Robust Solutions for Geographic Resource Allocation Problems

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Mehdi Behroozi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILISOPHY

Adviser: Dr. John Gunnar Carlsson

August, 2016

Acknowledgements

Throughout my PhD life at the University of Minnesota in the past few years I have earned the help and support of several people who kindly contributed to my experience. First of all my sincere appreciation goes to my adviser, Professor John Gunnar Carlsson, to whom I am indebted forever; I could not have done any of this without his guidance. When it comes to showing gratitude towards John, I can not be grateful enough for all great opportunities he has provided me. His brilliant insights, astonishing leadership, impressive hard working attitudes, and amazing skills in coordinating mentorship and friendship, showed me a far and bright horizon to set my professional goals. His pleasurable and intimate research group provided me a unique opportunity to nourish my creativity, productivity, and efficiency. Besides all these, he has always treated me like a friend. I would leave his research group after graduation but his friendship is something that I will never leave and I feel privileged by that deeply in my heart.

I would also like to thank the committee members Professor Shuzhong Zhang, Professor Ravi Janardan, Professor Zizhuo Wang, and Professor Qie He. Either in his amazing course or in several research problems where I had the pleasure and privilege to benefit from his insightful ideas and intuitions, Prof. Zhang has been a great source of inspiration for me. Working with him helped me to realize that what it looks like to be an accomplished academic. I have also learned a lot in the wonderful courses presented by Prof. Janardan and took the advantage of them together with many of his valuable insights in my research. In addition to that, the inspirations that I have attained from him regarding being accurate and disciplined are priceless for me. I was also privileged to take the amazing course of Prof. Wang which was very insightful for me and opened up my eyes to some new domains of research and expanded my expertise. I would also like to truly thank Prof. He for his support, priceless comments, and helpful insights which helped to significantly improve the quality of this research. I also appreciate the tremendous amount of efforts that the faculties of University of Minnesota have put for me to teach me and elaborate me to a higher level of thoughtfulness and understanding.

I would also like to thank my labmates and collaborators: Fan Jia, Raghuveer Devulapalli, Mikael Quist, Ying Li, Siyuan Song, and Xiang Li and other ISyE comrades at the University of Minnesota and Ye Wang and Xiangfei Meng at the University of Southern California for the stimulating discussions, and for all the fun we have had in the last few years. I appreciate the aid and support of all of my friends who have been with me during both happiness and hardship and made my time during my Ph.D. years more interesting and joyful and gave me a wonderful sense of exhilaration.

Last but not the least, I would like to express my deepest gratitude to my parents and my brothers whose unwavering love, unconditional support, and constant encouragement, throughout the ups and downs of my graduate studies and my entire life in general helped me to focus on my research and accomplish my Ph.D. journey.

Dedication

With the deepest feelings in my heart I dedicate this dissertation:

To,

My Parents,

who have held me dear over the years and I hold them dear forever!

Abstract

This thesis describes different ways to use robust optimization concepts and techniques in problems that arise in geographic resource allocation. Many problems in geographic resource allocation deal with uncertainty just like many other domains. Some of them, like the *k-centers problem*, are naturally defined in a *minimax* fashion, and some others can be treated under uncertainty where we seek robust solutions. Most geographic resource allocation problems can be settled in one or more of the categories, such as *location problems, segmentation (partitioning) problems, assignment problems, routing problems,* and *backbone network design problems*. In all such problems there are parameters that can be unknown in practice. It is sensible to define an uncertainty set for the unknown parameter based on some crude knowledge about that unknown parameter and then to treat the uncertainty like some deterministic variability of the values of the parameter, followed by ultimately solving the problem as that parameter is another variable in a higher dimension.

For such problems in geographic resource allocation, we take a robust approach to tackle the uncertainty. Depending on the problem and also geometry of the uncertainty set, the robust optimization model can be tractable or difficult to solve. We deal with both cases in this thesis where we combine elements from *computational geometry, geometric probability theory, vector space optimization,* and *topology*, to either solve the problem to optimality or develop fast algorithms to settle with an approximation solution. We also present a divide and conquer type of approach using geometric partitioning to solve robust optimization problems. In a generic robust optimization problem, if the uncertainty set is an infinite set (which is the case in most practical situations), then we will have an *infinite* or *semi-infinite dimensional optimization problem* since the model will have infinite number of constraints. We describe the drawbacks of the current approaches to solving such problems and their inability to obtain reasonable solutions for some special but common and practical cases, like clustered data, and then we show that our approach makes these problems easy to solve.

Contents

Acknowledgements **i**

Dedication **iii**

Abstract **iv**

List of Figures **viii**

1 Introduction **1**

1.1 An Overview on Optimization in GRA Problems 1

1.2 Optimizing GRA Problems under Uncertainty 6

1.3 Contributions of this Dissertation 7

1.3.1 Robust Network Design and Facility Location 7

1.3.2 Geometric Partitioning and Distributionally Robust Optimization . 9

2 Geometric Partitioning and Robust Ad-hoc Network Design **12**

2.1 Introduction 12

2.1.1 Related work 14

2.1.2 Notational Conventions 15

2.2 Imposing equal area and convexity 15

2.2.1 Analysis of Algorithm 1 18

2.3 The unconstrained version of $(*)$ 22

2.3.1 An approximation algorithm for problem $(*)$ 24

2.3.2 Analysis of Algorithm 4 29

2.4 Problem $(*)$ with a convexity constraint 32

2.5 Computational experiments 36

3 Continuous k-Centers Problem in a Convex Polygon **39**
3.1 An approximation algorithm for the continuous k-centers problem in a convex polygon . . . 40
3.2 Analysis of Algorithm 6 . . . 43

4 Worst-Case Demand Distributions in VRP under Limited Information **47**
4.1 Introduction . . . 47
4.2 Preliminaries . . . 50
4.3 Worst-case temporal demand distributions . . . 52
4.4 Worst-case scenarios for the capacitated VRP . . . 57
4.5 Worst-case distributions with moment information . . . 61
4.5.1 The impact of second-moment information . . . 62
4.5.2 A computational experiment . . . 66

5 The Distributionally Robust TSP: A Data-Driven Model **72**
5.1 Introduction . . . 73
5.1.1 Related work . . . 75
5.1.2 Notational conventions . . . 79
5.2 Preliminaries . . . 80
5.3 Worst-case distributions with Wasserstein distance constraints . . . 82
5.3.1 Comparison with other approaches . . . 83
5.3.2 Structure of the solution to (5.3) . . . 84
5.3.3 Variations and extensions . . . 88
5.4 Solving (5.3) efficiently . . . 90
5.4.1 Districting Service Region . . . 94
5.4.2 Selecting the distance parameter t . . . 95
5.5 Computational experiments . . . 97
5.5.1 Varying values of n . . . 97
5.5.2 A districting experiment with road network data . . . 99

6 Conclusion **110**
6.1 Future Work . . . 111
6.1.1 Extensions for Robust Network Design . . . 111
6.1.2 Extensions for Distributionally Robust VRP . . . 112
6.1.3 Extensions for Distributionally Robust Optimization . . . 112
6.1.4 Robust Optimization for Other Transportation Network Models . . . 112

Bibliography **114**

Appendix A. Additional Proofs **130**

A.1 Proof of Theorem 12 130
A.2 Proof of Theorem 16 132
A.3 Proof of Theorem 13 134
A.4 Proof of Lemma 21 135
A.5 Proof of Theorem 24 136
A.6 Proof of Proposition 25 138
A.7 Proof of Proposition 26 143
A.8 Proof of Theorem 27 143
A.9 Proof of Theorem 29 145
A.10 Proof of Lemma 33 145
A.11 Proof of Theorem 34 147

Appendix B. Additional Analyses **150**

B.1 Probabilistic analysis of the capacitated VRP 150
B.2 Analysis of Algorithm 8 151

List of Figures

1.1 Examples of different backbone networks for a given point set shown over the induced Voronoi partition. 5

1.2 Partitioning a region to maintain a robust communication network between vehicles. An initial region (a rectangle) is given (1.2a) and we partition it into n sub-regions (1.2b), and one vehicle will be assigned to each sub-region. This partition should be constructed in a way that, regardless of the positions of vehicles at any time, the communication network between all the vehicles (obtained by connecting any two vehicles within a distance r of each other) is connected as shown in (1.2c). 8

1.3 An example of TSP tour of *i.i.d.* random points. 10

1.4 An example of the worst case distribution for the length of TSP tour in a unit box. The dots in (1.4a) show an empirical distribution driven by a sample of size 10. The dashed lines show the optimal transport mapping between the ground truth distribution and the empirical distribution. Figure (1.4b) shows a surface plot of the worst case distribution. 11

2.1 Input and output to problem $(*)$. We are given a convex region C and an integer n (which is equal to 19 in this case), and our objective is to partition C into n sub-regions $R_1, \ldots, R_n$; we desire a partition whose *connectivity radius* is as small as possible. For the partition shown in 2.1b, the connectivity radius r is indicated in 2.1c; for any tuple $\mathbf{x} \in \prod_{i=1}^{n} R_i$, the graph $G_r(\mathbf{x})$ is connected. 14

2.2 Input and output to Algorithm 1 for the case where we desire $n = 19$ sub-regions of equal area. In 2.2b, we use a vertical line to divide the region into two pieces whose areas are 9/19 and 10/19 of the original area. In 2.2c, we use vertical lines to further subdivide these two pieces into four pieces whose areas are 4/19, 5/19, 5/19, and 5/19 of the original area (from left to right). In 2.2d, we use vertical and horizontal lines to divide these pieces into even more pieces whose areas are all either 2/19 or 3/19 of the original area. Figure 2.2e shows the output of Algorithm 1, which consists of $n = 19$ sub-regions of equal area. 15

2.3 In 2.3a, we see that the area of the right triangle defined by points $(1/4, 0)$, $(1, 0)$, and $(1/4, h)$ is $3h/8$ and the area of the trapezoid defined by points $(0, 0)$, $(1/4, 0)$, $(1/4, h)$, and $(0, h')$ is $7h/24$. In 2.3b, the area of the trapezoid defined by points $(0, 0)$, $(1/4, 0)$, $(1/4, h_1)$, and $(0, h')$ is $(5h_1 - h_2)/16$, the area of the trapezoid defined by points $(1/4, 0)$, $(3/4, 0)$, $(3/4, h_2)$, and $(1/4, h_1)$ is $(h_1 + h_2)/4$, and the area of the trapezoid defined by points $(3/4, 0)$, $(1, 0)$, $(1, h'')$, and $(3/4, h_2)$ is $(5h_2 - h_1)/16$. 18

2.4 "Shifting" the polygon C onto the horizontal axis. 19

2.5 When the input region C is a right triangle with area 1 whose leftmost angle is $\theta = 10°$, the leftmost sub-region always acts as a bottleneck for the overall connectivity radius. 21

2.6 The output of Algorithm 2 where $(p_1, q_1, p_2, q_2) = (4, 3, 6, 5)$ and ℓ is as indicated. We subdivide the rectangle into two grids, one of which has 4×3 rectangles and one of which has 6×5 rectangles; the width and height of these two grids is determined by ℓ. 25

2.7 An execution of Algorithm 4 with $n = 26$. We will eventually construct a partition consisting of n sub-regions R_i, in which $R_1, \ldots, R_{n-1}$ are each a single point, and R_n their complement. Thus, we will construct $n - 1$ rectangles and place a point in each of their centers. In 2.7a, we orient the imput region C to have its diameter be horizontal. We then compute the axis-aligned bounding box Q, as shown in 2.7b, which happens to have $w/h = 1.62$, which tells us that $p_0 = \lfloor\sqrt{w(n-1)/h}\rfloor = 6$ and $q_0 = \lfloor\sqrt{h(n-1)/w}\rfloor = 3$. Algorithm 4 then constructs various potential point arrangements by positioning two grids alongside each other, four of which are shown in 2.7c through 2.7f; for example, in 2.7d, we break Q into $p = p_0 + 1 = 7$ columns and we have $q = \lfloor(n-1)/p\rfloor = 3$; this gives $s = 4$, so that we break Q into two grids, one of which consists of $(p-s) \times q^{'} = 3 \times 3$ rectangles and the other consisting of $s \times (q+1) = 4 \times 4$ rectangles, thus giving $n - 1$ rectangles in total. Note that, in each of the four placements, there is always one rectangle $\boxdot_j$ whose associated point $c_j^{'}$ is off-center, as specified in Algorithm 3; this is done to ensure that the connectivity radius of the points placed in Algorithm 4 is at most equal to the maximum of the connectivity radii of each of the two halves (and is only added to make the upcoming proof of the approximation ratio easier). It turns out that the arrangement shown in 2.7f has the smallest connectivity radius of all the options of Algorithm 4. Thus, we project all points from 2.7f onto the original polygon C in 2.7g, and 2.7h shows the final output of our algorithm: the first $n - 1 = 25$ sub-regions are the points that are shown, and the n-th sub-region is their complement. 26

2.8 The approximation ratios realized by Algorithm 4, for $n \in \{2, \ldots, 20\}$ and $w \in [\sqrt{2}, \frac{4}{3}\sqrt{n})$. These are obtained by dividing the connectivity radius for the output solution by the maximum of the two lower bounds (2.1) and (2.2). Note that it would have sufficed to compute these ratios for the smaller interval $w \in [\sqrt{2}, \sqrt{n})$, by our previous analysis; we merely compute the ratio for this slightly longer interval for purposes of clarity. 31

2.9 An execution of Algorithm 5 with $n = 59$. We start with a convex polygon, which has been oriented in 2.9a to have its diameter be horizontal, embedded in bounding box Q. Algorithm 5 then constructs various potential convex partitions consisting of either a single grid or a pair of grids; three such partitions are shown in 2.9b through 2.9d. The partition in 2.9b has the smallest connectivity radius, and thus the output partition is shown in 2.9e. Note that there are 11 empty sub-regions, as indicated by the crossed-out boxes. Thus, the final output partition consists of 48 sub-regions. 33

2.10 The approximation ratios realized by Algorithm 5, for $n \in \{2, \ldots, 32\}$ and $w \in [\sqrt{2}, \frac{3}{5}\sqrt{n})$ and the maximum areas of the sub-regions that are output. 35

2.11 The ratio of the connectivity radii of our three algorithms. Figure 2.11a shows the ratio of the maximum radius produced by Algorithm 4 to the maximum radius produced by Algorithm 1. Figure 2.11b shows the same ratios between Algorithms 5 and 1. 38

3.1 The optimal solutions to the 5-center problem in a rectangle, which depend on the dimensions of said rectangle as shown above; the configurations are described formally in [1]. These are necessary for the "very particular edge case" described in the end of Algorithm 6. 42

3.2 Two executions of Algorithm 6 with $k = 31$ and $k = 15$. We start with a convex polygon with a horizontal diameter and a bounding box Q in 3.2a. We then apply the rectangular partitioning Algorithm 2 for various input values, and the best set of inputs is shown in 3.2b; the two arrows indicate that all rectangles have the same diagonal lengths. Figure 3.2c then shows the final output. Figures 3.2d through 3.2f show the same thing for a different polygon and $k = 15$, except that we also encounter the "very particular edge case" described in the end of Algorithm 6 in which we place 5 points according to Figure 3.1. 42

3.3 The approximation ratios realized by Algorithm 6, for $k \in \{6, \ldots, 30\}$ and $w \in [\sqrt{2}, \sqrt{k})$. These are obtained by dividing the objective value of the output solution (i.e. half of the largest diagonal of all the rectangles) by the maximum of the two lower bounds from Theorem 15. 44

3.4 The value of the left-hand side of (3.3), for $q_0 \in \{2, \ldots, 7\}$ and $k \geq \max\{q_0^2, 31\}$. It is entirely straightforward (albeit tedious) to verify algebraically that the ratio is always less than 1.99. 46

4.1 The profit function $f(n)$ as defined in (4.2). 54

4.2 Surface plots of $f^*(\cdot)$ with decreasing values of t from left to right. The depot is indicated by the black square and darker color values correspond to higher densities. 60

4.3 Surface plots of $f^*(\cdot)$ for $\mu = (0.75, 25)$, a depot located at $(0.33, 0.66)$, and varying values of Σ and t. In 4.3a-4.3d, we fix $\Sigma = \begin{pmatrix} 0.075 & -0.02 \\ -0.02 & 0.075 \end{pmatrix}$ and we have $t \in \{0.01, 0.1, 1, 10\}$; in 4.3e-4.3h, we fix $t = 1$ and let Σ be a scalar multiple of $\Sigma_0 = \begin{pmatrix} 0.06 & 0 \\ 0 & 0.095 \end{pmatrix}$. 63

4.4 The worst-case cost OBJ to problem (4.12) for $s \in (0, A/4\pi]$ and $A = 1$. The curve above shows the relationship between the spatial covariance of demand (measured through the single diagonal element s) and the worst-case cost of problem (4.12). 65

4.5 Figure (4.5a) shows the population density of Ramsey County, Minnesota, the mean and covariance matrices of the population density, and four post offices. In practice, of course, we do not know the true demand density, and thus (4.5b) shows the input to our various optimization procedures. 67

4.6 Two power diagram partitions of the service region $\mathcal{R}$. The weight associated with the bottom right sub-region is larger in (4.6a) than in (4.6b). 69

4.7 The plot above shows the workloads that are induced by Algorithm 8 for $t \in (0, 20]$, measured as the *maximum workload* of the four sub-regions R_i (rather than, say, the sum of the workloads in the four sub-regions R_i). For each value of t, we naturally require a separate run of the algorithm (we performed 100 runs overall). To the right of the dashed line (i.e. for $t \geq 1$), we plot the various upper and lower bounds to our problem, and to the left of the dashed line (i.e. for $t < 1$), we plot essentially the same quantities, but multiplied through by t in all cases. This is because all costs (the worst-case cost, the actual cost, and the best possible cost) explode towards infinity as $t \to 0$ and thus we find this to be a more useful measure of the relative costs. 71

5.1 Figure 5.1a shows a Wasserstein distance problem between two univariate distributions μ_1 and μ_2. Figures 5.1b-5.1d show that a Wasserstein mapping can be thought of as an infinite-dimensional generalization of a bipartite matching; here μ_1 and μ_2 are shown in 5.1b and 5.1c, and 5.1d shows a bipartite matching between a large number of samples collected from μ_1 and μ_2. Figures 5.1e-5.1g show an interpretation of a Wasserstein distance problem when μ_1 is a smooth density and μ_2 is atomic. The two distributions are shown in 5.1e, and 5.1f shows the solution to an assignment problem between a large number of samples from μ_1 and the atomic distribution μ_2; a side consequence is that the Lagrange multipliers of this assignment induce a partition of the region $\mathcal{R}$, each of whose cells are associated with one of the elements of μ_2. Figure 5.1g shows this partition together with the atomic distribution μ_2; each cell contains $1/n$ of the mass of the density, which is to be transported to the point contained within it. By using previous results [2], these dashed curves are computationally easy to compute. If we let R_i denote the cell associated with each point x_i in the atomic distribution and f denote the density, then the Wasserstein distance is $\sum_i \iint_{R_i} f(x)\|x - x_i\|\, dA$. 74

5.2 The above point sets in the unit square are extremely clustered and one would expect that their TSP tour should be short. However, because their sample mean and covariance matrix are the same as that of the uniform distribution, any robust methodology that uses only mean and covariance information will fail to recognize the clustering, thereby incurring significant overestimation. 77

5.3 Two views of an example of $f^*(x)$ as described in Theorem 41, where there are $n = 10$ points and the sub-region D_j is the lower quarter of the unit square. At left, the shading represents f^* and the dashed lines indicate the optimal Wasserstein map between f^* and $\hat{f}$; the Dirac delta functions are indicated by the thick black circles in both images. 95

5.4 Threshold values of t for significance levels between 0 and 1, where $n = 100$ points are sampled in the unit square. For example, at $1 - \beta = 0.9$, we have $t = 0.102$; this means that, if 100 samples are drawn from any distribution f in the unit square, then there is at least a 90% probability that $\mathcal{D}(f, \hat{f}) \leq 0.102$. 97

5.5 Figure 5.5a shows the worst-case costs that are computed during the 99×10 executions of our algorithm; the gray plots indicate the results obtained from individual samples and the thick line indicates the sample averages of the 10 trials for fixed n. Figure 5.5b shows the same data set, only we plot the worst-case costs as a function of the Wasserstein distance threshold t rather than a function of n; the gray points indicate individual experiments and the dark points again indicate the sample averages of the 10 trials for fixed n. For purposes of comparison, Figure 5.5c shows the estimates of $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$ obtained when one uses a uniform kernel density estimator; that is, if we draw n samples $x_1, \ldots, x_n$ from $\bar{f}$, then we define an estimator $\tilde{f}$ by setting $\tilde{f}(x) = \frac{1}{C} \sum_i \mathbb{1}(\|x - x_i\| \leq r)$, where r is a "bandwidth" parameter and C is a normalization constant. As in the preceding two figures, the gray plots indicate the results from individual samples, the thick lines indicate sample averages of 10 trials for fixed n, and the dashed line indicates the true value of $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$; furthermore, as indicated, we used 5 different values of r between 0.03 and 0.3; note that the estimate $\iint_{\mathcal{R}} \sqrt{\tilde{f}(x)}\, dA$ is highly sensitive to the choice of r. 98

5.6 Locations of 1704 crime reports filed in Los Angeles County during the first week of July, obtained from the website [3]. 99

5.7 The lengths of the TSP tours of n points sampled in Los Angeles County, where point-to-point distances are induced by a road network. In all three diagrams, the upper plot corresponds to uniformly distributed points and the lower plot corresponds to points that are sampled from the CrimeMapping.com website [3]. For each value of $n \in \{1, \ldots, 500\}$ and for each of the two sampling strategies (uniform or non-uniform), we perform 10 independent experiments in which n points are sampled and their TSP tour is calculated using Concorde [4]. Figure 5.7a shows the tour lengths (the thin lines) together with the best fit of these tour lengths to a curve of the form $C\sqrt{n}$, where we have $C \approx 159$ for uniformly sampled points and $C \approx 64$ for non-uniformly distributed points. Figure 5.7b shows a close-up of this plot for $n \in \{1, \ldots, 21\}$, where we can see that the fitted curve underestimates the true tour lengths when n is small. A better fit for these small values of n, as shown in Figure 5.7c, is to set $C \approx 185$ for uniformly sampled points and $C \approx 77$ for non-uniformly distributed points (as an aside, for small values of n, it is clearly common sense to fit a curve of the form $C\sqrt{n-1}$, since the TSP tour of a single point has length 0). This establishes that, provided one has an approximate estimate of the number of points n, the square root approximation is indeed a reasonable one. 100

5.8 The TSP tour of a collection of points uniformly sampled in the unit square, with varying metrics depending on quadrants. The dashed lines in the paths in the upper left quadrant correspond to the smaller of the two directions (horizontal or vertical) between points, which is relevant because the ℓ_∞ distance is used. 102

5.9 The shading in Figure 5.9a indicates the values of β_i associated with each of the square patches (lighter patches correspond to higher values of β_i). Figure 5.9b is a histogram of these same values; note that a handful of these values are actually *lower* than the current estimate [5] of the Euclidean TSP coefficient $\beta \approx 0.7124$; this appears to be a combination of statistical noise and an artifact of the Google Maps API [6] that was used to compute driving distances; for example, two locations belonging to the same venue (e.g., two stores on opposite ends of a large shopping mall) report a driving distance of 0. 103

5.10 Figure 5.10a is a Voronoi partition, that is, a power diagram with all weights w_i equal (each district consists of those points that are closer to their associated landmark point than the others). Figures 5.10b and 5.10c present the two power diagrams obtained by increasing and decreasing the weight associated with the shaded cell. 104
5.11 Locations of 4 police stations associated with the 4 largest cities in Los Angeles County. 105
5.12 The power diagram districts obtained according to the four partitioning criteria. 107
5.13 The workloads associated with the four districts and the four partitioning criteria from Figure 5.12, interpreted as follows: the colors of the various plots correspond to the workloads in the district of the same color (e.g. the magenta plots correspond to the magenta district, which belongs to Santa Clarita). The left set of plots corresponds to the workloads that result when we design districts according to the Wasserstein partitioning criterion: more precisely, for each value of t in the range shown, we construct a weight vector $\mathbf{w}^*$ such that the worst-case workloads in problem (5.16) are all equal. Thus, different values of t correspond to different values of $\mathbf{w}^*$, and thereby different partitions. The left-hand plot shows the true workloads for each of the districts as t varies; the maximum value of t of 18.3 km is calculated according to equation (5.18) with a 90% confidence level with $n = 50$ samples. The three sets of stem plots on the right show the workloads in each of the four districts that are obtained when one partitions according to the first three criteria of Section 5.5.2. The Wasserstein partitioning criterion consistently produces districts whose workloads are more balanced than those of the other three criteria, even though the Wasserstein partitions are constructed using a small number of samples, whereas the other three methods are actually permitted to make complete use of all 1704 sample points (the mean-covariance partitioning scheme uses exact knowledge of the mean and covariance of the data points). Surprisingly, we found that the the mean-covariance robust partitioning method is by far the *worst* of the three. . . . 108
5.14 1000 uniformly sampled points in Los Angeles County. 108
5.15 The power diagram districts obtained according to the four partitioning criteria, and the workloads that result therein, for the case where all samples are uniformly distributed. In this case, all four partitioning criteria perform more or less the same. 109

A.1 Figure A.1a shows a triangulation of a convex polygon; the origin is common to all triangles. Figure A.1b shows a collection of circular sectors, each having radius $r_0 = \max_{x \in \mathcal{R}} \|x\|$ and area equal to the triangle that they intersect. Figure A.1c shows a simple shearing operation that makes a triangle isosceles, and Figure A.1d shows the action of the map $M(\cdot, \cdot)$. 138

A.2 A division of the unit square $\mathcal{R}$ into $N^2 = 16$ grid cells. The larger square S_{10} has side length $1/N + 2/N^3$and contains s_{10}. 148

Chapter 1

Introduction

Introduction

This dissertation is devoted to the study of *geographic resource allocation (GRA) problems*. In such a problem, our goal is to minimize or maximize an objective function that is defined over a geographic region. The distinguishing attribute of such problems is that spatial properties such as distance, shape, perimeter, area, diameter, connectedness, fatness, convexity, or star-convexity will appear in these problems either in the objective function or as a constraint. The concept of geography is a fundamental one in many domains such as transportation, facility location, supply chain management, surveillance, reconnaissance, and territory division, among many others; indeed, the fields of geometry and optimization have enjoyed a long history of cross-pollination, from the statement of *Dido's isoperimetric problem* by the ancient Greeks [7], to Newton's solution of the *Brachistochrone problem* [8], to the (much more recent) use of convexity properties towards linear-time algorithms for low-dimensional linear programs [9]. In order to solve such problems, it is necessary to adopt an interdisciplinary approach that combines tools from optimization theory and computational geometry, and sometimes even real and complex analysis and topology.

