + \frac{1}{2} \log \pi + \mathcal{O}(\frac{1}{x})$ as $x \to \infty$.*

Lemma 8. *Let $l > 0$ and let $\mathcal{D} \subset \mathbb{R}^{dn}$ denote the set of all n-tuple ($\mathbf{u}_1, \ldots, \mathbf{u}_n$) of points in $\mathbb{R}^d$ such that $\sum_{i=1}^n \| \mathbf{u}_i \| \leq l$. The volume of $\mathcal{D}$, Vol($\mathcal{D}$), satisfies*

$$\text{Vol}(\mathcal{D}) = \left(\frac{2\pi^{d/2}}{\Gamma(d/2)} \right)^n \frac{\Gamma(d)^n}{\Gamma(dn+1)} \cdot l^{dn} \tag{4.1}$$

For $d = 2$, this equation reduces to

$$\text{Vol}(\mathcal{D}) = \frac{(2\pi l^2)^n}{(2n)!}$$

Proof. This proof is straightforward:

$$\text{Vol}(\mathcal{D}) = \int_{\mathcal{B}_d(l)} \int_{\mathcal{B}_d(l-\|\mathbf{u}_n\|)} \cdots \int_{\mathcal{B}_d(l-\sum_{i=2}^n \|\mathbf{u}_i\|)} 1 d\mathbf{u}_1 d\mathbf{u}_2 \ldots d\mathbf{u}_n$$

and apply Theorem 5. □

Corollary 8.1. *Let $l > 0$ and $\mathcal{D}' \subset \mathbb{R}^{dn}$ be the set of all n-tuples $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of points in $\mathbb{R}^d$ such that $\| \mathbf{x}_1 \| + \sum_{i=2}^n \| \mathbf{x}_i - \mathbf{x}_{i-1} \| \leq l$. The volume of $\mathcal{D}'$, Vol($\mathcal{D}'$) satisfies*

$$\text{Vol}(\mathcal{D}') = \left(\frac{2\pi^{d/2}}{\Gamma(d/2)} \right)^n \frac{\Gamma(d)^n}{\Gamma(dn+1)} \cdot l^{dn}$$

Proof. Apply Cavalieri's principle to Theorem 8. □

Lemma 9. *Let X_0 be the origin in $\mathbb{R}^d$ and let $X_1, \ldots, X_n$ be a collection of independent, uniform samples drawn from a region $\mathcal{R}$ of unit volume in $\mathbb{R}^d$. Then*

$$\text{Pr(TSP(}X_0, X_1, \ldots, X_n) \leq l) \leq \Gamma(n+1) \cdot \left(\frac{2\pi^{d/2}}{\Gamma(d/2)} \right)^n \frac{\Gamma(d)^n}{\Gamma(dn+1)} \cdot l^{dn}$$

For $d = 2$, this inequality reduces to

$$\text{Pr(TSP(}X_0, X_1, \ldots, X_n) \leq l) \leq n! \frac{(2\pi l^2)^n}{(2n)!}$$

Proof. We can regard the samples $X_1, \ldots, X_n$ as being a single sample drawn uniformly from $\mathcal{R}^{dn}$. Using Corollary 8.1, it is easy to see that the probability $\| X_1 \| + \sum_{i=2}^{n} \| X_i - X_{i-1} \| \leq l$ should be equal to $\left(\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)^n \frac{\Gamma(d)^n}{\Gamma(dn+1)} \cdot l^{dn}$. Since Vol($\mathcal{D}' \cap \mathcal{R}^n$)$\leq$ Vol($\mathcal{D}'$), the formulation becomes:

$$\Pr(\| X_1 \| + \sum_{i=2}^{n} \| X_i - X_{i-1} \| \leq l) \leq \left(\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)^n \frac{\Gamma(d)^n}{\Gamma(dn+1)} \cdot l^{dn}$$

Finally, we note that there exist $n! = \Gamma(n+1)$ different permutations of $X_1, \ldots, X_n$, which we multiply by the right-hand side of the above. $\square$

Theorem 10 (Uniform sparse subset TSP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of area A in $\mathbb{R}^2$. Let $L(X_1, \ldots, X_n; pn)$ denote the length of the shortest tour that visits pn of the points $X_1, \ldots, X_n$, for fixed $0 < p < 1$. We have*

$$\liminf_{n\to\infty} \frac{L(X_1, \ldots, X_n; pn)}{p\sqrt{n}} \geq \lambda_1 \sqrt{A}$$

with probability one, where $\lambda_1 = 0.2935$.

Proof. Let E_n be the event that $L(X_1, \ldots, X_n; pn) \leq bp\sqrt{An}$ for fixed b. Using Theorem 9 we have:

$$\Pr(E_n) \leq \binom{n}{\lceil pn \rceil} \Gamma(\lceil pn \rceil + 1) \cdot \left(\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)^{\lceil pn \rceil} \frac{\Gamma(d)^{\lceil pn \rceil}}{\Gamma(d\lceil pn \rceil + 1)} \cdot (bp\sqrt{An})^{d\lceil pn \rceil}$$

Here, $\binom{n}{\lceil pn \rceil}$ represent the number of possible subsets choosing $\lceil pn \rceil$ points from n. Since we are only interested in the case $d = 2$ in this paper, the above equation reduces to

$$\begin{aligned}\Pr(E_n) &\leq \binom{n}{\lceil pn \rceil} \Gamma(\lceil pn \rceil + 1) \cdot \left(\frac{2\pi}{\Gamma(1)}\right)^{\lceil pn \rceil} \frac{\Gamma(2)^{\lceil pn \rceil}}{\Gamma(2\lceil pn \rceil + 1)} \cdot (bp\sqrt{An})^{2\lceil pn \rceil} \\ &= \binom{n}{\lceil pn \rceil} \frac{\Gamma(\lceil pn \rceil + 1)}{\Gamma(2\lceil pn \rceil + 1)} (2\pi b^2 p^2 An)^{\lceil pn \rceil}\end{aligned}$$

Since

$$\binom{n}{\lceil pn \rceil} = \frac{n!}{\lceil pn \rceil!(n - \lceil pn \rceil)!} = \frac{\Gamma(n+1)}{\Gamma(\lceil pn \rceil + 1)\Gamma(n - \lceil pn \rceil + 1)}$$

Combining it into the above equation, we can get:

$$\begin{aligned}\Pr(E_n) &\leq \frac{\Gamma(n+1)}{\Gamma(\lceil pn \rceil + 1)\Gamma(n - \lceil pn \rceil + 1)} \frac{\Gamma(\lceil pn \rceil + 1)}{\Gamma(2\lceil pn \rceil + 1)} (2\pi b^2 p^2 An)^{\lceil pn \rceil} \\ &= \frac{\Gamma(n+1)}{\Gamma(2\lceil pn \rceil + 1)\Gamma(n - \lceil pn \rceil + 1)} (2\pi b^2 p^2 An)^{\lceil pn \rceil}\end{aligned}$$

It is easy to verify that $\Gamma(x+1) \leq \Gamma(\lceil x \rceil + 1) \leq \Gamma((x+1)+1) = (x+1)\Gamma(x+1) \leq (x+1)\Gamma(\lceil x \rceil + 1)$. The above formulation becomes:

$$\Pr(E_n) \leq \frac{(n - pn)\Gamma(n+1)}{\Gamma(2pn + 1)\Gamma(n - pn + 1)} (2\pi b^2 p^2 An)^{\lceil pn \rceil}$$

Taking $\log$ of both sides and applying Theorem 7, we have:

$$\log \Pr(E_n) \leq (p\log(1-p) + 2p\log b - p\log 2 + p\log\pi + p\log A - \log(1-p) + p)n$$
$$-\frac{1}{2}\log p + \frac{1}{2}\log n + \frac{1}{2}\log(1-p) - \log 2 - \frac{1}{2}\log\pi + \mathcal{O}(\frac{1}{n})$$

Since n dominates in the above formulation, when $n \to \infty$, if we want $\log \Pr(E_n) \to -\infty$, we should require that its coefficient be negative, i.e.:

$$p\log(1-p) + 2p\log b - p\log 2 + p\log\pi + p\log A - \log(1-p) + p < 0$$
$$\log(1-p) + 2\log b - \log 2 + \log\pi + \log A - \frac{1}{p}\log(1-p) + 1 < 0$$
$$\log b^2 < \frac{1-p}{p}\log(1-p) + \log\frac{2}{A\pi e}$$
$$b < \sqrt{\left(\frac{2}{A\pi e}(1-p)^{(1-p)/p}\right.}$$

This is a convex function and increasing in $p \in (0,1)$, so we can get:

$$\lim_{p\to 0^+}\sqrt{\left(\frac{2}{A\pi e}(1-p)^{(1-p)/p}\right.} \geq \frac{\sqrt{2A\pi}}{e} > 0.2935\sqrt{A} = \lambda_1\sqrt{A}$$

This guarantees that $\sum_{n=1}^{\infty} \Pr(E_n) < \infty$ because $\Pr(E_n) \leq a^{-n}$ for some $a > 1$. Applying Theorem 2, we get:

$$\lambda_1 \leq \liminf_{n\to\infty}\frac{L(X_1,\ldots,X_n;pn)}{p\sqrt{An}}$$

with probability one, which completes the proof. □

Corollary 10.1. *(Tour length in a subset with uniform demand) Let $X_1\ldots,X_n$ be independent uniform samples drawn from a compact region $\mathcal{R}\subset\mathbb{R}^2$ with area A and let $\mathcal{S}\subset\mathcal{R}$ with Area($\mathcal{S}$) = q. Then*

$$\liminf_{n\to\infty}\frac{L(X_1,\ldots,X_n\cap\mathcal{S};pn)}{p\sqrt{n}} \geq \begin{cases}\lambda_1\sqrt{A} & \text{if } p\leq q\\ \infty & \text{otherwise}\end{cases}$$

where $\lambda_1 = 0.2935$.

Proof. If $p > q$, then the Law of Large Numbers shows that $|X_1,\ldots,X_n\cap\mathcal{S}|/n \to q$ as $n\to\infty$ with probability one, so $L(X_1,\ldots,X_n\cap\mathcal{S};pn)$ does not exist.

If $p\leq q$, it is obvious that

$$L(X_1,\ldots,X_n\cap\mathcal{S};pn) \geq L(X_1,\ldots,X_n;pn)$$

Now we apply Theorem 10 to conclude the proof. □

4.2 Preliminaries related to the CVRP

This section presents some preliminary results that are necessary for our analysis of the Cumulative Capacitated VRP (CCVRP). They are fairly standard, albeit lengthy, and are merely a probabilistic limiting interpretation of a seminal result due to [43]:

Theorem 11 ([43]). *Let $\{X_1, \ldots, X_n\}$ be a set of customers (points in a Euclidean Plane). The optimal length of the shortest tour that visits all these n points with the vehicle capacity c, written as VRP*$(X_1, \ldots, X_n; c)$, satisfies the following:*

$$\max\{\frac{2n}{c}\bar{r}, \mathit{TSP}(X_1, \ldots, X_n)\} \leq \mathit{VRP}^*(X_1, \ldots, X_n; c) \leq 2\left\lceil\frac{n}{c}\right\rceil\bar{r} + (1 - \frac{1}{c})\mathit{TSP}(X_1, \ldots, X_n)$$

The purpose of this section is to apply Theorem 11 to randomly distributed demand in order to discern its asymptotic behavior. The upper and lower bounds of interest are as follows:

Claim 1. Let $f(x)$ be a density function with compact support $\mathcal{R} \subset \mathbb{R}^2$. Consider a VRP with capacities c that vary relative to n via the relationship $c = k_c\sqrt{n}$. We have

$$\limsup_{n\to\infty} \frac{\mathsf{VRP}(X_1, \ldots, X_n; k_c\sqrt{n})}{\sqrt{n}} \leq \iint_{x\in\mathcal{R}} (\beta_2\sqrt{f(x)} + \frac{2}{k_c}f(x)\|x\|_2)dx$$

and

$$\liminf_{n\to\infty} \frac{\mathsf{VRP}(X_1, \ldots, X_n; k_c\sqrt{n})}{\sqrt{n}} \geq \frac{1}{2}\iint_{x\in\mathcal{R}} (\beta_2\sqrt{f(x)} + \frac{2}{k_c}f(x)\|x\|_2)dx$$

Proof. See Section 4 of [24]. Note that the upper and lower bounds are within a factor of 2 of one another. □

4.2.1 Sparse Subset Vehicle Routing Problem (SSVRP)

For this section, we let VRP$(X_1, \ldots, X_n; c; pn)$ denote the length of the shortest tour that visits pn of points from $X_1, \ldots, X_n$ with capacity limit c. To save wear and tear on floors and ceilings, when pn is non-integer, we round it up.

Theorem 12 (Tour length from a step density). *Let $\phi(x) \equiv \sum_{i=1}^{s} a_i\mathbb{1}(x \in \boxdot_i)$ be a step density function with compact support $\mathcal{R}$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area($\mathcal{R}$) = 1). Consider capacity $c = k_c\sqrt{n}$. For all fixed $0 < p < 1$, we have:*

$$\liminf_{n\to\infty} \frac{\mathit{VRP}(X_1, \ldots, X_n; k_c\sqrt{n}; pn)}{\sqrt{n}} \geq \frac{1}{2}\int_{x\in\mathcal{R}} (\lambda_1\sqrt{\phi(x)} + \frac{2}{k_c}\phi(x)\|x\|_2)\mathbb{1}(N(x) \geq p)dx$$

with probability one, where

$$N(x) = \mathit{Pr}(\nu(X) \leq \nu(x)) = \iint_{x':\nu(x')\leq\nu(x)} \phi(x')dx'$$

and

$$\nu(x) \equiv \iint_{x \in \mathcal{R}} \left(\lambda_1 \sqrt{\phi(x)} + \frac{2}{k_c} \phi(x) \|x\|_2 dx \right.$$

Proof. Let $m = pn$, so that we have two cases:

1. $m \leq k_c\sqrt{n}$: If this is the case, this problem immediately becomes solvable as a Sparse Subset TSP problem because of the triangle inequality. Since we are taking about $n \to \infty$, we can pay attention to the second case.

2. $m > k_c\sqrt{n}$: Under this case, we have:

$$\begin{aligned}
\mathsf{VRP}(X_1, \ldots, X_n; k_c\sqrt{n}; m) &= \min_{|S|=m} \mathsf{VRP}(S; k_c\sqrt{n}) \\
&= \min_{|S|=m} \sum_{i=1}^{s} \left(\mathsf{VRP}(S \cap \boxdot_i; k_c\sqrt{n}) \right. \\
&= \min_{\mathbf{q} \in \mathbf{Q}} \sum_{i=1}^{s} \mathsf{VRP}(S \cap \boxdot_i; k_c\sqrt{n}; q_i)
\end{aligned}$$

where $\mathbf{q}$ is a vector denoting the number of points from each $\boxdot_i$ that are selected:

$$\mathbf{Q} = \{\mathbf{q} \in \mathbb{Z}_+^m : \sum_{i=1}^{s} \left(q_i = m, \quad q_i \leq |S \cap \boxdot_i|, \quad \forall i \right\}$$

Combined with Theorem 11, we can get:

$$\begin{aligned}
\liminf_{n\to\infty} \mathsf{VRP}(X_1, \ldots, X_n; k_c\sqrt{n}; m) &= \liminf_{n\to\infty} \min_{\mathbf{q} \in \mathbf{Q}} \sum_{i=1}^{s} \left(\mathsf{VRP}(S \cap \boxdot_i; k_c\sqrt{n}; q_i) \right. \\
&\geq \liminf_{n\to\infty} \min_{\mathbf{q} \in \mathbf{Q}} \sum_{i=1}^{s} \left(\max\{\mathsf{TSP}(\mathcal{S} \cap \boxdot_i, q_i), \tfrac{2}{k_c\sqrt{n}} \sum_{i: i \in \mathcal{S} \cap \boxdot_i} \left(r_i \right\} \right. \\
&= \liminf_{n\to\infty} \min_{\mathbf{q} \in \mathbf{Q}} \sum_{i=1}^{s} \left(\max\{\mathsf{TSP}(\mathcal{S} \cap \boxdot_i, q_i), \tfrac{2q_i}{k_c\sqrt{n}} \overline{r}_{\mathcal{S} \cap \boxdot_i} \right\} \\
&\geq \liminf_{n\to\infty} \min_{\mathbf{q} \in \mathbf{Q}} \sum_{i=1}^{s} \left(\max\{\mathsf{TSP}(\mathcal{S} \cap \boxdot_i, q_i), \tfrac{2q_i}{k_c\sqrt{n}} \overline{r}_{\boxdot_i} \right\} \\
&\geq \tfrac{1}{2} \liminf_{n\to\infty} \min_{\mathbf{q} \in \mathbf{Q}} \left(\sum_{i=1}^{s} \mathsf{TSP}(\mathcal{S} \cap \boxdot_i, q_i) + \sum_{i=1}^{s} \tfrac{2q_i}{k_c\sqrt{n}} \overline{r}_{\boxdot_i} \right)
\end{aligned}$$

where $\overline{r}_{\boxdot_i}$ is the average distance to depot for $\boxdot_i$ and $\overline{r}_{S \cap \boxdot_i}$ is the average distance to depot for those picked point within $\boxdot_i$. Using Definition 3.2.1, we can reconstruct:

$$\mathsf{TSP}(S \cap \boxdot_i) = \frac{1}{\sqrt{a_i}} \mathsf{TSP}(S' \cap \boxdot_i')$$

for all subset S and points in each $\boxdot_i$, $\forall i$ follows uniform distribution. It is easy to get:

$$\begin{aligned}
\min_{\mathbf{q}\in\mathbf{Q}} \sum_{i=1}^{s} \mathsf{TSP}(S\cap\boxdot_i; q_i) &= \min_{\mathbf{q}\in\mathbf{Q}} \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{TSP}(S'\cap\boxdot'_i; q_i) \\
&\geq \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{TSP}(S'\cap\boxdot'_i; \widetilde{q}_i) \\
\min_{\mathbf{q}\in\mathbf{Q}} \sum_{i=1}^{s} \frac{2q_i}{k_c\sqrt{n}}\overline{r}_{\boxdot_i} &\geq \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{2\widetilde{q}_i}{k_c\sqrt{n}}\overline{r}_{\boxdot_i}
\end{aligned}$$

Where $\widetilde{\mathbf{Q}}$ is a “lower bounding set ” of $\mathbf{Q}$ defined as follows: fix ϵ and let $\xi(t) = \epsilon\lfloor t/\epsilon\rfloor$, which in particular tells us that $0 \leq t-\xi(t) \leq \epsilon$ for all t. The set $\widetilde{\mathbf{Q}}$ is the image of $(\lfloor n\xi(q_1/n)\rfloor, \ldots, \lfloor n\xi(q_s/n)\rfloor)$ for all feasible vectors $\mathbf{q} \in \mathbf{Q}$. The detail is shown in [22].

Combined with the result in corollary 10.1, we can get:

$$\begin{aligned}
\liminf_{n\to\infty} \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{TSP}(S'\cap\boxdot'_i; \widetilde{q}_i) &\geq \liminf_{n\to\infty} \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{TSP}(S'\cap\boxdot'_i; \widetilde{q}_i) \\
&\geq \min_{\mathbf{t}\in p\Delta^{s-1}} \liminf_{n\to\infty} \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{TSP}(S'\cap\boxdot'_i; (t_i-\epsilon)n) \\
\limsup_{n\to\infty} \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{2\widetilde{q}_i}{k_c\sqrt{n}}\overline{r}_{\boxdot_i} &\geq \liminf_{n\to\infty} \min_{\widetilde{\mathbf{q}}\in\widetilde{\mathbf{Q}}} \sum_{i=1}^{s} \frac{2\widetilde{q}_i}{k_c\sqrt{n}}\overline{r}_{\boxdot_i} \\
&\geq \min_{\mathbf{t}\in p\Delta^{s-1}} \liminf_{n\to\infty} \sum_{i=1}^{s} \frac{2(t_i-\epsilon)\sqrt{n}}{k_c}\overline{r}_{\boxdot_i}
\end{aligned}$$

It is easy to show that $\widetilde{q}_i \geq (t_i-\epsilon)n$ by $\widetilde{q}_i = \lfloor n\epsilon\lfloor \frac{q_i}{n\epsilon}\rfloor\rfloor$ and $t = t_i = q_i/n$ in $0 \leq t-\xi(t) \leq \epsilon$. Now by corollary 10.1, we have

$$\liminf_{n\to\infty} \frac{\mathsf{TSP}(S'\cap\boxdot'_i; (t_i-\epsilon)n)}{\sqrt{n}} \geq \begin{cases} \lambda_1(t_i-\epsilon) & \text{if } t_i-\epsilon \leq 1/s \\ \infty & \text{otherwise} \end{cases}$$

and so ultimately, as this is a minimization problem, we can bound it below in terms of the fraction of points in each cell:

$$\begin{aligned}
\min_{\mathbf{t}} \quad & \tfrac{1}{2}[\lambda_1\sqrt{n}\sum_{i=1}^{s}\tfrac{1}{\sqrt{a_i}}(t_i-\epsilon) + \tfrac{2\sqrt{n}}{k_c}\sum_{i=1}^{n}\overline{r}_{\boxdot_i}(t_i-\epsilon)] \\
\text{s.t} \quad & \sum_{i=1}^{n} t_i = p \\
& 0 \leq t_i \leq 1/s \qquad \forall i
\end{aligned} \tag{4.2}$$

For clarity, we can rewrite the objective function as

$$\frac{1}{2}\sum_{i=1}^{s}\left(\frac{\lambda_1\sqrt{n}}{\sqrt{a_i}} + \frac{2\sqrt{n}\overline{r}_{\boxdot_i}}{k_c}\right)(t_i-\epsilon)$$

By monotonicity, it is easy to see that the optimal solution is achieved by setting $t_1 = \cdots = t_{\lfloor ps\rfloor} = 1/s$ and $t_{\lceil ps\rceil} = p-\lfloor ps\rfloor/s$ (see e.g. exercises 4.8(e) of Boyd [19]). We can disregard the $t_{\lfloor ps\rfloor}$ term for notational convenience and use the fact that $x \in \cup_{i=1}^{\lfloor ps\rfloor}\boxdot_i$ if and only if $N(x) \geq \lfloor ps\rfloor/s$. Therefore, the objective function of eq. (4.2) is at least

$$
\begin{aligned}
\frac{1}{2}\sum_{i=1}^{\lfloor ps\rfloor}\left(\frac{\lambda_1\sqrt{n}}{\sqrt{a_i}}+\frac{2\sqrt{n}\bar{r}_{\square_i}}{k_c}\right)(1/s-\epsilon) &= \frac{1}{2}\Big[\frac{\sqrt{n}}{s}\sum_{i=1}^{\lfloor ps\rfloor}\left(\frac{\lambda_1}{\sqrt{a_i}}+\frac{2\bar{r}_{\square_i}}{k_c}\right)-\epsilon\sum_{i=1}^{\lfloor ps\rfloor}\left(\frac{\lambda_1\sqrt{n}}{\sqrt{a_i}}+\frac{2\sqrt{n}\bar{r}_{\square_i}}{k_c}\right)\Big] \\
&\geq \frac{1}{2}\Big[\frac{\sqrt{n}}{s}\sum_{i=1}^{\lfloor ps\rfloor}\left(\frac{\lambda_1}{\sqrt{a_i}}+\frac{2\bar{r}_{\square_i}}{k_c}\right)-\epsilon\sum_{i=1}^{s}\left(\frac{\lambda_1\sqrt{n}}{\sqrt{a_i}}+\frac{2\sqrt{n}\bar{r}_{\square_i}}{k_c}\right)\Big] \\
&= \frac{\sqrt{n}}{2}\Big(\lambda_1\int_{\mathcal{R}}\sqrt{\phi(x)}\mathbb{1}(N(x)\geq p)dx+\frac{2}{k_c}\int_{\mathcal{R}}\phi(x)\|x\|_2\mathbb{1}(N(x)\geq p)dx\Big) \\
&\quad -\frac{\epsilon}{2}\sqrt{n}\Big(\lambda_1 s\int_{\mathcal{R}}\sqrt{\phi(x)}dx+\frac{2}{k_c}\int_{\mathcal{R}}\phi(x)\|x\|_2dx\Big) \\
&= \frac{\sqrt{n}}{2}\int_{\mathcal{R}}\Big(\lambda_1\sqrt{\phi(x)}+\frac{2}{k_c}\phi(x)\|x\|_2\Big)\mathbb{1}(N(x)\geq p)dx \\
&\quad -\frac{\epsilon\sqrt{n}}{2}\int_{\mathcal{R}}\Big(\lambda_1 s\sqrt{\phi x}+\frac{2}{k_c}\phi(x)\|x\|_2\Big)dx
\end{aligned}
$$

which completes the proof.

$\square$

Theorem 13 (Tour length from a general distribution)**.** *Let f, $\mathcal{R}$ be as in the notational conventions. The capacity satisfies $c = k_c\sqrt{n}$ With probability one, we have*

$$
\liminf_{n\to\infty}\frac{\textit{VRP}(X_1,\ldots,X_n;k_c\sqrt{n};pn)}{\sqrt{n}} \geq \frac{1}{2}\int_{x\in\mathcal{R}}\Big(\lambda_1\sqrt{f(x)}+\frac{2}{k_c}f(x)\|x\|_2\Big)\mathbb{1}(\Upsilon(x)\geq p)dx
$$

where

$$
\Upsilon(x) = \textit{Pr}(\upsilon(X)\leq\upsilon(x)) = \iint_{x':\upsilon(x')\leq\upsilon(x)}\upsilon(x')dx'
$$

and

$$
\upsilon(x) \equiv \iint_{x\in\mathcal{R}}\lambda_1\sqrt{f(x)}+\frac{2}{k_c}f(x)\|x\|_2dx
$$

Proof. Let ϕ be the approximation of f from theorem 12. By a standard coupling argument (for example the γ coupling of [51]), there is a joint distribution for random variables (X,Y) such that X has density f, Y has density ϕ and then $\Pr(X\neq Y)\leq\epsilon$ for any ϵ. We have

$$
\begin{aligned}
VRP(X_1,\ldots,X_n;k_c\sqrt{n};pn) &> L(X_1,\ldots,X_n : X_i = Y_i;k_c\sqrt{n};pn) \\
&= VRP(Y_1,\ldots,Y_n : X_i = Y_i;k_c\sqrt{n};pn) \\
&\geq VRP(Y_1,\ldots,Y_n;k_c\sqrt{n};pn) - VRP(Y_1,\ldots,Y_n : X_i\neq Y_i;k_c\sqrt{n};pn) - \mathcal{O}(1)
\end{aligned}
$$

$$
VRP(Y_1,\ldots,Y_n : X_i\neq Y_i;k_c\sqrt{n};pn) \leq \frac{2\lceil p\sqrt{n}\rceil}{k_c}\bar{r}+\Big(1-\frac{1}{k_c\sqrt{n}}\Big)TSP(Y_1,\ldots,Y_n : X_i\neq Y_i;k_c\sqrt{n};pn)
$$

Since the convergence result in [75] is almost surely for $TSP(Y_1, \ldots, Y_n : X_i \neq Y_i; k_c\sqrt{n}; pn)$, we have

$$\limsup_{n\to\infty} \frac{TSP(Y_1, \ldots, Y_n : X_i \neq Y_i; k_c\sqrt{n}; pn)}{\sqrt{n}} \leq \frac{\alpha_2\sqrt{\text{Area}(\mathcal{R})\epsilon pn}}{\sqrt{n}}$$

Combining them, we get:

$$\begin{aligned}
\liminf_{n\to\infty} \tfrac{VRP(X_1,\ldots,X_n;k_c\sqrt{n};pn)}{\sqrt{n}} &\geq \liminf_{n\to\infty} \tfrac{VRP(Y_1,\ldots,Y_n;k_c\sqrt{n};pn)}{\sqrt{n}} - \tfrac{VRP(Y_1,\ldots,Y_n:X_i\neq Y_i;k_c\sqrt{n};pn)}{\sqrt{n}} \\
&\geq \tfrac{1}{2}\int_{x\in\mathcal{R}} (\lambda_1\sqrt{\phi(x)} + \tfrac{2}{k_c}\phi(x)\|x\|_2)\mathbb{1}(N(x)\geq p)dx \\
&\quad - \left(\tfrac{2\epsilon p}{k_c}\bar{r} + (1-\tfrac{1}{k_c\sqrt{n}})\alpha_2\sqrt{\text{Area}(\mathcal{R})\epsilon pn}\right) \\
&= \quad \tfrac{\lambda_1}{2}\int_{x\in\mathcal{R}} \sqrt{\phi(x)}\mathbb{1}(N(x)\geq p)dx - (1-\tfrac{1}{k_c\sqrt{n}})\alpha_2\sqrt{\text{Area}(\mathcal{R})\epsilon p}\Big) \\
&\quad + \tfrac{1}{k_c}\int_{x\in\mathcal{R}} \phi(x)\|x\|_2\mathbb{1}(N(x)\geq p)dx - 2\epsilon p\bar{r}\Big)
\end{aligned} \tag{4.3}$$

We can select our approximation Υ, N arbitrarily closely so that

$$\begin{aligned}
\epsilon &\geq \max_x |\mathbb{1}(\Upsilon\geq p) - \mathbb{1}(N\geq p)| + \int_{\mathcal{R}} |\phi - f|dx \\
&\geq \int \phi|\mathbb{1}(\Upsilon\geq p) - \mathbb{1}(N\geq p)| + \int |\phi - f|\mathbb{1}(\Upsilon\geq p) \\
&\geq \int |\phi\mathbb{1}(N\geq p) - \phi\mathbb{1}(\Upsilon\geq p) + \phi\mathbb{1}(\Upsilon\geq p) - f\mathbb{1}(\Upsilon > p)| \\
&= \int |\phi\mathbb{1}(N\geq p) - f\mathbb{1}(\Upsilon\geq p)|
\end{aligned}$$

and furthermore, we have

$$\begin{aligned}
\sqrt{\text{Area}(\mathcal{R})p\epsilon} &\geq \int \sqrt{|\phi\mathbb{1}(N\geq p) - f\mathbb{1}(\Upsilon\geq p)|} \geq |\int \sqrt{\phi}\mathbb{1}(N\geq p) - \int \sqrt{f}\mathbb{1}(P\geq p)| \\
\Rightarrow \quad \int \sqrt{\phi}\mathbb{1}(N\geq p) &\geq \int \sqrt{f}\mathbb{1}(\Upsilon\geq p) - \sqrt{p\epsilon\text{Area}(\mathcal{R})} \\
\epsilon p\bar{r} &\geq \int |\phi\mathbb{1}(N\geq p) - f\mathbb{1}(\Upsilon\geq p)|\|x\|_2 \\
\Rightarrow \quad \int \phi\mathbb{1}(N\geq p)\|x\|_2 &\geq \int f\mathbb{1}(\Upsilon\geq p)\|x\|_2 - \epsilon p\bar{r}
\end{aligned}$$

Therefore, we ultimately find that

$$\begin{aligned}
&\tfrac{\lambda_1}{2}\int_{x\in\mathcal{R}} \sqrt{\phi(x)}\mathbb{1}(N(x)\geq p)dx - (1-\tfrac{1}{k_c\sqrt{n}})\alpha_2\sqrt{\text{Area}(\mathcal{R})\epsilon p} \\
\geq\ &\tfrac{\lambda_1}{2}\int_{x\in\mathcal{R}} \sqrt{f(x)}\mathbb{1}(\Upsilon(x)\geq p)dx - \sqrt{p\epsilon\text{Area}(\mathcal{R})}\Big) - (1-\tfrac{1}{k_c\sqrt{n}})\alpha_2\sqrt{\text{Area}(\mathcal{R})\epsilon p} \\
=\ &\tfrac{\lambda_1}{2}\int_{x\in\mathcal{R}} \sqrt{f(x)}\mathbb{1}(\Upsilon(x)\geq p)dx - \underbrace{(\frac{\lambda_1}{2} + (1-\frac{1}{k_c\sqrt{n}}))\sqrt{\text{Area}(\mathcal{R})\epsilon p}}_{(*)} \\
&\int_{x\in\mathcal{R}} \phi(x)\|x\|_2\mathbb{1}(N(x)\geq p)dx - 2\epsilon p\bar{r} \\
\geq\ &\int_{x\in\mathcal{R}} f(x)\mathbb{1}(\Upsilon(x)\geq p)\|x\|_2 dx - \epsilon p\bar{r} - 2\epsilon p\bar{r} \\
=\ &\int_{x\in\mathcal{R}} f(x)\mathbb{1}(\Upsilon(x)\geq p)\|x\|_2 dx - \underbrace{3\epsilon p\bar{r}}_{(**)}
\end{aligned}$$

Where (*) and (**) shrink to 0 by choosing sufficiently small values of ϵ.

Putting them into eq. (4.3), we have

$$\liminf_{n\to\infty} \frac{VRP(X_1,\ldots,X_n;k_c\sqrt{n};pn)}{\sqrt{n}} \quad = \quad \frac{1}{2}\int_{x\in\mathcal{R}} (\lambda_1\sqrt{f(x)} + \frac{2}{k_c}f(x)\|x\|_2)\mathbb{1}(\Upsilon(x) \geq p)dx$$

which completes the proof.

$\square$

5 Analysis of the Cumulative TSP when $n \to \infty$

This chapter presents a comprehensive probabilistic analysis on the upper and lower bounds for the Cumulative TSP where the number of points tends to infinity. The analysis begins by considering uniform distributions, followed by non-uniform distributions. The objective of this analysis is to provide a thorough understanding of the behavior of the bounds and their relationship to the distribution of the points, which is a crucial step towards the development of efficient algorithms for large-scale problems. The results of this analysis have significant implications for a range of applications, from logistics and transportation to network optimization and facility location.

5.1 The "most dense to Least dense" Rule Explained

In analyzing the non-uniform case of CTSP, we will repeatedly refer to a routing strategy that we call the "most dense to least dense". It proceeds as follows:

From Theorem 3, we know that step density function ϕ can be used as the approximation of f, due to a standard coupling argument. Define $\Psi : \mathcal{R} \to \mathbb{R}^2$ be the union of all Ψ_i, where $\{\Psi_i : \boxdot_i \to \boxdot'_i | \Psi_i(y) = \sqrt{a_i} y + \xi_i\}$. ξ_i is chosen to make:

- Each $\boxdot'_i$ disjoint;
- Area($\boxdot'_i$) $= \frac{1}{s}$ for all i;
- Points in $\boxdot'_i$ are uniform distribution;

The rule is shown in Figure 5.1. The idea is that the tour should start in the darkest (i.e., densest and smallest) areas to lighter (larger and sparser) areas.

Compared to the traditional TSP problem, this assumption is a strong one, which is decided by the different objective functions.

$$\text{TSP:} \quad \min \quad \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} X_{ij}$$

$$\text{CTSP:} \quad \min \quad \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} Y_{ij}$$

Where

$$X_{ij} = \begin{cases} 1, & \text{if the repairman travels arc (i, j)} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} n - k + 1, & \text{if arc (i, j) appears in position k on the Hamiltonian tour} \\ 0, & \text{if arc (i, j) is not used} \end{cases}$$

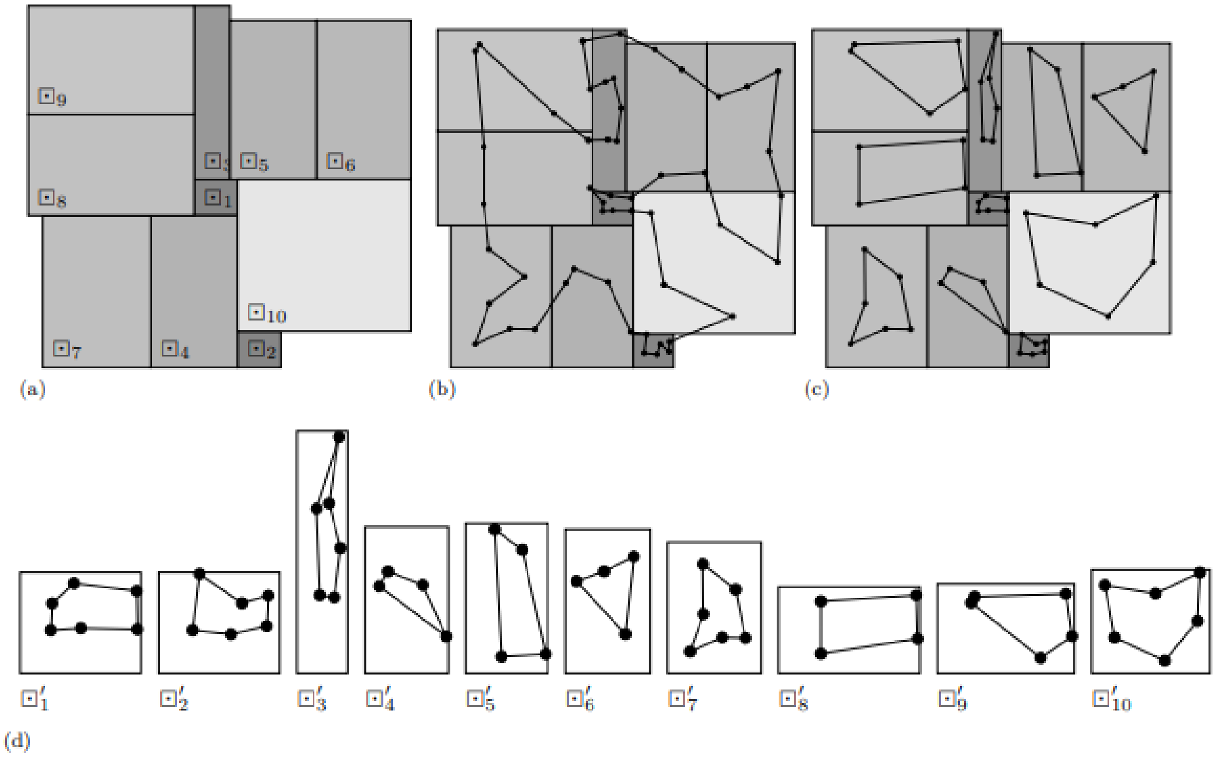

Figure 5.1: Figure (a) shows a step function ϕ that presumably satisfies the conditions (darker - i.e., denser- regions are smaller, reflecting the assumption that a_iArea($\boxdot_i$) = 1/s) for all i) (b) shows the TSP tour of a collection of independent samples $Y_1, \ldots, Y_n$ of ϕ whose length differs from that of a collection of tour within each component (c) by a constant. (d) shows the rescaled component $\boxdot_i' = \Psi(\boxdot_i)$ and the points $Y_1', \ldots, Y_n'$

While the TSP is order-invariant and allows for multiple optimal tours, the CTSP introduces the additional constraint of time, with the cost of an arc depending on its position in the tour. As such, minimizing the cost of the tour requires considering both arc costs and position, prioritizing the placement of arcs with smaller costs earlier in the tour. In the context of continuous approximation formulas, the tour should be constructed by traveling from the most dense to the least dense points.

5.2 Upper Bounds for CTSP

The upper bounding of the CTSP involves the key concept that the waiting time of each customer, represented by the distance travelled by the vehicle until it reaches the customer, should not exceed the waiting time of the previously visited customer. This approach ensures an effective upper bounding strategy of the CTSP, with each customer being visited within an appropriate time frame.

5.2.1 $X_i \overset{i.i.d}{\sim} \mathcal{U}$

Theorem 14 (Upper bound; Uniform CTSP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of $\mathcal{R}$ with area A in $\mathbb{R}^2$. We have*

$$\limsup_{n\to\infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} \leq \frac{\beta_2\sqrt{A}}{2} \tag{5.1}$$

with probability one where $\beta_2 \approx 0.714$.

Proof. The main objective of this proof is to establish a feasible solution for upper bounding the routing problem, where the optimal value of the problem should not exceed this feasible solution. To achieve this, we consider the TSP tour as a feasible tour, where the earlier position of a customer carries more weight in the CTSP objective. Since the optimal starting point is unknown, we take each node as a starting node and compute the average cost as the upper bound to avoid worst-case scenarios. The aggregate feasible solution costs are presented in the table 5.1, demonstrating the effectiveness of this approach.

Starting Position	Cumulative Tour Value
Starting at X_1	$nr_1 + (n-1)d_{12} + (n-2)d_{23} + \cdots + d_{(n-1)n} + 0d_{n0}$
Starting at X_2	$nr_2 + (n-1)d_{23} + (n-2)d_{34} + \cdots + d_{n1} + 0d_{12}$
$\ldots$	$\ldots$
Starting at X_n	$nr_n + (n-1)d_{n1} + (n-2)d_{12} + \cdots + d_{(n-2)(n-1)} + 0d_{(n-1)n}$

Table 5.1: Feasible Starting positions

Adding them together, we have:

$$\begin{aligned} &n\sum_{n} r_i + [(n-1)+(n-2)+\cdots+0]\text{TSP}(X_1, \ldots, X_n) \\ =&n\sum_{i=1}^{n} r_i + \frac{n(n-1)}{2}\text{TSP}(X_1, \ldots, X_n) \end{aligned}$$

As the minimum is smaller than the average, combined with Theorem 1, the upper bound will be:

$$\begin{aligned} \limsup_{n\to\infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} &\leq \limsup_{n\to\infty} \frac{\sum_{i:x_i\in\mathcal{R}} r_i}{n^{3/2}} + \frac{n-1}{2n^{3/2}}\text{TSP}(X_1, \ldots, X_n) \\ &= \frac{\beta_2\sqrt{A}}{2} + \lim_{n\to\infty} \frac{\overline{r}}{\sqrt{n}} - \frac{\beta_2\sqrt{A}}{2n} \\ &= \frac{\beta_2\sqrt{A}}{2} \end{aligned}$$

For the second line, as $n \to \infty$, $\frac{1}{\sqrt{n}} \to 0$ and $\frac{1}{n} \to 0$, which completes the proof. $\square$

5.2.2 $X_i \overset{i.i.d}{\sim} \mathcal{P}$

Theorem 15 (Upper bound; Step function CTSP). *Let $X_1, \dots, X_n$ be independent samples, where X_i follows the step density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R} \subset \mathbb{R}^2$ such that $a_1 \geq \dots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). Suppose that $\Pi(x)$ is as defined in Lemma 3. We have*

$$\limsup_{n \to \infty} \frac{CumL(X_1, \dots, X_n)}{n^{3/2}} \leq \beta_2 \iint_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx \tag{5.2}$$

with probability one where $\beta_2 \approx 0.714$.