1.1 An Overview on Optimization in GRA Problems

Most geographic resource allocation problems can be said to fall in one (or more) of the following categories:

1. Location Problems
2. Segmentation (Partitioning) Problems
3. Assignment Problems

4. Routing Problems

5. Backbone Network Design Problems

In all these categories, the concepts and techniques of convex optimization and computational geometry are highly interconnected.

Location Problems In *location problems* we aim to find optimal locations for some facilities in a region with respect to some constraints. In other words, we intend to find the best locations for, say, k or p, facilities, in order to minimize (maximize) an objective function. Some of the very well-studied problems in this category include:

- *k-medians problems*: Find k locations to place facilities so as to minimize the average (sum of) distances between demand points and their nearest facility.
- *k-means problems*: Find k locations to place facilities so as to minimize the average (sum of) squared distances between demand points and their nearest facility.
- *k-centers problems*: Find k locations to place facilities so as to minimize the maximum distance between demand points and their nearest facility.
- *p-dispersion problems*: Find p locations to place facilities so as to maximize the minimum distance between any pair of facilities.

The k-medians problem has roots back into the classical *Weber problem* [10] and is proven to be NP-hard [11]. Several approximation algorithms and PTASes have been developed for this problem; see for example [12, 13, 14, 15, 16, 17].

The difference between k-means problems and k-medians problems is in the cost function they use, i.e. the way each one measures point distances (sums of distances versus sums of squared distances). The k-means problem, *a.k.a.* the k-means clustering problem, has been studied in many fields including geometry and machine learning. In machine learning, *k-means clustering algorithms* are powerful tools in clustering data [18]. The k-means problem is NP-hard even when the point set lies in two dimensions [19, 20]. The venerated Lloyd's algorithm [21] is the most commonly used heuristic for solving this problem, although there are also several approximation algorithms such as [22, 23] and PTASes as in [24, 25] for this problem.

The k-centers problem, introduced by Hakimi [26] is known to be NP-hard [27, 28]. Furthermore it is shown that for the k-centers problem with a distance function satisfying a triangle inequality, there does not exist an approximation algorithm with an approximation

factor of $2-\epsilon$ for any $\epsilon > 0$, unless P = NP and also that one must assume triangle inequality in order to be able to approximate the optimal solution unless P=NP [29]. A simple greedy factor-2 approximation algorithm is presented in [30]. Another factor-2 approximation algorithm for solving this problem is developed by Hochbaum and Shmoys [31] and is summarized in Chapter 5 of [28]. In Chapter 3 of this dissertation, we duscuss more about a particular version of the k-centers problem.

The p-dispersion problem has a duality relationship with the $(p-1)$-centers problem [32] and has several variants [33, 34, 35]; the version stated above is the most common version and conventionally represents the problem. This problem is also proven to be NP-hard [36, 37]. There exist several approximation algorithms for different versions of p-dispersion problem including [38, 39], for example.

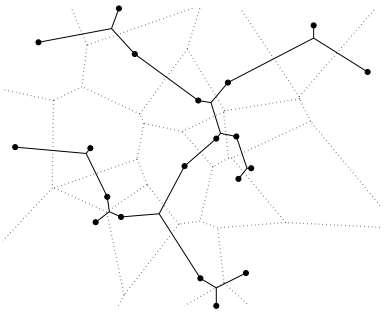
Segmentation (Partitioning) Problems In segmentation problems we are given a territory and we intend to partition that territory into some sub-regions in order to minimize (maximize) a cost (utility) function. The input in this problem is a geographic region R and possibly a distribution f. The distribution f can represent for example population density, distribution of crimes or other incidents, distribution of a resource or a natural phenomena, or distribution of locations of likely targets. There is also a set of given points in R that may represent facilities, resources, distribution depots, or delivery centers, for example. We want to partition R into n disjoint sub-regions $R_1, ..., R_n$ such that a certain objective function composed by the distribution f and some other features like cost or utility is optimized. We may also need to enforce some shape constraints, such as requiring all sub-regions to be simply connected (not having holes), connected, fat, convex, or star-convex, for example. Another important constraint here could be balancing the workload (resources, population, or area, for example) among the sub-regions. If $\boldsymbol{p}$ denotes the given set of points and x represents any point in R and we define a cost function $c(x, \boldsymbol{p})$, then a typical objective function can be the expected cost over the region R that can be shown with $\int_R f(x)c(x, \boldsymbol{p})\, dA$. This problem is addressed in [40, 41, 42, 2, 43, 44, 45], among many others.

Assignment Problems A problem that often accompanies the segmentation problem is the assignment problem. In an assignment problem, we are given a set of clients and a limited amount of one or multiple resources, and we want to allocate the resources to clients so that a certain objective function is minimized, and at the same time, we make this allocation as evenly as possible or as close as possible to a given quota. A natural strategy for this problem would be using the methods of partitioning used in the segmentation problems

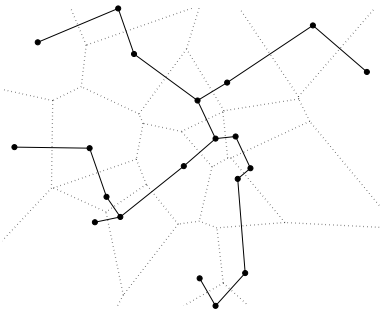
to divide the total amount of resources into different pieces and give each client a piece. This problem is in the company of [46, 47, 48, 49, 50], for example.

Routing Problems Many problems with an underlying geographic component can be considered as a *routing problem*. The goal of a routing problem is to select a path, or collection of paths, going through a set of points that could be demand points or clients or facilities, etc, so as to minimize total cost or total traversed distance. Some of these problems are computationally easy to solve like *shortest path problem*, *Canadian traveller problem*, or the *Chinese postman problem* and some of them are considerably more difficult, such as the *traveling salesman problem* (TSP) or *vehicle routing problem* (VRP) and its variants, and the *k-chinese postman problem*. In some cases we are interested in solving these problems for specific instances and in some cases we are interested in asymptotic analysis when the problem dimensions become large. As an iconic example for this category of problems, in TSP we are given a set of points and cost function defined on any pair of these points, which is usually the distance between two points, and we are asked to find a tour that starting and finishing at a specified point visits all other points exactly once so as to minimize the total cost incurred by traversing this tour. There is a rich literature about routing problems; some of the works closer to our interest include [51, 5, 52, 16, 53, 54].

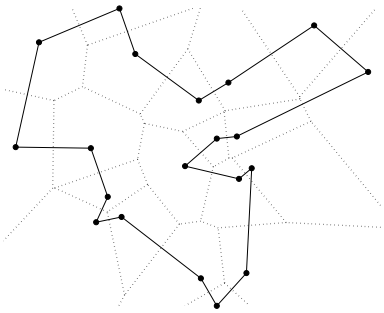
Backbone Network Design Problems In some GRA problems we have a system which is composed of a set of facilities, distribution centers, depots (possibly having multiple "tiers" or levels), and we need to build a network among the components of the system. The role of this network is to provide a support mechanism for the components of the system to ensure they will keep working continuously in fulfilling their mission. Examples of this support could be: providing a communication channel between components, transportation of goods from one facility to another to satisfy excess of demand, delivering products to the franchises of a company, transportation of materials and products in a supply chain, and etc. Such networks are usually designed one layer above the clients or demand points and they have a supportive role to make sure the products or services reach the customers without any interruption, and for this reason by an analogy of human neural network they are called backbone networks. Examples of such networks are spanning trees, Steiner trees, hub-and-spoke networks, TSP tours (Hamiltonian cycles), star networks, and so on. An illustration of these network structures is shown in Figure 1.1. Various types of such problems have been considered in [55, 56, 44, 46, 57].



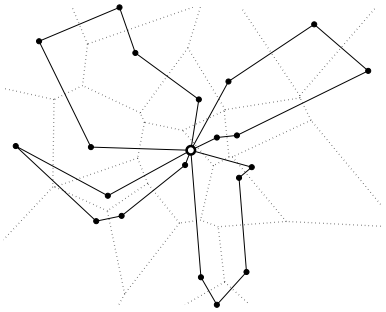
(a) Steiner Tree



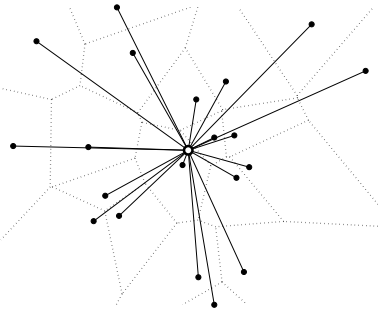
(b) MST



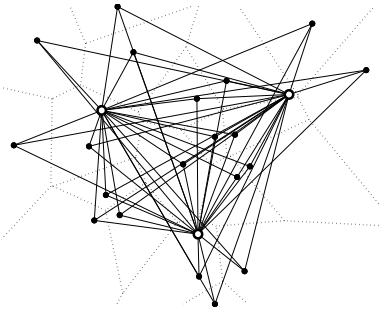
(c) TSP Tour (Hamiltonian Cycle)



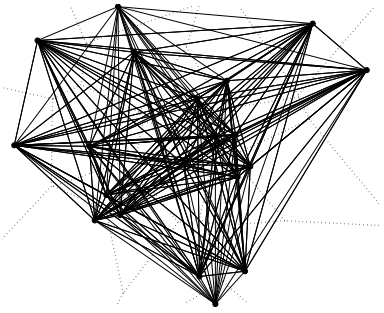
(d) VRP Tours



(e) Star Shape



(f) Complete Bipartite Graph



(g) Complete Graph

Figure 1.1: Examples of different backbone networks for a given point set shown over the induced Voronoi partition.

1.2 Optimizing GRA Problems under Uncertainty

In many problems with a geometric or geographic nature, there is some *uncertainty* about some of the parameters or characteristics of the problem. In such situations it is natural to desire some sort of *robustness* against the uncertainty. There are several ways to handle uncertainty within an optimization framework. One could simply ignore the uncertainty by choosing a nominal value for those unknown parameters. This approach usually requires some sensitivity analysis after solving the problem. Another drawback of this method is that a nominal value is not a good representative for an unknown parameter in most cases for the same reason that average is not an efficient representative of a the values of a random variable, and can lead to a class of systematic errors which is called the *flaw of averages* by noted consultant Sam L. Savage [58]. This is actually nothing but a century old Jensen's inequality. If our function is not linear and is actually strictly convex (concave) in the unknown parameter, which happens in many situations, and we use the average of our best guesses for the value of the parameter, then the evaluation of the function at the average value is less (greater) than the average of evaluations of the function at each of the single guesses. This shows by choosing nominal value for the unknown parameter we may incur a possibly huge error with respect to the ground-truth value of the parameter and increase the risk of our decision.

Another common way to handle uncertainty is to assume that the uncertainty has a stochastic nature, and hence apply techniques of *stochastic programming*. This approach relies on two major assumptions, namely the stochastic nature of the uncertainty and our knowledge about the true distribution of the uncertain parameters, neither of which is necessarily the case in reality.

In contrast to traditional methods, *robust optimization* is able to work with a vague knowledge about the unknown parameter to handle uncertainty. Consider an optimization problem with only one uncertain parameter. In a robust approach we only need to assume that the unknown parameter belongs to a given *uncertainty set*. Then, we can represent the uncertainty as deterministic variability in the value of the parameter itself and/or in the solution(s) of the problem. For a detailed discussion on robust optimization see the venerated paper of Ben-Tal and Nemirovsky [59].

We can already see, by definition, some essence of robustness in some of the problems with geometric or geographic nature. For instance, the k-centers problem is a robust optimization of facility locations since it is a *minimax* problem with uncertainty over the location of demand points.

Generally, in GRA problems, if we encounter uncertainty about some parameters, robust

optimization can serve a useful purpose; see for example [60, 61, 62, 63, 64, 65, 66]. On the other hand, if the problem is defined over a contiguous geographic region, it is highly possible that the resulting problem may have an infinite or semi-infinite dimensional formulation. Such problems are sometimes tractable using techniques of geometric partitioning, applied in [2, 67], for example.

1.3 Contributions of this Dissertation

Two main contributions of our research is the use of partitioning techniques in robust optimization and conversely to use robust optimization concepts and techniques in problems that arise in geographic resource allocation.

In the following two sections, we present three new geometric problems and we show new robust approaches to solve them. These two sections establish a basis for our discussions in the next four chapters, which are assigned to our achieved results throughout this research.

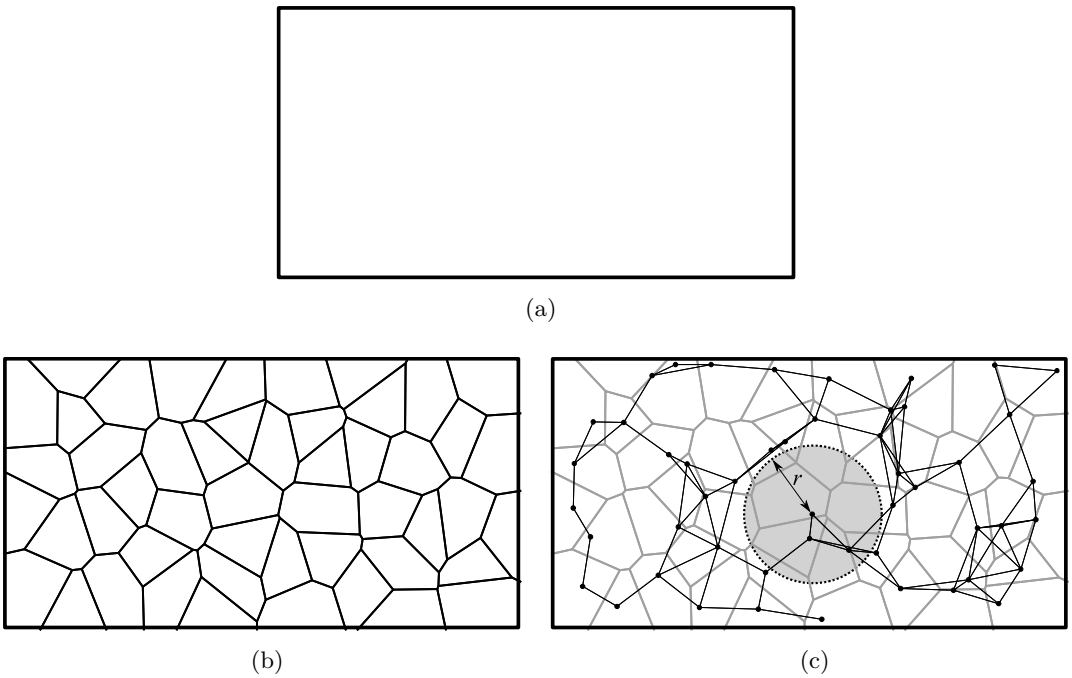
1.3.1 Robust Network Design and Facility Location

It seems that the term *robust network design* first appeared in [68], although work in this field remained without any connection to robust optimization and [69] appears to be the first paper establishing that connection. There exists a large body of work on this subject such as [70, 71, 72, 73, 74, 75] among others ranging from spanning tree networks and routing networks to other hub-and-spoke networks and from virtual private networks (VPNs) to telecommunication networks.

Backbone network design is an extremely broad field. Many real-world problems are concerned with coordinating vehicles, agents, robots, UAVs, or drones via an underlying backbone communications network. Examples of such applications are police surveillance missions, military reconnaissance missions, military drones operations, retailer delivery systems, mobile wireless sensor networks, cooperative robotics, and modular robotics among so many others.

In designing such networks, a desired property is to maintain "connectivity" of communication at any time. In other words, if we say two mobile objects can communicate with each other if their distance is less than some radius r, we want to have all the mobile objects connected to each other in a way that any two of them could communicate with each other (possibly through other intermediate objects) at any time.

At the same time, a small communication radius r is beneficial since it helps the quality of the communications and reduces the amount of noise. The optimization problem of



(a) (b) (c)

Figure 1.2: Partitioning a region to maintain a robust communication network between vehicles. An initial region (a rectangle) is given (1.2a) and we partition it into n sub-regions (1.2b), and one vehicle will be assigned to each sub-region. This partition should be constructed in a way that, regardless of the positions of vehicles at any time, the communication network between all the vehicles (obtained by connecting any two vehicles within a distance r of each other) is connected as shown in (1.2c).

designing such network is indeed a *robust* optimization problem since the mobile objects have to be connected regardless of their current or future position.

Suppose that this network with n mobile objects has to cover a convex region to do one of the above operations. In order to ensure that the entire region is covered we deploy a partitioning strategy in which we partition the region into n sub-regions and assign one mobile object to each sub-region, so each object will be only responsible to cover its own sub-region. An example of such connected network is shown in Figure 1.2. In Chapter 2 we discuss the robust network design in details and we present three approximation algorithms for different cases with different set of assumptions. These algorithms first appeared in our article [76].

Based on the results and algorithms developed in Chapter 2, we develop a new approximation algorithm for a restricted version of the continuous k-centers problem where the underlying region is a convex polygon, which is presented in Chapter 3 and first appeared in our article [76]. This problem is introduced by Suzuki and Drezner [77] and is equivalent to covering the convex polygon with k identical circles with the smallest possible radius. The

authors of [78, 79] considered a 2-centers problem over a convex polygon, and the paper [80] presents a factor 1.89 approximation algorithm for an even more restricted problem where the centers have to be on the *boundary* of a convex polygon. We conjecture that, with a slight modification, we can use the same algorithm to solve the k-medians problem approximately. Our k-centers algorithm might be also useful for solving continuous p-dispersion problem because of its duality relationship with the $(p-1)$-centers problem [32].

1.3.2 Geometric Partitioning and Distributionally Robust Optimization

A *distributionally robust optimization problem* is a problem in which one must make a decision that is optimal when considered against a set of probability distributions. By way of comparison, in stochastic optimization, one is concerned with problems of the form

$$\underset{\boldsymbol{x}}{\text{minimize}}\, F(\boldsymbol{x}, \mu)$$

where μ is a given probability measure and $F(\cdot, \cdot)$ is an objective function, which typically might involve an expectation, quantile, or supremum, for example. A distributionally robust optimization problem takes the form

$$\underset{\boldsymbol{x}}{\text{minimize}} \max_{\mu \in \mathcal{D}} F(\boldsymbol{x}, \mu)$$

where $\mathcal{D}$ is a set of probability distributions satisfying some criteria, such as a given support region or a given mean and covariance matrix. The inner problem for fixed values of $\boldsymbol{x}$, "maximize$_{\mu \in \mathcal{D}}$ $F(\boldsymbol{x}, \mu)$", is often infinite-dimensional because we are searching over a space of probability distributions on some (possibly infinite) domain; one typically considers $\mathcal{D}$ to be a subset of the Banach space consisting of Borel measures on the support region. The concept of geometry is an important one because of the different ways that one can characterize the set of distributions $\mathcal{D}$; for example, if $F(\boldsymbol{x}, \mu)$ is linear in μ (such as an expectation operator) and $\mathcal{D}$ consists of a finite set of linear constraints on μ, then *Carathéodory's theorem* [81] establishes that there exists an optimal measure μ^* consisting of a finite set of points.

An important application of distributionally robust optimization is in transportation science, and more specifically in this dissertation, the *Traveling Salesman Problem* (TSP) and the *Vehicle Routing Problem* (VRP). The length of the TSP tour for *i.i.d.* samples of a distribution $f(\cdot)$ is of theoretical interest for asymptotic analysis [51, 82, 83, 5]. An example of such large TSP tours is shown in Figure 1.3.

A commonly used way to define the set $\mathcal{D}$ is to impose some constraints on the moments

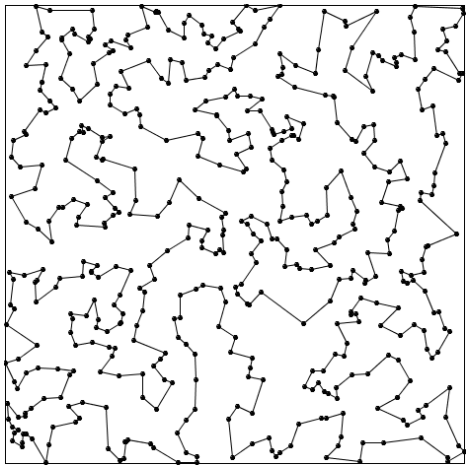
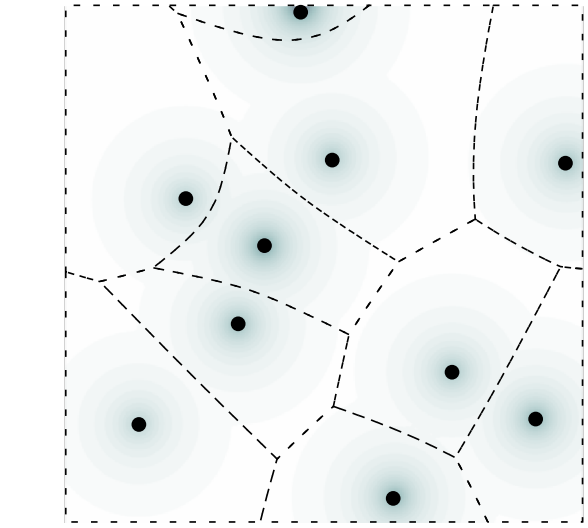


Figure 1.3: An example of TSP tour of *i.i.d.* random points.

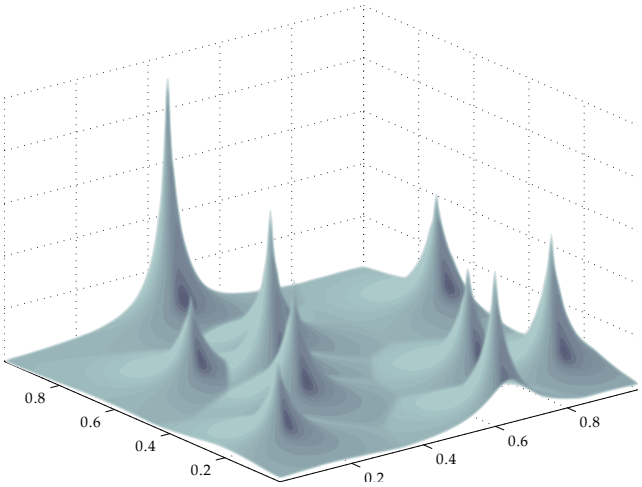
of the distribution like mean and covariance, see e.g. [84, 85, 86]. In the field of logistics and transportation, the authors of [65] used a mean-and-covariance-based approach to find the worst-case distribution for the length of TSP tour. We use a slightly different approach to find the worst-case demand distribution for the length of VRP tour which has more complications that a TSP tour. This is presented in Chapter 4 and first published in our article [87].

Besides the fact that we normally do not have enough information about the mean and covariance of an unknown distribution, a major drawback for this method is that it is often not faithful to the original problem when the ground truth distribution is clustered or multimodal. In section 5.3.1 we will show an example of such a situation that arises in geography.

A natural but less studied way of defining the set $\mathcal{D}$ is to enforce the worst case distribution to be "close enough" to an empirical distribution driven by drawing samples from the ground truth distribution with regards to some metrics like the *Wasserstein distance*, a.k.a. the *earth mover's distance* introduced in [88] in Russian and later in [89]. Using this method, we can find the worst case distribution in the asymptotic analysis of the length of large TSP tours. An example of the worst case distributions that we find by this method is presented in Figure 1.4. Using this method we can overcome clustered distributions and also we usually obtain concise and continuous worst case distributions, which is not frequently achieved using other approaches. This method is discussed in Chapter 5. To the best of our knowledge, our use of the Wasserstein distance in such an application, that first appeared in our article [90], is the first of its kind.



(a)



(b)

Figure 1.4: An example of the worst case distribution for the length of TSP tour in a unit box. The dots in (1.4a) show an empirical distribution driven by a sample of size 10. The dashed lines show the optimal transport mapping between the ground truth distribution and the empirical distribution. Figure (1.4b) shows a surface plot of the worst case distribution.

Chapter 2

Geometric Partitioning and Robust Ad-hoc Network Design

Geometric Partitioning and Robust Ad-hoc Network Design

In this chapter we develop geometric algorithms to provide robust solutions for a backbone network design problem. The goal is to design a decentralized communication network among a set of objects over a geographic convex region in order to minimize the maximum connectivity radius that we need to cover the entire region and keep the objects connected to each other. If the network has to be static, i.e. the shape of the network comprised of the position of the objects located at the vertices and the distance between objects or in fact the length of the edges are not allowed to change over time, then the problem becomes a location problem where we seek the best locations for the vertices of the network to minimize the connectivity radius. Otherwise, if the network is dynamic, say the objects move around, the problem becomes much harder since we encounter uncertainty about the position of the objects at each time. In this chapter we consider designing such dynamic networks in a robust way so as the communication connectivity is retained regardless of the position of the objects and the connectivity radius is minimized.

2.1 Introduction

Our problem can be motivated in the context of multi-vehicle coordination. Suppose that there are n vehicles that must provide service or surveillance to a convex region C. Further suppose that any two vehicles can communicate with each other if the distance between them is less than some given threshold radius r. In order to ensure that all vehicles be able to communicate with one another (possibly through intermediate vehicles), it is natural

to desire that the vehicles be configured in such a way that the communication network between them be *connected*. At the same time, in order to simplify operations and ensure that the entire region is covered, a natural strategy is to divide the region C into n sub-regions $R_1, \ldots, R_n$, such that each of the n vehicles is assigned to one of the n sub-regions.

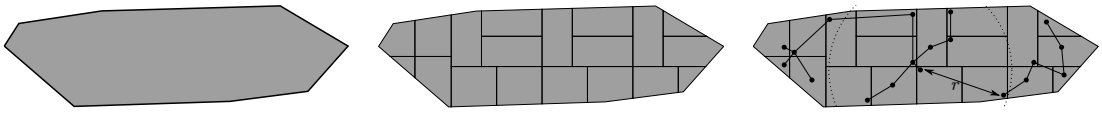
Before stating our problem formally, we find it useful to introduce some notational conventions. Given a set of points $\mathbf{x} = \{x_1, \ldots, x_n\} \subset C$ and a threshold radius r, we let $G_r(\mathbf{x})$ denote the graph whose nodes $\{1, \ldots, n\}$ correspond to the points $x_1, \ldots, x_n$, and whose edges (i, j) correspond to those pairs of points (i, j) such that $\|x_i - x_j\| \leq r$, where $\|\cdot\|$ denotes the Euclidean norm. Given a *partition* $R_1, \ldots, R_n$ of C (that is, a set of sub-regions R_i such that $\bigcup_{i=1}^n R_i = C$ and $R_i \cap R_j = \emptyset$ for all $i \neq j$), we let $\prod_{i=1}^n R_i$ denote the *Cartesian product* of those sub-regions, so that $\mathbf{x} = (x_1, \ldots, x_n) \in \prod_{i=1}^n R_i$ if and only if $x_i \in R_i$ for each $i \in \{1, \ldots, n\}$.