Proof. As in the uniform case, we begin by finding a feasible solution. For each $\boxdot_i$, we can reconstruct $\boxdot_i$ to $\boxdot_i'$ using a function $\psi_i(x) = \sqrt{a_i}x + \xi_i$ to make:

- Each $\boxdot_i'$ disjoint;
- Area$(\boxdot_i') = \frac{1}{s}$ for all i;
- Points in $\boxdot_i'$ are uniform distribution;

Basic scaling arguments tell us that TSP$(\boxdot_i) = \frac{1}{\sqrt{a_i}}TSP(\boxdot_i')$
Since now each point in this area is uniformly distributed, we can apply the BHH theorem:

$$\lim_{n \to \infty} \frac{\text{TSP}(\boxdot_i')}{\sqrt{n}} = \beta_2 \sqrt{\frac{1}{s}} \sqrt{\frac{1}{s}} = \frac{\beta_2}{s} \Rightarrow \lim_{n \to \infty} \text{TSP}(\boxdot_i) = \frac{1}{\sqrt{a_i}} \text{TSP}(\boxdot_i') = \frac{\beta_2}{s\sqrt{a_i}}$$

Since each $\boxdot_i$ is related to a_i, one of the possible ways to travel to all points is to follow "most dense to least dense" rule, which merely means that we travel from $\boxdot_1, \dots, \boxdot_s$. (From the definition of our step function, the densest part of the distribution is equivalent to the smallest area $\boxdot$, which is $\boxdot_1$). Using Theorem 4, for point $X_k \in \boxdot_j$, its waiting time should be less than or equal to :

$$\text{TSP}(\boxdot_1) + \text{TSP}(\boxdot_2) + \dots + \text{TSP}(\boxdot_j) = \sum_{i=1}^{j} \frac{\beta_2}{s} \sqrt{\frac{n}{a_i}} = \beta_2 \sqrt{n} \iint_{x' : \phi(x') \geq \phi(x)} \sqrt{\phi(x')} dx'$$

as $n \to \infty$.
Using the Law of Large Numbers,

$$|X_1, \dots, X_n \cap \boxdot_i|/n \to \frac{1}{s} \Rightarrow |X_1, \dots, X_n \cap \boxdot_i| \to \frac{n}{s}$$

We find that, as $n \to \infty$, each set will have $\frac{n}{s}$ points. Therefore, combining both terms together, we can get the cumulative waiting time in each $\boxdot_i$ for all i:

- cumulative waiting time in $\boxdot_1 \leq \frac{n}{s}L(\boxdot_1)$
- cumulative waiting time in $\boxdot_2 \leq \frac{n}{s}$[L$(\boxdot_1)$+L$(\boxdot_2)$].
- ...
- cumulative waiting time in $\boxdot_s \leq \frac{n}{s}$[L$(\boxdot_1) + \dots +L(\boxdot_s)$].

Summing all entries, we find:

$$
\begin{aligned}
CumL(X_1,\ldots,X_n) &\leq \sum_{j=1}^{s} \frac{n}{s} \sum_{i=1}^{j} \mathsf{L}(\square_i) = n \sum_{j=1}^{s} \frac{a_j}{a_j s} \sum_{i=1}^{j} \mathsf{L}(\square_i) \\
&= n \iint_{x\in\mathcal{R}} \phi(x) \sum_{i=1}^{j} \mathsf{L}(\square_i) dx
\end{aligned}
$$

and therefore, from our previous formulation, we have:

$$
\begin{aligned}
\limsup_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2}} &\leq \beta_2 \int_{x\in\mathcal{R}} \phi(x) \iint_{x':\phi(x')\geq\phi(x)} \sqrt{\phi(x')}dx'dx \\
&= \beta_2 \int_{x\in\mathcal{R}} \iint_{x':\phi(x')\geq\phi(x)} \phi(x)\sqrt{\phi(x')}dx'dx \\
&= \beta_2 \int_{x'\in\mathcal{R}} \iint_{x:\phi(x)\leq\phi(x')} \phi(x)\sqrt{\phi(x')}dxdx' \\
&= \beta_2 \int_{x'\in\mathcal{R}} \sqrt{\phi(x')} \left(\int_{x:\phi(x)\leq\phi(x')} f(x)dx \right) dx' \\
&= \beta_2 \iint_{x'\in\mathcal{R}} \sqrt{\phi(x')}\Pi(x')dx' \\
&= \beta_2 \iint_{x\in\mathcal{R}} \sqrt{\phi(x)}\Pi(x)dx
\end{aligned}
$$

which completes the proof. □

Theorem 16 (Upper bound; non-uniform CTSP). *Let $X_1,\ldots,X_n$ be independent samples from a region of $\mathcal{R}\subset\mathbb{R}^2$, where X_i follows the density function f, $\forall i$. $P(x)$ is as defined in Lemma 3. We have*

$$
\lim_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2}} \leq \beta_2 \iint_{x\in\mathcal{R}} \sqrt{f(x)}P(x)dx \tag{5.3}
$$

with probability one where $\beta_2 \approx 0.714$.

Proof. The key point is the same as what we did in step density function: Theorem 15. Based on Theorem 3, f can be approximated using ϕ. Based on the "Most dense to Least dense " Rule, for point x, the amount of time he/she has to wait is:

$$
\beta_2\sqrt{n} \int_{x':f(x')\geq f(x)} \sqrt{f(x')}dx'
$$

When traveling all n points, the length is

$$
\begin{aligned}
&\beta_2\sqrt{n} \int_{x\in\mathcal{R}} f(x) \iint_{x':f(x')\geq f(x)} \sqrt{f(x')}dx'dx \\
=&\beta_2\sqrt{n} \int_{x\in\mathcal{R}} \iint_{x':f(x')\geq f(x)} f(x)\sqrt{f(x')}dx'dx
\end{aligned}
$$

$$
\begin{aligned}
&= \beta_2 \sqrt{n} \int_{x' \in \mathcal{R}} \int_{x : f(x) \leq f(x')} f(x) \sqrt{f(x')}\, dx\, dx' \\
&= \beta_2 \sqrt{n} \int_{x' \in \mathcal{R}} \sqrt{f(x')} \left(\int_{x : f(x) \leq f(x')} f(x)\, dx \right) dx' \\
&= \beta_2 \sqrt{n} \int_{x' \in \mathcal{R}} \sqrt{f(x')} P(x')\, dx' \\
&= \beta_2 \sqrt{n} \int_{x \in \mathcal{R}} \sqrt{f(x)} P(x)\, dx
\end{aligned}
$$

Each points' traveling time should be less than or equal to the total tour length, so we get:

$$
\lim_{n \to \infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} \leq \beta_2 \iint_{x \in \mathcal{R}} \sqrt{f(x)} P(x)\, dx
$$

which completes the proof. $\square$

5.3 Lower Bound for CTSP

This section presents a lower bounding approach for the tour length of CTSP, which involves dividing the optimal tour of n points into m consecutive sets and obtaining lower bounds for each point by considering the shortest possible tour that visits n/m points out of n points to travel. The starting time for each set is determined by adding the travelling time of all previous visited sets. By summing up the lower bounds for each point, we can obtain a lower bound for the CTSP tour length.

5.3.1 $X_i \overset{i.i.d}{\sim} \mathcal{U}$

Theorem 17 (Lower bound; Uniform CTSP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of area $\mathcal{R}$ in $\mathbb{R}^2$. We have*

$$
\liminf_{n \to \infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{\lambda_1}{2} \sqrt{A}
$$

with probability one, where $\lambda_1 = 0.2935$.

Proof. Fix $p \in (0, 1)$. Divide all n points into $\frac{1}{p}$ sets. As introduced in Section 4, let $L(X_1, \ldots, X_n; pn)$ denote the length of the shortest tour that visits pn points out of $X_1, \ldots, X_n$ Each customer's waiting time in the ith set should be at least $(i-1)L(X_1, \ldots, X_n; pn)$. Since each set has pn points, then the cumulative waiting time in this set is $(i-1)pnL(X_1, \ldots, X_n; pn)$. Based on this analysis, the total waiting time should be at least:

$$
\begin{aligned}
CumL(X_1, \ldots, X_n) &\geq \sum_{i=1}^{\frac{1}{p}} (i-1) pnL(X_1, \ldots, X_n; pn) \\
&= \frac{1}{2}\left(\frac{1}{p} - 1\right) nL(X_1, \ldots, X_n; pn)
\end{aligned}
$$

Applying Theorem 10, we see that

$$
\liminf_{n \to \infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{1}{2}(1-p)\lambda_1 \sqrt{A}
$$

This holds for all p, and $\frac{1}{2}(1-p)\lambda_1\sqrt{A}$ increases as p decreases. The tightest lower bound will be reached when we choose $p \to 0^+$. Therefore,

$$\liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2}} \geq \frac{\lambda_1}{2}\sqrt{A}$$

which completes the proof. □

5.3.2 $X_i \overset{i.i.d}{\sim} \mathcal{P}$

Theorem 18 (Lower bound; step function CTSP). *Let $X_1,\ldots,X_n$ be independent samples, where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R}$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). $\Pi(x)$ is defined in Theorem 3. We have*

$$\liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2}} \geq \lambda_1 \iint_{x\in\mathcal{R}} \sqrt{\phi(x)}\Pi(x)dx$$

with probability one, where $\lambda_1 = 0.2935$.

Proof. For each $\boxdot_i$, using the same mapping procedure as shown in Section 5.1 to reconstruct $\mathsf{L}(\boxdot_i) = \frac{1}{\sqrt{a_i}}\mathsf{L}(\boxdot_i')$. Applying the same reasoning as in the uniform analysis, suppose that we want to visit pn points out of n in total as cheaply as possible. Suppose we pick $p_i n$ points in each $\boxdot_i'$ so that $\sum_{i=1}^{s} p_i = p$. Using the result in Corollary 10.1, we see that

$$\mathsf{L}(X_1,\ldots,X_n;pn) = \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}}\mathsf{L}(\boxdot_i';p_i n) \geq \lambda_1\sqrt{n}\sum_{i=1}^{s}\frac{p_i}{\sqrt{a_i}}$$

as $n \to \infty$, where $\mathsf{L}(\boxdot_i';p_i n)$ represents that picking $p_i n$ points from $\boxdot_i'$

It is straightforward to verify that the tightest lower bound can be achieved if we visit as many points as possible in the denser part (where a_i is large). Therefore, based on the Law of Large Numbers: $|X_1,\ldots,X_n \cap \boxdot'|/n \to \frac{1}{s}$ as $n\to\infty$, the maximum number of points we can visit in each $\boxdot_i'$ is $\frac{n}{s}$.

The optimal p_i values that minimize the above expression are to set $p_1 = \cdots = p_{\lfloor ps\rfloor} = 1/s$ and $p_{\lceil ps\rceil} = p - \lfloor ps\rfloor/s$. Under this assignment, the equation above changes to:

$$\begin{aligned}\lambda_1\sqrt{n}\sum_{i=1}^{s}\frac{p_i}{\sqrt{a_i}} &= \lambda_1\sqrt{n}\Big(\sum_{i=1}^{\lfloor ps\rfloor}\frac{p_i}{\sqrt{a_i}} + \frac{p_{\lceil ps\rceil}}{\sqrt{a_{\lceil ps\rceil}}}\Big) \geq \lambda_1\sqrt{n}\sum_{i=1}^{\lfloor ps\rfloor}\frac{p_i}{\sqrt{a_i}} = \lambda_1\sqrt{n}\sum_{i=1}^{\lfloor ps\rfloor}\frac{1}{s\sqrt{a_i}}\\ &= \lambda_1\sqrt{n}\int_{x\in\mathcal{R}}\sqrt{\phi(x)}\mathbb{1}(\Pi(X)\geq\lceil ps\rceil/s)dx\\ &\geq \lambda_1\sqrt{n}\int_{x\in\mathcal{R}}\sqrt{\phi(x)}\mathbb{1}(\Pi(X)\geq p)dx\end{aligned}$$

as $n\to\infty$. Using the Law of Large Numbers, $|X_1,\ldots,X_n \cap \boxdot_i|/n \to \frac{1}{s} \Rightarrow |X_1,\ldots,X_n\cap\boxdot_i| \to \frac{n}{s}$. Finally, we observe that these bounds enable us to state the following:

- cumulative waiting time in $\boxdot_1 \geq 0$.
- cumulative waiting time in $\boxdot_2 \geq \frac{n}{s}\mathsf{L}(\boxdot_1,pn) \geq \frac{n}{s}\mathsf{L}(X_1,\ldots,X_n,pn)$.

- ...
- cumulative waiting time in $\boxdot_s \geq \frac{n}{s}\mathsf{L}(\cup_{i=1}^{s-1}\boxdot_i, pn) \geq \frac{n}{s}\mathsf{L}(X_1, \ldots, X_n, pn)$.

$$\begin{aligned}
CumL(X_1, \ldots, X_n) &\geq \sum_{j=1}^{s} \frac{n}{s}\mathsf{L}(X_1, \ldots, X_n, pn) = n \sum_{j=0}^{s} \frac{1}{s}\mathsf{L}(X_1, \ldots, X_n, pn) \\
&= n \int_0^1 \mathsf{L}(X_1, \ldots, X_n, pn) dp \\
&\geq n \int_0^1 \lambda_1 \sqrt{n} \int\limits_{x \in \mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq p) dx dp \\
&= \lambda_1 n^{3/2} \int_0^1 \int\limits_{x \in \mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq p) dx \\
&= \lambda_1 n^{\frac{3}{2}} \int_{x \in \mathcal{R}} \int_0^1 \sqrt{\phi(x)} \mathbb{1}(\Pi(x) \geq p) dp dx \\
&= \lambda_1 n^{\frac{3}{2}} \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \left(\int_0^1 \mathbb{1}(\Pi(x) \geq p) dp \right) dx \\
&= \lambda_1 n^{\frac{3}{2}} \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx \\
\Rightarrow \quad \liminf_{n \to \infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} &\geq \lambda_1 \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx
\end{aligned}$$

□

The preceding step function analysis yields the following result by standard coupling arguments:

Theorem 19 (Lower bound; non-uniform CTSP). *Let $X_1, \ldots, X_n$ be independent samples drawn from a region of $\mathcal{R} \subset \mathbb{R}^2$, where X_i follows the density function f, $\forall i$. $P(x)$ is defined in Lemma 3. We have*

$$\liminf_{n \to \infty} \frac{CumL(X_1, \ldots, X_n)}{n^{3/2}} \geq \lambda_1 \iint\limits_{x \in \mathcal{R}} \sqrt{f(x)} P(x) dx$$

with probability one, where $\lambda_1 = 0.2935$.

6 Analysis of the Cumulative CVRP when $n \to \infty$

In this chapter, we present a comprehensive probabilistic analysis of the upper and lower bounds for the cumulative capacitated vehicle routing problem (CCVRP) as the number of points tends to infinity. Starting with uniform distributions, we then extend our analysis to non-uniform distributions. We begin by examining the scenario where there is only one vehicle, and in the following chapter, we will generalize our approach to consider multiple vehicles. This analysis will provide a deeper understanding of the performance of the algorithms designed to solve the CCVRP, which is an essential problem in the field of operations research.

6.1 Choice of c

The core idea to find the necessary bounds for CCVRP is the same as our approach for the CTSP. In Chapter 5, we use a probabilistic bound for the TSP (the BHH theorem), and for this section, it is necessary to find a related bound for the CVRP.

The bound that we will use for CVRP is proven in Section 4.2. In a nutshell, we will find that local cost at a point x can be approximated by $\beta_2\sqrt{f(x)} + \frac{2\sqrt{n}}{c}f(x)\|x\|_2$. Since we are interested in asymptotic behavior, we will assume that c varies relative to n via the relationship $c = k_c\sqrt{n}$. This is because if $c = o(\sqrt{n})$, then the radial cost term $\frac{2\sqrt{n}}{c}f(x)\|x\|_2$ dominates, and if $c = \omega(\sqrt{n})$, then the TSP term $\beta_2\sqrt{f(x)}$ dominates; both of these are easier problems to analyze.

As in our preceding analysis, the key for our analysis of the non-uniform CCVRP relies on an approximation of the true density f with a step function:

Lemma 20 (Approximation with a step function). *Let f be a probability density function with compact support $\mathcal{R} \subset \mathbb{R}^2$ whose level set have Lebesgue measure zero. Define*

$$\Upsilon(x, c, k_p) = Pr(\upsilon(X, c, k_p) \le \upsilon(x, c, k_p)) = \iint_{x':\upsilon(x',c,k_p)\le\upsilon(x,c,k_p)} f(x')dx'$$

Where

$$\upsilon(x, c, k_p) = k_p\sqrt{f(x)} + \frac{2\sqrt{n}}{c}f(x)\|x\|_2$$

For any $\epsilon > 0$, there exists a step density function $\phi(x) := \sum_{i=1}^{s} a_i \mathbb{1}(x \in \square_i)$ and corresponding

$$I(x, c, k_p) = Pr(\iota(X, c, k_p) \le \iota(x, c, k_p)) = \iint_{x':\iota(x',c,k_p)\le\iota(x,c,k_p)} \phi(x')dx'$$

Where

$$\iota(x, c, k_p) = k_p\sqrt{\phi(x)} + \frac{2\sqrt{n}}{c}\phi(x)\|x\|_2$$

such that the following conditions hold:

1. $\int_{\mathcal{R}} |\phi(x) - f(x)| dx \leq \epsilon$,
2. $|\mathbb{1}(\Upsilon(X,c) \geq p) - \mathbb{1}(I(X,c) \geq p)| \leq \epsilon \quad \forall x \in \mathcal{R}$,
3. *All of the components of ϕ have the same mass, i.e. a_iArea$(\boxdot_i) = 1/s$.*

Proof. The proof is essentially identical to the step function approximation from Theorem 3 and we omit it for brevity; we merely replace $\sqrt{f(x)}$ with $\upsilon(x, c, k_p)$. ☐

The functions $\Upsilon(\cdot)$ and $\upsilon(\cdot)$ serve as substitutes for the functions $P(x)$ and $\sqrt{f(x)}$ which we used in our analysis of the CTSP; equivalently, the functions $I(\cdot)$ and $\iota(\cdot)$ serve as substitutes for $\Pi(x)$ and $\sqrt{\phi(x)}$, i.e., the step function approximations of $P(x)$ and $\sqrt{f(x)}$.

The necessary machinery to bound the CCVRP under the assumption that $c = k_c\sqrt{n}$ follows nearly identical logic to our analysis of the CTSP; in the interest of brevity, we will present our bounds here without additional proofs.

6.2 Upper Bound for CCVRP

Theorem 21 (Upper bound; uniform CCVRP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from region $\mathcal{R}$ in $\mathbb{R}^2$, whose area is A and vehicle capacity is $k_c\sqrt{n}$. We have*

$$\limsup_{n\to\infty} \frac{CumL(X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \leq \frac{\overline{r}}{k_c} + \frac{\beta_2\sqrt{A}}{2}$$

Proof. Inspired by the Upper bound in Theorem 11:

- We construct a feasible solution coming from Optimal Tour Partition (OPT)[43] to this routing problem whose value is at most equal to the upper bound.
- Cumulative length of the n solutions comes from two parts: leaving and entering depot plus tour length.

We start from shortest travelling salesman's tour, breaking it into $l = \lceil \frac{\sqrt{n}}{k_c} \rceil$ disjoint segments so that each segment cannot contain more than $k_c\sqrt{n}$ customers, and connecting endpoints of each segment to the depot.

To improve on this result, we start from an arbitrary orientation of the travelling salesman's tour, repeating the above construction by moving the endpoints of the original l paths 1, 2, $\ldots$ up to n positions in the direction of this orientation. The value of the best resulting solution will be less than the average value founded.

W.L.O.G, suppose that $\{X_1, X_2, \ldots, X_n\}$ are the shortest travelling salesman's tour we consider. For leaving and coming back to the depot, the summation is

$$\begin{aligned}
&\left\{n + 2[(n - k_c\sqrt{n}) + \cdots + (n - (\frac{\sqrt{n}}{k_c} - 1)k_c\sqrt{n})]\right\} \sum r_i \\
=&[n + n(\frac{\sqrt{n}}{k_c} - 1)] \sum r_i \\
=&\frac{n^{3/2}}{k_c} \sum r_i
\end{aligned}$$

The sum of cumulative tour length is the similar as what we did in Theorem 14:

1. For the starting sub-tour $L(X_1,\ldots,X_c)$, its cumulative tour length is $(n-1)d_{12}+(n-2)d_{23}+\cdots+1d_{(n-1)n}+0d_{n1}$
2. For the starting sub-tour $L(X_2,\ldots,X_c+1)$, its cumulative tour length is $(n-1)d_{23}+(n-2)d_{34}+\cdots+1d_{n1}+0d_{12}$
3. $\ldots$
4. For the starting sub-tour $L(X_n,\ldots,X_{c-1})$, its cumulative tour length is $(n-1)d_{n1}+(n-2)d_{12}+\cdots+1d_{(n-2)(n-1)}+0d_{(n-1)n}$

The sum of all the terms above is $\frac{n(n-1)}{2}L(X_1,\ldots,X_n)$
Combining them together, the average value we found is:

$$\begin{aligned}\lim_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}} &\le \frac{\bar{r}}{k_c}+\frac{\beta\sqrt{A}}{2}-\frac{\beta\sqrt{A}}{2n}\\ &\le \frac{\bar{r}}{k_c}+\frac{\beta\sqrt{A}}{2}\end{aligned}$$

which completes the proof.

$\square$

Theorem 22 (Upper bound; Step function CCVRP). *Let $X_1,\ldots,X_n$ be independent samples where X_i follows the density function $\phi(x)=\sum_{i=1}^{s}a_i\mathbb{1}(x\in\boxdot_i)$ with compact support $\mathcal{R}\subset\mathbb{R}^2$ such that $a_1\ge\cdots\ge a_s$ and a_iArea$(\boxdot_i)=\frac{1}{s}$ for all i (so that Area$(\mathcal{R})=1$). Suppose capacity is $k_c\sqrt{n}$ and that $\iota(x,k_c\sqrt{n},\beta_2)$ and $I(x,k_c\sqrt{n},\beta_2)$ are as defined in Lemma 20. We have*

$$\limsup_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}}\le\iint_{x\in\mathcal{R}}\iota(x,k_c\sqrt{n},\beta_2)I(x,k_c\sqrt{n},\beta_2)dx$$

Proof. As in the uniform case, we begin by finding a feasible solution. For each $\boxdot_i$, we can reconstruct $\boxdot_i$ to $\boxdot_i'$. The details are shown in Theorem 15 Proof.