We now state our problem formally: given a convex planar region C with area A and an integer n, we are interested in partitioning C into sub-regions $R_1, ..., R_n$ in such a fashion that for any $\mathbf{x} = (x_1, \ldots, x_n) \in \prod_{i=1}^n R_i$ the graph $G_r(\mathbf{x})$ is connected for some radius r. We would like to minimize r among all such partitions. We can write our optimization problem as

$$\underset{R_1,\ldots,R_n,r}{\text{minimize}} \quad r \quad s.t. \tag{$*$}$$

$$\begin{aligned} G_r(\mathbf{x}) \text{ is connected} \qquad & \forall \mathbf{x} \in \prod_{i=1}^n R_i \,, \\ \bigcup_{i=1}^n R_i &= C \,, \\ R_i \cap R_j &= \emptyset \,. \end{aligned}$$

For a given partition, the minimum radius r that guarantees connectivity of all tuples $\mathbf{x} \in \prod_{i=1}^n R_i$ is called the *connectivity radius* associated with that partition. Inputs and outputs to problem $(*)$ are shown in Figure 2.1. This formulation does not impose any additional constraints on the shapes or sizes of the sub-regions R_i; it is of course sensible (from a practical standpoint) to add additional such requirements, such as convexity of the R_i's.



(a) (b) (c)

Figure 2.1: Input and output to problem $(*)$. We are given a convex region C and an integer n (which is equal to 19 in this case), and our objective is to partition C into n sub-regions $R_1, \ldots, R_n$; we desire a partition whose *connectivity radius* is as small as possible. For the partition shown in 2.1b, the connectivity radius r is indicated in 2.1c; for any tuple $\mathbf{x} \in \prod_{i=1}^{n} R_i$, the graph $G_r(\mathbf{x})$ is connected.

2.1.1 Related work

This work combines two related issues that commonly arise in multi-vehicle coordination, namely *geometric partitioning* and *ad-hoc network design*, and as such there exist two distinct bodies of work from which it stems. Geometric partitioning has emerged as a useful tool for allocating vehicles in a territory efficiently: for example, the papers [50, 91, 43, 53, 54, 46, 44] all describe various ways to balance the workloads of a fleet of vehicles (or facilities) by dividing the service region into sub-regions and localizing vehicles (or facilities) to those sub-regions. The paper [40] gives an algorithm for dividing a convex region into "fat" sub-regions of equal area, and in fact, Section 2.2 of this chapter employs essentially the same algorithm for our problem. As we will later make clear, the constraint that all sub-regions have equal area can lead to situations in which the connectivity radius is quite high due to a "bottleneck" region at the periphery of the service region; this motivates our new algorithms in Sections 2.3 and 2.4. In addition, as a side consequence, our algorithm introduced in Section 2.3 can also be used to describe an approximation algorithm for the classical *k-centers problem* in a convex polygon with approximation factor 1.99; this is notable because it is known that there does not exist any approximation algorithm for the k-centers problem in a general metric space with a constant less than 2 unless P=NP; see for example [29].

The model for connectivity in this chapter (namely, that two vehicles can communicate with each other if the distance between them is below the threshold r) is a common one and there exists a large body of work dealing with various problems under such an assumption [92, 93, 94, 95, 96]. Generally speaking, the problems of interest often involve *decentralized* approaches to solving combinatorial problems along such a network, such as finding a connected dominating set, a minimal covering of the service region, or the shortest path between a pair of points.

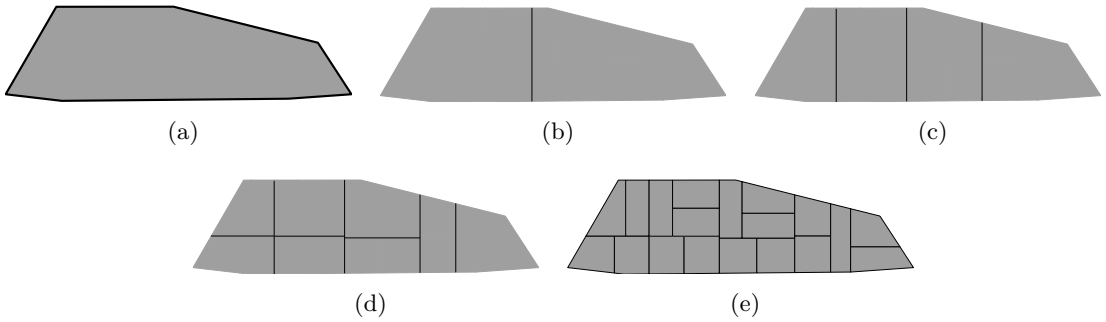


Figure 2.2: Input and output to Algorithm 1 for the case where we desire $n = 19$ sub-regions of equal area. In 2.2b, we use a vertical line to divide the region into two pieces whose areas are 9/19 and 10/19 of the original area. In 2.2c, we use vertical lines to further subdivide these two pieces into four pieces whose areas are 4/19, 5/19, 5/19, and 5/19 of the original area (from left to right). In 2.2d, we use vertical and horizontal lines to divide these pieces into even more pieces whose areas are all either 2/19 or 3/19 of the original area. Figure 2.2e shows the output of Algorithm 1, which consists of $n = 19$ sub-regions of equal area.

2.1.2 Notational Conventions

Throughout this chapter we use the following notational conventions: Area(D) denotes the area of a region D. The *width* and *height* of a region are defined as the width and height of the minimum-area axis-aligned bounding rectangle of D and are denoted by width(D) and height(D). The *aspect ratio* of a rectangle R is the ratio of the length of the longer side of R to the the length of the shorter side of R and is written as AR(R). The approximation ratio of each algorithm, i.e. the ratio of the upper bound to lower bound on the connectivity radius r, will be denoted Rat.

2.2 Imposing equal area and convexity

We begin by considering problem ($*$) where we impose an additional constraint that all sub-regions R_i have equal area and be convex. The algorithm we use is due to [40] (which considers the closely related problem of partitioning C into "fat" sub-regions). This algorithm is quite simple: we rotate C so that its diameter is horizontal, and we then recursively divide C with either a horizontal line or a vertical line, depending on which of the two directions results in a "fatter" shape. This is described formally in Algorithm 1 and illustrated in Figure 2.2; we make a minor modification to the algorithm in order to improve our bound by applying a new Lemma 4. We find it helpful to introduce two lower bounds to our problem stated below:

Input: A convex polygon C and an integer n.

Output: A partition of C into n convex sub-regions, each having area $\text{Area}(C)/n$.

Note: In the very first call of this algorithm – i.e. not the recursive calls – the input region C should be oriented so that its diameter is horizontal.

if $n = 1$ **then**
  return C;
else
  Set $n_1 = \lfloor n/2 \rfloor$ and $n_2 = \lceil n/2 \rceil$;
  Let w denote the width of C and h the height;
  if $w \geq h$ **then**
    With a vertical line, divide C into two pieces $R_1^{'}$ and $R_2^{'}$ with area $\frac{n_1}{n} \cdot \text{Area}(C)$ on the right and $\frac{n_2}{n} \cdot \text{Area}(C)$ on the left. Let $w^{'} = \max\{\text{width}(R_1^{'}), \text{width}(R_2^{'})\}$;
    With a vertical line, divide C into two pieces $R_1^{''}$ and $R_2^{''}$ with area $\frac{n_1}{n} \cdot \text{Area}(C)$ on the left and $\frac{n_2}{n} \cdot \text{Area}(C)$ on the right. Let $w^{''} = \max\{\text{width}(R_1^{''}), \text{width}(R_2^{''})\}$;
    if $w^{'} \leq w^{''}$ **then**
      Set $R_1 = R_1^{'}$ and $R_2 = R_2^{'}$;
    else
      Set $R_1 = R_1^{''}$ and $R_2 = R_2^{''}$;
    end
  else
    With a horizontal line, divide R into two pieces $R_1^{'}$ and $R_2^{'}$ with area $\frac{n_1}{n} \cdot \text{Area}(C)$ on the top and $\frac{n_2}{n} \cdot \text{Area}(C)$ on the bottom. Let $h^{'} = \max\{\text{height}(R_1^{'}), \text{height}(R_2^{'})\}$;
    With a horizontal line, divide R into two pieces $R_1^{''}$ and $R_2^{''}$ with area $\frac{n_1}{n} \cdot \text{Area}(C)$ on the bottom and $\frac{n_2}{n} \cdot \text{Area}(C)$ on the top. Let $h^{''} = \max\{\text{height}(R_1^{''}), \text{height}(R_2^{''})\}$;
    if $h^{'} \leq h^{''}$ **then**
      Set $R_1 = R_1^{'}$ and $R_2 = R_2^{'}$;
    else
      Set $R_1 = R_1^{''}$ and $R_2 = R_2^{''}$;
    end
  end
  return $\mathsf{EqualAreaPartition}(R_1, n_1) \cup \mathsf{EqualAreaPartition}(R_2, n_2)$;
end

Algorithm 1: Algorithm $\mathsf{EqualAreaPartition}(C, n)$ is due to [40]; it takes as input a convex polygon C and a positive integer n. We have made a minor modification to that of [40] in that we select a partition by comparing $w^{'}$ and $w^{''}$ (or $h^{'}$ and $h^{''}$).

Theorem 1. *Let C be an input to problem $(*)$ with an additional constraint that all sub-regions be convex and have equal area. Then $r^* \geq \sqrt{2A/\pi n}$.*

Proof. Suppose that $R_1, \ldots, R_n$ is a partition of C into convex pieces of equal area and that r is sufficiently large that the graph $G_r(\mathbf{x})$ is connected for all $\mathbf{x} \in \prod_{i=1}^{n} R_i$. For any sub-region R_i and any point $x_i \in R_i$, it must be the case that there exists some *other* sub-region R_j such that R_j is entirely contained within a ball B_i of radius r centered at x_i (if this were not the case, then $G_r(\mathbf{x})$ would not be connected because we could isolate point x_i from the rest of the nodes). However, by the convexity assumption, R_i and R_j must be linearly separable, say by line ℓ. We therefore find that R_j must lie on one side of ℓ (say in half-plane H_ℓ), with x_i lying on the other side, and therefore R_j must be contained in $B_i \cap H_\ell$. Since $\text{Area}(B_i \cap H_\ell) \leq \pi r^2/2$ and $\text{Area}(R_j) = A/n$ by the equal-area assumption, we have

$$\begin{aligned} \frac{\pi r^2}{2} \geq \text{Area}(B_i \cap H_\ell) \geq \text{Area}(R_j) &= A/n \\ \implies r &\geq \sqrt{\frac{2A}{\pi n}} \end{aligned}$$

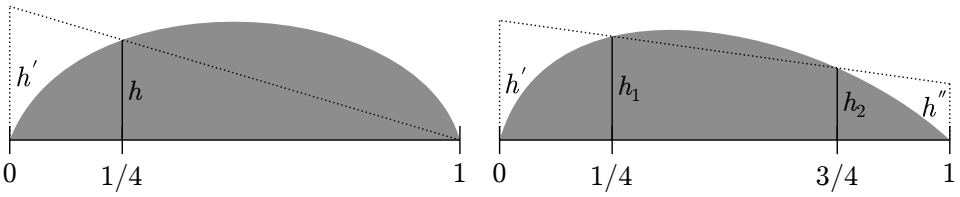
as desired. □

Theorem 2. *Let C be an input to problem $(*)$ with an additional constraint that all sub-regions be convex and have equal area and assume that C is oriented so that its diameter is horizontal. Let H_1 denote a vertically oriented half-space that cuts off exactly $\text{Area}(C)/n$ of C on the left and let H_2 denote a vertically oriented half-space that cuts off exactly $\text{Area}(C)/n$ of C on the right. Let w_1 and w_2 denote the widths of $C \cap H_1$ and $C \cap H_2$ respectively. Then $r^* \geq \max\{w_1, w_2\}$.*

Proof. Assume without loss of generality that $w_1 \geq w_2$. Consider a convex equal-area partition $R_1, \ldots, R_n$ and assume without loss of generality that the leftmost point of C, which we will call x_1, is contained in R_1. Since R_1 is convex, it is easy to see that $\text{Area}(R_1 \cap H_1) > 0$. We now claim that for any other region R_i, it must be possible to select a point $x_i \in R_i$ such that $\|x_1 - x_i\| \geq w_1$, which will complete the proof. This is straightforward: if there existed a region R_i such that this claim did not hold, then R_i would have to lie entirely to the left of H_1 (i.e. $R_i = R_i \cap H_1$). This is impossible because then

$$\text{Area}(C \cap H_1) \geq \text{Area}(R_1 \cap H_1) + \text{Area}(R_i \cap H_1) = \text{Area}(R_1 \cap H_1) + \text{Area}(R_i) > \text{Area}(C)/n$$

which contradicts our construction of H_1. □



(a) (b)

Figure 2.3: In 2.3a, we see that the area of the right triangle defined by points $(1/4, 0)$, $(1, 0)$, and $(1/4, h)$ is $3h/8$ and the area of the trapezoid defined by points $(0, 0)$, $(1/4, 0)$, $(1/4, h)$, and $(0, h')$ is $7h/24$. In 2.3b, the area of the trapezoid defined by points $(0, 0)$, $(1/4, 0)$, $(1/4, h_1)$, and $(0, h')$ is $(5h_1 - h_2)/16$, the area of the trapezoid defined by points $(1/4, 0)$, $(3/4, 0)$, $(3/4, h_2)$, and $(1/4, h_1)$ is $(h_1+h_2)/4$, and the area of the trapezoid defined by points $(3/4, 0)$, $(1, 0)$, $(1, h'')$, and $(3/4, h_2)$ is $(5h_2 - h_1)/16$.

2.2.1 Analysis of Algorithm 1

We will now show that Algorithm 1 solves $(*)$ (subject to equal-area and convexity constraints) within a factor of 7.31. We shall make use of several pre-existing results from [40], which will be cited when appropriate.

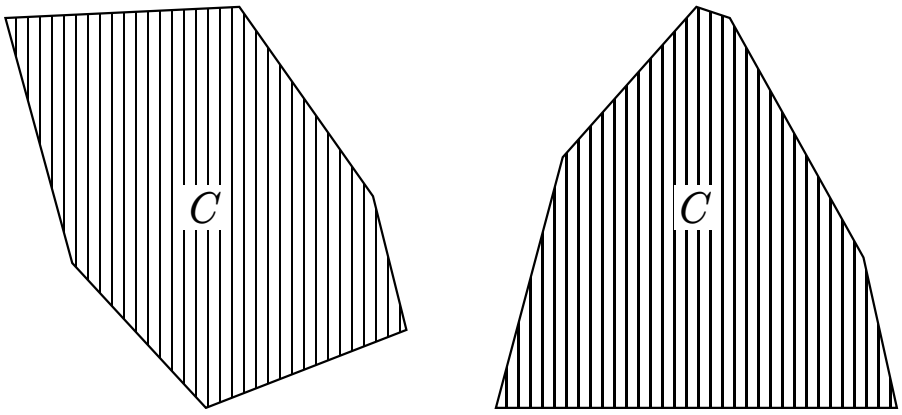
Claim 3. Let $f(\cdot) : [0, 1] \to \mathbb{R}^+$ be a concave function such that $\int_0^1 f(t)\,dt = 1$. Then $\int_0^{1/4} f(t)\,dt < 1/2$ and $\int_{1/4}^{3/4} f(t)\,dt \geq 1/2$.

Proof. Let $h = f(1/4)$. As evident by Figure 2.3a, we see that certainly $\int_{1/4}^1 f(x)\,dx \geq 3h/8$, and therefore $\int_0^{1/4} f(t)\,dt \leq 1 - 3h/8$. It is also evident from Figure 2.3a that $\int_0^{1/4} f(t)\,dt \leq 7h/24$. Combining these two upper bounds we see that $\min\{1 - 3h/8, 7h/24\} \leq 7/16 < 1/2$ for all $h > 0$, which completes the proof of the first claim.

The second claim is similar; let $h_1 = f(1/4)$ and let $h_2 = f(3/4)$. Figure 2.3b shows that $\int_0^{1/4} f(t)\,dt + \int_{3/4}^1 f(t)\,dt \leq (h_1 + h_2)/4$, and therefore $\int_{1/4}^{3/4} f(t)\,dt \geq 1 - (h_1 + h_2)/4$. It is also evident from Figure 2.3b that $\int_{1/4}^{3/4} f(t)\,dt \geq (h_1 + h_2)/4$. Combining these upper bounds we see that $\max\{1 - (h_1 + h_2)/4, (h_1 + h_2)/4\} \geq 1/2$ for all $h_1, h_2 > 0$, which completes the proof of the second claim. □

Lemma 4. *Let C be a convex polygon of width w. Let ℓ denote a vertical line that cuts off a fraction α of the area of C to the left and $1 - \alpha$ to the right, and let ℓ' denote a vertical line that cuts off $1 - \alpha$ to the left and α to the right. Let C_1 and C_2 denote the left and right pieces induced by ℓ and let C_1' and C_2' denote the left and right pieces induced by ℓ'. Then if $\alpha \in [1/3, 2/3]$, it must be the case that either*

$$\text{width}(C_1), \text{width}(C_2) \in [w/4, 3w/4]$$



(a) (b)

Figure 2.4: "Shifting" the polygon C onto the horizontal axis.

or

$$\text{width}(C_1'), \text{width}(C_2') \in [w/4, 3w/4]\,.$$

Proof. Assume without loss of generality that $w = 1$ and that $\alpha \leq 1/2$. First, we "shift" C onto the horizontal axis, so that C can equivalently be represented as a concave function $f : [0,1] \to \mathbb{R}^+$ as shown in Figure 2.4. We can also assume that $\int_0^1 f(t)\,dt = 1$ since we can scale $f(\cdot)$ arbitrarily in the vertical direction, i.e. replace $f(\cdot)$ by $af(\cdot)$ for some positive scalar a. It will therefore suffice to show that, for any such concave function, if we have

$$\int_0^c f(t)\,dt = \alpha \text{ and } \int_d^1 f(t)\,dt = \alpha\,,$$

then either $c \in [1/4, 3/4]$ or $d \in [1/4, 3/4]$. This follows from Claim 3. In particular, we merely have to rule out the two possibilities that $c, d < 1/4$ or that $c < 1/4$ and $d > 3/4$ (the case $c, d > 3/4$ is taken care of by symmetry).

1. Suppose for a contradiction that $c, d < 1/4$. Then by increasing α, we see that c must move to the right and d must move to the left, and therefore when we set $\alpha = 1/2$ we find that $c = d < 1/4$. Thus we have $\int_0^{1/4} f(x)\,dx > 1/2$, a contradiction of the first statement of Claim 3.

2. Suppose for a contradiction that $c < 1/4$ and $d > 3/4$. This implies that $\int_0^{1/4} f(x)\,dx > \alpha$ and $\int_{3/4}^1 f(x)\,dx > \alpha$. This would then imply that $\int_{1/4}^{3/4} f(x)\,dx < 1/3$, a contradiction of the second statement of Claim 3. This completes the proof.

□

Lemma 5. *For any sub-region R_i that is output by Algorithm 1, we have* $\text{Area}(R_i) \geq \text{width}(R_i) \cdot \text{length}(R_i)/2$.

Proof. This is Corollary 4.2 of [40] (which actually proves a stronger result, namely that the above also holds for any *intermediate* sub-region obtained throughout the execution of Algorithm 1, not just the final output $R_1, \ldots, R_n$). □

Lemma 6. *Suppose that R_i is a sub-region output from Algorithm 1 such that* $\text{AR}(R_i) > 4$. *Then all cuts leading to R_i were vertical.*

Proof. This is essentially Corollary 4.6 of [40], which deals with the special case where $n = 2^k$ so that Algorithm 1 always divides sub-regions in half. By applying Lemma 4, we are able to generalize this result to arbitrary n (the paper [40] does indeed consider partitioning for general n, and their results hold for the case where $\text{AR}(R_i) > 6$). □

Lemma 7. *Suppose that R_i is a sub-region output from Algorithm 1 such that* $\text{AR}(R_i) > 4$. *Then there exists a unique leftmost sub-region or a unique rightmost sub-region R_{i^*} such that* $\text{AR}(R_{i^*}) \geq \text{AR}(R_j)$ *for all sub-regions R_j.*

Proof. This follows from the same proof as Claim 4.8 of [40]. □

Combining the above lemmas, we can now determine the overall approximation guarantee of Algorithm 1:

Theorem 8. *Algorithm 1 is a* 7.31 *approximation algorithm for problem (∗) subject to the additional constraint that all sub-regions R_i be convex and have equal area. Its running time is $\mathcal{O}((m+n)\log n)$, where m is the number of edges of the input region.*

Proof. The running time of Algorithm 1 is given in Section 4.1 of [40]. The approximation ratio depends on the maximum aspect ratio of any of the sub-regions, $\max_i \text{AR}(R_i)$:

- If $\max_i \text{AR}(R_i) > 4$, then we apply Lemma 7 to consider the region R_{i^*} whose aspect ratio is as large as possible (which must be the leftmost or rightmost sub-region); let $z = \text{AR}(R_{i^*})$. It is then easy to verify that $\text{diam}(R_i) \leq \sqrt{2A(z + 1/z)/n}$ for all i (this is just the diagonal length of a rectangle with aspect ratio z and area $2A$; the "$2A$" term arises apropos of Lemma 5). It also follows immediately that the radius of connectivity for our problem is at most twice that, i.e. that for any tuple $\mathbf{x} \in \prod_{i=1}^{n} R_i$, the graph $G_r(\mathbf{x})$ is connected, with $r = 2\sqrt{2A(z + 1/z)/n}$. Since $z > 4$ and R_{i^*} is either the leftmost or rightmost sub-region, we see that Theorem 2 gives us a lower

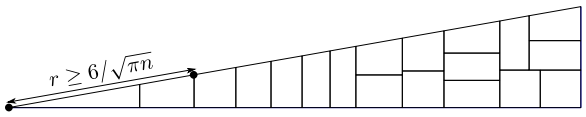


Figure 2.5: When the input region C is a right triangle with area 1 whose leftmost angle is $\theta = 10°$, the leftmost sub-region always acts as a bottleneck for the overall connectivity radius.

bound that is simply the width of R_{i^*}, which is at least $\sqrt{Az/n}$. We therefore find that

$$\mathsf{Rat} \leq \frac{\mathrm{UB}}{\mathrm{LB}} = \frac{2\sqrt{2A(z+1/z)/n}}{\sqrt{Az/n}} = 2\sqrt{2(1+1/z^2)} < \sqrt{17/2} \approx 2.9155$$

since $z > 4$.

- If $\max_i \mathrm{AR}(R_i) \leq 4$, then we must have $\mathrm{diam}(R_i) \leq \sqrt{17A/2n}$ (this is, again, the diagonal length of a rectangle with aspect ratio 4 and area $2A$; the "$2A$" term arises apropos of Lemma 5). It follows immediately that the radius of connectivity for the output of our algorithm is at most twice that, i.e. for any tuple $\mathbf{x} \in \prod_{i=1}^{n} R_i$, the graph $G_r(\mathbf{x})$ is connected, with $r = \sqrt{34A/n}$. Applying the lower bound of Theorem 1, we find that

$$\mathsf{Rat} \leq \frac{\mathrm{UB}}{\mathrm{LB}} = \frac{\sqrt{34A/n}}{\sqrt{2A/\pi n}} = \sqrt{17\pi} < 7.31\,.$$

This completes the proof. □

One of the practical drawbacks to Algorithm 1 is the interaction between the objective of minimizing the connectivity radius (which is in a sense a *maximum* distance taken over all sub-regions R_i) and the constraint that all sub-regions have equal area. Consider for example Figure 2.5, in which the input region is a right triangle with area 1 whose leftmost angle is $\theta = 10°$. We therefore find that, for sufficiently large n, it will always be the case that $r^* \geq \sqrt{\frac{2}{\theta \cdot \frac{\pi \text{ rad}}{180°} \cdot n}} = 6/\sqrt{\pi n}$, and one can plainly see that leftmost region acts as a bottleneck for the problem because all other sub-regions are much closer to their neighbors. Thus, it is natural to consider alternative problem formulations in which we relax the equal-area constraint (or the convexity constraint). We will next consider the fully unconstrained version of $(*)$, which naturally brings other issues of its own (and which we will subsequently rectify).

2.3 The unconstrained version of $(*)$

We began the preceding section by introducing relevant lower bounds for $(*)$ when convexity and equal area were imposed. We will do the same for this section, where we remove these restrictions:

Theorem 9. *The optimal solution r^* to problem $(*)$ must satisfy*

$$r^* \geq \sqrt{\frac{A}{\pi + (n-1)(\pi/3 + \sqrt{3}/2)}}\,. \tag{2.1}$$

In order to prove Theorem 9, we require two simple lemmas:

Lemma 10. *For any partition $R_1, \ldots, R_n$ of C, there exists an n-tuple of points $\mathbf{x} \in \prod_{i=1}^n R_i$ and an index i^* such that $\|x_{i^*} - x_j\| \geq \sqrt{A/\pi n}$ for all $j \neq i^*$.*

Proof. Given any partition $R_1, \ldots, R_n$, select the n-tuple $\mathbf{x}$ arbitrarily and center an open ball B_i of radius $\sqrt{A/\pi n}$ at each element x_i. If there exists an index $\bar{i}$ such that $B_{\bar{i}}$ is disjoint from all other balls, then clearly $i^* = \bar{i}$ and we are done. If no such element exists, then $\text{Area}(\bigcup_i B_i) < \sum_i \text{Area}(B_i) = A$ and therefore the balls B_i do not cover C. We can therefore select a point $x^* \in C \setminus \bigcup_i B_i$, so that $\|x^* - x_i\| \geq \sqrt{A/\pi n}$ for all i. Since $R_1, \ldots, R_n$ is a partition of C we know that $x^* \in R_{i^*}$ for some sub-region R_{i^*}. Setting $x_{i^*} = x^*$ completes the proof. □

Proof of Theorem 9. Suppose that $R_1, \ldots, R_n$ is a partition of C and that r is sufficiently large such that the graph $G_r(\mathbf{x})$ is connected for all $\mathbf{x} \in \prod_{i=1}^n R_i$. Select any n-tuple $\mathbf{x} \in \prod_{i=1}^n R_i$ arbitrarily and center a ball of radius r at each element x_i. Following the same reasoning as the proof of Lemma 10, we see that the balls B_i must cover the region C. Furthermore, since $G_r(\mathbf{x})$ is connected, we know that $G_r(\mathbf{x})$ must have a spanning tree T. For notational simplicity, suppose without loss of generality that $1, \ldots, n$ is a *pre-order traversal* [97] of T, so that for any index k, the subgraph of T associated with nodes $1, \ldots, k$ is also a tree (and therefore connected). For any k, we will let T_k denote this subtree (so that $T = T_n$). Consider the leaf node $n \in G_r(\mathbf{x})$ of T and the portion of C that is *uniquely* associated with x_n, i.e. $(B_n \setminus \bigcup_{j=1}^{n-1} B_j) \cap C$. Since node n is within a distance r of at least one other node j^*, we see that portion uniquely associated with x_n is at most $(\pi/3 + \sqrt{3}/2)r^2$ (due to a direct computation of the area between two disks, i.e. a circular lens), whence

$$\text{Area}\left(\left(B_n \setminus \bigcup_{j=1}^{n-1} B_j\right) \cap C\right) \leq \text{Area}\left(B_n \setminus \bigcup_{j=1}^{n-1} B_j\right) \leq \text{Area}(B_n \setminus B_{j^*}) \leq (\pi/3 + \sqrt{3}/2)r^2\,.$$

Deleting node n (i.e. point x_n) and the region $B_n \setminus \bigcup_{j=1}^{n-1} B_j$, we now have a smaller region $C^{'}$ containing points $x_1, \ldots, x_{n-1}$ which are connected by the edges of T_{n-1}. It again must follow that the portion of $C^{'}$ that is uniquely associated with leaf node $n-1$ must satisfy

$$\text{Area}\left(\left(B_{n-1} \setminus \bigcup_{j=1}^{n-2} B_j\right) \cap C\right) \leq \text{Area}\left(B_{n-1} \setminus \bigcup_{j=1}^{n-2} B_j\right) \leq (\pi/3 + \sqrt{3}/2)r^2\,.$$

We can apply this process iteratively to find that, for each of the nodes $n, n-1, \ldots, 2$, the area uniquely associated with that node must be at most $(\pi/3 + \sqrt{3}/2)r^2$. Our proof is complete by observing that, after the $n-1$ leaf deletions are completed, we have a single remaining point x_1, which must obviously satisfy $\text{Area}(B_1) \leq \pi r^2$. Summing these together, we see that

$$A = \text{Area}(C) = \sum_{i=0}^{n-2} \text{Area}\left(\left(B_{n-i} \setminus \bigcup_{j=1}^{n-i-1} B_j\right) \cap C\right) + \text{Area}(B_1 \cap C) \leq \pi r^2 + (n-1)(\pi/3 + \sqrt{3}/2)r^2$$

from which (2.1) follows. □

Note that Theorem 9 does not take the *shape* of C into account; it depends only on the area A. We find it necessary to introduce a second lower bound that applies when C is long and skinny:

Theorem 11. *The optimal solution r^* to problem $(*)$ must satisfy*

$$r^* \geq d/n\,, \tag{2.2}$$

where d is the diameter *of C.*

Proof. This is similar in spirit to the proof of Theorem 9, restricted to one dimension. Again, suppose that $R_1, \ldots, R_n$ is a partition of C and that r is sufficiently large such that the graph $G_r(\mathbf{x})$ is connected for all $\mathbf{x} \in \prod_{i=1}^{n} R_i$. Select any n-tuple $\mathbf{x} \in \prod_{i=1}^{n} R_i$ arbitrarily and center a ball of radius r at each element x_i with . We again see that the balls B_i must cover the region C, and in particular, the balls B_i must cover the longest line segment s in C (whose length is by definition the diameter d of C). Project each point x_i onto s, obtaining a new n-tuple $\mathbf{x}^{'} = (x_1^{'}, x_2^{'}, \ldots, x_n^{'})$. Similarly, project each ball B_i onto s, obtaining line segments $s_1^{'}, \ldots, s_n^{'}$ of length at most $2r$. Since convex projection is a *nonexpansive* mapping (see e.g. [98]), we know that the graph $G_r(\mathbf{x}^{'})$ must also be connected.

Assume without loss of generality that s is aligned with the horizontal axis and that the points $(x_1^{'}, \ldots, x_n^{'})$ are sorted in order from left to right. Further assume that $x_n^{'}$ is

the rightmost endpoint of s (which we are free to do because the initial tuple $\mathbf{x}$ was chosen arbitrarily anyway), which implies that $\text{length}(s_n^{'}) \leq r$. As in the proof of Theorem 9, consider the portion of s that is uniquely associated with $s_1^{'}$, i.e. $s_1^{'} \setminus \bigcup_{j=2}^{n} s_j^{'}$. Since we know that $\|x_1^{'} - x_2^{'}\| \leq r$ by connectivity of $G_r(\mathbf{x}^{'})$, we find that

$$\text{length}\left(s_1^{'} \setminus \bigcup_{j=2}^{n} s_j^{'}\right) \leq r$$

and, deleting $x_1^{'}$ and $s_1^{'}$, we find that

$$\text{length}\left(s_2^{'} \setminus \bigcup_{j=3}^{n} s_j^{'}\right) \leq r$$

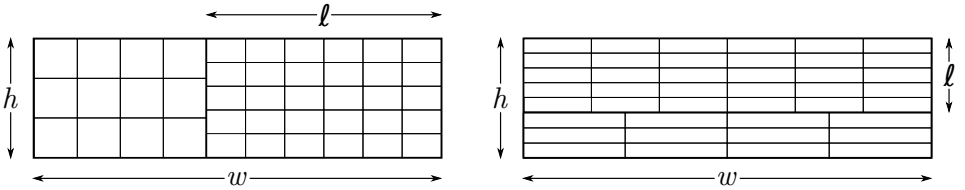
and so on and so forth. We ultimately conclude that

$$d = \text{length}(s) = \sum_{i=1}^{n-1} \text{length}\left(s_i^{'} \setminus \bigcup_{j=i+1}^{n} s_j^{'}\right) + \text{length}(s_n^{'}) \leq nr$$

from which the desired result follows. □

2.3.1 An approximation algorithm for problem (∗)

In this section we give a factor 2.77 approximation algorithm for problem (∗). As a side consequence, it turns out that this algorithm can also be used to give a factor 1.99 approximation algorithm for the continuous *k-centers problem* in a convex polygon, which we will elaborate on in Section 3.1; this is notable because it is known that there does not exist any approximation algorithm for the k-centers problem in a general metric space with a constant less than 2 unless P=NP; see for example [29]. In a nutshell, our algorithm is extremely simple: we first build a bounding box of C that is aligned with its diameter. Next, we attempt to divide this bounding box, which we will call Q, into $n-1$ rectangles that are as "square" as possible (it will turn out that all of the sub-regions that our algorithm produces will be single points, with the exception of the n-th sub-region, which is the remainder of C). If w and h are the width and height of Q respectively, then a good starting point is to set $p_0 = \lfloor\sqrt{w(n-1)/h}\rfloor$ and $q_0 = \lfloor\sqrt{h(n-1)/w}\rfloor$, and then divide Q into a rectangular $p_0 \times q_0$ grid. Because of the floor functions, it is likely that $p_0 q_0 < n-1$, which means that we have some "leftover" grid cells at our disposal. We can insert these additional grid cells by either adding them to the columns or the rows of the existing grid. After this step, we let the centers of these $n-1$ rectangles be the first $n-1$ sub-regions of our partition of C,



(a) "flag" = `VERTICAL` (b) "flag" = `HORIZONTAL`

Figure 2.6: The output of Algorithm 2 where $(p_1, q_1, p_2, q_2) = (4, 3, 6, 5)$ and ℓ is as indicated. We subdivide the rectangle into two grids, one of which has 4×3 rectangles and one of which has 6×5 rectangles; the width and height of these two grids is determined by ℓ.

and we let R_n be the remaining area of R left over.

Formally, our algorithm is based on a simple scheme which we call subroutine RectanglePartition, in which we subdivide an axis-aligned rectangle Q into k smaller rectangles by juxtaposing a pair of rectangular grids. This is described in Algorithm 2 and sketched in Figure 2.6. Another subroutine that builds off of this, which we call PointPlacement, applies Algorithm 2 recursively and determines an "optimal" way to merge two grids together; this is described in Algorithm 3. Finally, our 2.77 approximation algorithm for problem $(*)$ simply applies Algorithm 3 for select values of (p_1, q_1, p_2, q_2) and outputs a partition $R_1, \ldots, R_n$ in which the first $n-1$ sub-regions R_i are each equal to a single point c_i, and the final sub-region R_n is the complement of this, $R_n = C \setminus (\{c_1\} \cup \cdots \cup \{c_{n-1}\})$. This is described formally in Algorithm 4 and sketched in Figure 2.7.

Input: An axis-aligned rectangle Q, having dimensions $w \times h$, integers p_1, q_1, p_2, q_2, a positive number ℓ, and a "flag" equal to `VERTICAL` or `HORIZONTAL`.
Output: A partition of Q into $p_1q_1 + p_2q_2$ rectangles.
if *"flag" is* `VERTICAL` **then**
  Let Q_1 be the left half of Q, having dimensions $(w - \ell) \times h$;
  Let Q_2 be the right half of Q, having dimensions $\ell \times h$;
  Break Q_1 into a $p_1 \times q_1$ rectangular grid, and call the rectangular cells $\boxdot_1, \ldots, \boxdot_{p_1q_1}$;
  Break Q_2 into a $p_2 \times q_2$ rectangular grid, and call the rectangular cells $\boxdot_{p_1q_1+1}, \ldots, \boxdot_{p_1q_1+p_2q_2}$;
else
  Let Q_1 be the bottom half of Q, having dimensions $w \times (h - \ell)$;
  Let Q_2 be the top half of Q, having dimensions $w \times \ell$;
  Break Q_1 into a $p_1 \times q_1$ rectangular grid, and call the rectangular cells $\boxdot_1, \ldots, \boxdot_{p_1q_1}$;
  Break Q_2 into a $p_2 \times q_2$ rectangular grid, and call the rectangular cells $\boxdot_{p_1q_1+1}, \ldots, \boxdot_{p_1q_1+p_2q_2}$;
end
return $\boxdot_1, \ldots, \boxdot_{p_1q_1+p_2q_2}$;

Algorithm 2: Algorithm RectanglePartition$(Q, p_1, q_1, p_2, q_2, \ell,$ "flag") decomposes rectangle Q into a pair of rectangular grids.

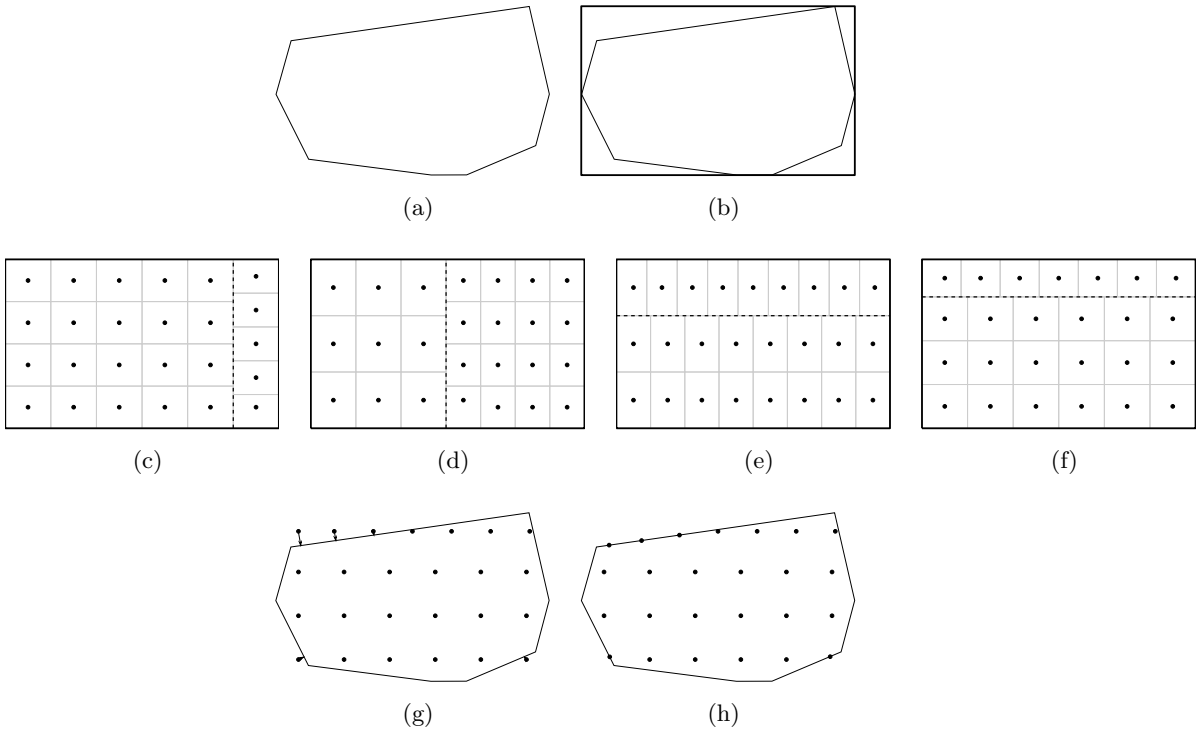


Figure 2.7: An execution of Algorithm 4 with $n = 26$. We will eventually construct a partition consisting of n sub-regions R_i, in which $R_1, \ldots, R_{n-1}$ are each a single point, and R_n their complement. Thus, we will construct $n-1$ rectangles and place a point in each of their centers. In 2.7a, we orient the imput region C to have its diameter be horizontal. We then compute the axis-aligned bounding box Q, as shown in 2.7b, which happens to have $w/h = 1.62$, which tells us that $p_0 = \lfloor\sqrt{w(n-1)/h}\rfloor = 6$ and $q_0 = \lfloor\sqrt{h(n-1)/w}\rfloor = 3$. Algorithm 4 then constructs various potential point arrangements by positioning two grids alongside each other, four of which are shown in 2.7c through 2.7f; for example, in 2.7d, we break Q into $p = p_0 + 1 = 7$ columns and we have $q = \lfloor(n-1)/p\rfloor = 3$; this gives $s = 4$, so that we break Q into two grids, one of which consists of $(p-s) \times q^{'} = 3 \times 3$ rectangles and the other consisting of $s \times (q+1) = 4 \times 4$ rectangles, thus giving $n-1$ rectangles in total. Note that, in each of the four placements, there is always one rectangle $\boxdot_j$ whose associated point $c_j^{'}$ is off-center, as specified in Algorithm 3; this is done to ensure that the connectivity radius of the points placed in Algorithm 4 is at most equal to the maximum of the connectivity radii of each of the two halves (and is only added to make the upcoming proof of the approximation ratio easier). It turns out that the arrangement shown in 2.7f has the smallest connectivity radius of all the options of Algorithm 4. Thus, we project all points from 2.7f onto the original polygon C in 2.7g, and 2.7h shows the final output of our algorithm: the first $n-1 = 25$ sub-regions are the points that are shown, and the n-th sub-region is their complement.

Input: An axis-aligned rectangle Q, having dimensions $w \times h$, integers p_1, q_1, p_2, q_2, and a "flag" equal to `VERTICAL` or `HORIZONTAL`.
Output: A collection of $p_1q_1 + p_2q_2$ points in Q.
Translate Q so that its bottom left corner is the origin;
if *"flag" is* `VERTICAL` **then**
  Set $L = \{wp_2/(p_1 + p_2)\,,\, hp_2/q_1\}$;
  Remove any elements $\ell \in L$ that are not between 0 and w;
  `/* The first value of` L `will give grids whose cells have the same (horizontal) width; the second value will give grids in which the (vertical) height of one set of cells is equal to the (horizontal) width of the other */`
else
  Set $L = \{hq_2/(q_1 + q_2)\,,\, wq_2/p_1\}$;
  Remove any elements $\ell \in L$ that are not between 0 and h;
  `/* The first value of` L `will give grids whose cells have the same (vertical) height; the second value will give grids in which the (vertical) height of one set of cells is equal to the (horizontal) width of the other */`
end
Set $r = \infty$;
for $\ell \in L$ **do**
  Let $\boxdot_1^{'}, \ldots, \boxdot_{p_1q_1+p_2q_2}^{'}$ = RectanglePartition$(Q, p_1, q_1, p_2, q_2, \ell,$ "flag");
  Let $c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$ be the centers of $\boxdot_1^{'}, \ldots, \boxdot_{p_1q_1+p_2q_2}^{'}$;
  Let Q_1 and Q_2 be the two large rectangles that divided Q in the execution of RectanglePartition;
  if *"flag" is* `VERTICAL` **then**
    Let $c_i^{'}$ be the center of the bottom right sub-rectangle $\boxdot_i$ of Q_1;
    Let $c_j^{'}$ be the center of the bottom left sub-rectangle $\boxdot_j$ of Q_2;
    Move $c_j^{'}$ to be vertically aligned with $c_i^{'}$;
  else
    Let $c_i^{'}$ be the center of the top left sub-rectangle $\boxdot_i$ of Q_1;
    Let $c_j^{'}$ be the center of the bottom left sub-rectangle $\boxdot_j$ of Q_2;
    Move $c_j^{'}$ to be horizontally aligned with $c_i^{'}$;
  end
  Let $r^{'}$ be the connectivity radius of $c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
  if $r^{'} < r$ **then**
    Set $r = r^{'}$;
    Set $c_1, \ldots, c_{p_1q_1+p_2q_2} = c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
  end
end
return $c_1, \ldots, c_{p_1q_1+p_2q_2}$;

Algorithm 3: Algorithm PointPlacement$(Q, p_1, q_1, p_2, q_2,$ "flag") places $p_1q_1 + p_2q_2$ points inside rectangle Q at the centers of a collection of rectangles.

Input: A convex polygon C and an integer n.
Output: A partition of C into n sub-regions that solves problem $(*)$ within a factor of 2.77.
Rotate C so that its diameter is aligned with the x-axis;
Let Q be the axis-aligned bounding box of C;
Set $p_0 = \lfloor\sqrt{w(n-1)/h}\rfloor$;
Set $q_0 = \lfloor\sqrt{h(n-1)/w}\rfloor$;
Set $r = \infty$;
for $p \in \{p_0 - 1, p_0, p_0 + 1\}$ **do**
    `/* break` Q `into` p `columns, and then subdivide these columns as evenly as possible */`
    Set $q = \lfloor(n-1)/p\rfloor$;
    if $p, q \geq 1$ **then**
        Set $s = (n-1) - pq$, so that $n - 1 = pq + s = (p-s)q + s(q+1)$;
        Let $(c_1^{'}, \ldots, c_{n-1}^{'}) =$ PointPlacement$(p - s, q, s, q + 1,$ `VERTICAL`$)$;
        Let $r^{'}$ be the connectivity radius of $c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
        if $r^{'} < r$ **then**
            Set $r = r^{'}$;
            Set $c_1, \ldots, c_{p_1q_1+p_2q_2} = c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
        end
    end
end
for $q \in \{q_0 - 1, q_0, q_0 + 1\}$ **do**
    `/* break` Q `into` q `rows, and then subdivide these rows as evenly as possible */`
    Set $p = \lfloor(n-1)/q\rfloor$;
    if $p, q \geq 1$ **then**
        Set $s = (n-1) - pq$, so that $n - 1 = pq + s = p(q-s) + (p+1)s$;
        Let $(c_1, \ldots, c_{n-1}) =$ PointPlacement$(p, q - s, p + 1, s,$ `HORIZONTAL`$)$;
        Let $r^{'}$ be the connectivity radius of $c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
        if $r^{'} < r$ **then**
            Set $r = r^{'}$;
            Set $c_1, \ldots, c_{p_1q_1+p_2q_2} = c_1^{'}, \ldots, c_{p_1q_1+p_2q_2}^{'}$;
        end
    end
end
Project each point c_i onto the polygon C (if it is not already inside C);
Let $R_i = \{c_i\}$ for $i \in \{1, \ldots, n-1\}$ and let $R_n = C \setminus (\{c_1\} \cup \cdots \cup \{c_{n-1}\})$;
return $R_1, \ldots, R_n$;

Algorithm 4: Algorithm RegionPartition(C, n) partitions convex polygon C into n sub-regions.

2.3.2 Analysis of Algorithm 4

This section is devoted to a proof of the following theorem:

Theorem 12. *Algorithm 4 solves problem (∗) within a factor of* 2.77*. Its running time is* $\mathcal{O}(m + n \log m)$*, where* m *is the number of edges of the input region.*

The running time is easy to verify: We require $\mathcal{O}(m)$ time to find the minimum bounding box of C and we require $\mathcal{O}(n)$ time to run Algorithm 2. We can project the $n-1$ points inside C in $\mathcal{O}(n \log m)$ time by using a point-in-polygon algorithm [99]. We will make some simplifying observations: first, we can disregard the step in Algorithm 4 where the points $c_i^{'}$ are projected onto C because convex projection is always non-expansive [98]; in other words, $\|c_i^{'} - c_j^{'}\| \geq \|c_i - c_j\|$ for all i, j. Secondly, we will assume for simplicity that C has area 1; by construction, this means that the bounding box of Q must have an area between 1 and 2 (this is because the diameter of C is positioned horizontally, i.e. aligned with the long side of box Q). We will assume without loss of generality that Q has area 2 (clearly, larger boxes can only have larger point-to-point distances). Thus, we will say that Q has width w and height $h = 2/w$, with $w \geq h$, i.e. $w \geq \sqrt{2}$. Our approximation ratio will be completely characterized by w and n. Finally, we note that our algorithm always produces a union of two rectangular grids, with a single point that is offset from the others. It is straightforward to verify that the connectivity radius of this union is always at most the connectivity radii of the two halves. Thus, in certain situations below, we will occasionally construct a *single* rectangular grid having the desired approximation ratio (possibly having fewer points than we are permitted), which must therefore be no worse than the union of rectangular grids that Algorithm 4 produces.

We next observe that our algorithm certainly attains the desired approximation ratio whenever $w \geq \sqrt{n}$. This is because one of the configurations produced is to simply divide Q into an $(n-1) \times 1$ grid of identical rectangles, each having width $w/(n-1)$ and height $h = 2/w$. Each of the $n-1$ points $c_i^{'}$ is exactly $w/(n-1)$ away from its left and right neighbors, and therefore the connectivity radius of $R_1, \ldots, R_{n-1}$ is at most $w/(n-1)$. Furthermore, any point $x_n \in R_n$ must be within a distance of

$$\frac{1}{2}\sqrt{\left(\frac{w}{n-1}\right)^2 + \left(\frac{2}{w}\right)^2}$$

to one of the points $c_i^{'}$ (this is just half the diagonal of any of the $n-1$ rectangles). Therefore, the approximation ratio is at most the maximum of these two distances, divided by the lower

bound of w/n (from Theorem 11) :

$$\begin{aligned}
\mathsf{Rat} \leq \frac{\max\left\{\frac{w}{n-1}\,,\,\frac{1}{2}\sqrt{\left(\frac{w}{n-1}\right)^2+\left(\frac{2}{w}\right)^2}\right\}}{w/n} &= \max\left\{\frac{n}{n-1}\,,\,\frac{\frac{1}{2}\sqrt{\left(\frac{w}{n-1}\right)^2+\left(\frac{2}{w}\right)^2}}{w/n}\right\} \\
&\leq \max\left\{2\,,\,\frac{\frac{1}{2}\left[\left(\frac{w}{n-1}\right)+\left(\frac{2}{w}\right)\right]}{w/n}\right\} \\
&\leq \max\left\{2\,,\,\frac{\frac{1}{2}\left[\left(\frac{w}{n-1}\right)+\left(\frac{2}{\sqrt{n}}\right)\right]}{w/n}\right\} \\
&\leq \max\left\{2\,,\,\frac{n}{2(n-1)}+1\right\} \leq 2 < 2.77 \text{ for all } n \geq 2
\end{aligned}$$

as desired. We are therefore permitted to assume that $w < \sqrt{n}$ throughout.

We will next verify computationally that the approximation ratio of 2.77 holds for $n \leq 20$. Note that for any fixed n, we can consider only those values of w in the finite interval $[\sqrt{2}, \sqrt{n})$. These ratios are shown in the plot in Figure 2.8. Since all values on that plot are less than 2.77, we obtain the desired result, and we are therefore free to assume that $w < \sqrt{n}$ and that $n \geq 21$. Note that we can now improve the result from the previous paragraph to conclude that our approximation ratio holds whenever $w \geq 0.61\sqrt{n}$. This is because we can safely assume that $n \geq 21$, and because the connectivity radius associated with an $(n-1) \times 1$ grid of identical rectangles satisfies

$$\begin{aligned}
\mathsf{Rat} \leq \frac{\max\left\{\frac{w}{n-1}\,,\,\frac{1}{2}\sqrt{\left(\frac{w}{n-1}\right)^2+\left(\frac{2}{w}\right)^2}\right\}}{w/n} &= \max\left\{\frac{n}{n-1}\,,\,\frac{\frac{1}{2}\sqrt{\left(\frac{w}{n-1}\right)^2+\left(\frac{2}{w}\right)^2}}{w/n}\right\} \\
&\leq \max\left\{\frac{21}{20}\,,\,\frac{\frac{1}{2}\sqrt{\left(\frac{21}{20}\right)^2\left(\frac{w}{n}\right)^2+\left(\frac{2}{w}\right)^2}}{w/n}\right\} \\
&= \max\left\{\frac{21}{20}\,,\,\frac{\sqrt{441w^4+1600n^2}}{40w^2}\right\} \\
&\leq \max\left\{\frac{21}{20}\,,\,\frac{\sqrt{441w^4+1600(w/0.61)^4}}{40w^2}\right\} < 2.77
\end{aligned}$$

as desired. Thus, we will now assume that $n \geq 21$ and $w < 0.61\sqrt{n}$.

The assumption that $w < 0.61\sqrt{n}$ and that $n \geq 21$ implies that the number of rows, q_0,

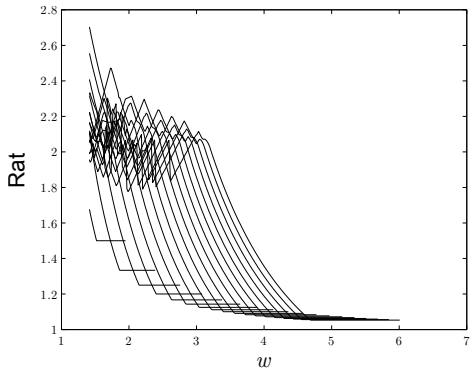


Figure 2.8: The approximation ratios realized by Algorithm 4, for $n \in \{2, \ldots, 20\}$ and $w \in [\sqrt{2}, \frac{4}{3}\sqrt{n})$. These are obtained by dividing the connectivity radius for the output solution by the maximum of the two lower bounds (2.1) and (2.2). Note that it would have sufficed to compute these ratios for the smaller interval $w \in [\sqrt{2}, \sqrt{n})$, by our previous analysis; we merely compute the ratio for this slightly longer interval for purposes of clarity.

that are used in Algorithm 4 is

$$q_0 = \left\lfloor \sqrt{\frac{h(n-1)}{w}} \right\rfloor = \left\lfloor \sqrt{\frac{2(n-1)}{w^2}} \right\rfloor = \left\lfloor \sqrt{\frac{2n}{w^2} \cdot \frac{n-1}{n}} \right\rfloor \geq \left\lfloor \sqrt{\frac{2n}{w^2} \cdot \frac{20}{21}} \right\rfloor > \lfloor 2.27 \rfloor = 2\,;$$

We can divide Q into $\lfloor (n-1)/q_0 \rfloor \times q_0$ sub-rectangles and place a point c_i' in the center of each of these. Note that by definition of the floor function, we always have

$$q_0 \leq \sqrt{\frac{h(n-1)}{w}} = \sqrt{\frac{2(n-1)}{w^2}} < q_0 + 1\,,$$

or equivalently

$$\frac{\sqrt{2(n-1)}}{q_0+1} < w \leq \frac{\sqrt{2(n-1)}}{q_0}\,. \tag{2.3}$$

The distance between any vertical neighbors is obviously $h/q_0 = 2/(q_0 w)$ and the distance between any horizontal neighbors is at most

$$\frac{w}{\lfloor (n-1)/q_0 \rfloor} \leq \frac{w}{(n-1)/q_0 - 1} = \frac{q_0 w}{(n-1) - q_0}$$

and therefore, applying lower bound (2.1), the approximation ratio is at most

$$\mathsf{Rat} \leq \frac{\max\left\{ \frac{2}{q_0 w}\,,\, \frac{q_0 w}{(n-1)-q_0} \right\}}{\sqrt{\frac{1}{\pi + (n-1)(\pi/3 + \sqrt{3}/2)}}}\,.$$

Assume that $q_0 \geq 3$; we address the case where $q_0 = 2$ in Section A.1 of the appendices.