Starting from $\boxdot_1$ and the cumulative waiting time in each $\boxdot_i$ for all i is:

- cumulative waiting time in $\boxdot_1\le\frac{n}{s}$VRP$(\boxdot_1;k_c\sqrt{n})$.
- cumulative waiting time in $\boxdot_2\le\frac{n}{s}[$VRP$(\boxdot_1;k_c\sqrt{n})+$VRP$(\boxdot_2;k_c\sqrt{n})]$.
- $\ldots$
- cumulative waiting time in $\boxdot_1\le\frac{n}{s}[$VRP$(\boxdot_1;k_c\sqrt{n})+\cdots+VRP(\boxdot_s;k_c\sqrt{n})]$.

Summing all entries, we find:

$$\begin{aligned}CumL(X_1,\ldots,X_n;k_c\sqrt{n})\le\sum_{j=1}^{s}\frac{n}{s}\sum_{i=1}^{j}\text{VRP}(\boxdot_i;k_c\sqrt{n})&=n\sum_{j=1}^{s}\frac{a_j}{a_j s}\sum_{i=1}^{j}\text{VRP}(\boxdot_i;k_c\sqrt{n})\\ &=n\iint_{x\in\mathcal{R}}\phi(x)\sum_{i=1}^{j}\text{VRP}(\boxdot_i;k_c\sqrt{n})dx\end{aligned}$$

and therefore, combined the result in Remark 1, we have:

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}} &\le \int_{x\in\mathcal{R}}\phi(x)\iint_{x':\iota(x',k_c\sqrt{n},\beta_2)\ge\iota(x,k_c\sqrt{n},\beta_2)}\iota(x',k_c\sqrt{n},\beta_2)dx'dx\\
&= \int_{x\in\mathcal{R}}\iint_{x':\iota(x',k_c\sqrt{n},\beta_2)\ge\iota(x,k_c\sqrt{n},\beta_2)}\phi(x)\iota(x',k_c\sqrt{n},\beta_2)dx'dx\\
&= \int_{x'\in\mathcal{R}}\iint_{x:\iota(x,k_c\sqrt{n},\beta_2)\le\iota(x',k_c\sqrt{n},\beta_2)}\phi(x)\iota(x',k_c\sqrt{n},\beta_2)dxdx'\\
&= \int_{x'\in\mathcal{R}}\iota(x',k_c\sqrt{n},\beta_2)\left(\iint_{x:\iota(x,k_c\sqrt{n},\beta_2)\le\iota(x',k_c\sqrt{n},\beta_2)}\phi(x)dx\right)dx'\\
&= \iint_{x'\in\mathcal{R}}\iota(x',k_c\sqrt{n},\beta_2)I(x',k_c\sqrt{n},\beta_2)dx'\\
&= \iint_{x\in\mathcal{R}}\iota(x,k_c\sqrt{n},\beta_2)I(x,k_c\sqrt{n},\beta_2)dx
\end{aligned}
$$

which completes the proof. □

The preceding step function analysis yields the following result by standard coupling arguments:

Theorem 23 (Upper bound; non-uniform CCVRP). *Let $X_1,\ldots,X_n$ be independent samples drawn from a region of $\mathcal{R}\subset\mathbb{R}^2$, where X_i follows the density function f, $\forall i$. Suppose capacity is $k_c\sqrt{n}$ and that $\upsilon(x,k_c\sqrt{n},\beta_2)$ and $\Upsilon(x,k_c\sqrt{n},\beta_2)$ are as defined in Lemma 20. We have*

$$\limsup_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}}\le\iint_{x\in\mathcal{R}}\upsilon(x,k_c\sqrt{n},\beta_2)\Upsilon(x,k_c\sqrt{n},\beta_2)dx$$

6.3 Lower Bound for CCVRP

Theorem 24 (Lower bound; uniform CCVRP). *Let $X_1,\ldots,X_n$ be independent uniform samples drawn from a region of area A in $\mathbb{R}^2$ and vehicle capacity is $k_c\sqrt{n}$. We have*

$$\liminf_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}}\ge\max\{\frac{1}{2}\lambda_1\sqrt{A},\frac{1}{k_c}\bar{r}\}$$

Proof. From Theorem 11, it is easy to see that the optimal CVRP should be larger than or equal to both TSP result and leaving and returning depot distances. This conclusion can apply to CCVRP as well, which is the optimal CCVRP should be larger than or equal to both CTSP result and leaving and returning depot distances summation.

From Theorem 17, we can get:

$$\liminf_{n\to\infty}\frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}}\ge\liminf_{n\to\infty}\frac{CumL(X_1,\ldots,X_n)}{n^{3/2}}\ge\frac{1}{2}\lambda_1\sqrt{A}$$

Consider the subset $\mathcal{X}_j=\{X_{j_1},X_{j_2},\ldots,X_{j_q}\}$ of customers, visited by the jth tour. Ignoring any tours before this tour, It's easy to know the time for this tour is at least $2\max_{X_i\in\mathcal{X}_j}r_i$

To make it easy to illustrate, we define:

$$\begin{aligned} \widetilde{r}_j &\equiv \sum_{X_i \in \mathcal{X}_j} r_i \\ r_{mj} &\equiv \max_{X_i \in \mathcal{X}_j} r_i \end{aligned}$$

We have

$$\begin{aligned} r_{mj} &= \max_{X_i \in \mathcal{X}_j} r_i \geq \frac{\sum_{X_i \in \mathcal{X}_j} r_i}{|\mathcal{X}_j|} \geq \frac{1}{k_c\sqrt{n}}\widetilde{r}_j \\ r_{mj} &\leq \max_{i=1,\ldots,n} \{r_i\} \end{aligned}$$

So the total time is :

$$\begin{aligned} CumL(X_1,\ldots,X_n;k_c\sqrt{n}) &\geq 2\sum_{i=1}^{\sqrt{n}/k_c}(n-(i-1)k_c\sqrt{n})r_{mi} \\ &= 2(\frac{n^{3/2}}{k_c} - \frac{n}{2}(\frac{\sqrt{n}}{k_c}-1))r_{mi} \\ &= \frac{n^{3/2}}{k_c}r_{mi} + nr_{mj} \\ &\geq \frac{n^{3/2}}{k_c}\bar{r} + n\bar{r} \\ \Rightarrow \liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}} &\geq \frac{1}{k_c}\bar{r} \end{aligned}$$

which completes the proof. □

Theorem 25 (Lower bound; Step function CCVRP). *Let $X_1,\ldots,X_n$ be independent samples where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R} \subset \mathbb{R}^2$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). Suppose capacity is $k_c\sqrt{n}$ and that $\iota(x,k_c\sqrt{n},\lambda_1)$ and $I(x,k_c\sqrt{n},\lambda_1)$ are as defined in Lemma 20. We have*

$$\liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2}} \geq \frac{1}{2}\iint_{x\in\mathcal{R}} \iota(x,k_c\sqrt{n},\lambda_1)I(x,k_c\sqrt{n},\lambda_1)dx$$

Proof. Similar as uniform case, for each $\boxdot_i$, we can reconstruct $\boxdot_i$ to $\boxdot_i'$. The details are shown in Theorem 15 Proof. With the result in Theorem 12, each $\boxdot_i$ cumulative waiting time (set the start time in $\boxdot_i$ as zero) satisfies:

- cumulative waiting time in $\boxdot_1 \geq 0$.
- cumulative waiting time in $\boxdot_2 \geq \frac{n}{s}VRP(\boxdot_1;k_c\sqrt{n};pn) \geq \frac{n}{s}VRP(X_1,\ldots,X_n;k_c\sqrt{n};pn)$.
- …
- cumulative waiting time in $\boxdot_s \geq \frac{n}{s}VRP(\cup_{i=1}^{s-1}\boxdot_i;k_c\sqrt{n};pn) \geq \frac{n}{s}VRP(X_1,\ldots,X_n;k_c\sqrt{n};pn)$.

$$
\begin{aligned}
CumL(X_1,\dots,X_n) &\geq \sum_{j=1}^{s} \frac{n}{s}\text{VRP}(X_1,\dots,X_n;k_c\sqrt{n};pn) = n\sum_{j=0}^{s}\frac{1}{s}\text{VRP}(X_1,\dots,X_n;k_c\sqrt{n};pn) \\
&= n\int_0^1 \text{VRP}(X_1,\dots,X_n;k_c\sqrt{n};pn)dp \\
\liminf_{n\to\infty}\frac{CumL(X_1,\dots,X_n;k_c\sqrt{n})}{n^{3/2}} &\geq \int_0^1 \frac{1}{2}\int_{x\in\mathcal{R}} (\lambda_1\sqrt{\phi(x)} + \tfrac{2}{k_c}\phi(x)\|x\|_2)\mathbb{1}(N(x)\geq p)dxdp \\
&= \frac{1}{2}\int_0^1 \int_{x\in\mathcal{R}} \iota(x,k_c\sqrt{n},\lambda_1)\mathbb{1}(N(x)\geq p)dxdp \\
&= \frac{1}{2}\int_{x\in\mathcal{R}} \iota(x,k_c\sqrt{n},\lambda_1)\int_0^1 \mathbb{1}(N(x)\geq p)dpdx \\
&= \frac{1}{2}\int_{x\in\mathcal{R}} \iota(x,k_c\sqrt{n},\lambda_1)I(x,k_c\sqrt{n},\lambda_1)dx
\end{aligned}
$$

which completes the proof. □

The preceding step function analysis yields the following result by standard coupling arguments:

Theorem 26 (Lower bound; non-uniform CCVRP). *Let $X_1,\dots,X_n$ be independent samples drawn from a region of $\mathcal{R}\subset\mathbb{R}^2$, where X_i follows the density function f, $\forall i$. Suppose capacity is $k_c\sqrt{n}$ and that $\upsilon(x,k_c\sqrt{n},\lambda_1)$ and $\Upsilon(x,k_c\sqrt{n},\lambda_1)$ are as defined in Lemma 20. We have*

$$
\liminf_{n\to\infty}\frac{CumL(X_1,\dots,X_n;k_c\sqrt{n})}{n^{3/2}} \geq \frac{1}{2}\iint_{x\in\mathcal{R}} \upsilon(x,k_c\sqrt{n},\lambda_1)\Upsilon(x,k_c\sqrt{n},\lambda_1)dx
$$

7 Multiple Vehicle Cumulative Routing Problem: m-CTSP

In this chapter, we aim to expand our analysis from the single-vehicle case to multiple vehicles in the context of cumulative vehicle routing problems. Our focus is on providing a probabilistic analysis of upper and lower bounds for the multiple vehicles' cumulative traveling salesman problem (m-CTSP) and the multiple vehicles' cumulative capacitated vehicle routing problem (m-CCVRP), taking into consideration the scenario where the number of points tends to infinity. Our analysis will begin with the uniform distribution case and subsequently move on to non-uniform distributions.

7.1 Upper Bound for m-CTSP

7.1.1 $X_i \overset{i.i.d}{\sim} \mathcal{U}$

Theorem 27 (Upper bound; Uniform m-CTSP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of $\mathcal{R}$ with area A in $\mathbb{R}^2$. There exist m vehicles in total. We have*

$$\limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \le \frac{\beta_2\sqrt{A}}{2m^2} \tag{7.1}$$

with probability one where $\beta_2 \approx 0.714$.

Proof. Divide the region A into m equal-area sub-regions, such that each vehicle visits one sub-region. The optimal value of this problem should be less than or equal to the value generated by this feasible solution.

For each sub-region, we can use Theorem 14 to finish the proof. The number of each sub-region's points converge to $\frac{n}{m}$ when $n \to \infty$. Since all vehicles start tour at the same time, each vehicle's cumulative traveling time is equal to this cumulative problem's traveling time:

$$\begin{aligned} \limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{(n/m)^{3/2}} &\le \frac{\beta_2}{2}\sqrt{\frac{A}{m}} \\ \Rightarrow \limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} &\le \frac{\beta_2\sqrt{A}}{2m^2} \end{aligned}$$

□

7.1.2 $X_i \overset{i.i.d}{\sim} \mathcal{P}$

Theorem 28 (Upper bound; Step function m-CTSP). *Let $X_1, \ldots, X_n$ be independent samples, where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R} \subset \mathbb{R}^2$ such that $a_1 \ge$*

$\cdots \geq a_s$ *and* a_i*Area*$(\boxdot_i) = \frac{1}{s}$ *for all* i *(so that Area*$(\mathcal{R}) = 1$*). Suppose that* $\Pi(x)$ *is as defined in Lemma 3. There exist* m *vehicles in total. We have*

$$\limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \leq \frac{\beta_2}{m^2} \iint_{x\in\mathcal{R}} \sqrt{\phi(x)}\Pi(x)dx \tag{7.2}$$

with probability one where $\beta_2 \approx 0.714$.

Proof. As in the uniform case, we begin by finding a feasible solution. For each $\boxdot_i$, we can reconstruct $\boxdot_i$ to $\boxdot_i'$ using a function $\psi_i(x) = \sqrt{a_i}x + \xi_i$ to make:

- Each $\boxdot_i'$ disjoint;
- Area$(\boxdot_i') = \frac{1}{s}$ for all i;
- Points in $\boxdot_i'$ are uniform distribution;

Basic scaling argument tells us that TSP$(\boxdot_i) = \frac{1}{\sqrt{a_i}}TSP(\boxdot_i')$

Similar as uniform case proof before, we can divide each cell $\boxdot_i'$ into m sub-cell and every vehicle travels one sub-cell. Since now each point in the sub-cell is uniformly distributed, we can apply the BHH theorem:

$$\lim_{n\to\infty} \frac{\text{TSP}(\boxdot_{i';m})}{\sqrt{n}} = \beta_2\frac{1}{sm} \Rightarrow \lim_{n\to\infty} \text{TSP}(\boxdot_i) = \frac{1}{\sqrt{a_i}}\text{TSP}(\boxdot_i') = \frac{\beta_2}{sm\sqrt{a_i}}$$

where $\boxdot_{i';m}$ represents the mth sub-cell related to $\boxdot_{i'}$ cell.

All m vehicles start at the same time, so one vehicle's TSP can represent this region's TSP. Each $\boxdot_i$ is related to a_i, one of the possible ways to travel to all points is to follow "most dense to least dense" rule, which merely means that we travel from $\boxdot_1, \ldots, \boxdot_s$. (From the definition of our step function, the densest part of the distribution is equivalent to the smallest area $\boxdot$, which is $\boxdot_1$). Using Theorem 4, for point $X_l \in \boxdot_j$, its waiting time should be less than or equal to :

$$\text{TSP}(\boxdot_{1,m}) + \text{TSP}(\boxdot_{2,m}) + \cdots + \text{TSP}(\boxdot_{j,m}) = \sum_{i=1}^{j} \frac{\beta_2}{sm}\sqrt{\frac{n}{a_i}} = \frac{\beta_2}{m}\sqrt{n} \iint_{x':\phi(x')\geq\phi(x)} \sqrt{\phi(x')}dx'$$

as $n \to \infty$.

Using the Law of Large Numbers,

$$|X_1, \ldots, X_n \cap \boxdot_{i,m}|/n \to \frac{1}{sm} \Rightarrow |X_1, \ldots, X_n \cap \boxdot_i| \to \frac{n}{sm}$$

We find that, as $n \to \infty$, each set's sub-region will have $\frac{n}{sm}$ points. Therefore, combining both terms together, we can get the cumulative waiting time in each $\boxdot_i$ for all i:

- cumulative waiting time in $\boxdot_{1,m} \leq \frac{n}{sm}L(\boxdot_1)$
- cumulative waiting time in $\boxdot_{2,m} \leq \frac{n}{sm}$[L$(\boxdot_{1,m})$+L$(\boxdot_{2,m})$].
- ...
- cumulative waiting time in $\boxdot_{s,m} \leq \frac{n}{sm}$[L$(\boxdot_{1,m}) + \cdots +L(\boxdot_{s,m})$].

Summing all entries, we find:

$$
\begin{aligned}
CumL(m; X_1, \ldots, X_n) &\leq \sum_{j=1}^{s} \frac{n}{sm} \sum_{i=1}^{j} \mathsf{L}(\boxdot_{i,m}) = \frac{n}{m} \sum_{j=1}^{s} \frac{a_j}{a_j s} \sum_{i=1}^{j} \mathsf{L}(\boxdot_{i,m}) \\
&= \frac{n}{m^2} \iint_{x \in \mathcal{R}} \phi(x) \sum_{i=1}^{j} \mathsf{L}(\boxdot_i) dx
\end{aligned}
$$

and therefore, from our previous formulation, we have:

$$
\begin{aligned}
\limsup_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} &\leq \frac{\beta_2}{m^2} \int_{x \in \mathcal{R}} \phi(x) \iint_{x' : \phi(x') \geq \phi(x)} \sqrt{\phi(x')} dx' dx \\
&= \frac{\beta_2}{m^2} \int_{x \in \mathcal{R}} \iint_{x' : \phi(x') \geq \phi(x)} \phi(x) \sqrt{\phi(x')} dx' dx \\
&= \frac{\beta_2}{m^2} \int_{x' \in \mathcal{R}} \iint_{x : \phi(x) \leq \phi(x')} \phi(x) \sqrt{\phi(x')} dx dx' \\
&= \frac{\beta_2}{m^2} \int_{x' \in \mathcal{R}} \sqrt{\phi(x')} \left(\int_{x : \phi(x) \leq \phi(x')} f(x) dx \right) dx' \\
&= \frac{\beta_2}{m^2} \iint_{x' \in \mathcal{R}} \sqrt{\phi(x')} \Pi(x') dx' \\
&= \frac{\beta_2}{m^2} \iint_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx
\end{aligned}
$$

which completes the proof. □

Theorem 29 (Upper bound; non-uniform m-CTSP). *Let $X_1, \ldots, X_n$ be independent samples from a region of $\mathcal{R} \subset \mathbb{R}^2$, where X_i follows the density function f, $\forall i$. $P(x)$ is as defined in Lemma 3. There exist m vehicles in total. We have*

$$
\lim_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \leq \frac{\beta_2}{m^2} \iint_{x \in \mathcal{R}} \sqrt{f(x)} P(x) dx \tag{7.3}
$$

with probability one where $\beta_2 \approx 0.714$.

Proof. The key point is the same as what we did in step density function: Theorem 28. Based on Theorem 3, f can be approximated using ϕ. Again, since we have m vehicles, each vehicle will handle $\frac{n}{m}$ points as $n \to \infty$. Based on “Most dense to Least dense ” Rule, for point x, the amount of time he/she has to wait is:

$$
\beta_2 \frac{\sqrt{n}}{m} \iint_{x' : f(x') \geq f(x)} \sqrt{f(x')} dx'
$$

When traversing $\frac{n}{m}$ points, the length is

$$
\beta_2 \frac{\sqrt{n}}{m} \int_{x \in \mathcal{R}} f(x) \iint_{x' : f(x') \geq f(x)} \sqrt{f(x')} dx' dx
$$

$$
\begin{aligned}
&= \beta_2 \frac{\sqrt{n}}{m} \int_{x \in \mathcal{R}} \int_{x': f(x') \geq f(x)} f(x) \sqrt{f(x')} dx' dx \\
&= \beta_2 \frac{\sqrt{n}}{m} \int_{x' \in \mathcal{R}} \int_{x: f(x) \leq f(x')} f(x) \sqrt{f(x')} dx dx' \\
&= \beta_2 \frac{\sqrt{n}}{m} \int_{x' \in \mathcal{R}} \sqrt{f(x')} \left(\int_{x: f(x) \leq f(x')} f(x) dx \right) dx' \\
&= \beta_2 \frac{\sqrt{n}}{m} \int_{x' \in \mathcal{R}} \sqrt{f(x')} P(x') dx' \\
&= \beta_2 \frac{\sqrt{n}}{m} \int_{x \in \mathcal{R}} \sqrt{f(x)} P(x) dx
\end{aligned}
$$

Each points' traveling time should be less than or equal to the total tour length and each tour length, there exists $\frac{n}{m}$ points, so we get:

$$
\lim_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \leq \frac{\beta_2}{m^2} \int_{x \in \mathcal{R}} \sqrt{f(x)} P(x) dx
$$

which completes the proof. □

7.2 Lower Bound for m-CTSP

7.2.1 $X_i \overset{i.i.d}{\sim} \mathcal{U}$

Theorem 30 (Lower bound; Uniform m-CTSP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of area $\mathcal{R}$ with area A in $\mathbb{R}^2$. There exist m vehicles in total. We have*

$$
\liminf_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{\lambda_1}{2m^2} \sqrt{A}
$$

with probability one, where $\lambda_1 = 0.2935$.

Proof. Each vehicle will travel $\frac{n}{m}$ as $n \to \infty$, so we can fix $p \in (0, 1)$. Divide all n points into $\frac{m}{p}$ sets. As introduced in Section 4, let $L(X_1, \ldots, X_n; pn/m)$ denote the length of the shortest tour that visits pn/m points out of $X_1, \ldots, X_n$ Each customer's waiting time in the ith set should be at least $(i-1)L(X_1, \ldots, X_n; pn)$. Since each set has pn/m points, then the cumulative waiting time in this set is $(i-1)\frac{pn}{m}L(X_1, \ldots, X_n; pn)$. Based on this analysis, the total waiting time for each vehicle should be at least:

$$
\begin{aligned}
CumL(m; X_1, \ldots, X_n) &\geq \sum_{i=1}^{\frac{1}{p}} (i-1) \frac{pn}{m} L(X_1, \ldots, X_n; pn/m) \\
&= \frac{1}{2} (\frac{1}{p} - 1) \frac{n}{m} L(X_1, \ldots, X_n; pn)
\end{aligned}
$$

Applying Theorem 10, we see that

$$
\liminf_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{1}{2}(1-p) \frac{1}{m^2} \lambda_1 \sqrt{A}
$$

This holds for all p, and $\frac{1}{2m^2}(1-p)\lambda_1\sqrt{A}$ increases as p decreases. The tightest lower bound will be reached when we choose $p \to 0^+$. Therefore,

$$\liminf_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{\lambda_1}{2m^2}\sqrt{A}$$

which completes the proof. □

7.2.2 $X_i \overset{i.i.d}{\sim} \mathcal{P}$

Theorem 31 (Lower bound; step function m-CTSP). *Let $X_1, \ldots, X_n$ be independent samples, where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R}$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). $\Pi(x)$ is defined in Theorem 3. There exist m vehicles in total. We have*

$$\liminf_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{\lambda_1}{m^2} \iint_{x\in\mathcal{R}} \sqrt{\phi(x)}\Pi(x)dx$$

with probability one, where $\lambda_1 = 0.2935$.