The first term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as small as possible, which occurs when $w = \sqrt{2(n-1)}/(q_0+1)$; the above ratio is then equal to

$$\begin{aligned}
\frac{\frac{2}{q_0 w}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{2}{q_0\sqrt{2(n-1)}/(q_0+1)}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\
&= \frac{q_0+1}{q_0}\cdot\frac{\sqrt{(9\sqrt{3}+6\pi)(n-1)+18\pi}}{3\sqrt{n-1}} \\
&\leq \frac{4}{3}\cdot\frac{\sqrt{(9\sqrt{3}+6\pi)(n-1)+18\pi}}{3\sqrt{n-1}} < 2.77 \text{ for } n \geq 21\,.
\end{aligned}$$

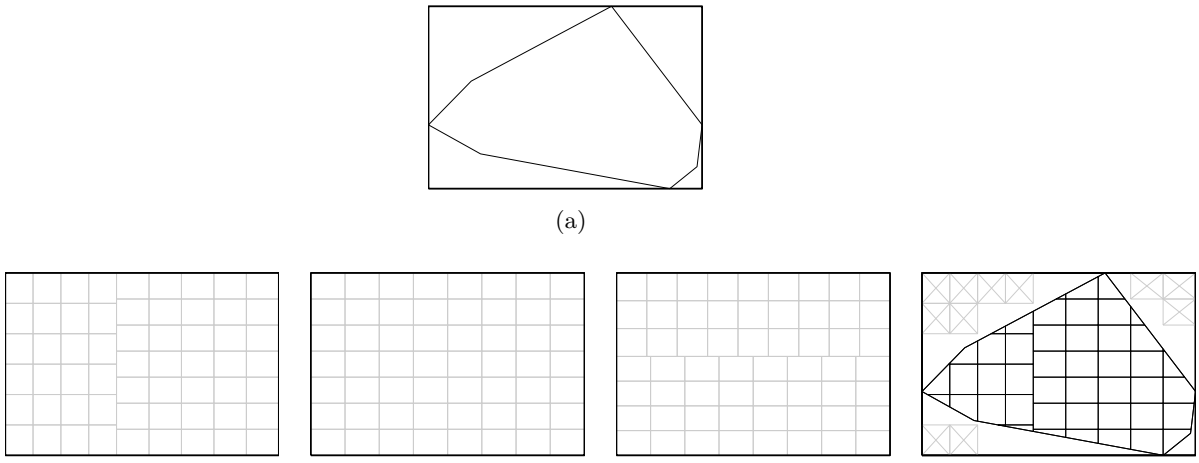
The second term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as large as possible, which occurs when $w = \sqrt{2(n-1)}/q_0$; the above ratio is then equal to

$$\begin{aligned}
\frac{\frac{q_0 w}{(n-1)-q_0}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{\sqrt{2(n-1)}}{(n-1)-q_0}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\
&\leq \frac{\frac{\sqrt{2(n-1)}}{(n-1)-\sqrt{n-1}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\
&= \frac{\sqrt{3}}{3}\cdot\frac{\sqrt{(3\sqrt{3}+2\pi)(n-1)+6\pi}}{\sqrt{n-1}-1} < 2.77 \text{ for all } n \geq 21\,,
\end{aligned}$$

where we have used the fact that $q_0 \leq \sqrt{n-1}$, which holds because we have $h \leq w$ and we define $q_0 = \lfloor\sqrt{h(n-1)/w}\rfloor$. This completes the proof for the case where $q_0 \neq 2$ and we refer the reader to Section A.1 of the appendices for the remaining analysis.

2.4 Problem $(*)$ with a convexity constraint

Although Algorithm 4 does not have the same "bottleneck" problem as Algorithm 1 (as described in the end of Section 2.2), it of course has deficiencies of its own: first of all, the sub-regions $R_1, \ldots, R_n$ have irregular shapes, and secondly, the areas of these sub-regions are extremely unbalanced (all of the area is allocated to the single region R_n except for a finite set of $n-1$ points). Thus, we propose Algorithm 5, which remedies both of these issues by producing sub-regions that are all convex and that all have areas that do not exceed $\frac{22}{9}A/n = 2.\bar{4}A/n$ (recall that Algorithm 1 produced convex sub-regions whose areas were equal to A/n). As in Section 2.3, Algorithm 5 is based on a sequence of calls to the RectanglePartition subroutine, Algorithm 2, and is sketched in Figure 2.9. This section is



(a) (b) (c) (d) (e)

Figure 2.9: An execution of Algorithm 5 with $n = 59$. We start with a convex polygon, which has been oriented in 2.9a to have its diameter be horizontal, embedded in bounding box Q. Algorithm 5 then constructs various potential convex partitions consisting of either a single grid or a pair of grids; three such partitions are shown in 2.9b through 2.9d. The partition in 2.9b has the smallest connectivity radius, and thus the output partition is shown in 2.9e. Note that there are 11 empty sub-regions, as indicated by the crossed-out boxes. Thus, the final output partition consists of 48 sub-regions.

devoted to a proof of the following theorem:

Theorem 13. *Algorithm 5 is a* 5.94 *approximation algorithm for problem (∗) with an additional constraint that all sub-regions R_i be convex. In addition, all sub-regions output by Algorithm 5 have area of at most $\frac{22}{9}A/n = 2.\bar{4}A/n$. Its running time is $\mathcal{O}(m + n \log m)$, where m is the number of edges of the input region.*

The claimed running time is true for the same reasons as in the proof of Theorem 12. This proof turns out to be much shorter than the preceding results. We will use the lower bounds from Theorems 9 and 11. The following result bounds the output of Algorithm 5 from above:

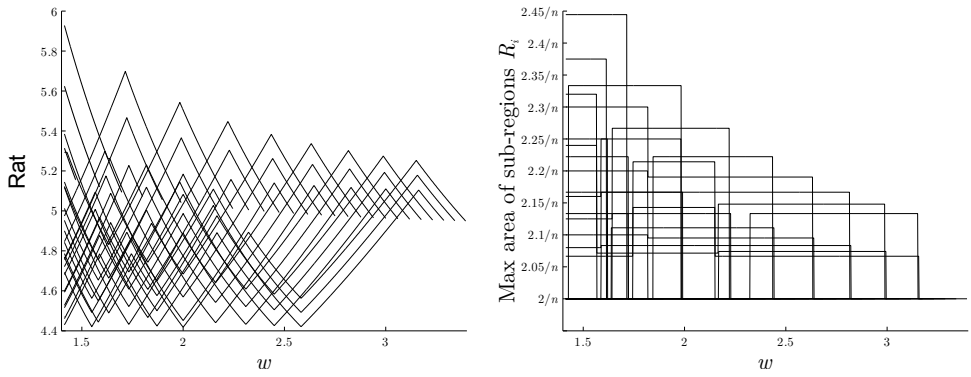
Lemma 14. *Let C be a convex polygon and let $R_1, \ldots, R_n$ be the output of Algorithm 5. If $R_1, \ldots, R_n$ is obtained by intersecting a single $p \times q$ grid with C, then the connectivity radius r satisfies*

$$r \leq \max\left\{\sqrt{(2w/p)^2 + (h/q)^2}\,,\ \sqrt{(w/p)^2 + (2h/q)^2}\right\},$$

and if $R_1, \ldots, R_n$ is obtained by merging two grids together, with sub-rectangles having

Input: A convex polygon C and an integer n.
Output: A partition of C into at most n convex sub-regions that solves problem $(*)$ within a factor of 5.94.
Rotate C so that its diameter is aligned with the x-axis;
Let Q be the axis-aligned bounding box of C;
Set $p_0 = \lfloor \sqrt{wn/h} \rfloor$;
Set $q_0 = \lfloor \sqrt{hn/w} \rfloor$;
Set $r = \infty$;
for $p \in \{p_0, p_0 + 1\}$ **do**
    /* break Q into p columns, and then subdivide these columns as evenly as possible */
    Set $q = \lfloor n/p \rfloor$;
    if $p, q \geq 1$ **then**
        /* Try a simple subdivision into identical rectangles first */
        Let $\boxdot_1, \ldots, \boxdot_{pq}$ be a subdivision of Q into a $p \times q$ grid (with each cell having dimensions $w/p \times h/q$);
        Let r' be the connectivity radius of $\boxdot_1, \ldots, \boxdot_{pq}$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $R_i = \boxdot_i \cap C$ for each $i \in \{1, \ldots, pq\}$;
        end
        /* Try using the RectanglePartition subroutine */
        Set $s = n - pq$, so that $n = pq + s = (p - s)q + s(q + 1)$;
        Set $\ell = \frac{ws(q+1)}{pq+s}$;
        Let $\boxdot_1, \ldots, \boxdot_n =$ RectanglePartition$(Q, p - s, q, s, q + 1, \ell,$ VERTICAL$)$;
        /* By construction of ℓ, all boxes $\boxdot_i$ have equal area */
        Let r' be the connectivity radius of $\boxdot_1, \ldots, \boxdot_n$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $R_i = \boxdot_i \cap C$ for each $i \in \{1, \ldots, n\}$;
        end
    end
end
for $q \in \{q_0, q_0 + 1\}$ **do**
    /* break Q into q rows, and then subdivide these rows as evenly as possible */
    Set $p = \lfloor n/q \rfloor$;
    if $p, q \geq 1$ **then**
        /* Try a simple subdivision into identical rectangles first */
        Let $\boxdot_1, \ldots, \boxdot_{pq}$ be a subdivision of Q into a $p \times q$ grid (with each cell having dimensions $w/p \times h/q$);
        Let r' be the connectivity radius of $\boxdot_1, \ldots, \boxdot_{pq}$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $R_i = \boxdot_i \cap C$ for each $i \in \{1, \ldots, pq\}$;
        end
        /* Try using the RectanglePartition subroutine */
        Set $s = n - pq$, so that $n = pq + s = p(q - s) + (p + 1)s$;
        Set $\ell = \frac{hs(p+1)}{pq+s}$;
        Let $\boxdot_1, \ldots, \boxdot_n =$ RectanglePartition$(Q, p, q - s, p + 1, s, \ell,$ HORIZONTAL$)$;
        Let r' be the connectivity radius of $\boxdot_1, \ldots, \boxdot_n$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $R_i = \boxdot_i \cap C$ for each $i \in \{1, \ldots, n\}$;
        end
    end
end
return $R_1, \ldots, R_n$;

Algorithm 5: Algorithm ConvexPartition(C, n) partitions convex polygon C into at most n convex sub-regions.



(a) (b)

Figure 2.10: The approximation ratios realized by Algorithm 5, for $n \in \{2, \ldots, 32\}$ and $w \in [\sqrt{2}, \frac{3}{5}\sqrt{n})$ and the maximum areas of the sub-regions that are output.

dimensions $w_i \times h_i$ for $i \in \{1, 2\}$, then r satisfies

$$r \leq \max\left\{\sqrt{(2w_i)^2 + h_i^2}\,,\ \sqrt{w_i^2 + (2h_i)^2}\,,\ \sqrt{(w_1 + w_2)^2 + (h_1 + h_2)^2}\right\} \text{ for } i \in \{1, 2\}.$$

Proof. This follows from basic facts about convexity; the third term in the second inequality, $\sqrt{(w_1 + w_2)^2 + (h_1 + h_2)^2}$, guarantees that the two grids are connected to each other. □

Our proof of Theorem 13 now follows: as in all of the previous analyses, we assume that C has area 1 and that Q has area 2. We first observe that our approximation ratio certainly holds whenever $w \geq \frac{3}{5}\sqrt{n}$, because the connectivity radius of an $n \times 1$ grid is at most $\sqrt{(2w/n)^2 + h^2}$ and therefore the approximation ratio is

$$\begin{aligned} \mathsf{Rat} \leq \frac{\sqrt{(2w/n)^2 + h^2}}{w/n} &= \frac{\sqrt{(2w/n)^2 + (2/w)^2}}{w/n} \\ &= \frac{2\sqrt{w^4 + n^2}}{w^2} \leq \frac{2\sqrt{w^4 + \frac{625}{81}w^4}}{w^2} < 5.94 \end{aligned}$$

as desired (and certainly, each sub-region R_i has area equal to $2/n < \frac{22}{9n}$). We will next verify computationally that the approximation ratio of 5.94 holds and that all sub-regions have area of at most $\frac{22}{9n}$ for $n \leq 32$ (we choose a large threshold value of n solely to make this proof as short as possible). Note that for any fixed n, we can consider only those values of w in the finite interval $[\sqrt{2}, \frac{3}{5}\sqrt{n})$. These ratios are shown in the plot in Figure 2.10. Since all values on that plot are less than 5.94, we obtain the desired result, and we are therefore free to assume that $w < \frac{3}{5}\sqrt{n}$ and that $n \geq 33$.

We next consider the case where $q_0 = \lfloor\sqrt{hn/w}\rfloor \geq 4$, or equivalently, $w \leq \sqrt{2n}/4$.

Suppose that we partition Q into a $\lfloor n/q_0 \rfloor \times q_0$ grid; the connectivity radius of such a grid is certainly bounded above by that of a $p_0 \times q_0$ grid because $p_0 \geq \lfloor n/q_0 \rfloor$ by construction. If we use a $p_0 \times q_0$ grid, then we know that each of the (identical) grid cells has an aspect ratio of at most $5/4$ because $p_0, q_0 \geq 4$. The union of any two adjacent rectangles therefore has an aspect ratio of at most $5/2$. Thus, the connectivity radius of such a configuration is at most equal to the diagonal length of this union of rectangles, which by routine calculations is at most

$$r \leq \frac{\sqrt{290}/5}{\sqrt{p_0 q_0}} < \frac{3.41}{\sqrt{p_0 q_0}}\,.$$

By construction, we have $n = p_0 q_0 + s$, with $s \leq p_0 + q_0$. Thus, the approximation ratio is at most

$$\begin{aligned}
\mathsf{Rat} &\leq \frac{\frac{\sqrt{290}/5}{\sqrt{p_0 q_0}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\
&\leq \frac{\frac{\sqrt{290}/5}{\sqrt{p_0 q_0}}}{\sqrt{\frac{1}{\pi+(p_0 q_0+p_0+q_0-1)(\pi/3+\sqrt{3}/2)}}} \\
&= \sqrt{\left(\frac{58\pi}{15}+\frac{29\sqrt{3}}{5}\right)(1+1/p_0+1/q_0)+\frac{29(4\pi-3\sqrt{3})}{15 p_0 q_0}} \\
&\approx \sqrt{22.194\,(1+1/p_0+1/q_0)+\frac{14.249}{p_0 q_0}} \leq \sqrt{22.194\,(1+1/4+1/4)+\frac{14.249}{4\cdot 4}} \approx 5.85 < 5.94
\end{aligned}$$

as desired. If we partition Q into a $\lfloor n/q_0 \rfloor \times q_0$ grid (as originally postulated), then the area of each rectangle is

$$\frac{2}{\lfloor n/q_0 \rfloor q_0} \leq \frac{2}{(n/q_0-1)q_0} = \frac{2}{n-q_0} \leq \frac{2}{n-\sqrt{n}} < \frac{22}{9n}$$

because we have $q_0^2 \leq n$ and $n \geq 33$.

Our proof is therefore complete if we consider the interval $\sqrt{2n}/4 < w < \frac{3}{5}\sqrt{n}$ and assume that $n \geq 33$. As we explain in Section A.3 of the appendices, we can achieve the desired approximation ratio by using a grid of dimensions $\lfloor n/3 \rfloor \times 3$ when $\sqrt{2n}/4 \leq w < \sqrt{n/3}$, and a grid of dimensions $n \times 1$ when $\sqrt{n/3} \leq w < \frac{3}{5}\sqrt{n}$, which completes the proof.

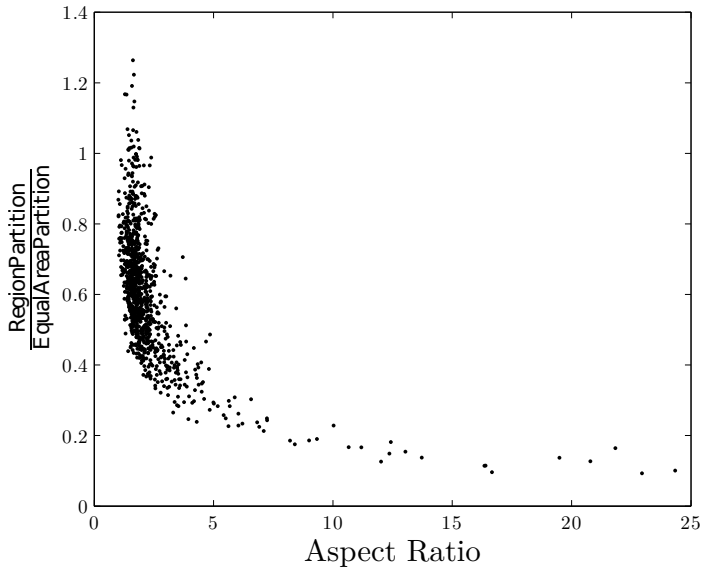
2.5 Computational experiments

In order to compare our three algorithms, this section presents the results of a simple computational experiment using randomly generated input regions C. The following procedure

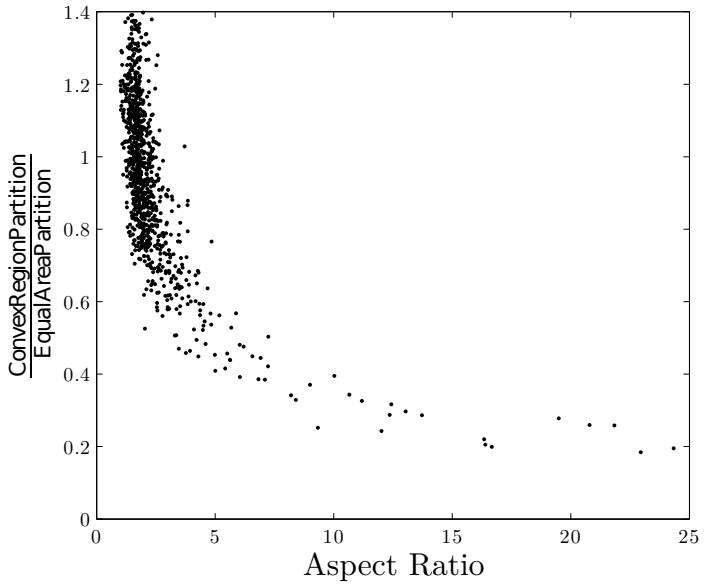
was performed 1000 times:

1. Let n, the desired number of sub-regions, be drawn uniformly between 2 and 60.
2. Let C be the convex hull of m points selected uniformly at random in the unit square, with m also drawn uniformly between 3 and 10 inclusively.
3. Run Algorithm 1, Algorithm 4, and Algorithm 5 on C.
4. For each of the three algorithm outputs, perform the following steps 1000 times:
   (a) Sample $x_1, \ldots, x_n$ uniformly at random from $R_1, \ldots, R_n$.
   (b) Let r be the connectivity radius of $x_1, \ldots, x_n$.
5. Let r_1 be the largest connectivity radius that was ever obtained in the 1000 samples associated with Algorithm 1.
6. Let r_2 be the largest connectivity radius that was ever obtained in the 1000 samples associated with Algorithm 4.
7. Let r_3 be the largest connectivity radius that was ever obtained in the 1000 samples associated with Algorithm 5.

At the end of this procedure, we have 1000 samples of the form (r_1, r_2, r_3). Figures 2.11a and 2.11b show the ratios r_2/r_1 and r_3/r_1 respectively, as a function of the aspect ratio of C (to be precise, the aspect ratio of the *bounding box* of C). The results are not surprising: first, we see that Algorithm 4 outperforms Algorithm 1 in nearly all cases (since nearly all of the data points have $r_2/r_1 < 1$), with the exception of a small number of samples that occur when the aspect ratio of C is nearly 1. Figure 2.11b shows that Algorithm 1 may outperform Algorithm 5 somewhat when the aspect ratio of C is very close to 1, but there is again an unmistakable trend in favor of Algorithm 5 as the aspect ratio increases. This is unsurprising because it reflects precisely the phenomenon that motivated us to introduce Algorithm 5 in the first place, namely the "bottleneck" phenomenon that occurs with input shapes that are long and skinny, which is shown in Figure 2.5.



(a)



(b)

Figure 2.11: The ratio of the connectivity radii of our three algorithms. Figure 2.11a shows the ratio of the maximum radius produced by Algorithm 4 to the maximum radius produced by Algorithm 1. Figure 2.11b shows the same ratios between Algorithms 5 and 1.

Chapter 3

Continuous k-Centers Problem in a Convex Polygon

Continuous k-Centers Problem in a Convex Polygon

Finding the optimal locations of geographic assets plays a key role in the geographic resource allocation problems. Besides the category of location problems, many problems in other categories mentioned in section 1.1 are somehow integrated or inter-correlated with location theory. In this chapter we consider the *k-centers problem*, which is one of the classical problems in location theory. This problem is by definition a robust optimization problem where the uncertainty set is defined over the location of demand points.

In the *k-centers problem*, we are given a domain $\mathcal{D}$ equipped with a distance function $\delta(\cdot, \cdot)$ and our objective is to place k points $x_1, \ldots, x_k$ in $\mathcal{D}$ to minimize the maximum distance between any point $x \in \mathcal{D}$ and its nearest neighbor x_i; that is, the problem is

$$\underset{\{x_1,\ldots,x_k\}\subset\mathcal{D}}{\text{minimize}} \max_{x\in\mathcal{D}} \min_{i} \delta(x, x_i)\,. \qquad (**)$$

Although we can define other sorts of uncertainty for this problem by introducing new parameters, e.g. the number of outlier centers, we stick with the basic definition of the problem. In this chapter we will present a fast approximation algorithm for the special case where the domain $\mathcal{D}$ is a convex polygon.

3.1 An approximation algorithm for the continuous k-centers problem in a convex polygon

In this section for the special case where $\mathcal{D}$ is a convex polygon C in the plane and $\delta(\cdot,\cdot)$ is the Euclidean norm, we show how the rectangular partitioning scheme of Algorithm 2 can be applied to design an approximation algorithm with factor 1.99. This is notable because it is known that for the problem $(**)$ in a general metric space, there does not exist any approximation algorithm with a constant less than 2 unless P=NP and the assumption of triangle inequality is necessary if the problem $(**)$ is to be approximated unless P=NP; see for example [29]. We require two extremely simple lower bounds:

Lemma 15. *The optimal objective value r^* to problem $(**)$, where $\mathcal{D}$ is a convex polygon C with area A and diameter d and the distance function $\delta(\cdot,\cdot)$ is the Euclidean norm, satisfies*

$$r^* \geq \max\left\{\sqrt{\frac{A}{\pi k}}\,,\ \frac{d}{2k}\right\}.$$

Proof. This is straightforward: if $x_1^*,\ldots,x_k^*$ are an optimal solution, then if we center a ball of radius r^* at each x_i^*, we must cover all of C, therefore $A \leq k\pi(r^*)^2$. The second bound arises from the observation that the longest line segment in C (whose length is d) must be covered by the k points, and each point is capable of covering a length of at most $2r^*$, whence $d \leq 2kr^*$. □

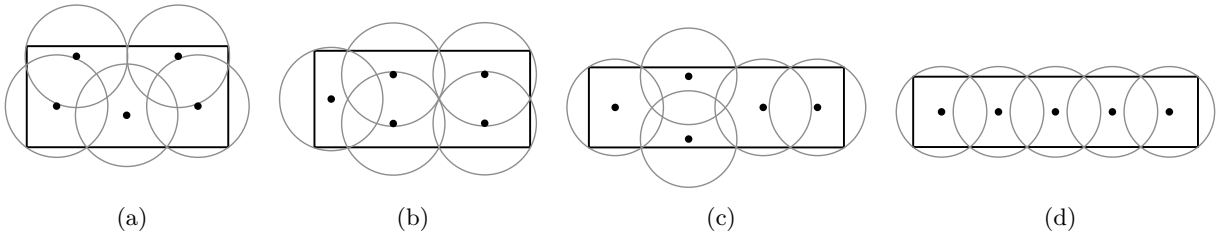
The intuition behind the approximation factor of 1.99 is as follows: assume without loss of generality that $A = 1$, and recall from before that we can enclose C in a box Q whose area is at most 2. Suppose that we somehow divide Q into k rectangles, each having area equal to $2/k$, and suppose that the aspect ratios of these rectangles (i.e. the ratios of the long side to the short side) do not exceed 2. If we place points $x_1,\ldots,x_k$ at the centers of these rectangles, then the distance between any point $x \in C \subseteq Q$ and its nearest center x_i is at most half of the diagonal of these rectangles, which (by our bounded aspect ratio) is at most $\frac{1}{2}\sqrt{(1/\sqrt{k})^2 + (2/\sqrt{k})^2} = \frac{1}{2}\sqrt{5/k}$. On the other hand, our first lower bound says that $r^* \geq \sqrt{1/(\pi k)}$ and thus the approximation ratio is at most

$$\mathsf{Rat} \leq \frac{\frac{1}{2}\sqrt{5/k}}{\sqrt{1/(\pi k)}} \approx 1.982 < 1.99\,.$$

Our algorithm is described in Algorithm 6 and sketched in Figure 3.2.

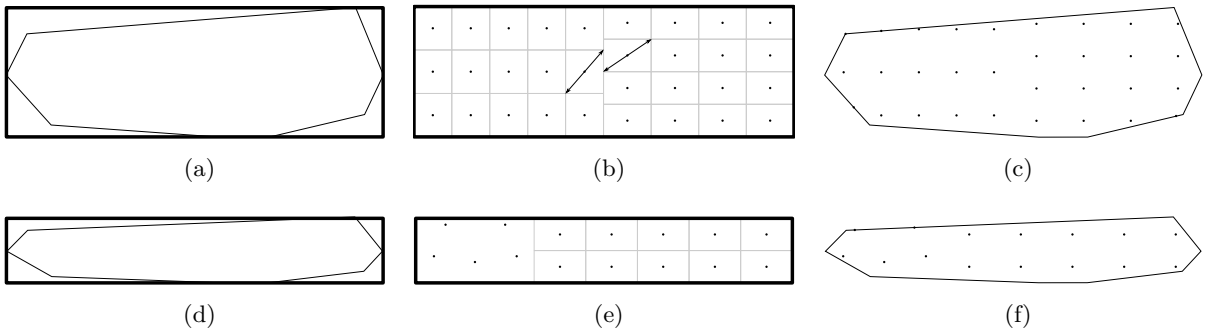
Input: A convex polygon C and an integer $k \geq 6$.
Output: The locations of k points inside C to solve the continuous k-centers problem in C within a factor of 1.99.
Rotate C so that its diameter is aligned with the x-axis;
Let Q be the axis-aligned bounding box of C;
Set $p_0 = \lfloor \sqrt{wk/h} \rfloor$;
Set $q_0 = \lfloor \sqrt{hk/w} \rfloor$;
Set $r = \infty$;
for $p \in \{p_0 - 1, p_0, p_0 + 1\}$ **do**
    /* break Q into p columns, and then subdivide these columns as evenly as possible */
    Set $q = \lfloor k/p \rfloor$;
    if $p, q \geq 1$ **then**
        Set $s = k - pq$, so that $k = pq + s = (p - s)q + s(q + 1)$;
        Let ℓ be the solution to $\left(\frac{w-\ell}{p-s}\right)^2 + \left(\frac{h}{q}\right)^2 = \left(\frac{\ell}{s}\right)^2 + \left(\frac{h}{q+1}\right)^2$ that satisfies $0 \leq \ell \leq w$; if no such ℓ exists, let $\ell = w$;
        Let $\boxdot_1, \ldots, \boxdot_k = \mathsf{RectanglePartition}(Q, p - s, q, s, q + 1, \ell, \texttt{VERTICAL})$;
        /* By construction of ℓ, all boxes have the same diagonal length if $\ell < w$ */
        Let $(x_1', \ldots, x_k')$ be the centers of the boxes $\boxdot_i$;
        Let $r' = \max_{x \in C} \min_i \|x - x_i'\|$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $x_1, \ldots, x_k = x_1', \ldots, x_k'$;
        end
    end
end
for $q \in \{q_0 - 1, q_0, q_0 + 1\}$ **do**
    /* break Q into q rows, and then subdivide these rows as evenly as possible */
    Set $p = \lfloor k/q \rfloor$;
    if $p, q \geq 1$ **then**
        Set $s = k - pq$, so that $k = pq + s = p(q - s) + (p + 1)s$;
        Let ℓ be the solution to $\left(\frac{w}{p}\right)^2 + \left(\frac{h-\ell}{q-s}\right)^2 = \left(\frac{w}{p+1}\right)^2 + \left(\frac{\ell}{s}\right)^2$ that satisfies $0 \leq \ell \leq h$; if no such ℓ exists, let $\ell = h$;
        Let $\boxdot_1, \ldots, \boxdot_k = \mathsf{RectanglePartition}(Q, p, q - s, p + 1, s, \ell, \texttt{HORIZONTAL})$;
        /* By construction of ℓ, all boxes have the same diagonal length if $\ell < h$ */
        Let $(x_1', \ldots, x_k')$ be the centers of the boxes $\boxdot_i$;
        Let $r' = \max_{x \in C} \min_i \|x - x_i\|$;
        if $r' < r$ **then**
            Set $r = r'$;
            Set $x_1, \ldots, x_k = x_1', \ldots, x_k'$;
        end
    end
end
if $q_0 = 1$ *and* k *is odd* **then**
    /* This is a very particular edge case that requires special attention */
    Set $\ell = 11.08/\sqrt{k} - 6.10/w$ and divide Q into two rectangles Q_1 and Q_2 with dimensions $\ell \times h$ and $(w - \ell) \times h$ respectively;
    Place $(x_1', \ldots, x_k')$ as follows: put 5 points in Q_1 according to Figure 3.1 and place $k - 5$ points in a $(k - 5)/2 \times 2$ grid in Q_2;
    Let $r' = \max_{x \in C} \min_i \|x - x_i\|$;
    if $r' < r$ **then**
        Set $r = r'$;
        Set $x_1, \ldots, x_k = x_1', \ldots, x_k'$;
    end
end
Project each point x_i onto the polygon C (if it is not already inside C);
return $x_1, \ldots, x_k$;

Algorithm 6: Algorithm $\mathsf{KCenters}(C, k)$ places k points inside C.



(a) (b) (c) (d)

Figure 3.1: The optimal solutions to the 5-center problem in a rectangle, which depend on the dimensions of said rectangle as shown above; the configurations are described formally in [1]. These are necessary for the "very particular edge case" described in the end of Algorithm 6.



(a) (b) (c)

(d) (e) (f)

Figure 3.2: Two executions of Algorithm 6 with $k = 31$ and $k = 15$. We start with a convex polygon with a horizontal diameter and a bounding box Q in 3.2a. We then apply the rectangular partitioning Algorithm 2 for various input values, and the best set of inputs is shown in 3.2b; the two arrows indicate that all rectangles have the same diagonal lengths. Figure 3.2c then shows the final output. Figures 3.2d through 3.2f show the same thing for a different polygon and $k = 15$, except that we also encounter the "very particular edge case" described in the end of Algorithm 6 in which we place 5 points according to Figure 3.1.

3.2 Analysis of Algorithm 6

This section is similar to Section 2.3.2 and is devoted to a proof of the following theorem:

Theorem 16. *Algorithm 6 solves the continuous k-centers problem in a convex polygon C within a factor of* 1.99 *for any $k \geq 6$. Its running time is $\mathcal{O}(m + n \log m)$, where m is the number of edges of the input region.*

The claimed running time is true for the same reasons as in the proof of Theorem 12. We make the same simplifying observations as in Section 2.3.2, namely that:

- We disregard the projection of the points x_i' onto onto C.
- We assume that C has area 1.
- We assume that C is oriented so that its diameter is horizontal and that its bounding box Q has area 2, and hence Q has width w and height $h = 2/w$, with $w \geq h$, i.e. $w \geq \sqrt{2}$.

We next observe that our algorithm certainly attains the desired approximation ratio whenever $w \geq \frac{3}{2}\sqrt{k}$. This is because one of the configurations produced is to simply divide Q into a $k \times 1$ grid of identical rectangles, each having width w/k and height $h = 2/w$, and place each x_i' in the middle of these. Any point $x \in C \subseteq Q$ must be within a distance of

$$\frac{1}{2}\sqrt{\left(\frac{w}{k}\right)^2 + \left(\frac{2}{w}\right)^2}$$

to one of the points x_i' and therefore (using the lower bound $r^* \geq w/(2k)$) the approximation ratio is at most

$$\begin{aligned} \mathsf{Rat} &\leq \frac{\frac{1}{2}\sqrt{\left(\frac{w}{k}\right)^2 + \left(\frac{2}{w}\right)^2}}{w/(2k)} \\ &\leq \frac{\frac{1}{2}(w/k + 2/w)}{w/(2k)} = 1 + 2k/w^2 < 1.99 \end{aligned}$$

as desired. By using the lower bound $r^* \geq 1/\sqrt{\pi k}$, our algorithm also attains the 1.99 approximation whenever $\sqrt{k} \leq w \leq 2\sqrt{k}$ because the same $k \times 1$ configuration has an

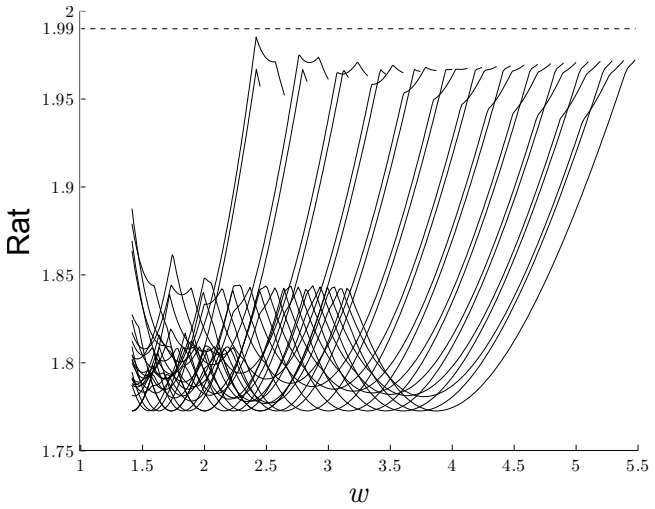


Figure 3.3: The approximation ratios realized by Algorithm 6, for $k \in \{6, \ldots, 30\}$ and $w \in [\sqrt{2}, \sqrt{k})$. These are obtained by dividing the objective value of the output solution (i.e. half of the largest diagonal of all the rectangles) by the maximum of the two lower bounds from Theorem 15.

approximation ratio of

$$\begin{aligned} \mathsf{Rat} &\leq \frac{\frac{1}{2}\sqrt{\left(\frac{w}{k}\right)^2 + \left(\frac{2}{w}\right)^2}}{1/\sqrt{\pi k}} \\ &= \frac{\sqrt{\pi}}{2} \cdot \sqrt{t^2 + 4/t^2} < 1.99 \end{aligned}$$

where we set $t = w/\sqrt{k}$. Thus, from now on, we will assume that $w < \sqrt{k}$.

We will next verify computationally that the approximation ratio of 1.99 holds for $k \leq 30$. Note that for any fixed k, we can consider only those values of w in the finite interval $[\sqrt{2}, \sqrt{k})$. These ratios are shown in the plot in Figure 3.3. Since all values on that plot are less than 1.99, we obtain the desired result, and we are therefore free to assume that $w < \sqrt{k}$ and that $k \geq 31$.

The remainder of our analysis is a case-by-case study where we consider the different values of $q_0 = \lfloor \sqrt{hk/w} \rfloor = \lfloor \sqrt{2k}/w \rfloor$. Because $w \geq h$, it is of course always true that $q_0 \leq p_0$. Suppose that $q_0 \geq 8$, whence $p_0 \geq 8$; we can partition Q into $p_0 q_0$ rectangles, and each of these rectangles has an aspect ratio of at most $(q_0 + 1)/q_0 \leq 9/8$. Thus, each rectangle has area $2/(p_0 q_0)$ and aspect ratio at most $9/8$, and therefore each rectangle's diagonal is at most $\frac{1}{6}\sqrt{\frac{145}{p_0 q_0}}$. The distance from any point $x \in C \subseteq Q$ to the center of the rectangle containing it is at most half of this. By construction, we know that $k = p_0 q_0 + s$,

where $s \leq p_0 + q_0$. Thus, our approximation ratio is at most

$$\begin{aligned} \mathsf{Rat} &\leq \frac{\frac{1}{12}\sqrt{\frac{145}{p_0 q_0}}}{1/\sqrt{\pi k}} = \frac{\frac{1}{12}\sqrt{\frac{145}{p_0 q_0}}}{1/\sqrt{\pi(p_0 q_0 + s)}} \\ &\leq \frac{\frac{1}{12}\sqrt{\frac{145}{p_0 q_0}}}{1/\sqrt{\pi(p_0 q_0 + p_0 + q_0)}} < 1.779\sqrt{1 + \frac{1}{p_0} + \frac{1}{q_0}} < 1.99 \end{aligned}$$

since $p_0, q_0 \geq 8$. Thus, we can assume from now on that $q_0 \in \{0, \ldots, 7\}$. In fact, since we have already concluded that our ratio holds when $w \geq \sqrt{k}$, we see that the case where $q_0 = 0$ is already taken care of (this is because $q_0 = 0 \iff \sqrt{2k}/w < 1 \iff w > \sqrt{2k}$). Therefore, we will now consider the case where $q_0 \in \{1, \ldots, 7\}$ and $n \geq 31$, which will complete the proof. We will assume that $q_0 \geq 2$ and address the case where $q_0 = 1$ in Section A.2 of the appendices.

If $q_0 \geq 2$, we will show that the desired approximation ratio holds when we divide Q into a grid consisting of $\lfloor k/q_0 \rfloor \times q_0$ rectangles or a grid of $\lfloor k/(q_0+1) \rfloor \times (q_0+1)$ rectangles. If we use a $\lfloor k/q_0 \rfloor \times q_0$ grid, then each sub-rectangle has dimensions

$$\frac{w}{\lfloor k/q_0 \rfloor} \times \frac{h}{q_0} = \frac{w}{\lfloor k/q_0 \rfloor} \times \frac{2}{q_0 w}$$

and half of the diagonal of each sub-rectangle (which, as we have already seen, is the largest distance between a point $x \in C \subseteq Q$ and its nearest neighbor x_i) is

$$\begin{aligned} \frac{1}{2}\sqrt{\left(\frac{w}{\lfloor k/q_0 \rfloor}\right)^2 + \left(\frac{2}{q_0 w}\right)^2} &\leq \frac{1}{2}\sqrt{\left(\frac{w}{k/q_0 - 1}\right)^2 + \left(\frac{2}{q_0 w}\right)^2} \\ &= \sqrt{\frac{q_0^2 w^2}{4(k-q_0)^2} + \frac{1}{q_0^2 w^2}} = \sqrt{\frac{t^2}{2kq_0^2} + \frac{kq_0^2}{2(k-q_0)^2 t^2}} \end{aligned} \quad (3.1)$$

where we define $t = \sqrt{2k}/w$, which must satisfy $q_0 \leq t < q_0 + 1$. We similarly find that the $\lfloor k/(q_0+1) \rfloor \times (q_0+1)$ grid gives rectangles whose half-diagonals are at most

$$\frac{1}{2}\sqrt{\left(\frac{w}{\lfloor k/(q_0+1) \rfloor}\right)^2 + \left\lceil \frac{2}{(q_0+1)w} \right\rceil^2} \leq \sqrt{\frac{t^2}{2k(q_0+1)^2} + \frac{k(q_0+1)^2}{2(k-q_0-1)^2 t^2}}\,. \quad (3.2)$$

We are therefore interested in showing that the minimum of (3.1) and (3.2) is bounded above by $1.99/\sqrt{\pi k}$ for $q_0 \in \{2, \ldots, 7\}$ and all $k \geq 31$. Suppose that q_0 and k are fixed, and consider (3.1) and (3.2) as functions of t; since each of these is the square root of a convex function in t, they are both maximized when t is as large or as small as possible, i.e. $t = q_0$ or $t = q_0 + 1$. The function of t defined by taking the minimum of (3.1) and (3.2) must

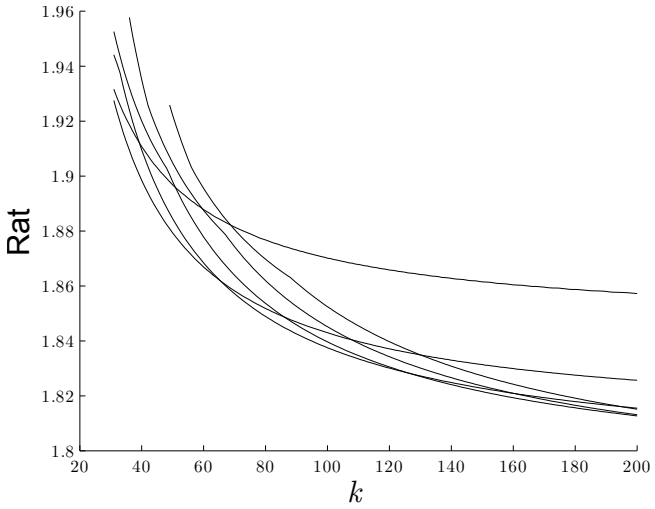


Figure 3.4: The value of the left-hand side of (3.3), for $q_0 \in \{2, \ldots, 7\}$ and $k \geq \max\{q_0^2, 31\}$. It is entirely straightforward (albeit tedious) to verify algebraically that the ratio is always less than 1.99.

be maximized either at one of these values, or at a value of t such that (3.1) and (3.2) are equal. Such values of t can be computed analytically because both terms are of the form $\sqrt{\alpha t^2 + \beta/t^2}$ for constant terms α, β. Thus, we can restrict ourselves to a finite set of values $t \in T$, and our goal is to show that

$$\frac{\min\left\{\sqrt{\frac{t^2}{2kq_0^2} + \frac{kq_0^2}{2(k-q_0)^2t^2}}\,,\ \sqrt{\frac{t^2}{2k(q_0+1)^2} + \frac{k(q_0+1)^2}{2(k-q_0-1)^2t^2}}\right\}}{1/\sqrt{\pi k}} \leq 1.99$$

for all $k \geq 31$, $q_0 \in \{2, \ldots, 7\}$, and all $t \in [q_0, q_0 + 1)$. By the preceding argument, we can remove the dependency on t, so that we must show that

$$\max_{t \in T} \frac{\min\left\{\sqrt{\frac{t^2}{2kq_0^2} + \frac{kq_0^2}{2(k-q_0)^2t^2}}\,,\ \sqrt{\frac{t^2}{2k(q_0+1)^2} + \frac{k(q_0+1)^2}{2(k-q_0-1)^2t^2}}\right\}}{1/\sqrt{\pi k}} \leq 1.99 \tag{3.3}$$

which now depends only on q_0 and k (since the finite set T is determined by q_0 and k). Since there are only 6 different values of q_0 that are of interest, we therefore are left with 6 different univariate functions of k alone, and these are shown in Figure 3.4. It is entirely straightforward (albeit tedious) to verify algebraically that the desired results hold, and we omit this for brevity. This completes the proof for the case where $q_0 \neq 1$ and we refer the reader to Section A.2 of the appendices for the remaining analysis.

Chapter 4

Worst-Case Demand Distributions in VRP under Limited Information

Worst-Case Demand Distributions in VRP under Limited Information

In Chapter 3 we studied a classical location problem. In many cases a location problem is followed by a routing problem like vehicle routing problem (VRP) or travelling salesman problem (TSP), since the located facilities are supposed to provide services to some demand points and that usually requires designing a service network that can have different configurations like a star shape, TSP tour, mTSP (multiple TSP) tour, VRP tour or a mixture thereof. We study vehicle routing problem in this chapter and travelling salesman problem in next chapter.

A recent focal point in research on the vehicle routing problem is the issue of robustness in which customer demand is uncertain. In this chapter, we consider the case where our knowledge about the demand distribution is limited to the information about the moments of the distribution. We conduct a theoretical analysis of the demand distributions whose induced workloads are as undesirable as possible. We study two common variations of VRP in a continuous approximation setting: the first is the VRP with time windows, and the second is the capacitated VRP, in which regular returns to the vehicle's point of origin are required.

4.1 Introduction

Since its original formulation in 1959, two of the primary features that have distinguished the *vehicle routing problem* (VRP) from the *travelling salesman problem* (TSP) have been the introduction of *capacities* on vehicles that originate from a central depot [100] and

the imposition of *time windows* that constrain the times when customers can be visited [101]. Not surprisingly, the imposition of such constraints presents a major obstacle in obtaining solutions to a problem instance, both in terms of the added computational burden of finding solutions and in the overall quality of the solution itself. As identified in [102], one useful feature of the capacitated VRP is that one can actually describe the additional cost somewhat concretely:

> Any solution for the capacitated VRP has two cost components; the first component is proportional to the total "radial" cost between the depot and the customers. The second component is proportional to the "circular" cost; the cost of traveling between customers. This cost is related to the cost of the optimal traveling salesman tour. It is well known [51] that, for large N, the cost of the optimal traveling salesman tour grows like $\sqrt{N}$, while the total radial cost between the depot and the customers grows like N.... Therefore, it is intuitive that when the number of customers is large enough the first cost component will dominate the optimal solution value.

The additional cost due to time windows is more difficult to quantify, although insights can be made under certain assumptions [103]:

> Imagine an extreme case, where only a tiny fraction of the customers have very stringent...time window constraints. Because, as we shall see, the distance travelled increases with [the square root of the number of time windows], the total system cost may be large because of the requirements of very few customers.

This chapter addresses VRP from the perspective of a *continuous approximation model*: we assume that a fleet of vehicles must provide service to a contiguous planar geographic region, and our goal is to quantify precisely the role that vehicle capacities and time windows play in the worst-case workloads of the vehicles. We assume that customer demands are independently sampled from a (possibly unknown) demand distribution, and study the asymptotic behavior of the worst-case distributions as the number of customers becomes large. In this sense, our work is philosophically similar to (for example) [104], which analytically determines trade-offs between transportation and inventory costs, [105], which shows how to route emergency relief vehicles to beneficiaries in a time-sensitive manner, and [106], which describes a simple geometric model for determining the optimal mixture of a fleet of vehicles that perform distribution. The basic premise of the continuous approximation paradigm is that one replaces combinatorial quantities that are difficult to compute

with simpler mathematical formulas, which (under certain conditions) provide accurate estimations of the desired quantity[107, 108, 109, 110]. Such approximations exist for many combinatorial problems, such as the travelling salesman problem [51, 111], facility location [112, 27, 11], and any *subadditive Euclidean functional* such as a minimum spanning tree, Steiner tree, or matching [113, 83, 82]. In our computational districting experiment, an approximation of this kind is used as the first level of an optimization problem in which we design service zones that are associated with different vehicles. for example. Our study of the VRP with time windows adopts similar assumptions to those of [103, 114], namely, that the service period is divided into a collection of pre-specified intervals of equal duration. One might contrast this model with other approaches like [115], which assumes that time windows are independently drawn from an arbitrary probability measure. Our study of the capacitated VRP makes extensive use of upper and lower bounds derived in [116, 117, 118], as well as seminal results on the TSP that can be found in [51, 113, 83].

A more recent focal point in research on VRP and its variants is the isssue of *robustness* in which one seeks a policy that performs as well as possible against all possible realizations of demand that are compatible with some set of observations or initial conditions. Robust methodologies for the capacitated VRP were first introduced in the paper [63], which adapts the methodology of [59] to solve problems in which customer demands and travel times are uncertain; the goal is to find vehicle routes that meet all feasibility requirements in the worst-case scenario, which occurs precisely when all customer demands and travel times attain their worst-case realizations simultaneously. In most models of the robust VRP, one has a pre-defined ambiguity region and seeks a set of routes that is as good as possible with respect to all of the outcomes; this ambiguity region is usually described as a finite collection of scenarios or a polyhedral set [119, 120, 121, 122, 123, 63], although the recent paper [124] adopts a "robust mean-variance" approach that minimizes a weighted sum of the average cost and the variance of a route when sampled over many scenarios. In our problem, we are concerned with robustness in the *distributional* sense [125]: we seek the distribution of demand for which the expected cost of a tour is as high as possible, while remaining consistent with some observed data samples or some parameters derived thereof. The most closely related result to our work is [65], which determines the worst-case spatial demand distribution for the TSP when the first and second moments are fixed. Our work can be seen as a generalization of these principles to the cases where vehicles have capacities and time window constraints.

Our present work uses the notion of robustness to study the negative consequences of fluctuation in demand for delivery services, in either a spatial or temporal sense. Demand

fluctuation is of particular concern for emerging delivery services such as Good Eggs, DoorDash, BiteSquad, and Caviar [126, 127, 128, 129], which face extremely high volatility in demand due to seasonality and the time-sensitive nature of the requests they satisfy [130, 131]:

> Our business is highly dependent on diner behavior patterns that we have observed over time. In our metropolitan markets, we generally experience a relative increase in diner activity from September to April and a relative decrease in diner activity from May to August. In addition, we benefit from increased order volume in our campus markets when school is in session and experience a decrease in order volume when school is not in session, during summer breaks and other vacation periods. Diner activity can also be impacted by colder or more inclement weather, which typically increases order volume, and warmer or sunny weather, which typically decreases order volume. Seasonality will likely cause fluctuations in our financial results on a quarterly basis. In addition, other seasonality trends may develop and the existing seasonality and diner behavior that we experience may change or become more extreme.

In total, this chapter makes the following contributions: Section 4.3 analyzes the vehicle routing problem with time windows, characterizing the worst-case distributions that can arise when demand varies over a specified time horizon. Section 4.4 deals with the capacitated VRP, and Section 4.5 extends this analysis to more sophisticated models in which we have information about the mean or covariance of the demand distribution and describes some computational experiments.

4.2 Preliminaries

We make the following notational conventions in this chapter: given a point set X, the *star network* of X is written $\mathsf{SN}(X)$ and consists of the network in which each point in X is connected to some central "depot" point (the location of this central point will be made clear from context). Vehicle capacities are either denoted by the letter c, indicating that a vehicle can visit c destinations before returning to its depot, or by the capacity coefficient t, which satisfies the relationship $c = t\sqrt{|X|}$; this is a standard and useful representation, as can be seen in Section 4.2 of [132] or the paper [117]. A TSP tour of a set of points will be denoted by $\mathsf{TSP}(X)$. A capacitated VRP tour of a set of points is written $\mathsf{VRP}(X)$, where we suppress the capacity in the interest of notational brevity. Finally, we say that $f(x) \in o(g(x))$ if $\lim_{x\to\infty} f(x)/g(x) = 0$, we say that $f(x) \sim g(x)$ if $\lim_{x\to\infty} f(x)/g(x) = 1$,

we say that $f(x) \in \mathcal{O}(g(x))$ if $f(x) \leq \gamma g(x)$ for some $\gamma > 0$ and all x exceeding some threshold x_0, and we say that $f(x) \in \Omega(g(x))$ if $f(x) \geq \gamma g(x)$ for some $\gamma > 0$ and all x exceeding some threshold x_0.

To approximate the length of a TSP tour of a collection of points, we will use the well-known *BHH Theorem* [51], which says that the length of an optimal TSP tour of a set of points follows a law of large numbers:

Theorem 17. *Suppose that $X = \{X_1, X_2, \dots\}$ is a sequence of random points i.i.d. according to a probability density function $f(\cdot)$ defined on a compact planar region $\mathcal{R}$. Then with probability one, the length of $\mathsf{TSP}(X)$ satisfies*

$$\lim_{N\to\infty} \frac{\text{length}(\mathsf{TSP}(X))}{\sqrt{N}} = \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$$

where β is a constant and $\bar{f}(\cdot)$ represents the absolutely continuous part of $f(\cdot)$.

It is additionally known that $0.6250 \leq \beta \leq 0.9204$ and estimated that $\beta \approx 0.7124$; see [133, 51]. Theorem 31 can also be expressed deterministically, removing any assumptions about the distribution of the points X_i; see for example [134, 135]:

Theorem 18. *There exists a constant α satisfying the following: if $X = \{X_1, X_2, \dots\}$ is any sequence of points contained in a compact planar region $\mathcal{R}$ with area 1, then*

$$\limsup_{N\to\infty} \frac{\mathsf{TSP}(X_1, \dots, X_N)}{\sqrt{N}} \leq \alpha\,.$$

Furthermore, it is also true that $(4/3)^{1/4} \leq \alpha < 1.392$.

The paper [136] actually proves a stronger result that these worst-case point sets must be uniformly distributed in an asymptotic sense.

The following result from [118] gives upper and lower bounds for a capacitated VRP tour:

Theorem 19. *For any set X of demand points serviced by a fleet of vehicles with capacity c that originate from a single depot, we have*

$$\max\left\{\frac{2}{c}\,\text{length}(\mathsf{SN}(X)), \text{length}(\mathsf{TSP}(X))\right\} \leq \text{length}(\mathsf{VRP}(X)) \leq 2\left\lceil\frac{|X|}{c}\right\rceil \cdot \frac{\text{length}(\mathsf{SN}(X))}{|X|} + (1 - 1/c)\,\text{length}(\mathsf{TSP}(X))\,. \tag{4.1}$$

4.3 Worst-case temporal demand distributions

In this section, we will study the worst-case *temporal* distributions for a particular version of the vehicle routing problem with time windows. In order to make this problem tractable, we assume as in [103] that the service period is divided into $m \geq 2$ equally long time periods and that each customer is placed into one of these time periods, so that each period i has n_i customers. In order to isolate the temporal aspects of our problem, we will assume that all demand is uniformly distributed throughout the service region, which has area 1.