Proof. For each $\boxdot_i$, using the same mapping procedure as shown in Section 5.1 to reconstruct $\mathsf{L}(\boxdot_i) = \frac{1}{\sqrt{a_i}}\mathsf{L}(\boxdot_i')$. Applying the same reasoning as in the uniform analysis, suppose that we want to visit pn/m points out of n/m in total for each vehicle as cheaply as possible. Suppose we pick $p_i n/m$ points in each $\boxdot_i'$ so that $\sum_{i=1}^{s} p_i = p$. Using the result in Corollary 10.1, we see that

$$\mathsf{L}(X_1, \ldots, X_n; pn/m) = \sum_{i=1}^{s} \frac{1}{\sqrt{a_i}} \mathsf{L}(\boxdot_i'; p_i n/m) \geq \lambda_1 \sqrt{n} \sum_{i=1}^{s} \frac{p_i}{m\sqrt{a_i}}$$

as $n \to \infty$, where $\mathsf{L}(\boxdot_i'; p_i n/m)$ represents that picking $p_i n/m$ points from $\boxdot_i'$

It is straightforward to verify that the tightest lower bound can be achieved if we visit as many points as possible in the denser part (where a_i is large). Therefore, based on the Law of Large Numbers: $|X_1, \ldots, X_n \cap \boxdot'|/n \to \frac{1}{s}$ as $n \to \infty$, the maximum number of points we can visit in each $\boxdot_i'$ is $\frac{n}{s}$.

The optimal p_i values that minimize the above expression are to set $p_1 = \cdots = p_{\lfloor ps \rfloor} = 1/s$ and $p_{\lceil ps \rceil} = p - \lfloor ps \rfloor / s$. Under this assignment, the equation above changes to:

$$\begin{aligned}
\lambda_1 \frac{\sqrt{n}}{m} \sum_{i=1}^{s} \frac{p_i}{\sqrt{a_i}} &= \lambda_1 \frac{\sqrt{n}}{m} \left(\sum_{i=1}^{\lfloor ps \rfloor} \frac{p_i}{\sqrt{a_i}} + \frac{p_{\lceil ps \rceil}}{\sqrt{a_{\lceil ps \rceil}}}\right) \geq \lambda_1 \frac{\sqrt{n}}{m} \sum_{i=1}^{\lfloor ps \rfloor} \frac{p_i}{\sqrt{a_i}} = \lambda_1 \frac{\sqrt{n}}{m} \sum_{i=1}^{\lfloor ps \rfloor} \frac{1}{s\sqrt{a_i}} \\
&= \lambda_1 \frac{\sqrt{n}}{m} \iint_{x\in\mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq \lceil ps \rceil / s) dx \\
&\geq \lambda_1 \frac{\sqrt{n}}{m} \iint_{x\in\mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq p) dx
\end{aligned}$$

as $n \to \infty$. Using the Law of Large Numbers, $|X_1, \ldots, X_n \cap \boxdot_i|/n \to \frac{1}{s} \Rightarrow |X_1, \ldots, X_n \cap \boxdot_i| \to \frac{n}{s}$. Each vehicle only needs to travel $\frac{n}{sm}$ out of all $\frac{n}{s}$ points. Finally, we observe that these bounds enable us to state the following:

$$
\begin{aligned}
CumL(m; X_1, \ldots, X_n) &\geq \sum_{j=1}^{s} \tfrac{n}{ms} \mathsf{L}(X_1, \ldots, X_n, pn/m) = \tfrac{n}{m} \sum_{j=0}^{s} \tfrac{1}{s} \mathsf{L}(X_1, \ldots, X_n, pn/m) \\
&= \tfrac{n}{m} \int_0^1 \mathsf{L}(X_1, \ldots, X_n, pn/m) dp \\
&\geq \tfrac{n}{m} \int_0^1 \lambda_1 \tfrac{\sqrt{n}}{m} \int\limits_{x \in \mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq p) dx dp \\
&= \lambda_1 \tfrac{n^{3/2}}{m^2} \int_0^1 \int\limits_{x \in \mathcal{R}} \sqrt{\phi(x)} \mathbb{1}(\Pi(X) \geq p) dx \\
&= \lambda_1 \tfrac{n^{3/2}}{m^2} \int_{x \in \mathcal{R}} \int_0^1 \sqrt{\phi(x)} \mathbb{1}(\Pi(x) \geq p) dp dx \\
&= \lambda_1 \tfrac{n^{3/2}}{m^2} \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \left(\int_0^1 \mathbb{1}(\Pi(x) \geq p) dp \right) dx \\
&= \lambda_1 \tfrac{n^{3/2}}{m^2} \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx \\
\Rightarrow \quad \liminf_{n \to \infty} \tfrac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} &\geq \tfrac{\lambda_1}{m^2} \int_{x \in \mathcal{R}} \sqrt{\phi(x)} \Pi(x) dx
\end{aligned}
$$

□

The preceding step function analysis yields the following result by standard coupling arguments:

Theorem 32 (Lower bound; non-uniform m-CTSP). *Let $X_1, \ldots, X_n$ be independent samples drawn from a region of $\mathcal{R} \subset \mathbb{R}^2$, where X_i follows the density function f, $\forall i$. $P(x)$ is defined in Lemma 3. There exist m vehicles in total. We have*

$$
\liminf_{n \to \infty} \frac{CumL(m; X_1, \ldots, X_n)}{n^{3/2}} \geq \frac{\lambda_1}{m^2} \int\limits_{x \in \mathcal{R}} \sqrt{f(x)} P(x) dx
$$

with probability one, where $\lambda_1 = 0.2935$.

8 Multiple Vehicle Cumulative Routing Problem: m-CCVRP

In this chapter, we build upon our previous analysis by introducing vehicle constraints to our models. Specifically, we consider the cumulative capacitated vehicle routing problem (CCVRP) with the assumption that the relationship between the vehicle capacity c and the number of points n is given by $c = k_c\sqrt{n}$. We will investigate the probabilistic upper and lower bounds for this problem, focusing on both uniform and non-uniform distributions.

8.1 Upper Bound for m-CCVRP

Theorem 33 (Upper bound; uniform m-CCVRP)**.** *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from region $\mathcal{R}$ in $\mathbb{R}^2$, whose area is A and vehicle capacity is $k_c\sqrt{n}$. There exist m vehicles in total. We have*

$$\limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \leq \frac{1}{m^2}\left(\frac{\overline{r}}{k_c} + \frac{\beta_2\sqrt{A}}{2}\right)$$

Proof. Divide the region A into m equal-area sub-regions, such that each vehicle visits one sub-region. This problem reduces to CCVRP with region $\frac{A}{m}$. Using the result shown in Theorem 21 and proofs in m-CTSP, we can complete the proof. □

By reducing all cases to a CCVRP that restricts each vehicle to a sub-region, we obtain the non-uniform result as:

Theorem 34 (Upper bound; Step function m-CCVRP)**.** *Let $X_1, \ldots, X_n$ be independent samples where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R} \subset \mathbb{R}^2$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). Suppose capacity is $k_c\sqrt{n}$ and that $\iota(x, k_c\sqrt{n}, \beta_2)$ and $I(x, k_c\sqrt{n}, \beta_2)$ are as defined in Lemma 20. There exist m vehicles in total. We have*

$$\limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \leq \frac{1}{m^2} \iint_{x\in\mathcal{R}} \iota(x, k_c\sqrt{n}, \beta_2) I(x, k_c\sqrt{n}, \beta_2)\, dx$$

Theorem 35 (Upper bound; non-uniform m-CCVRP)**.** *Let $X_1, \ldots, X_n$ be independent samples drawn from a region of $\mathcal{R} \subset \mathbb{R}^2$, where X_i follows the density function f, $\forall i$. Suppose capacity is $k_c\sqrt{n}$ and that $\upsilon(x, k_c\sqrt{n}, \beta_2)$ and $\Upsilon(x, k_c\sqrt{n}, \beta_2)$ are as defined in Lemma 20. There exist m vehicles in total. We have*

$$\limsup_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \leq \frac{1}{m^2} \iint_{x\in\mathcal{R}} \upsilon(x, k_c\sqrt{n}, \beta_2) \Upsilon(x, k_c\sqrt{n}, \beta_2)\, dx$$

8.2 Lower Bound for m-CCVRP

Theorem 36 (Lower bound; uniform CCVRP). *Let $X_1, \ldots, X_n$ be independent uniform samples drawn from a region of area A in $\mathbb{R}^2$ and vehicle capacity is $k_c\sqrt{n}$. There exist m vehicles in total. We have*

$$\liminf_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; c)}{n^{3/2}} \geq \frac{1}{m^2} \max\{\lambda_1\sqrt{A}, \frac{1}{k_c}(2\bar{r} - \max_{i=1,\ldots,n}\{r_i\})\}$$

Proof. Divide the region A into m equally sub-regions, each vehicle visits one sub-region. This problem reduces to CCVRP with region $\frac{A}{m}$. Using the result shown in Theorem 24 and proofs in m-CTSP, we can complete the proof. □

By reducing all cases to sub-region CCVRP, we can get the non-uniform result as:

Theorem 37 (Lower bound; Step function m-CCVRP). *Let $X_1, \ldots, X_n$ be independent samples where X_i follows the density function $\phi(x) = \sum_{i=1}^{s} a_i \mathbb{1}(x \in \boxdot_i)$ with compact support $\mathcal{R} \subset \mathbb{R}^2$ such that $a_1 \geq \cdots \geq a_s$ and a_iArea$(\boxdot_i) = \frac{1}{s}$ for all i (so that Area$(\mathcal{R}) = 1$). Suppose capacity is $k_c\sqrt{n}$ and that $\iota(x, k_c\sqrt{n}, \lambda_1)$ and $I(x, k_c\sqrt{n}, \lambda_1)$ are as defined in Lemma 20. There exist m vehicles in total. We have*

$$\liminf_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \geq \frac{1}{2m^2} \iint_{x\in\mathcal{R}} \iota(x, k_c\sqrt{n}, \lambda_1) I(x, k_c\sqrt{n}, \lambda_1) dx$$

Theorem 38 (Lower bound; non-uniform m-CCVRP). *Let $X_1, \ldots, X_n$ be independent samples drawn from a region of $\mathcal{R} \subset \mathbb{R}^2$, where X_i follows the density function f, $\forall i$. Suppose capacity is $k_c\sqrt{n}$ and that $\upsilon(x, k_c\sqrt{n}, \lambda_1)$ and $\Upsilon(x, k_c\sqrt{n}, \lambda_1)$ are as defined in Lemma 20. There exist m vehicles in total. We have*

$$\liminf_{n\to\infty} \frac{CumL(m; X_1, \ldots, X_n; k_c\sqrt{n})}{n^{3/2}} \geq \frac{1}{2m^2} \iint_{x\in\mathcal{R}} \upsilon(x, k_c\sqrt{n}, \lambda_1) \Upsilon(x, k_c\sqrt{n}, \lambda_1) dx$$

9 Experimental Results

In this section, we present a comprehensive evaluation of the proposed bounds for simulated instances and real-world datasets. To illustrate the effectiveness of our approach, we provide several examples demonstrating the tightness of the bounds in simulated instances. Furthermore, we apply subsets of real data to predict the optimal waiting time. To obtain the optimal solutions, we utilize LKH-3, a heuristic method, for solving the cumulative Traveling Salesman Problem, and Google OR-Tools for solving the cumulative Vehicle Routing Problems. The experimental results provide insightful observations on the efficacy of our proposed bounds in practice.

9.1 Single Vehicle

9.1.1 Simulated Instances

In this section, we explore two different scenarios: regular and irregular regions. In the regular case, we generate instances from a uniform distribution and Normal distribution with variances $\sigma^2 = 490000$ and $\sigma^2 = 250000$, centered at $(500, 500)$ and $(750, 750)$, respectively, within a square area of side length 1000. For our first instance, we analyze cumulative TSP experiments with uniform sample points.

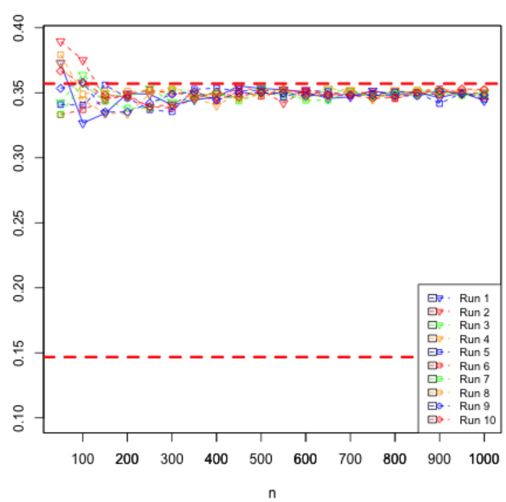

Figure 9.1: Ratio of uniform customers' cumulative waiting time to $\sqrt{A}n^{3/2}$

Figure 9.1 illustrates the results of the uniform distribution experiment, where we show the ratio of cumulative waiting time to $\sqrt{A}N^{3/2}$ for 10 runs of each $n \in \{50, 100, 150, \ldots, 1000\}$. The red dashed line represents the upper bound we proved before. The results show that as n increases, the ratio converges to a value close to $[0.34, 0.35]$, which is consistent with our upper bound.

For the non-uniform distribution experiment, we used a bi-variate Gaussian distribution. Figure 9.2(left) displays the results when the depot is located at the center of the square, while Figure 9.2(right) shows the results when the depot is located at an offset center. We also experimented with two different standard deviations, 500 and 700, as shown in the top and bottom panels, respectively. The results indicate that the variance only has an impact at the beginning when the data points are sparse. As the size of the data increases, having the same depot location, the results converge to the same range. The different depot locations converge to different ranges, but their difference is negligible.

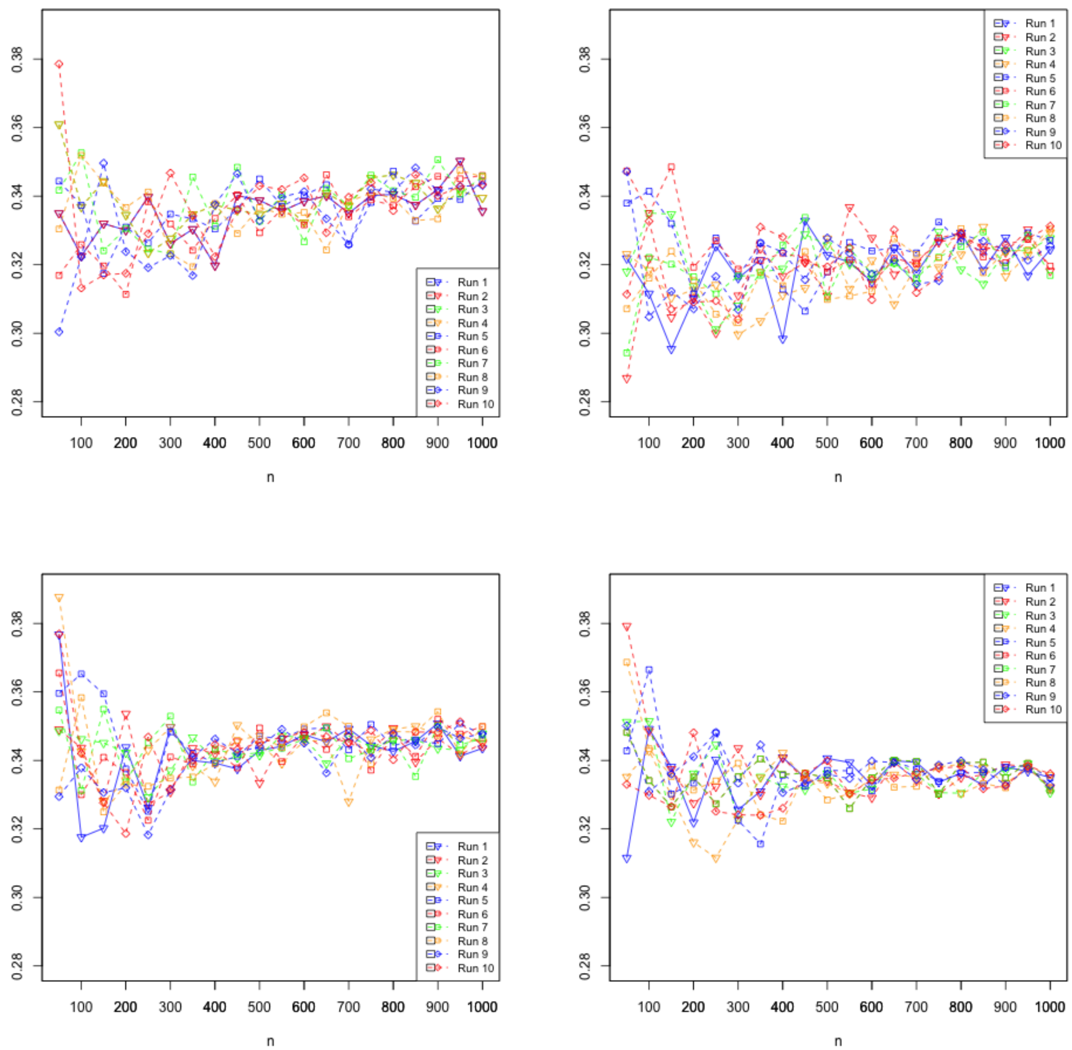

Figure 9.2: Ratio of normal customers' cumulative waiting time to $\sqrt{A}n^{3/2}$. (Top Left) figure depot is at (500, 500) with standard deviation 500. (Top Right) figure depot is at (750, 750) with standard deviation 500. (Bottom Left) figure depot is at (500, 500) with standard deviation 700. (Bottom Right) figure depot is at (750, 750) with standard deviation 700.

Another case we consider is an irregularly-shaped region. For this part, we shift our focus to two convex and one non-convex scenario. Both of them take a bi-variate Gaussian distribution. Figure 9.3 (left) shows the results of this experiment when the region is a circle, triangular and pentagram, with corresponding heatmaps (right). We observe that the ratio of cumulative waiting time is primarily determined by the density function of the sample points.

For the CCVRP with uniformly distributed sample points, we choose the vehicle capacity to be $\sqrt{n}$. The upper bound for this problem is $O(1.914n^{3/2})$. Figure 9.4 presents the results of the experiment, which shows the ratio of cumulative waiting time $\bar{r}n^{3/2}$ for 10 runs of each $n \in \{50, 100, \ldots, 1000\}$ for uniform sample points. We observe that the ratio converges to $[1.7, 1.8]$, which is consistent with the theoretical upper bound.

Figure 9.5 is the result when all points follow bi-variate Gaussian distribution. The left figures show the experiment when the depot is in the center of the square and the right figures have the depot location at an offset from the center. In addition, the top figures follow standard deviation 500 and the bottom figures have 700 instead. We can figure out that variance only has impact at the beginning when the data points are few. When we increase the size of data, having the same depot location, they will converge to the same ratio range. Different depot locations have impact on the ratio range.

Still same as CTSP, we did the experiments considering another convex region: circle and non-convex case: pentagram, whose points follow the bi-variate Gaussian distribution. The result is shown in Figure 9.6. It still presents that the ratio only relates to the density distribution.

Overall, the results of our experiments for both CTSP and CCVRP indicate that the ratio of cumulative waiting time for large n is close to the upper bounds we proved for uniform and non-uniform distributions. Furthermore, we have demonstrated that the upper bounds hold for different types of regions and distributions.

9.2 Multiple Vehicles

Instead of considering only one vehicle, we investigate the effect of having multiple vehicles in this section. Specifically, we consider four different cases where we have 2, 4, 8, and 16 vehicles, respectively. Our investigation aims to deepen our understanding of the m-CTSP and m-CCVRP under different scenarios and provide insights for practical applications.

9.2.1 Simulated Instances

As in the preceding single-vehicle situations, we still consider the vehicle with infinite and finite capacity respectively. For each scenario, both regular and irregular boundary will be considered. As a service region, we use a square area with side length 1000.

In this section, we extend our investigation to the Multi-Vehicle Capacitated Traveling Salesman Problem (m-CTSP) by examining experiments with both uniform and non-uniform sample points. The top left of Figure 9.7 illustrates the ratio of cumulative waiting time to $\sqrt{A}N^{3/2}$ for each $n \in \{50, 100, 150, \ldots, 1000\}$, where the points follow the uniform distribution. To assess the impact of varying numbers of vehicles, we consider four different cases: 2, 4, 8, or 16 vehicles in the system, with the single-vehicle case serving as the baseline for comparison. Remarkably, our findings demonstrate that, similar to the single-vehicle case, regardless of the number of vehicles in the system, the ratio converges to a constant value of $\frac{1}{m^2}[0.34, 0.35]$, where m denotes the number of vehicles. This result closely aligns with the upper bound we previously provided, further validating the correctness of our approach. Moreover, doubling the number of vehicles leads to a reduction of the cumulative waiting time to a quarter of its original value. To clarify this effect, we take the logarithm of the ratio results, shown in the top right of Figure 9.7. We further examine the tours of multiple vehicles by sampling 500 data points, which must be visited by one of four vehicles. The resulting heatmap and tour plan are depicted in the bottom of Figure 9.7. The tour results confirm the validity of our proof process, with the entire region divided into four equal sub-regions, and each vehicle assigned to take charge of one.

For the non-uniform distribution, we consider a bi-variate Gaussian distribution, with two cases: the depot location at the center (Figure 9.8) and the offset center (Figure 9.9), similar to the single-vehicle case. Our findings indicate that doubling the number of vehicles reduces the cumulative waiting time to a quarter of its initial value. Moreover, from the tour plans on the right in Figure 9.8 and Figure 9.9, we observe that minimizing the cumulative waiting time yields an optimal solution that follows our previously established rule, progressing from the most dense to the least dense regions.

In addition, we examine the effect of region irregularity on the performance of the m-CTSP algorithm, considering three different scenarios: two convex regions and one non-convex region. Each region is assumed to follow a bi-variate Gaussian distribution. Figure 9.10 presents the experimental results for circular, triangular, and pentagram regions, respectively. The outcomes demonstrate that the algorithm performs as expected in each scenario, thereby providing further evidence of the algorithm's robustness and versatility.

In the subsequent experiments, we aim to examine the performance of the Capacitated Vehicle Routing Problem (CVRP) with multiple vehicles, considering both uniform and non-uniform sample points. Similar to the CCVRP case, we adopt the square root of the number of nodes, denoted as $\sqrt{n}$, as the vehicle capacity.

For uniformly distributed points, Figure 9.11 verifies that double the number of vehicles will decrease the cumulative waiting time to a quarter.

Figure 9.12 is the result when all points follow bi-variate Gaussian distribution. We also did the experiments whose region is not regular, which is shown in Figure 9.13. Here, we only consider one convex (Circle) and one non-convex (pentagram) case.

9.2.2 Experiments with Road Network Data

The dataset we are using is driving times using HERE Maps API for 1500 points sampled in a rectangle in downtown Los Angeles. We consider CTSP and CCVRP respectively, using subsets of data, from 50 points to 1450 data points, to predict the cumulative waiting time. The predicted rule is as follows:

1. Randomly select n points from N data points where $|n| \in \{50, 100, \dots, 1500\}$.
2. If this is CCVRP case, calculate the average distance between picked points and depot.
3. Put previous equation in theorem to calculate ratios:
   (a) $ratio_{CTSP} = \frac{CumL(x_1,\dots,x_n)}{n^{3/2}}$
   (b) $ratio_{CCVRP} = \frac{CumL(x_1,\dots,x_n)}{\bar{r}n^{3/2}}$
4. Use ratio got from 3, multiple $N^{3/2}$ (or $\bar{R}N^{3/2}$) to predict the simulated cost for 1500 data points.
5. For the multiple vehicles' case, we need to multiple $\frac{N^{3/2}}{m^2}$ (or $\frac{\bar{R}N^{3/2}}{m^2}$), where m is the number of vehicles to predict the simulated cost for 1500 data points.

Figure 9.14 presents the results of our study. The bold red line represents the average cost when we travel to all 1500 data points under CTSP and CCVRP situations. From the left figure, we observe that as we increase the number of data points, the predicted cost becomes closer to the real cost. Moreover, the right-hand side of the figure shows that using only 100 data points is already sufficient to obtain predictions close to the real results.

Figure 9.15 is the result showing what's happening if we have multiple vehicles. Same, the bold red line represents the average cost when we travel to all 1500 data points under CTSP and CCVRP situations when we have one vehicle. Still, it verifies our result.

9.3 Managerial Insights

Based on the experiment results and continuous approximation we proved, we can get the following managerial insights:

- CTSP
  - When the demand is uniform, as shown from the experiment results that optimal CTSP is close to our upper bound. Since we generated the upper bound using TSP, this indicates that CTSP and TSP when the demand is uniform do not have much difference.
  - However, when the demand is non-uniform, the TSP can be dramatically worse than CTSP. As shown in our analysis process, the tour that we generated for CTSP needs to follow the "most dense to least dense" rule. However, the optimal TSP tour does not have any such notion in its objective, so there is no particular incentive to visit denser regions earlier in the tour. In fact, an extreme situation could occur when the TSP tour is travelled from the totally opposite direction, which is "Least dense to most dense", which can be arbitrarily poor.
  - Combined these two observations, we conclude that a good cumulative TSP tour is a good TSP tour, but not vice versa.
  - Analysis suggests that *districting strategies* are not viable for non-uniform CTSP (but they are popular for TSP), because all vehicles should move simultaneously from the densest regions to less dense regions (see [23] for more details).
- CCVRP
  - From both the experiment results and our analysis process, we figure out unlike CTSP, the depot location has significant impact on CCVRP.
  - For CCVRP, if capacities c are not of the same order as $\sqrt{n}$, then either the returns to the depot or driving between customers dominates
- Multiple Vehicle (m-CTSP and m-CCVRP)
  - As shown from the experiment result that we can easily divide the region into multiple small non-overlap sub-region and assign one vehicle to travel.
- Classical Vehicle Routing Problem vs Cumulative Vehicle Routing Problem
  - It is easy to see that the Cumulative Problem is at least as hard as the Classical Problem. For example, given a set of points on which we want to minimize the length of the TSP tour, one can augment the set of points with N points at "infinity" (for a large number N) (see BHH theorem results, $O(n)$), so that the cumulative problem on the augmented set of points will have to minimize the length of the TSP on the original set of points (see our result, $O(n^{3/2})$). This connection shows that the Cumulative problem is NP-hard even in the case where the metric space is a plane.

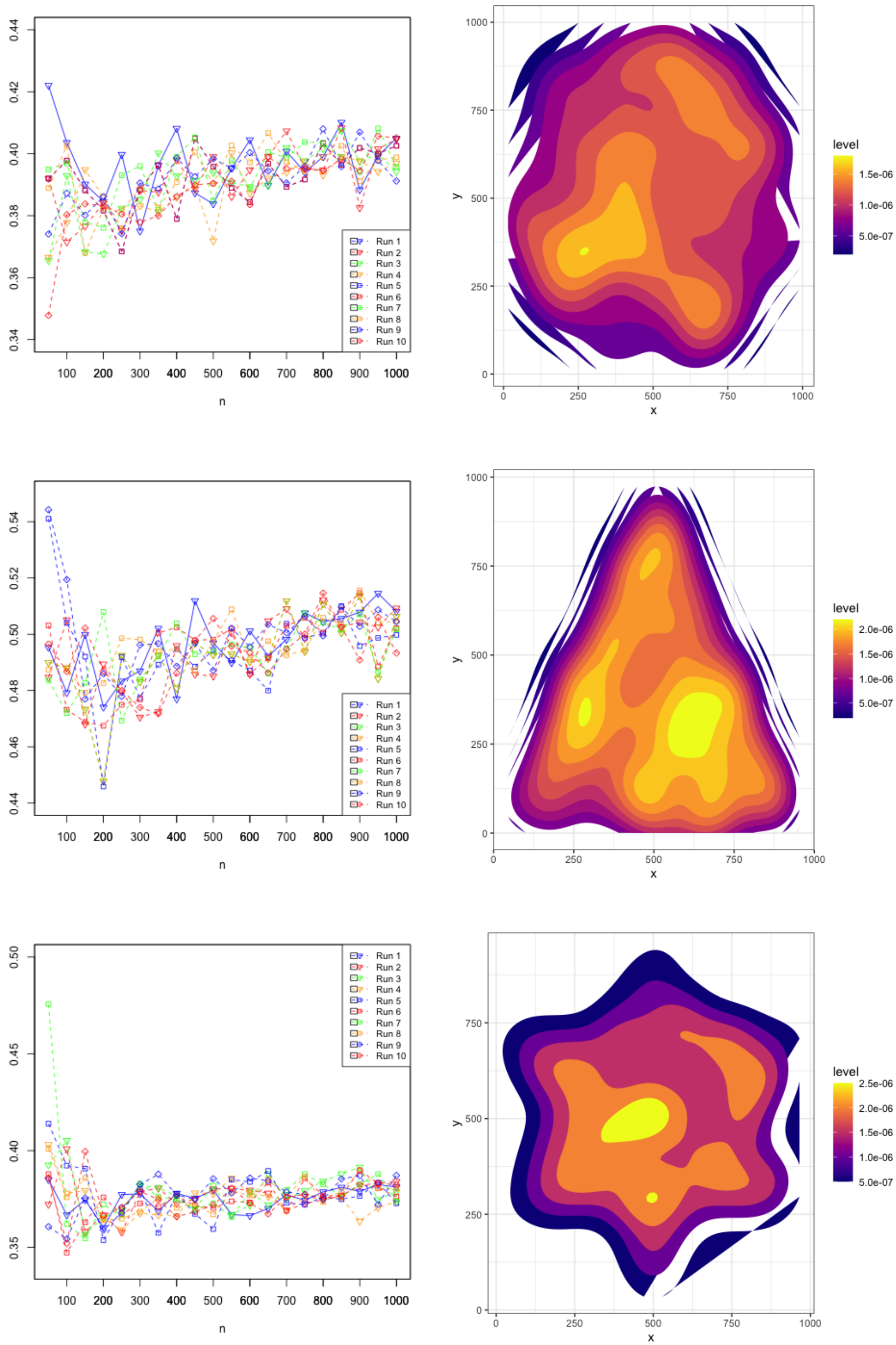

Figure 9.3: Ratio of normal customers' cumulative waiting time to $\sqrt{A}n^{3/2}$ (left) heatmap for corresponding region (right)

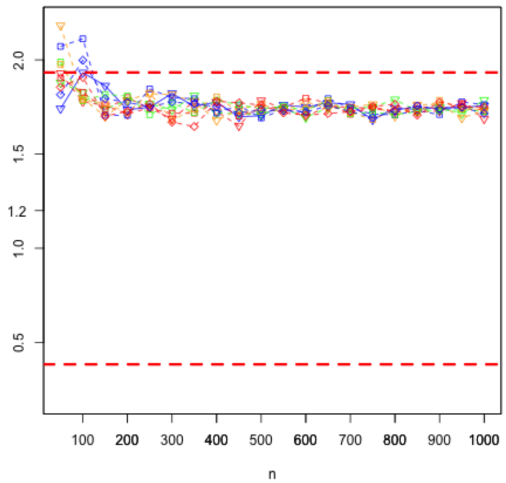

Figure 9.4: Ratio of uniform customers' cumulative waiting time to $\bar{r}n^{3/2}$

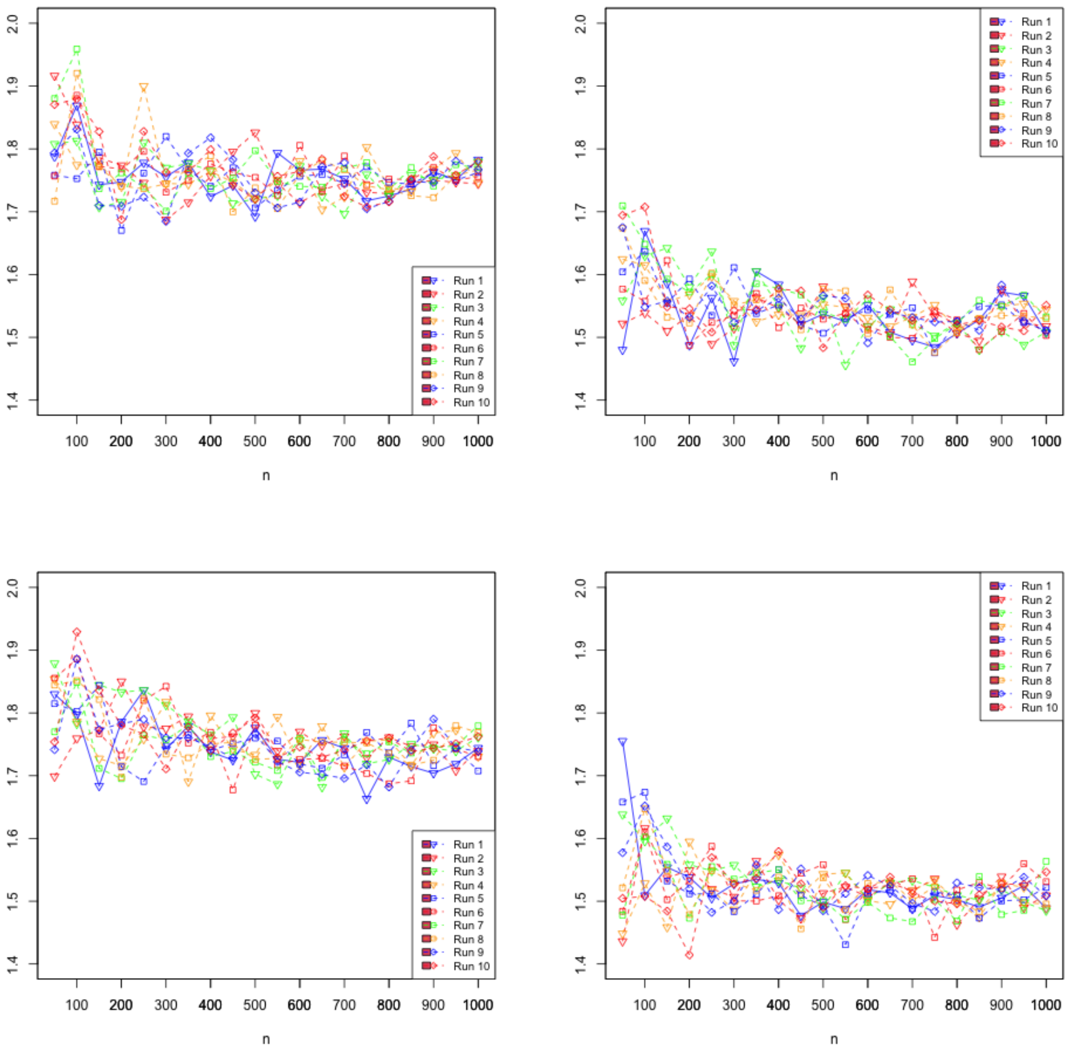

Figure 9.5: Ratio of normal customers' cumulative waiting time to $\bar{r}n^{3/2}$. (Top Left) figure depot is at (500, 500) with standard deviation 500. (Top Right) figure depot is at (750, 750) with standard deviation 500. (Bottom Left) figure depot is at (500, 500) with standard deviation 700. (Bottom Right) figure depot is at (750, 750) with standard deviation 700.

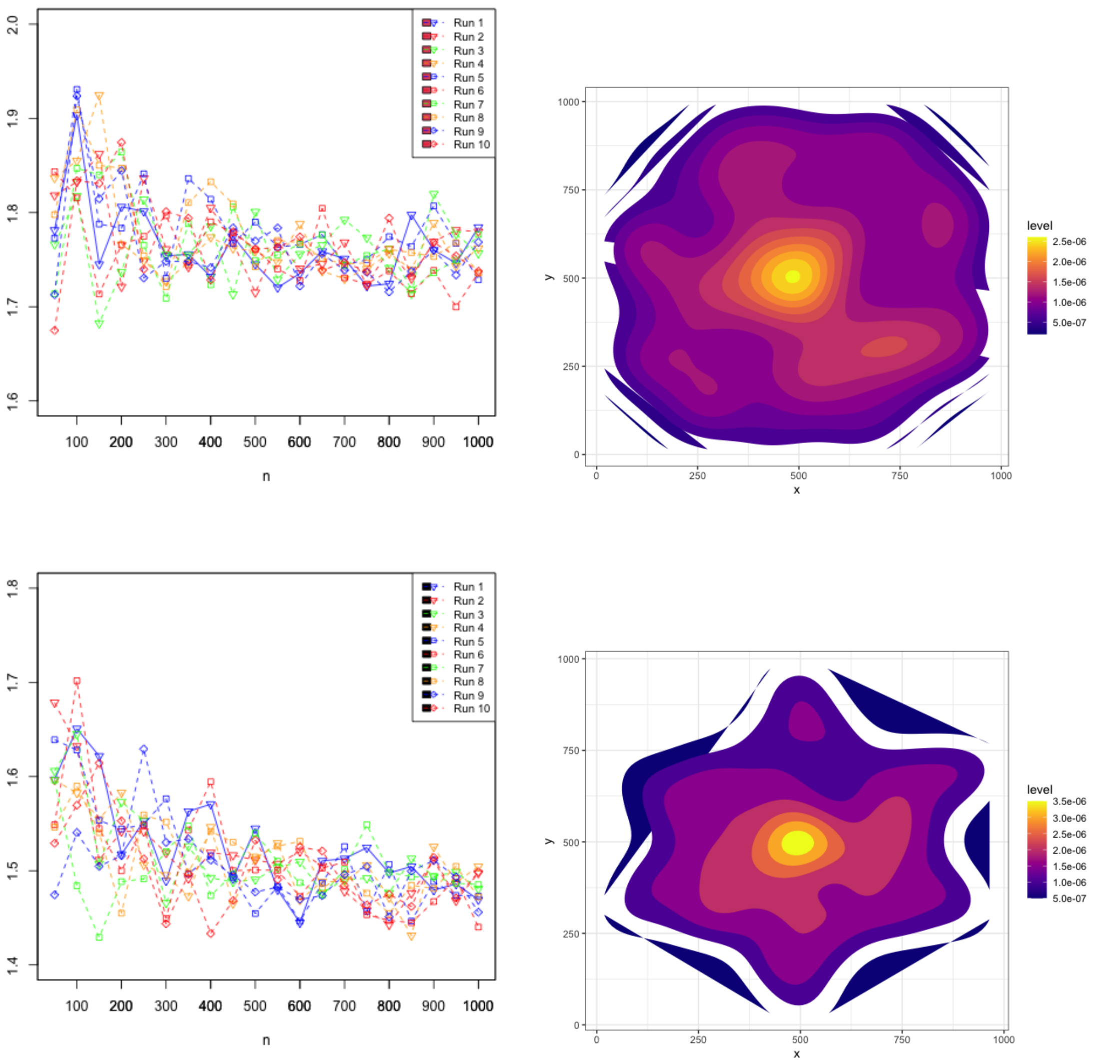

Figure 9.6: Ratio of normal customers' cumulative waiting time to $\bar{r}n^{3/2}$ (left) heatmap for Circle and Pentagram region (right)

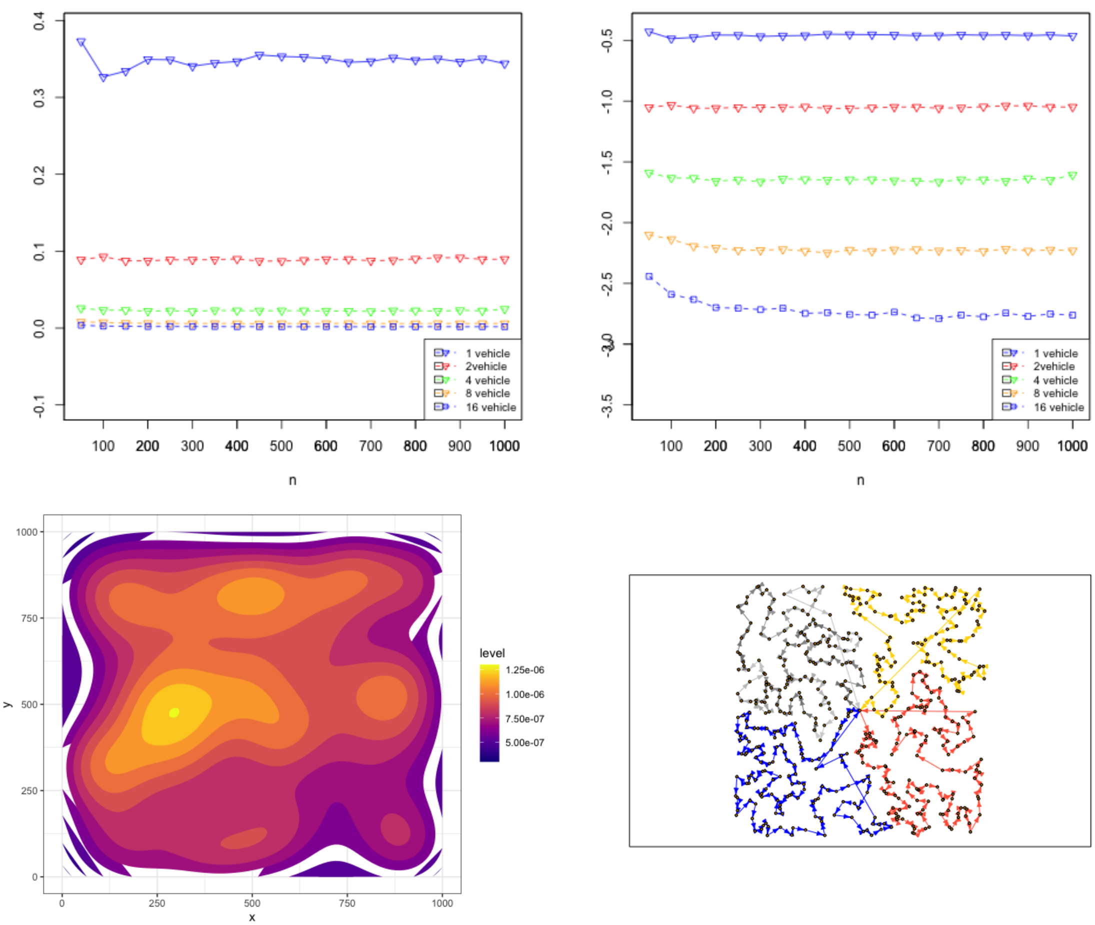

Figure 9.7: (Top Left) Ratio of uniform customers' cumulative waiting time to $\sqrt{A}n^{3/2}$. (Top Right) Log ratio of uniform customers' cumulative waiting time to $\sqrt{A}n^{3/2}$. (Bottom Left) 500 data points' heatmap for rectangular and uniform case. (Bottom Right) Tour schedule for 4 vehicles following m-CTSP.