Since each period has a finite length, there is a maximum distance ℓ that each vehicle can travel within each time window (equal to the speed of the vehicle multiplied by the duration of the time window), which thereby induces a limit on the number of customers that can be visited in that time period. Typically, delivery services charge a flat fee for each delivery [126, 127], and thus the revenue from providing service to customers is simply linearly proportional to the number of customers. In order to simplify our analysis, we will assume that one vehicle, capable of traversing a distance ℓ in each time period, provides service to the n_i customers (or as many of the n_i customers as possible) in each time period i. We assume that there is a fixed price p that the company charges for each order and that there is a price q that the company incurs per unit distance of travel. Thus, if the length of the TSP tour of these points $X_1, \ldots, X_{n_i}$ is less than ℓ, then the company receives a net profit of

$$pn_i - q \operatorname{length}(\mathsf{TSP}(X_1, \ldots, X_{n_i}))$$

in that time period. By applying Theorem 31, we can approximate $\operatorname{length}(\mathsf{TSP}(X_1, \ldots, X_{n_i})) \approx \beta\sqrt{n_i}$. If the length of the TSP tour exceeds ℓ – or equivalently, if $\beta\sqrt{n_i} > \ell$ – then the company's goal is to make as much money as possible (i.e. visit as many customers as possible), without violating this distance constraint. In order to study this scenario, the following theorem is useful:

Theorem 20. *Suppose that $\{X_1, \ldots, X_n\}$ are uniformly distributed in the unit square and let $\ell > 0$. Let N_ℓ denote the maximum number of points X_i that can be visited with a path of length ℓ. Then $\mathbf{E}(N_\ell) \in \Omega(\ell\sqrt{n})$ and $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ as $n \to \infty$. In particular, we have $\mathbf{E}(N_\ell) \geq \alpha^{-1}\ell\sqrt{n}$ and $\mathbf{E}(N_\ell) \leq 3.41\ell\sqrt{n}$ for sufficiently large n, where α is the constant from Theorem 18.*

Before proving this result, the following lemma is helpful:

Lemma 21. *Suppose that $\{X_1, \ldots, X_n\}$ are uniformly distributed in the unit square and let $\ell > 0$. Then*

$$\Pr(\text{length}(\mathsf{TSP}(X_1, \ldots, X_n)) \leq \ell) \leq \frac{n!}{(2n-2)!}(2\pi\ell^2)^{n-1}$$

as $n \to \infty$.

Proof. See Section A.4 of the online supplement. □

The proof of Theorem 20 follows:

Proof of Theorem 20. The fact that $\mathbf{E}(N_\ell) \in \Omega(\ell\sqrt{n})$ is straightforward: simply take a TSP tour of all n points, whose length L satisfies $L \leq \alpha\sqrt{n}$ as in Theorem 18. If we divide this tour into segments of length ℓ, we are then left with at most $\lceil \alpha\sqrt{n}/\ell \rceil$ segments which cumulatively touch all n points. If we select one of these segments uniformly at random, the expected number of points in the segment is at least $n/\lceil \alpha\sqrt{n}/\ell \rceil \sim \alpha^{-1}\ell\sqrt{n}$, as desired.

The proof that $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ is trickier: using Lemma 21, and letting $c = 3.41 > e\sqrt{\pi/2}$, we can write

$$\begin{aligned}
\mathbf{E}(N_\ell) &\leq c\ell\sqrt{n} + n\Pr(N_\ell \geq c\ell\sqrt{n}) \\
&= c\ell\sqrt{n} + n\Pr(\text{there exists a path through } \lceil c\ell\sqrt{n} \rceil \text{ of the } n \text{ points with length} \leq \ell) \\
&\leq c\ell\sqrt{n} + np_\ell \binom{n}{\lceil c\ell\sqrt{n} \rceil}
\end{aligned}$$

by the union bound, where p_ℓ is the probability that a uniformly distributed set of $\lceil c\ell\sqrt{n} \rceil$ points $\{X_1, \ldots, X_{\lceil c\ell\sqrt{n} \rceil}\}$ has a TSP tour whose length does not exceed ℓ. It will suffice to confirm that $np_\ell \binom{n}{\lceil c\ell\sqrt{n} \rceil} \to 0$ as $n \to \infty$. By Lemma 21, we can substitute for p_ℓ, giving

$$\begin{aligned}
np_\ell \binom{n}{\lceil c\ell\sqrt{n} \rceil} &\leq \frac{\lceil c\ell\sqrt{n} \rceil!}{(2\lceil c\ell\sqrt{n} \rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n} \rceil - 1} \cdot n \cdot \binom{n}{\lceil c\ell\sqrt{n} \rceil} \\
&\leq \frac{\lceil c\ell\sqrt{n} \rceil!}{(2\lceil c\ell\sqrt{n} \rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n} \rceil - 1} \cdot \frac{n^{\lceil c\ell\sqrt{n} \rceil + 1}}{\lceil c\ell\sqrt{n} \rceil!} \\
&= \frac{n^{\lceil c\ell\sqrt{n} \rceil + 1}}{(2\lceil c\ell\sqrt{n} \rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n} \rceil - 1} \\
&\leq \frac{n(2c\ell\sqrt{n} - 2)^{3/2}}{2\pi\sqrt{2\pi}\ell^2 e^2} \cdot \left[\frac{2\pi\ell^2 e^2 n}{(2c\ell\sqrt{n} - 2)^2}\right]^{c\ell\sqrt{n}},
\end{aligned}$$

where we have used Stirling's approximation [137, 138] in the last inequality. Note that the Stirling's inequality also dominates the approximation of the ceiling function. Since $c > e\sqrt{\pi/2}$, the bracketed quantity is eventually less than 1 for sufficiently large n, at which point the above quantity approaches 0 since the left term is polynomially increasing but the right term is super-polynomially decreasing. Thus, we conclude that $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ as desired, which completes the proof. □

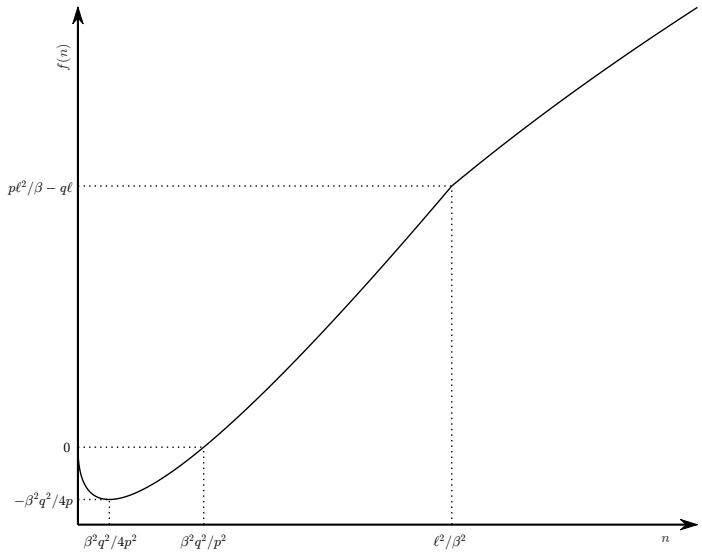


Figure 4.1: The profit function $f(n)$ as defined in (4.2).

The preceding theorem establishes that, if n points are sampled in a region with area 1 and $\ell < \beta\sqrt{n}$, then the maximum number of points that can be visited using a path of length ℓ is proportional to $\ell\sqrt{n}$. In order to impose continuity between the domains $\ell < \beta\sqrt{n}$ and $\ell \geq \beta\sqrt{n}$, we approximate this maximum number of points as $\ell\sqrt{n}/\beta$. Thus, we propose the following profit function when n points are sampled in a region with area 1 with p and q:

$$f(n) := \begin{cases} pn - q\beta\sqrt{n} & \text{if } \beta\sqrt{n} \leq \ell \\ p\ell\sqrt{n}/\beta - q\ell & \text{otherwise .} \end{cases} \tag{4.2}$$

The profit function $f(n)$ with the selection of $\ell = 3.5,\ p = 1,\ q = 4$ is shown in Figure 4.1, and is convex until $n = (\ell/\beta)^2$ (and concave thereafter), with a minimizer at $n = \frac{q^2\beta^2}{4p^2}$. We will assume that $\frac{\ell p}{\beta^2 q} \geq 1$, which implies that $f(n) \geq 0$ when $\beta\sqrt{n} = \ell$ (in other words, the company is profitable when the TSP length of demand points matches the limit ℓ exactly). If we assume that the aggregate demand over m time periods is a constant N, then the worst-case demand distribution over m time periods is then obtained by solving the optimization problem

$$\begin{aligned} \underset{n_1,\ldots,n_m}{\text{minimize}} \sum_{i=1}^{m} f(n_i) \qquad & s.t. \\ \sum_{i=1}^{m} n_i \quad &= \quad N \end{aligned} \tag{4.3}$$

for all $n_i \geq 0$. The following theorem characterizes these worst-case distributions; we either have a large amount of demand concentrated in a single time period, or uniform demand throughout all time periods:

Theorem 22. *The optimal solution $n_1^*, \ldots, n_m^*$ to problem (4.3) takes one of two forms:*

1. *Either $n_i^* = N/m$ for all n_i^*, or*
2. *A single largest entry (say n_1^*) satisfies*

$$n_1^* = (z^*)^2 > \ell^2/\beta^2\,,$$

where z^* is a solution to the quartic equation

$$4\beta^2p^2z^4 - 4\beta\ell p^2z^3 + (\ell^2p^2 + \beta^4mq^2 - \beta^4q^2 - 4N\beta^2p^2)z^2 + 4N\beta\ell p^2z - N\ell^2p^2 = 0\,,$$

and all other entries n_i^* satisfy

$$\frac{\beta^2q^2}{4p^2} < n_i^* = \frac{N - n_1^*}{m-1} < \frac{\beta^2q^2}{p^2}.$$

Furthermore, if

$$N \leq \frac{1}{4}\left(\frac{\ell}{\beta} + \frac{q\beta\sqrt{m}}{p}\right)^2,$$

then the optimal solution takes the first form, and if

$$N \geq \left(\frac{\ell}{\beta} + \frac{q\beta\sqrt{m}}{p}\right)^2,$$

the optimal solution takes the second form.

Proof. It will suffice to consider the equivalent problem

$$\begin{aligned} \underset{x_1,\ldots,x_m}{\text{minimize}} \sum_{i=1}^m g(x_i) \qquad & s.t. \\ \sum_{i=1}^m x_i \;&=\; c \end{aligned}$$

for all $x_i \geq 0$, where we define

$$g(x) := a \cdot \begin{cases} \frac{1}{b}\left(x - \sqrt{x}\right) & \text{if } x \leq b^2 \\ \sqrt{x} - 1 & \text{otherwise}\,; \end{cases}$$

this is evident by applying the transformation $a \mapsto q\ell$, $b \mapsto \frac{p\ell}{q\beta^2}$, $c \mapsto \frac{p^2 N}{q^2\beta^2}$ and $x_i \mapsto \frac{p^2 n_i}{q^2\beta^2}$. The two cases described above are equivalent to

1a. Either $x_i^* = c/m$ for all x_i^*, or

2a. A single largest entry (say x_1^*) satisfies

$$x_1^* = (z^*)^2 > b^2\,,$$

where z^* is a solution to the quartic equation

$$4z^4 - 4bz^3 + (b^2 + m - 4c - 1)z^2 + 4bcz - b^2c = 0\,,$$

and all other entries x_i^* satisfy

$$1/4 < x_i^* = (c - x_1^*)/(m-1) < 1.$$

Our assumption that $\frac{\ell p}{\beta^2 q} \geq 1$ is equivalent to an assumption that $b \geq 1$. First of all, it is obvious that there are no minimizers other than cases 1a and 2a; this is an immediate consequence of the fact that the function $g(\cdot)$ is convex on its first component and concave on its second. The quartic function in case 2a simply arises from equating derivatives in the first and second components, and the fact that $1/4 < x_i^* < 1$ (also in case 2a) is also implied by this (the derivative of the second component can only take values between 0 and $a/2b$). Our proof will be complete if we can show that case 1a is optimal when $c \leq \frac{1}{4}(b + \sqrt{m})^2$ and that case 2a is optimal when $c \geq (b + \sqrt{m})^2$. First, note that we have $1/4 < (c - x_1^*)/(m-1) < (c - b^2)/(m-1)$, which we can rearrange for c to obtain

$$b^2 + m/4 - 1/4 < c \leq \frac{1}{4}(b + \sqrt{m})^2\,.$$

We will also assume without loss of generality that $a = 1$.

If $c \leq \frac{1}{4}(b + \sqrt{m})^2$, then we will verify that case 1a is optimal by taking a lower bound of case 2a. Since case 2a must have $g(x_1^*) > b - 1$ and $g(x_i^*) > -1/4b$ for all other i, a valid lower bound is simply $b - 1 - \frac{m-1}{4b}$. We now want to show that case 1a is optimal, i.e. that

$$\left[b - 1 - \frac{(m-1)}{4b}\right] - \frac{1}{b}\left(c - \sqrt{cm}\right) \geq 0\,.$$

The left-hand side of the above is concave in c and is therefore minimized at the endpoints $c = b^2 + m/4 - 1/4$ and $c = 1/4(b + \sqrt{m})^2$; it is routine to verify that the desired inequality

holds by using the fact that $b \geq 1$ and $m \geq 2$.

If $c \geq (b + \sqrt{m})^2$, it will suffice to verify that case 1a is sub-optimal; this is immediate because we can simply set $x_1 = c$ and $x_i = 0$ otherwise, which would incur a cost of $\sqrt{c} - 1$, as opposed to a cost of $(c - \sqrt{cm})/b$. We then want to show that

$$\frac{1}{b}(c - \sqrt{cm}) - (\sqrt{c} - 1) = 1 - (1 + \sqrt{m}/b)\sqrt{c} + c/b \geq 0\,;$$

the left-hand side of the above is decreasing in m and is therefore minimized when m is as large as possible, namely when $c = (b + \sqrt{m})^2$, in which case the above expression is precisely equal to 1, which completes the proof. □

Theorem 22 distinguishes between two worst-case demand distributions: either the workloads are uniform across time windows, or there is one time window that contains the vast majority of the work, with all other time windows being non-profitable. One can determine which of these distributions is worst via the quantity $N(\ell/\beta + q\beta\sqrt{m}/p)^{-2}$.

4.4 Worst-case scenarios for the capacitated VRP

In this section, we assume that a fleet of vehicles must originate from a single depot in a service region and visit a set of demand points X. These vehicles are *capacitated*, in the sense that they must return to the depot after visiting a specified number of destinations c (in other words, we assume that the "loads", so to speak, at each of the points $x_i \in X$ are the same). Our goal is to describe the *spatial* distributions of demand that maximize the workloads of the vehicles that provide service.

To begin this section, we re-phrase Theorem 38 in a probabilistic sense as follows: suppose that demand points $\{X_1, \ldots, X_N\}$ are independent samples from a probability distribution $f(\cdot)$ defined on a compact subset $\mathcal{R}$ of the plane. Assume without loss of generality that the single depot point is the origin $(0, 0) \in \mathbb{R}^2$. It is then immediately clear that the average length of the star network $\mathsf{SN}(X)$ is simply $\mathbf{E}\,\mathrm{length}(\mathsf{SN}(X)) = N \iint_{\mathcal{R}} \|x\| f(x)\, dA$. We assume, as explained in the introduction, that our capacity constraints take the form $c = t\sqrt{N}$. We will consider the expected length of the VRP tour of points that are sampled from $f(\cdot)$. By exchanging the expectation and $\max\{\cdot, \cdot\}$ operators, we can express the

bound (5.12) as

$$\begin{aligned} & \max\left\{\frac{2\sqrt{N}}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA,\ \beta\sqrt{N}\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA+o(\sqrt{N})\right\} \\ \leq\ & \mathbf{E}\,\text{length}(\mathsf{VRP}(X)) \\ \leq\ & 2\left\lceil\frac{\sqrt{N}}{t}\right\rceil\iint_{\mathcal{R}}\|x\|f(x)\,dA+\left(1-\frac{1}{\sqrt{N}\cdot t}\right)\beta\sqrt{N}\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA+o(\sqrt{N})\,. \end{aligned}$$

Note that $\left\lceil\sqrt{N}/t\right\rceil$ is simply the number of trips needed to provide service. Since we are interested in the limiting behavior as $N\to\infty$, we can therefore write

$$\sqrt{N}\cdot\max\left\{\frac{2}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA,\ \beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right\}\lesssim\mathbf{E}\,\text{length}(\mathsf{VRP}(X))\lesssim\sqrt{N}\cdot\left(\frac{2}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA+\beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right) \tag{4.4}$$

where we have adopted the notation "$\lesssim$" to denote the "approximate" inequality, neglecting the lower-order terms (large values of N imply that $\left\lceil\sqrt{N}/t\right\rceil\approx\sqrt{N}/t$ and that $1/(\sqrt{N}\cdot t)\approx 0$). We are free to neglect the $o(\sqrt{N})$ terms since we are concerned with the worst-case demand distribution $f^*(\cdot)$ and we are simply assuming that N is large. Note that the two terms in the $\max\{\cdot\}$ operator on the left-hand side of the above represent the two sources of cost as identified in the introduction. If we search for distributions that maximize either of these expressions in isolation, the trade-off between these costs becomes clear: if we consider only the worst-case "radial" cost by maximizing $2/t\iint_{\mathcal{R}}\|x\|f(x)\,dA$, then the resulting distribution will simply be an atomic mass located at the farthest point from the depot, $\arg\max_{x\in\mathcal{R}}\|x\|$. Of course, since such a distribution has no continuous component whatsoever, the corresponding "circular" cost is zero. Similarly, if we consider only the worst-case "circular" cost by maximizing $\beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA$, it is not hard to verify (by a simple application of Jensen's inequality) that the resulting distribution is uniform on $\mathcal{R}$. The corresponding "radial" cost, while not equal to zero, is then $2/t\iint_R\|x\|\,dA$, substantially less than that obtained from the atomic distribution.

We are now interested in determining the distribution $f^*(\cdot)$ that maximizes $\mathbf{E}\,\text{length}(\mathsf{VRP}(X))$, which we will bound from above by maximizing the right-hand side of (4.4). The concave maximization problem given by

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}}\ \iint_{\mathcal{R}}\frac{2}{t}\|x\|f(x)+\beta\sqrt{f(x)}\,dA \qquad & s.t. \\ \iint_{\mathcal{R}}f(x)\,dA &= 1 \\ f(x) &\geq 0 \quad \forall x\in\mathcal{R}\,, \end{aligned} \tag{4.5}$$

with $f(\cdot)$ belonging to (for example) the Banach space L^2 over $\mathcal{R}$, can readily be approached using standard techniques of vector space optimization [139], and admits a Lagrangian dual problem in a single variable given by

$$\begin{aligned} \underset{\nu}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu - \frac{2}{t}\|x\|}\, dA + \nu \qquad & s.t. \\ \nu \;\; & \geq \;\; \frac{2}{t}\|x\| \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{4.6}$$

The optimality conditions for the dual problem, which say that

$$\iint_{\mathcal{R}} \frac{\beta^2}{4(\nu - \frac{2}{t}\|x\|)^2}\, dA = 1\,,$$

immediately imply the appropriate distributional form for $f^*(\cdot)$:

Theorem 23. *The optimal solution $f^*(\cdot)$ that solves (4.5) is of the form*

$$f^*(x) = \frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2}\,,$$

where $\nu^* \geq \frac{2}{t}\max_{x\in\mathcal{R}}\|x\|$ is the unique scalar such that $f^*(\cdot)$ integrates to one over the region.

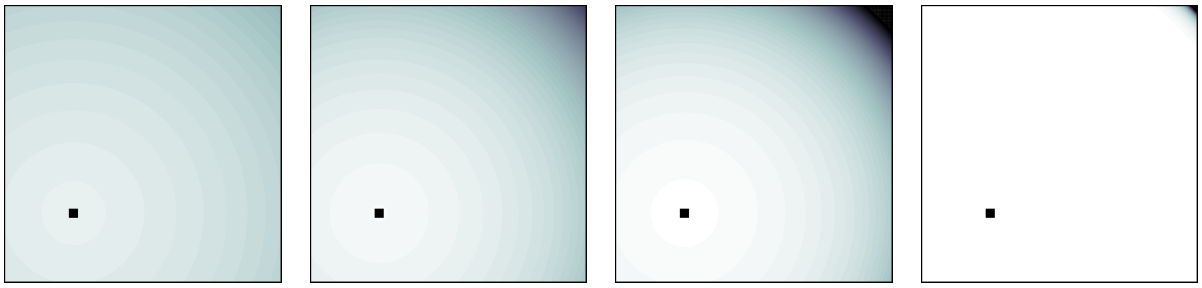
Proof. Thanks to weak duality (which says that any feasible solution to (4.5) has an objective value that is at most equal to that of a feasible solution to (4.6)), it will suffice to verify that the primal and dual objective values are the same when $f^*(\cdot)$ and ν^* are as defined above:

$$\begin{aligned} \iint_{\mathcal{R}} \frac{2}{t}\|x\| f^*(x) + \beta\sqrt{f^*(x)}\, dA &= \iint_{\mathcal{R}} \left(\frac{2}{t}\|x\| \left[\frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \right] + \beta \sqrt{\frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2}} \right) dA \\ &= \iint_{\mathcal{R}} \left(\frac{1}{4} \cdot \frac{\beta^2}{\left(\nu^* - \frac{2}{t}\|x\|\right)^2} \left(\nu^* - \frac{2}{t}\|x\| \right) + \frac{\nu^* \beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \right) dA \\ &= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\left(\nu^* - \frac{2}{t}\|x\|\right)^2} \left(\nu^* - \frac{2}{t}\|x\| \right) dA + \nu^* \underbrace{\iint_{\mathcal{R}} \frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2}\, dA}_{=1} \\ &= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* - \frac{2}{t}\|x\|}\, dA + \nu^* \end{aligned}$$

as desired. Figure 4.2 shows a collection of surface plots of the worst-case distribution $f^*(\cdot)$ on the unit square for varying values of t.

□

One of the salient attributes of the worst-case distribution $f^*(\cdot)$ as written above is that



(a) $t = 0.9$; $\nu^* = 2.5$ (b) $t = 0.5$; $\nu^* = 4.5$ (c) $t = 0.3$; $\nu^* = 7.5$ (d) $t = 0.1$; $\nu^* = 22.5$

Figure 4.2: Surface plots of $f^*(\cdot)$ with decreasing values of t from left to right. The depot is indicated by the black square and darker color values correspond to higher densities.

the presence of the square root in the objective of (4.5) establishes an inverse proportionality between the optimal solution $f^*(\cdot)$ and the *square* of the distance to the depot (with some additional additive and multiplicative weights from the dual variable ν^* and the parameter t). This same inverse proportionality is shared by the classical *geographic gravity model* [140, 141, 142], which is "the most common formulation of the spatial interaction method" [141] and has historically been used to model a wide variety of demographic phenomena such as population migration [143], spatial utility for retail stores [144], and trip distributions between cities [145]. This would appear to lend credibility to our expression $f^*(\cdot)$, inasmuch as it takes a form that closely matches that of distributions for related problems.

Note that in proving Theorem 23, we maximized the right-hand side of (4.4), which is an upper bound of $\mathbf{E}\,\mathrm{length}(\mathsf{VRP}(X))$. It is natural to wonder how much error we are incurring by maximizing something that is itself already an upper bound of the quantity of interest. To this end we offer the following:

Theorem 24. *For any convex planar region $\mathcal{R}$ that contains the origin and any scalar t, we have*

$$\frac{\max_{f(\cdot)} \frac{2}{t} \iint_{\mathcal{R}} \|x\| f(x)\, dA + \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA}{\max_{f(\cdot)} \max\left\{ \frac{2}{t} \iint_{\mathcal{R}} \|x\| f(x)\, dA, \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA \right\}} \leq 3/2\,, \tag{4.7}$$

where the maximization "$\max_{f(\cdot)}$" is taken over all probability distributions $f(\cdot)$ defined on $\mathcal{R}$.

Proof. See Section A.5 of the Online Supplement. Note that it is obvious that the above ratio is bounded above by 2 because the numerator is the sum of the two terms in the $\max\{\cdot,\cdot\}$ expression in the denominator. □

4.5 Worst-case distributions with moment information

Our recent article [65] studies the uncapacitated version of (4.5) (that is, the TSP), where we are presented with additional constraints on the first and second moments of the distribution $f(\cdot)$; we require that $\iint_{\mathcal{R}} xf(x)\,dA = \mu$ for some given $\mu \in \mathcal{R}$ and that $\iint_{\mathcal{R}} xx^T f(x)\,dA \preceq \Sigma + \mu\mu^T$ where $\Sigma \succ \mathbf{0}$ is a given covariance matrix. In practice, these values of μ and Σ would be obtained by observing historical data. In this section, we study the problem (4.5) in which these moment constraints are present, so that the new problem of interest is written as

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{R}} \frac{2}{t}\|x\| f(x) + \beta\sqrt{f(x)}\,dA \qquad & s.t. \\ \iint_{\mathcal{R}} xf(x)\,dA &= \mu \\ \iint_{\mathcal{R}} xx^T f(x)\,dA &\preceq \Sigma + \mu\mu^T \\ \iint_{\mathcal{R}} f(x)\,dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,, \end{aligned} \tag{4.8}$$

with $f(\cdot)$ belonging to (for example) the Banach space L^2 over $\mathcal{R}$, whose dual is

$$\begin{aligned} \underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^2,Q\in\mathbb{R}^{2\times 2}}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2}{\nu+\lambda^T x + x^T Qx - \frac{2}{t}\|x\|}\,dA + \nu + \lambda^T\mu + (\Sigma+\mu\mu^T)\bullet Q \qquad & s.t. \\ \nu + \lambda^T x + x^T Qx &\geq \frac{2}{t}\|x\| \quad \forall x \in \mathcal{R} \\ Q &\succeq \mathbf{0}\,, \end{aligned} \tag{4.9}$$

where $A \bullet B$ is the inner product of matrices A and B.