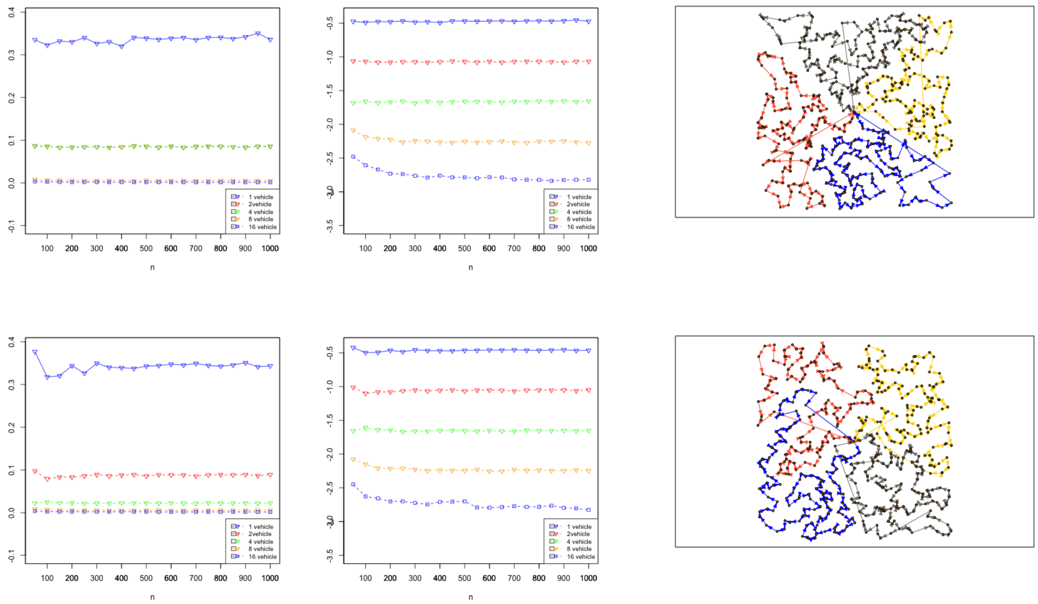

Figure 9.8: (Left) Ratio of uniform customers' cumulative waiting time to $\sqrt{A}n^{3/2}$. (Center) Log ratio of customers', following normal distribution, cumulative waiting time to $\sqrt{A}n^{3/2}$. (Right) Tour schedule for 4 vehicles following m-CTSP. (Top) Depot is at (500, 500) with standard deviation 500. (Bottom) Depot is at (500, 500) with standard deviation 700.

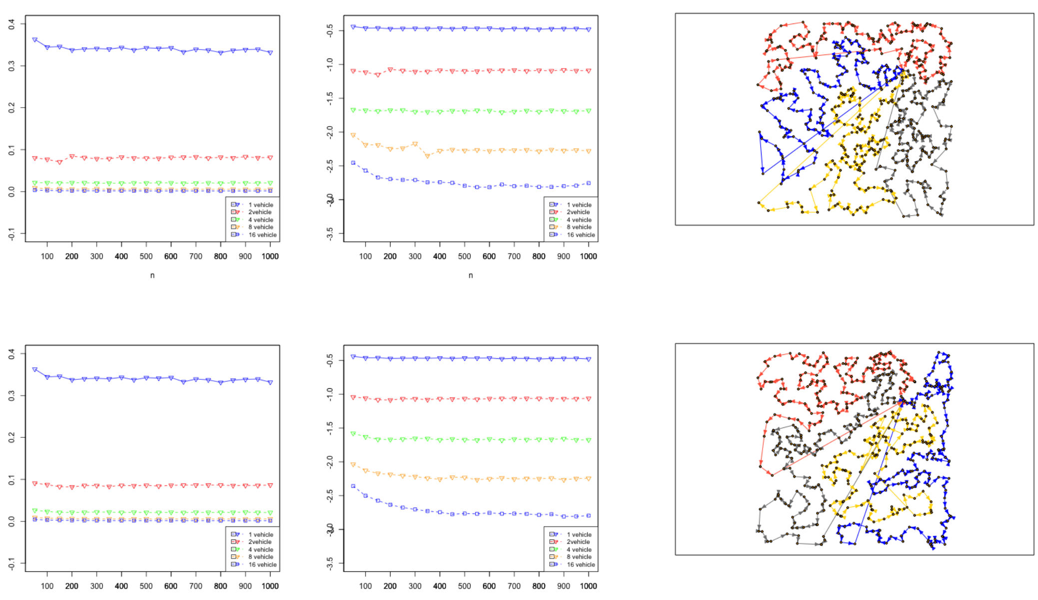

Figure 9.9: (Left) Ratio of uniform customers' cumulative waiting time to $\sqrt{A}n^{3/2}$. (Center) Log ratio of customers', following normal distribution, cumulative waiting time to $\sqrt{A}n^{3/2}$. (Right) Tour schedule for 4 vehicles following m-CTSP. (Top) Depot is at (750, 750) with standard deviation 500. (Bottom) Depot is at (750, 750) with standard deviation 700.

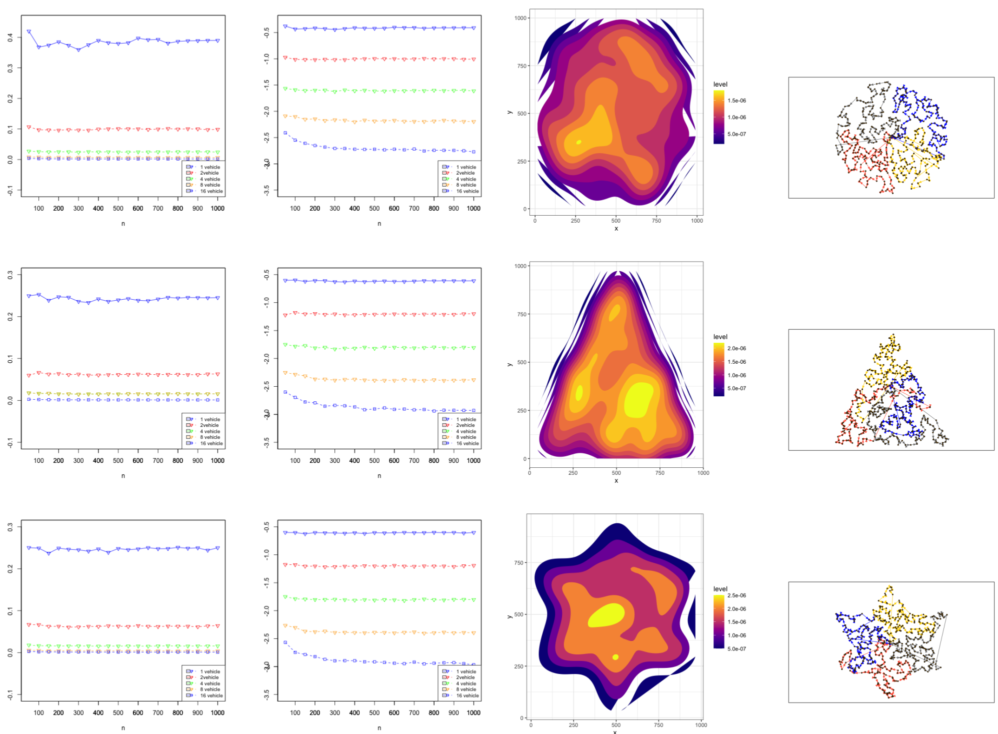

Figure 9.10: (Left 1) Ratio of customers', who follow the normal distribution, cumulative waiting time to $\sqrt{A}n^{3/2}$. (Left 2) Log ratio of customers', following normal distribution, cumulative waiting time to $\sqrt{A}n^{3/2}$. (Left 3) heatmap for different regions. (Left 4) Tour schedule for 4 vehicles following m-CTSP. (Top) Circle Region. (Middle) Triangular Region. (Bottom) Pentagram Region.

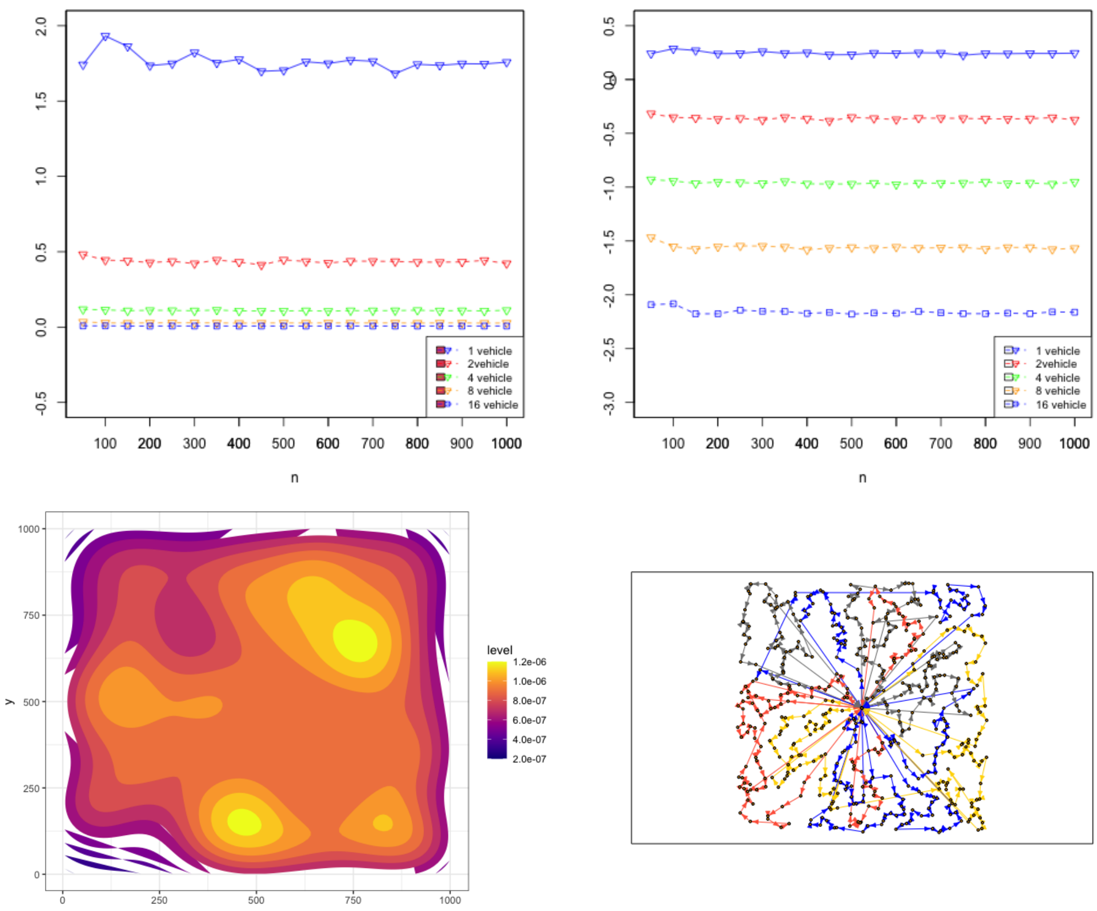

Figure 9.11: (Top Left) Ratio of uniform customers' cumulative waiting time to $\bar{r}n^{3/2}$. (Top Right) Log ratio of uniform customers' cumulative waiting time to $\bar{r}n^{3/2}$. (Bottom Left) 500 data points' heatmap for rectangular and uniform case. (Bottom Right) Tour schedule for 4 vehicles following m-CCVRP.

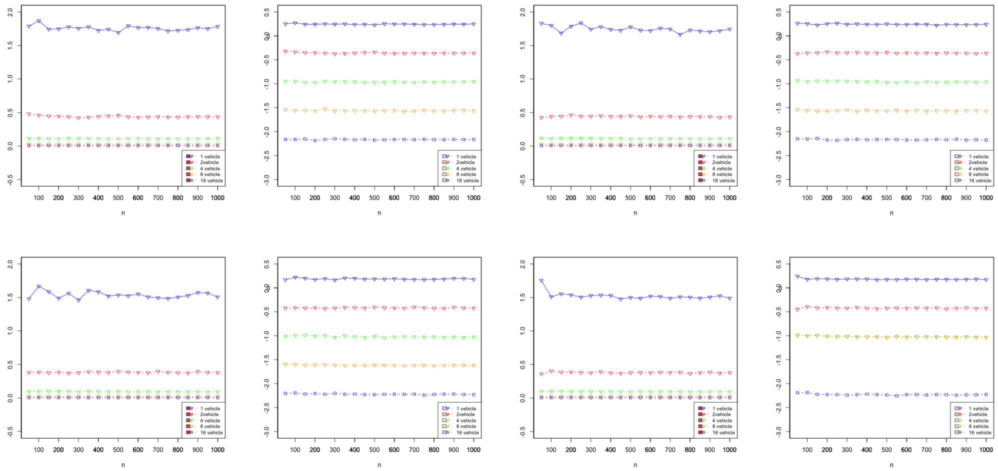

Figure 9.12: (1^{st} and 3^{rd} column) Ratio of customers', sampled from normal distribution. cumulative waiting time to $\bar{r}n^{3/2}$. 1^{st} column with standard deviation 500 and 3^{rd} column with standard deviation 700. (2^{nd} and 4^{th} column) Log ratio of customers' cumulative waiting time to $\bar{r}n^{3/2}$. (Top) Depot is at (500, 500). (Bottom) Depot is at (750, 750).

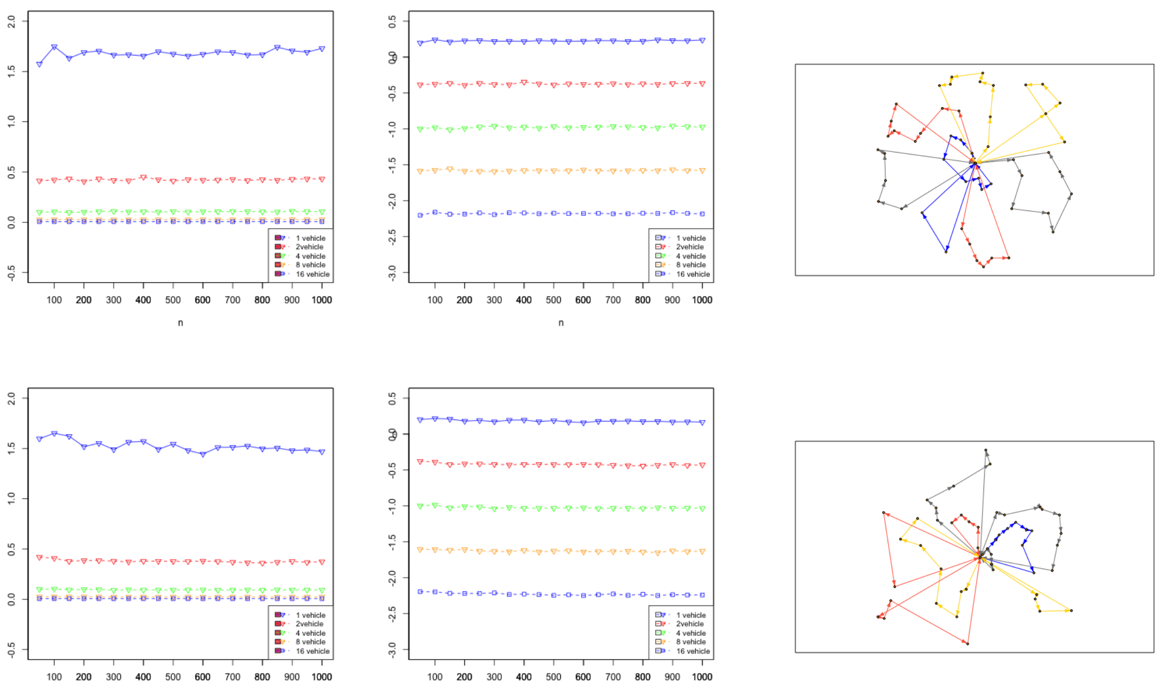

Figure 9.13: (Left) Ratio of customers', who follow the normal distribution, cumulative waiting time to $\bar{r}n^{3/2}$. (Center) Log ratio of customers', following normal distribution, cumulative waiting time to $\bar{r}n^{3/2}$. (Right) Tour schedule for 4 vehicles following m-CCVRP. (Top) Circle Region. (Bottom) Pentagram Region.

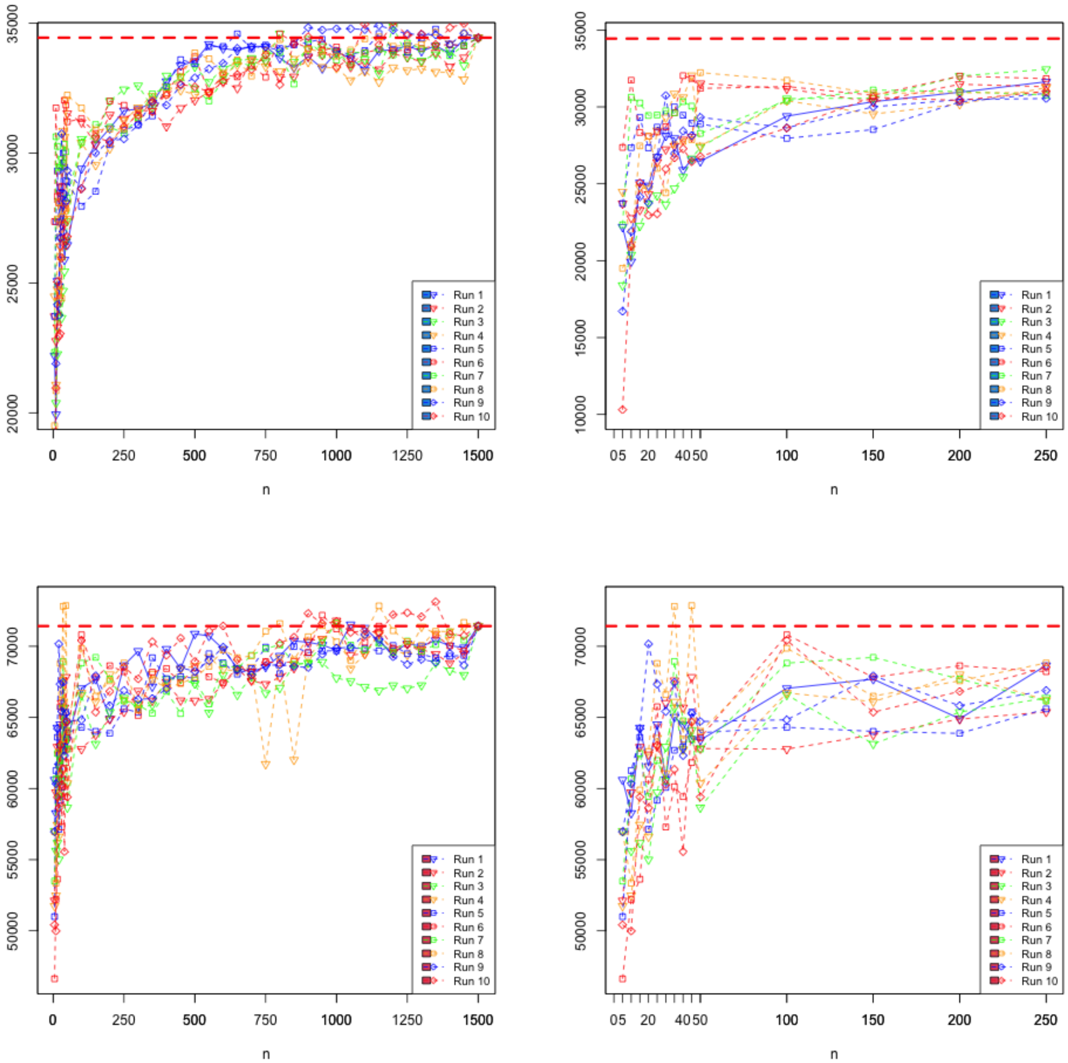

Figure 9.14: Prediction Cost: (Up) CTSP. (Bottom) CCVRP

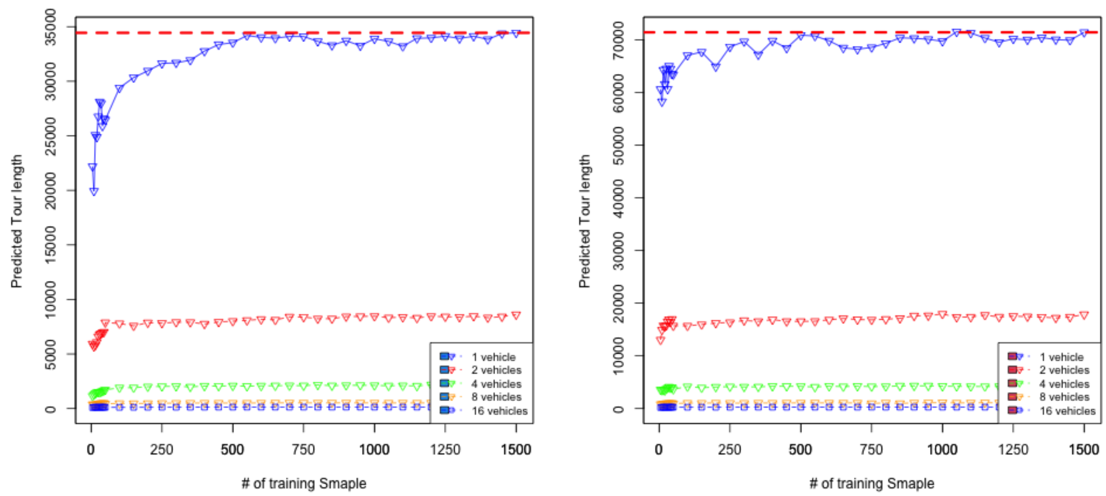

Figure 9.15: Prediction Cost: (Left) m-CTSP. (Right) m-CCVRP

10 Conclusion

We conducted asymptotic analysis of the cumulative tour problems of n points. For this problem, we repeatedly refer to a routing strategy that we call it "most dense to least dense". This rule forces that vehicles travel based on the probability density. Later, our simulated experiments show that the optimal solution closes to tour solution generated via this rule.

We start from Cumulative TSP, where $n \to \infty$ and without vehicle capacity's limit. We prove

$$\lambda_1 \leq \liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2} \int\limits_{x\in\mathcal{R}} \sqrt{f(x)}P(x)dx} \leq \limsup_{n\to\infty} \frac{CumL(X_1,\ldots,X_n)}{n^{3/2} \int\limits_{x\in\mathcal{R}} \sqrt{f(x)}P(x)dx} \leq \beta_2$$

where $f(x)$ is the probability density function and

$$P(x) := \mathrm{Pr}(f(X) \leq f(x)) = \iint\limits_{x':f(x')\leq f(x)} f(x')dx'$$

The results we provide tell us that the cumulative tour length $CumL(X_1,\ldots,X_n) = \mathcal{O}(n^{3/2})$.

In the next section, we added capacity constraints to the vehicles, which suggests a particular scaling factor for modelling vehicle capacity size. Considering capacity $c = \sqrt{n}$, as $n \to \infty$, we found that:

$$\frac{1}{2} \leq \liminf_{n\to\infty} \frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2} \int\limits_{x\in\mathcal{R}} \upsilon(x,k_c\sqrt{n},\lambda_1)\Upsilon(x,k_c\sqrt{n},\lambda_1)dx} \leq \limsup_{n\to\infty} \frac{CumL(X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2} \int\limits_{x\in\mathcal{R}} \upsilon(x,k_c\sqrt{n},\beta_2)\Upsilon(x,k_c\sqrt{n},\beta_2)dx} \leq 1$$

Definitions about $\upsilon(x,k_c\sqrt{n},\_)$ and $\Upsilon(x,k_c\sqrt{n},\_)$ are shown in 20.

The results provide a constant boundary.

Until now, the above results have a strong assumption that we only have a single vehicle, which is unrealistic. From Chapter 7, we extend our analysis from single vehicle to multiple vehicles. We divide the original region into multiple sub-regions and assign one vehicle to travel one sub-region. According to this strategy, for multiple CTSP case, we get

$$\frac{\lambda_1}{m^2} \leq \liminf_{n\to\infty} \frac{CumL(m;X_1,\ldots,X_n)}{n^{3/2} \int\limits_{x\in\mathcal{R}} \sqrt{f(x)}P(x)dx} \leq \limsup_{n\to\infty} \frac{CumL(m;X_1,\ldots,X_n)}{n^{3/2} \int\limits_{x\in\mathcal{R}} \sqrt{f(x)}P(x)dx} \leq \frac{\beta_2}{m^2}$$

For multiple CCVRP case, still, we consider the vehicle size is $\sqrt{n}$, we have

$$\frac{1}{2m^2} \leq \liminf_{n\to\infty} \frac{CumL(m;X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2} \int\limits_{x\in\mathcal{R}} \upsilon(x,k_c\sqrt{n},\lambda_1)\Upsilon(x,k_c\sqrt{n},\lambda_1)dx} \leq \limsup_{n\to\infty} \frac{CumL(m;X_1,\ldots,X_n;k_c\sqrt{n})}{n^{3/2} \int\limits_{x\in\mathcal{R}} \upsilon(x,k_c\sqrt{n},\beta_2)\Upsilon(x,k_c\sqrt{n},\beta_2)dx} \leq \frac{1}{m^2}$$

Both m-CTSP and m-CCVRP cases, the conclusions are similar to single vehicle, but when we double the vehicle size, the boundary will decrease to quarter.

In Chapter 9, we conducted numerical experiments on CTSP and CCVRP, considering single vehicle and multiple vehicles, using artificially generated data and real map data. The numerical results align with our theoretical analysis, we verify that our upper bounds are tight to the optimal solution. In addition, our results are suitable for any region, no matter convex or not. Depot locations may influence final converge result, but their difference can be negligible. In addition, we show that our analysis provides a good prediction when applied to simulations in the Euclidean plan and on road network data, where we are able to predict the total costs within 5 % of the ground truth.

References

[1] Emile HL Aarts, Jan HM Korst, and Peter JM van Laarhoven. “A quantitative analysis of the simulated annealing algorithm: A case study for the traveling salesman problem”. In: *Journal of Statistical Physics* 50 (1988), pp. 187–206.

[2] Foto Afrati, Stavros Cosmadakis, Christos H Papadimitriou, George Papageorgiou, and Nadia Papakostantinou. “The complexity of the travelling repairman problem”. In: *RAIRO-Theoretical Informatics and Applications-Informatique Théorique et Applications* 20.1 (1986), pp. 79–87.

[3] David Aldous and Maxim Krikun. “Percolating paths through random points :” in: (2005), pp. 1–28. URL: http://arxiv.org/abs/math/0509492.

[4] Mircea Ancău. “The optimization of printed circuit board manufacturing by improving the drilling process productivity”. In: *Computers & Industrial Engineering* 55.2 (2008), pp. 279–294.

[5] Sina Ansari, Mehmet Başdere, Xiaopeng Li, Yanfeng Ouyang, and Karen Smilowitz. “Advancements in continuous approximation models for logistics and transportation systems: 1996–2016”. In: *Transportation Research Part B: Methodological* 107 (2018), pp. 229–252.

[6] D Applegate, W Cook, DS Johnson, and NJA Sloane. “Using large-scale computation to estimate the Beardwood-Halton-Hammersley TSP constant”. In: *Presentation at* 42 (2010).

[7] Aaron Archer and David P Williamson. “Faster approximation algorithms for the minimum latency problem”. In: *SODA*. Vol. 3. 2003, pp. 88–96.

[8] Sanjeev Arora and George Karakostas. “Approximation schemes for minimum latency problems”. In: *SIAM Journal on Computing* 32.5 (2003), pp. 1317–1337.

[9] Giorgio Ausiello, Stefano Leonardi, and Alberto Marchetti-Spaccamela. “On salesmen, repairmen, spiders, and other traveling agents”. In: *Italian Conference on Algorithms and Complexity*. Springer. 2000, pp. 1–16.

[10] Roberto Baldacci, Paolo Toth, and Daniele Vigo. “Recent advances in vehicle routing exact algorithms”. In: *4OR* 5.4 (2007), pp. 269–298.

[11] Nikhil Bansal, Avrim Blum, Shuchi Chawla, and Adam Meyerson. “Approximation algorithms for deadline-TSP and vehicle routing with time-windows”. In: *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*. 2004, pp. 166–174.

[12] Jillian Beardwood, J. H. Halton, and J. M. Hammersley. “The shortest path through many points”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 55.4 (1959), pp. 299–327. ISSN: 14698064. DOI: 10.1017/S0305004100034095.

[13] Tolga Bektaş and Gilbert Laporte. “The pollution-routing problem”. In: *Transportation Research Part B: Methodological* 45.8 (2011), pp. 1232–1250.

[14] Patrizia Beraldi, Maria Elena Bruni, Demetrio Laganà, and Roberto Musmanno. “The risk-averse traveling repairman problem with profits”. In: *Soft Computing* 23.9 (2019), pp. 2979–2993.

[15] Livio Bertacco, Lorenzo Brunetta, and Matteo Fischetti. “The linear ordering problem with cumulative costs”. In: *European Journal of Operational Research* 189.3 (2008), pp. 1345–1357.

[16] Lucio Bianco, Aristide Mingozzi, and Salvatore Ricciardelli. “The traveling salesman problem with cumulative costs”. In: *Networks* 23 (1993), pp. 81–91.

[17] Robert G Bland and David F Shallcross. “Large travelling salesman problems arising from experiments in X-ray crystallography: a preliminary report on computation”. In: *Operations Research Letters* 8.3 (1989), pp. 125–128.

[18] Avrim Blum, Prasad Chalasani, Don Coppersmith, Bill Pulleyblank, Prabhakar Raghavan, and Madhu Sudan. “The minimum latency problem”. In: *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*. 1994, pp. 163–171.

[19] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[20] Maria Elena Bruni, Patrizia Beraldi, and Sara Khodaparasti. “A hybrid reactive GRASP heuristic for the risk-averse k-traveling repairman problem with profits”. In: *Computers & Operations Research* 115 (2020), p. 104854.

[21] Ann Melissa Campbell, Dieter Vandenbussche, and William Hermann. “Routing for relief efforts”. In: *Transportation Science* 42.2 (2008), pp. 127–145. ISSN: 15265447. DOI: 10.1287/trsc.1070.0209.

[22] John Gunnar Carlsson. “Continuous Approximation for Selection Routing Problems”. In: *Available at SSRN 4286865* (2021).

[23] John Gunnar Carlsson. “Dividing a territory among several vehicles”. In: *INFORMS Journal on Computing* 24.4 (2012), pp. 565–577.

[24] John Gunnar Carlsson and Mehdi Behroozi. “Worst-case demand distributions in vehicle routing”. In: *European Journal of Operational Research* 256.2 (2017), pp. 462–472.

[25] Ping Chen, Xingye Dong, and Yanchao Niu. “An Iterated Local Search Algorithm for the Cumulative Capacitated Vehicle Routing Problem”. In: *Technology for Education and Learning*. Ed. by Honghua Tan. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 575–581. ISBN: 978-3-642-27711-5.

[26] Nicos Christofides and Samuel Eilon. "Expected Distances in Distribution Problems". In: *OR* 20.4 (Oct. 1969), pp. 437–443. ISSN: 14732858. DOI: 10.2307/3008762.

[27] Jean-François Cordeau, Gilbert Laporte, Martin WP Savelsbergh, and Daniele Vigo. "Vehicle routing". In: *Handbooks in operations research and management science* 14 (2007), pp. 367–428.

[28] Carlos F Daganzo. "The Distance Traveled to Visit N Points with a Maximum of C Stops per Vehicle: An Analytic Model and an Application". eng. In: *Transportation science* 18.4 (1984), pp. 331–350. ISSN: 0041-1655. DOI: 10.1287/trsc.18.4.331.

[29] Carlos F. Daganzo. "The length of tours in zones of different shapes". In: *Transportation Research Part B* 18.2 (1984), pp. 135–145. ISSN: 01912615. DOI: 10.1016/0191-2615(84)90027-4.

[30] G. B. Dantzig and J. H. Ramser. "The Truck Dispatching Problem Author (s): G . B . Dantzig and J . H . Ramser Published by : INFORMS Stable URL : http://www.jstor.org/stable/2627477 REFERENCES Linked references are available on JSTOR for this article : You may need to log in to JSTOR t". In: *Management Science* 6.1 (1959), pp. 80–91. URL: https://www.jstor.org/stable/2627477.

[31] Thijs Dewilde, Dirk Cattrysse, Sofie Coene, Frits CR Spieksma, and Pieter Vansteenwegen. "Heuristics for the traveling repairman problem with profits". In: *Computers & Operations Research* 40.7 (2013), pp. 1700–1707.

[32] Karen M. Douglas and Robbie. M. Sutton. "Kent Academic Repository". In: *European Journal of Social Psychology* 40.2 (2010), pp. 366–374. ISSN: 1939-1285.

[33] Ekrem Duman and I Or. "Precedence constrained TSP arising in printed circuit board assembly". In: *International Journal of Production Research* 42.1 (2004), pp. 67–78.

[34] Samuel Eilon, Carl Donald Tyndale Watson-Gandy, Nicos Christofides, and Richard de Neufville. "Distribution management-mathematical modelling and practical analysis". In: *IEEE Transactions on Systems, Man, and Cybernetics* 6 (1974), p. 589.

[35] C-N Fiechter. "A parallel tabu search algorithm for large traveling salesman problems". In: *Discrete Applied Mathematics* 51.3 (1994), pp. 243–267.

[36] Matteo Fischetti, Gilbert Laporte, and Silvano Martello. "The delivery man problem and cumulative matroids". In: *Operations Research* 41.6 (1993), pp. 1055–1064.

[37] Anna Franceschetti, Ola Jabali, and Gilbert Laporte. "Continuous approximation models in freight distribution management". In: *Top* 25.3 (2017), pp. 413–433. ISSN: 18638279. DOI: 10.1007/s11750-017-0456-1.

[38] A Garcıéa, Javier Tejel, and Pedro Jodrá Esteban. "A note on the travelling repairman problem." In: *Pre-publicaciones del Seminario Matemático" Garcıéa de Galdeano"* 3 (2001), pp. 1–16.

[39] Maria Teresa Godinho, Luis Gouveia, and Pierre Pesneau. "Natural and extended formulations for the time-dependent traveling salesman problem". In: *Discrete Applied Mathematics* 164 (2014), pp. 138–153.

[40] Michel Goemans and Jon Kleinberg. “An improved approximation ratio for the minimum latency problem”. In: *Mathematical Programming* 82.1 (1998), pp. 111–124.

[41] Luis Gouveia and Stefan Voß. “A classification of formulations for the (time-dependent) traveling salesman problem”. In: *European Journal of Operational Research* 83.1 (1995), pp. 69–82.

[42] John Grefenstette, Rajeev Gopal, Brian Rosmaita, and Dirk Van Gucht. “Genetic algorithms for the traveling salesman problem”. In: *Proceedings of the first International Conference on Genetic Algorithms and their Applications*. Psychology Press. 2014, pp. 160–168.

[43] M. Haimovich and A. H.G. Rinnooy Kan. “Bounds and Heuristics for Capacitated Routing Problems.” In: *Mathematics of Operations Research* 10.4 (1985), pp. 527–542. ISSN: 0364765X. DOI: 10.1287/moor.10.4.527.

[44] Michael Held and Richard M Karp. “A dynamic programming approach to sequencing problems”. In: *Journal of the Society for Industrial and Applied mathematics* 10.1 (1962), pp. 196–210.

[45] Brian Kallehauge, Jesper Larsen, and Oli B.G. Madsen. “Lagrangian duality applied to the vehicle routing problem with time windows”. In: *Computers and Operations Research* 33.5 (2006), pp. 1464–1487. ISSN: 03050548. DOI: 10.1016/j.cor.2004.11.002.

[46] Liangjun Ke and Zuren Feng. “A two-phase metaheuristic for the cumulative capacitated vehicle routing problem”. In: *Computers and Operations Research* 40.2 (2013), pp. 633–638. ISSN: 03050548. DOI: 10.1016/j.cor.2012.08.020.

[47] Raphael Kramer, Jean-François Cordeau, and Manuel Iori. “Rich vehicle routing with auxiliary depots and anticipated deliveries: An application to pharmaceutical distribution”. In: *Transportation Research Part E: Logistics and Transportation Review* 129 (2019), pp. 162–174.

[48] Gilbert Laporte. “Fifty years of vehicle routing”. In: *Transportation science* 43.4 (2009), pp. 408–416.

[49] Anh Vu Le, Prabakaran Veerajagadheswar, Phone Thiha Kyaw, Mohan Rajesh Elara, and Nguyen Huu Khanh Nhan. “Coverage Path Planning Using Reinforcement Learning-Based TSP for hTetran—A Polyabolo-Inspired Self-Reconfigurable Tiling Robot”. In: *Sensors* 21.8 (2021), p. 2577.

[50] Shen Lin and Brian W Kernighan. “An effective heuristic algorithm for the traveling-salesman problem”. In: *Operations research* 21.2 (1973), pp. 498–516.

[51] Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.

[52] Jens Lysgaard and Sanne Wøhlk. “A branch-and-cut-and-price algorithm for the cumulative capacitated vehicle routing problem”. In: *European Journal of Operational Research* 236.3 (2014), pp. 800–810. ISSN: 03772217. DOI: 10.1016/j.ejor.2013.08.032.

[53] Glaydston Mattos Ribeiro and Gilbert Laporte. “An adaptive large neighborhood search heuristic for the cumulative capacitated vehicle routing problem”. In: *Computers and Operations Research* 39.3 (2012), pp. 728–735. ISSN: 03050548. DOI: 10.1016/j.cor.2011.05.005.

[54] Isabel Méndez-Dıéaz, Paula Zabala, and Abilio Lucena. “A new formulation for the traveling deliveryman problem”. In: *Discrete applied mathematics* 156.17 (2008), pp. 3223–3237.

[55] Yuichi Nagata. “New EAX crossover for large TSP instances”. In: *Parallel Problem Solving from Nature-PPSN IX: 9th International Conference, Reykjavik, Iceland, September 9-13, 2006, Proceedings*. Springer. 2006, pp. 372–381.

[56] Gordon F Newell. “Dispatching policies for a transportation route”. In: *Transportation Science* 5.1 (1971), pp. 91–105.

[57] Gordon Frank Newell. “Scheduling, location, transportation, and continuum mechanics: some simple approximations to optimization problems”. In: *SIAM Journal on Applied Mathematics* 25.3 (1973), pp. 346–360.

[58] Sandra Ulrich Ngueveu, Christian Prins, and Roberto Wolfler Calvo. “An effective memetic algorithm for the cumulative capacitated vehicle routing problem”. In: *Computers and Operations Research* 37.11 (2010), pp. 1877–1885. ISSN: 03050548. DOI: 10.1016/j.cor.2009.06.014.

[59] Pamela C Nolz, Karl F Doerner, Walter J Gutjahr, and Richard F Hartl. “A bi-objective metaheuristic for disaster relief operation planning”. In: *Advances in multi-objective nature inspired computing*. Springer, 2010, pp. 167–187.

[60] Włodzimierz Ogryczak. “Inequality measures and equitable approaches to location problems”. In: *European Journal of Operational Research* 122.2 (2000), pp. 374–391.

[61] Robert D Plante, Timothy J Lowe, and R Chandrasekaran. “The product matrix traveling salesman problem: an application and solution heuristic”. In: *Operations Research* 35.5 (1987), pp. 772–783.

[62] Masoud Rabbani, Hamed Farrokhi-Asl, and Bahare Asgarian. “Solving a bi-objective location routing problem by a NSGA-II combined with clustering approach: application in waste collection problem”. In: *Journal of Industrial Engineering International* 13.1 (2017), pp. 13–27.

[63] H Donald Ratliff and Arnon S Rosenthal. “Order-picking in a rectangular warehouse: a solvable case of the traveling salesman problem”. In: *Operations research* 31.3 (1983), pp. 507–521.

[64] Gerhard Reinelt. *The traveling salesman: computational solutions for TSP applications*. Vol. 840. Springer, 2003.

[65] Omar Rifki, Thierry Garaix, Christine Solnon, Omar Rifki, Thierry Garaix, Christine Solnon, Omar Rifki, Thierry Garaix, and Christine Solnon. “An asymptotic approximation of the traveling salesman problem with uniform non-overlapping time windows To cite this version : HAL Id : hal-03270043 An asymptotic approximation of the traveling salesman problem with uniform non-overlapping time windows”. In: (2021).

[66] Juan Carlos Rivera, H. Murat Afsar, and Christian Prins. “A multistart iterated local search for the multitrip cumulative capacitated vehicle routing problem”. In: *Computational Optimization and Applications* 61.1 (2015), pp. 159–187. ISSN: 15732894. DOI: 10.1007/s10589-014-9713-5.

[67] Juan Carlos Rivera, H. Murat Afsar, and Christian Prins. “Mathematical formulations and exact algorithm for the multitrip cumulative capacitated single-vehicle routing problem”. In: *European Journal of Operational Research* 249.1 (2016), pp. 93–104. ISSN: 03772217. DOI: 10.1016/j.ejor.2015.08.067.

[68] Amir Salehipour, Kenneth Sörensen, Peter Goos, and Olli Bräysy. “Efficient GRASP+ VND and GRASP+ VNS metaheuristics for the traveling repairman problem”. In: *4or* 9.2 (2011), pp. 189–209.

[69] David Simchi-Levi and Oded Berman. “Minimizing the total flow time of n jobs on a network”. In: *IIE TRANSACTIONS* 23.3 (1991), pp. 236–244.

[70] René Sitters. “The minimum latency problem is NP-hard for weighted trees”. In: *International conference on integer programming and combinatorial optimization*. Springer. 2002, pp. 230–239.

[71] J Michael Steele. *Probability theory and combinatorial optimization*. SIAM, 1997.

[72] J Michael Steele. “Subadditive Euclidean functionals and nonlinear growth in geometric probability”. In: *The Annals of Probability* (1981), pp. 365–376.

[73] Elias M Stein and Rami Shakarchi. *Real analysis*. Princeton University Press, 2009.

[74] M H J Webb. “Cost Functions in the Location of Depots for Multiple-Delivery Journeys”. In: *Journal of the Operational Research Society* 19.3 (1968), pp. 311–320. ISSN: 1476-9360. DOI: 10.1057/jors.1968.74.

[75] Eric W Weisstein. “Traveling Salesman Constants”. In: *https://mathworld. wolfram. com/* (2004).

[76] Bang Ye Wu. “Polynomial time algorithms for some minimum latency problems”. In: *Information Processing Letters* 75.5 (2000), pp. 225–229.

[77] Bang Ye Wu, Zheng-Nan Huang, and Fu-Jie Zhan. “Exact algorithms for the minimum latency problem”. In: *Information Processing Letters* 92.6 (2004), pp. 303–309.

[78] Angel Augusto Agudelo Zapata, Eduardo Giraldo Suarez, and Jairo Alberto Villegas Florez. “Application of vrp techniques to the allocation of resources in an electric power distribution system”. In: *Journal of Computational Science* 35 (2019), pp. 102–109.

Data Management Plan

Products of Research

The data that were collected consist of uniformly sampled points in a geographic region as well as lat/lng pairs sampled from Southern California. All origin-destination distances can be computed using OpenStreetMaps, Google Maps, or HERE Maps.

Data Format and Content

There are no files to share; all experiments can be reproduced using only the contents of this paper.

Data Access and Sharing

The general public can access the data from this paper by repeating the experiments that we conducted, which merely require a random number generator.

Reuse and Redistribution

No restrictions to report.