We again find that the optimality conditions for (4.9), which say that

$$\begin{aligned} \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2}{(\nu+\lambda^T x + x^T Qx - \frac{2}{t}\|x\|)^2}\,dA &= 1 \\ \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2 x}{(\nu+\lambda^T x + x^T Qx - \frac{2}{t}\|x\|)^2}\,dA &= \mu \\ \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2 xx^T}{(\nu+\lambda^T x + x^T Qx - \frac{2}{t}\|x\|)^2}\,dA &\preceq \Sigma + \mu\mu^T\,, \end{aligned}$$

describe the optimal form of $f^*(\cdot)$:

Proposition 25. *The optimal solution $f^*(\cdot)$ that solves (4.8) is of the form*

$$f^*(x) = \frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}, \tag{4.10}$$

where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, and $Q^* \succeq \mathbf{0}$ satisfy

$$\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\| \geq 0$$

everywhere on $\mathcal{R}$, and

$$\begin{aligned}
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA &= 1 \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA &= \mu \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x x^T}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA &\preceq \Sigma + \mu\mu^T .
\end{aligned}$$

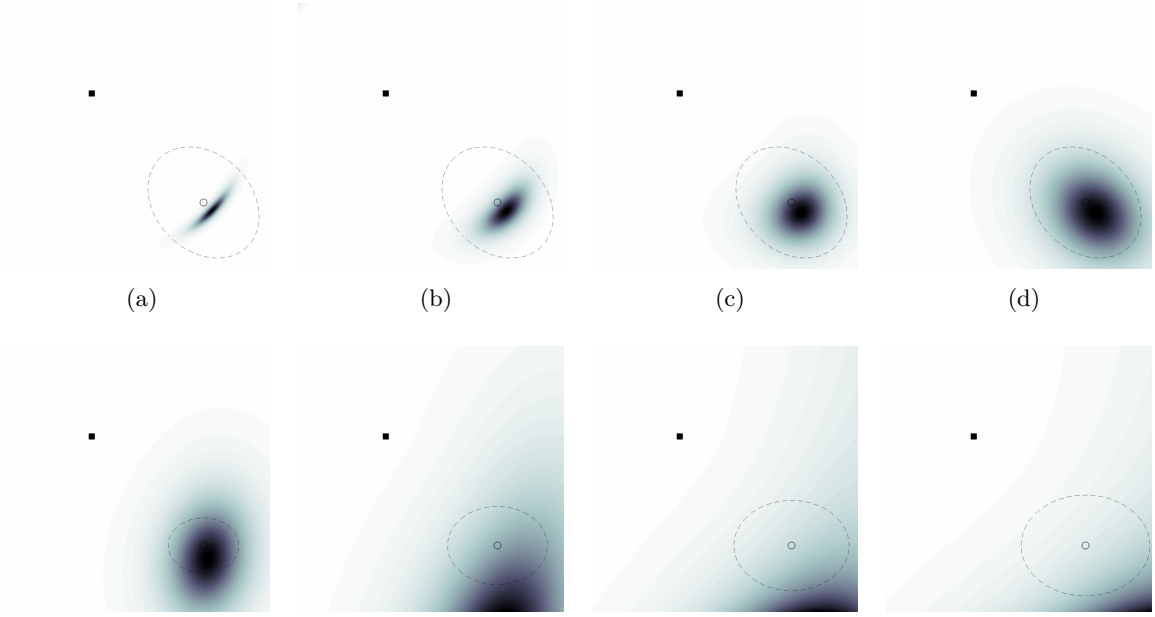
Proof. See Section A.6 of the online supplement for a rigorous proof (this is almost the same as in the paper [65]). Figure 4.3 shows a few plots of $f^*(\cdot)$ for varying values of μ, Σ, and t. □

One of the salient attributes of the worst-case distribution $f^*(\cdot)$, as can readily be observed in Figure 4.3, is that small values of t (as in 4.3a) result in worst-case distributions that are tightly clustered around the centroid. In addition, as we vary the covariance matrix in 4.3e-4.3h, the worst-case distribution also becomes more spread out.

4.5.1 The impact of second-moment information

In this section we describe some results that shed some light on the usefulness of second-moment information: it is clear that distributions with high VRP costs are those that are as "spread out" as possible, either in the sense of having large amounts of demand located away from the depot (when capacities are low and thus the "radial cost" is dominant) or in the sense of being distributed as uniformly as possible (when capacities are high and thus the "circular cost" is dominant). One would therefore expect that, in determining the worst-case cost of an unknown distribution, the covariance bound $\iint_{\mathcal{R}} xx^T f(x)\, dA \preceq \Sigma + \mu\mu^T$ should be particularly relevant.

In order to derive a rough understanding of the effect of this covariance bound, we consider a simple model in which $\mathcal{R}$ is a disk $\mathcal{D}$ with area A, Σ is a diagonal matrix whose



(a) (b) (c) (d)

(e) $\Sigma = 0.3 \cdot \Sigma_0$ (f) $\Sigma = 0.6 \cdot \Sigma_0$ (g) $\Sigma = 0.8 \cdot \Sigma_0$ (h) $\Sigma = \Sigma_0$

Figure 4.3: Surface plots of $f^*(\cdot)$ for $\mu = (0.75, 25)$, a depot located at $(0.33, 0.66)$, and varying values of Σ and t. In 4.3a-4.3d, we fix $\Sigma = \left(\begin{smallmatrix} 0.075 & -0.02 \\ -0.02 & 0.075 \end{smallmatrix}\right)$ and we have $t \in \{0.01, 0.1, 1, 10\}$; in 4.3e-4.3h, we fix $t = 1$ and let Σ be a scalar multiple of $\Sigma_0 = \left(\begin{smallmatrix} 0.06 & 0 \\ 0 & 0.095 \end{smallmatrix}\right)$.

entries are both a scalar s, and both the depot and the mean of the distribution μ are located at the origin. Purely for the ease of exposition, we find it easier to consider the radial cost and the circular cost independently of one another. The problem of maximizing the radial cost is given by

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{D}} \frac{2}{t}\|x\| f(x)\, dA \qquad & s.t. \\ \iint_{\mathcal{D}} x f(x)\, dA &= (0,0)^T \\ \iint_{\mathcal{D}} x x^T f(x)\, dA &\preceq \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} \\ \iint_{\mathcal{D}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{D}\,, \end{aligned} \tag{4.11}$$

and the following result is entirely unsurprising:

Proposition 26. *An optimal solution to problem (4.11) occurs when $f(\cdot)$ is a uniform mixture of two atomic masses at the points $(\pm r, 0)^T$, where $r = \min\{\sqrt{s}, \sqrt{A/\pi}\}$.*

Proof. See Section A.7 of the online supplement. Note that the size of the service region $\mathcal{D}$ has no effect on the solution to (4.11) provided that $s \leq A/\pi$. □

A more interesting problem is that of maximizing the circular cost, which is instead given by

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{R}} \beta\sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{D}} x f(x)\, dA &= (0,0)^T \\ \iint_{\mathcal{D}} x x^T f(x)\, dA &\preceq \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} \\ \iint_{\mathcal{D}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{D} \end{aligned} \tag{4.12}$$

whose dual problem is

$$\begin{aligned} \underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^2,Q\in\mathbb{R}^{2\times 2}}{\text{minimize}} \iint_{\mathcal{D}} \frac{1}{4}\cdot\frac{\beta^2}{\nu+\lambda^T x + x^T Q x}\, dA + \nu + s(q_{11}+q_{22}) \qquad & s.t. \\ \nu + \lambda^T x + x^T Q x &\geq 0 \quad \forall x \in \mathcal{D} \\ Q &\succeq \mathbf{0}\,. \end{aligned} \tag{4.13}$$

By exploiting the symmetry of $\mathcal{D}$ and using concavity of the objective function and convexity of the feasible region in (4.12), it is not hard to show that the first-moment constraint $\iint_{\mathcal{R}} x f(x)\, dA = (0,0)^T$ in (4.12) is redundant, and thus we can remove the variable λ from (4.13). It follows similarly that the optimal matrix Q must also be a diagonal matrix whose entries are both a scalar q, and we are therefore free to consider the simpler dual problem

$$\begin{aligned} \underset{\nu,q\in\mathbb{R}}{\text{minimize}} \iint_{\mathcal{D}} \frac{1}{4}\cdot\frac{\beta^2}{\nu+q\|x\|^2}\, dA + \nu + 2sq \qquad & s.t. \\ \nu, q &\geq 0\,. \end{aligned} \tag{4.14}$$

The optimal solution to (4.12) can be described by way of (4.14) as follows:

Theorem 27. *For fixed values of s, the cost* OBJ *of (4.12) satisfies*

$$\mathsf{OBJ} \sim \beta\sqrt{2\pi s \log A}$$

as $A \to \infty$.

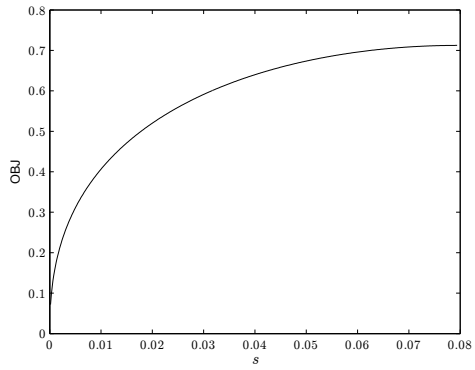


Figure 4.4: The worst-case cost OBJ to problem (4.12) for $s \in (0, A/4\pi]$ and $A = 1$. The curve above shows the relationship between the spatial covariance of demand (measured through the single diagonal element s) and the worst-case cost of problem (4.12).

Proof. See Section A.8 of the online supplement; the key idea is to set

$$\begin{aligned} q &= \frac{\beta}{4} \cdot \sqrt{\frac{2\pi \log A}{s}} \\ \nu &= \frac{\beta^2}{4} \cdot \sqrt{\frac{2\pi s}{\log A}} \end{aligned}$$

and look at the resulting asymptotic series in A. □

Figure 4.4 shows a plot of the cost OBJ as a function of the variance term s for $A = 1$. Notice that, when s is very small, the worst-case workload is close to 0 because all demand must then be tightly clustered around the depot. Moreover, for $s > A/4\pi$, the covariance constraint is not binding because the uniform distribution on the region $\mathcal{D}$ is already feasible for (4.12), and therefore OBJ is constant thereafter. The key insight that Theorem 27 offers us is a concise expression that relates the covariance of the demand distribution and the workload that vehicles incur. Applying Theorem 27, we can therefore conclude that a valid asymptotic bound for the worst-case VRP workload is therefore

$$\mathbf{E}\,\text{length}(\mathsf{VRP}(X)) \lesssim \sqrt{N} \cdot \left(\frac{2}{t} \min\{\sqrt{s}, \sqrt{A/\pi}\} + \beta\sqrt{2\pi s \log A} \right) .$$

Remark. If we want to consider arbitrary covariance matrices Σ, one approach is to use an alternate support region for which the integral in the objective function is easier to calculate.

In particular, one might consider the problem

$$\begin{aligned}
\underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{E}} \beta\sqrt{f(x)}\,dA \qquad & s.t. \\
\iint_{\mathcal{E}} x f(x)\,dA &= (0,0)^T \\
\iint_{\mathcal{E}} x x^T f(x)\,dA &\preceq \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \\
\iint_{\mathcal{E}} f(x)\,dA &= 1 \\
f(x) &\geq 0 \quad \forall x \in \mathcal{E}
\end{aligned}$$

where $\mathcal{E}$ is the ellipse that satisfies

$$\mathcal{E} := \left\{ (x_1, x_2) \in \mathbb{R}^2 : \left(\frac{x_1}{s_1}\right)^2 + \left(\frac{x_2}{s_2}\right)^2 \leq \frac{A}{\pi s_1 s_2} \right\},$$

which clearly has area A. The objective function of the above problem can obviously be converted to an integral over a disk with area A by applying the change of variables $x_1 \mapsto u_1\sqrt{s_1/s_2}$ and $x_2 \mapsto u_2\sqrt{s_2/s_1}$, so that the resulting integral is taken over a disk of area A in the (u_1, u_2) plane. The cost OBJ of the optimal solution (4.12) then satisfies

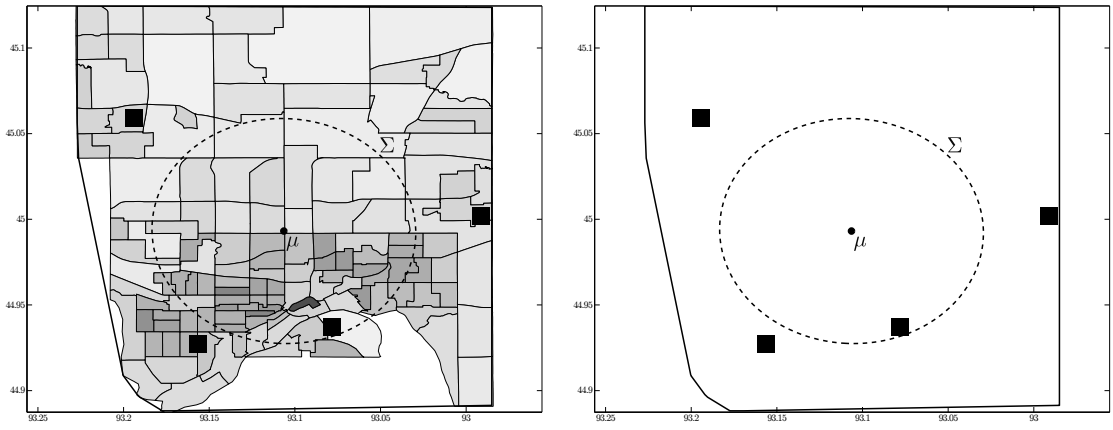
$$\mathsf{OBJ} \sim \sqrt{2\pi}\beta(s_1 s_2)^{1/4}\sqrt{\log A}\,.$$

4.5.2 A computational experiment

In this section we describe the results of a computational experiment applied to a data set obtained from the United States Census [146]. The region $\mathcal{R}$ is given by the convex hull of Ramsey County, Minnesota, and the mean μ and covariance matrix Σ are obtained from the population density thereof. In addition, we suppose that there are four vehicle depots, which are taken as the locations of four post offices; the complete setup of our experiment is shown in Figure 4.5, and was used previously to study the uncapacitated VRP in the paper [65]. Note that Figure (4.5b) shows the true input to our problem (i.e. the population density is not assumed to be known and we have only first and second moment information).

Theoretical preliminaries

In the following simulation, we will consider the problem of designing service districts for each of the four vehicle depots. This requires a minor generalization of Proposition 25



(a) (b)

Figure 4.5: Figure (4.5a) shows the population density of Ramsey County, Minnesota, the mean and covariance matrices of the population density, and four post offices. In practice, of course, we do not know the true demand density, and thus (4.5b) shows the input to our various optimization procedures.

for the case where we are interested in the distribution whose workload on a particular sub-region R_i is as large as possible; this problem is written as

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}} \iint_{R_i} \frac{2}{t}\|x-p_i\|f(x) + \beta\sqrt{f(x)}\,dA \qquad & s.t. \\ \iint_{\mathcal{R}} xf(x) &= \mu \\ \iint_{\mathcal{R}} xx^T f(x)\,dA &\preceq \Sigma + \mu\mu^T \\ \iint_{\mathcal{R}} f(x)\,dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,, \end{aligned} \tag{4.15}$$

which differs from problem (4.8) only in that we have a different domain of integration in the objective function.

Proposition 28. *The optimal solution $f^*(\cdot)$ that solves (4.15) is of the form*

$$f^*(x) = \frac{1}{4}\cdot\frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x-p_i\|\right]^2}\mathcal{I}(x\in R_i) + s^*\delta(x-\bar{x})\,,$$

where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, $\delta(\cdot)$ is the Dirac delta function, and $Q^* \succeq \mathbf{0}$ satisfy

$$\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x-p_i\| \geq 0$$

everywhere on $\mathcal{R}$, $\mathcal{I}(\cdot)$ denotes an indicator function, $\bar{x}$ is a point in $\mathcal{R}$, and s^* is a scalar.

Proof. This is almost identical to Proposition 25, which is proven in Section A.6 of the online supplement; the only difference is the domain of integration in the objective function. □

Recall also that problems (4.5), (4.8), and (4.15) are all designed to maximize the *upper* bound (i.e. the right-hand side) of the original expression (4.4). If we want any information about the solution quality of our approach, it is necessary that we also examine the *lower* bound thereof; this problem is given by

$$\begin{aligned} \underset{f(\cdot)}{\text{maximize}} \max\left\{ \iint_{R_i} \frac{2}{t}\|x-p_i\|f(x),\ \iint_{R_i} \beta\sqrt{f(x)}\,dA \right\} \qquad & s.t. \\ \iint_{\mathcal{R}} xf(x) &= \mu \\ \iint_{\mathcal{R}} xx^T f(x)\,dA &\preceq \Sigma + \mu\mu^T \\ \iint_{\mathcal{R}} f(x)\,dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,, \end{aligned} \tag{4.16}$$

whose optimal solution is characterized as follows:

Proposition 29. *The optimal solution $f^*(\cdot)$ that solves (4.16) takes either the form*

$$f_1^*(x) := \frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x\right]^2} \mathcal{I}(x \in R_i) + s^* \delta(x - \bar{x})$$

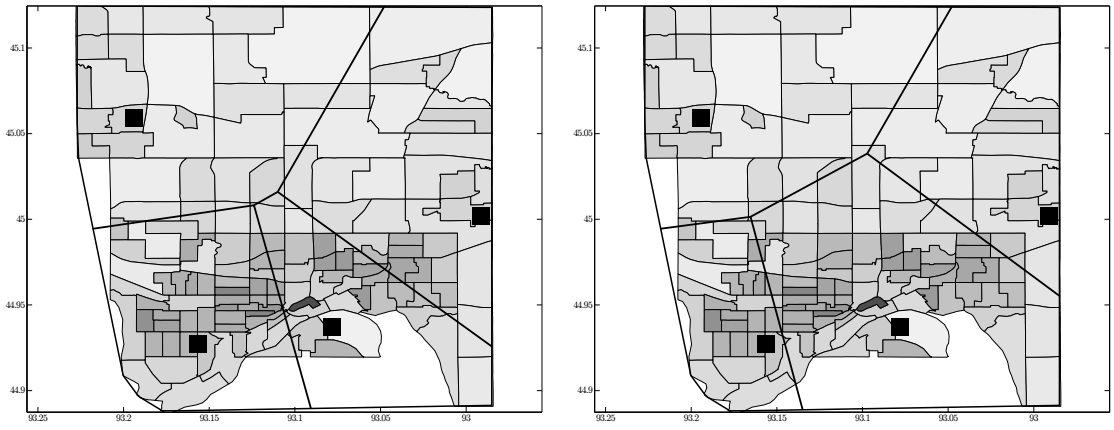
where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, and $Q^* \succeq \mathbf{0}$, or the form

$$f_2^*(x) := \sum_{i=1}^{7} s_i^* \delta(x - \bar{x}_i)\,,$$

i.e. a mixture of at most seven atomic distributions.

Proof. Obviously, we simply analyze problem (4.16) by isolating the two components of the objective function. The proof that the worst-case distribution takes the form of $f_1^*(\cdot)$ when we seek to maximize $\iint_{R_i} \beta\sqrt{f(x)}\,dA$ was already derived in [65] and the proof that the worst-case distribution takes the form $f_2^*(\cdot)$ when we seek to maximize $\iint_{R_i} \frac{2}{t}\|x-p_i\|f(x)$ is in Section A.9 of the online supplement. □

For the sake of notational compactness, from now on, we will let $\bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ and $\Psi(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ denote the objective values of problems (4.15) and (4.16) respectively.



(a) (b)

Figure 4.6: Two power diagram partitions of the service region $\mathcal{R}$. The weight associated with the bottom right sub-region is larger in (4.6a) than in (4.6b).

Computational results

Our objective here is to design service districts for each of the four vehicle depots whose worst-case workloads are as small as possible. In particular, we will focus on the problem of designing sub-regions R_i such that the upper bound of the maximum workload of all $n = 4$ sub-regions, i.e. $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$, is minimized. Our problem can be written as

$$\begin{aligned} \underset{R_1,\ldots,R_n}{\text{minimize}} \max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R}) \qquad & s.t. \\ \bigcup_{i=1}^{n} R_i &= \mathcal{R} \\ R_i \cap R_j &= \emptyset \quad \forall i \neq j\,. \end{aligned}$$

In order to accomplish this we decide to restrict the set of sub-regions R_i (which is, of course, an infinite-dimensional family) to the set of *power diagrams* associated with the depots [147]: given a vector of weights (w_1, w_2, w_3, w_4), we assign depot i (located at point p_i) to the sub-region R_i, defined as

$$R_i = \left\{ x \in \mathcal{R} : \|x - p_i\|^2 - w_i \leq \|x - p_j\|^2 - w_j \quad \forall j \right\}. \tag{4.17}$$

A power diagram is simply a Voronoi diagram in which one also has weights associated with the depots; it has the useful property that all of the regions R_i are polygonal and convex. Of course, when all weights are equal, the power diagram simply reduces to the

standard Voronoi diagram, and if we increase the weight associated with one of the depots p_i, the size of sub-region R_i will increase. Figure 4.6 shows two examples of power diagram partitions. In Section B.2 of the online supplement, Algorithm 8 describes a standard branch-and-bound procedure for determining weights w_i^* that approximately minimize $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$. We refer the reader to Figure 4.7, which shows the worst-case workloads associated with the sub-regions R_i together with the actual workloads, as induced by the true population density. The three bands shown there are interpreted as follows: the lightest shaded band corresponds to the worst-case cost and indicates the upper and lower bounds that are obtained from the optimal objective function cost to (4.15), which is an upper bound of the worst possible VRP cost, and the optimal objective function cost to (4.16), which is a lower bound of the worst possible VRP cost. In other words, the lightest shaded band shows $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ and $\max_i \underline{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$, when the sub-regions R_i are power diagrams that are obtained by Algorithm 8 in the online supplement. Similarly, the second shaded band corresponds to the actual cost associated with the partition obtained by Algorithm 8 and indicates the upper and lower bounds of the VRP cost of our algorithm, i.e. $\max_i \iint_{R_i} \frac{2}{t}\|x\| f(x) + \beta\sqrt{\bar{f}(x)}\, dA$ for the upper bound and $\max_i \max\left\{\iint_{R_i} \frac{2}{t}\|x\| f(x),\ \iint_{R_i} \beta\sqrt{\bar{f}(x)}\, dA\right\}$ for the lower bound, where $f(\cdot)$ represents the true population density as shown in Figure 4.5a. Finally, the darkest shaded band corresponds to the best possible actual costs (in terms of upper and lower bounds) that could be realized by partitioning the region optimally, which we obtained by using the algorithm in [2].

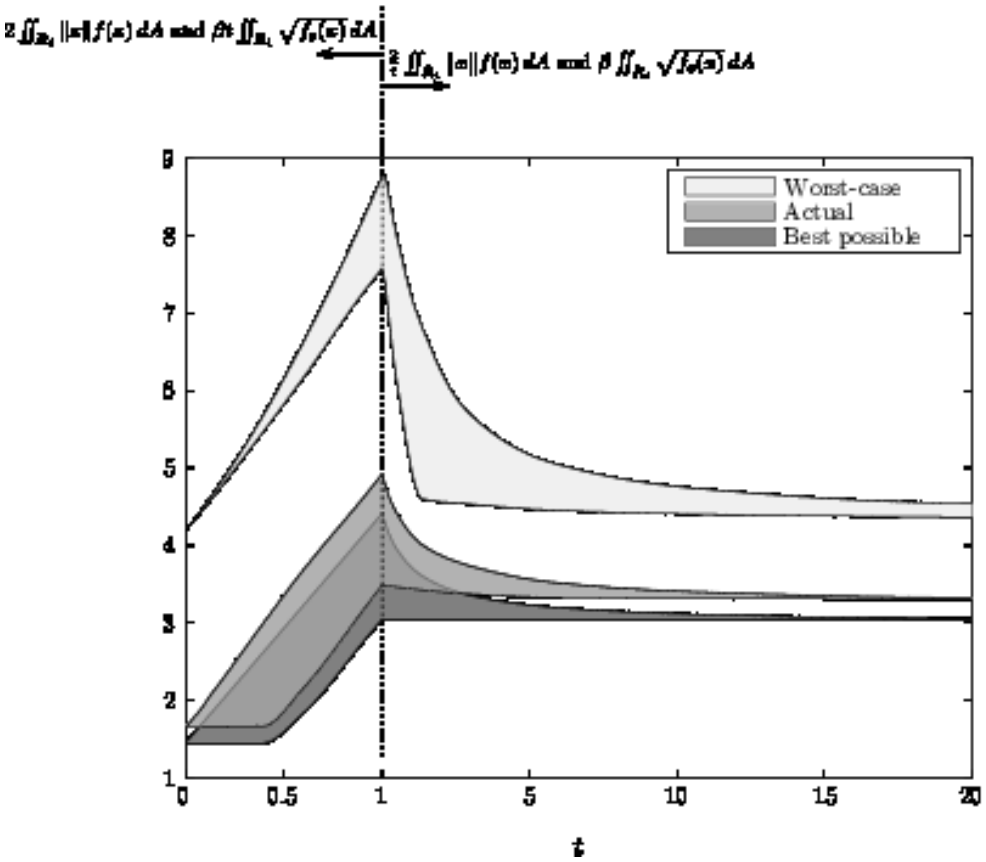


Figure 4.7: The plot above shows the workloads that are induced by Algorithm 8 for $t \in (0, 20]$, measured as the *maximum workload* of the four sub-regions R_i (rather than, say, the sum of the workloads in the four sub-regions R_i). For each value of t, we naturally require a separate run of the algorithm (we performed 100 runs overall). To the right of the dashed line (i.e. for $t \geq 1$), we plot the various upper and lower bounds to our problem, and to the left of the dashed line (i.e. for $t < 1$), we plot essentially the same quantities, but multiplied through by t in all cases. This is because all costs (the worst-case cost, the actual cost, and the best possible cost) explode towards infinity as $t \to 0$ and thus we find this to be a more useful measure of the relative costs.

Chapter 5

The Distributionally Robust TSP: A Data-Driven Model

The Distributionally Robust TSP: A Data-Driven Model The travelling salesman problem (TSP) is one of the most heavily studied optimization problems that has played a major role in theoretical fields like combinatorial optimization, and in practical fields like transportation. It is also one of the domains in which optimization and geometry have crossed paths.

One of the theoretically interesting aspects of TSP is the asymptotic analysis of the length of the TSP tour as the number of demand points becomes large. The problem in this chapter concerns a *robust TSP* for large number of *i.i.d.* random points where the ambiguity is about the distribution of these points.

The region of ambiguity, in recent research on the robust and stochastic travelling salesman problem and the vehicle routing problem, has been described in many different ways such as taking convex combinations of observed demand vectors or imposing constraints on the moments of the spatial demand distribution, out of which we studied the latter in Chapter 4. The reader is also encouraged to see [65] for additional insights on defining the ambiguity set based on fixed-mean and bounded covariance matrix.

A data-driven approach that has been used outside the transportation sector is the use of statistical metrics that describe a distance function between two probability distributions. In this chapter, we consider a distributionally robust version of the Euclidean travelling salesman problem in which we compute the worst-case spatial distribution of demand against all distributions whose *Wasserstein distance* to an observed demand distribution is bounded from above. This constraint allows us to circumvent common overestimation that

arises when other procedures are used, such as fixing the center of mass and the covariance matrix of the distribution. Numerical experiments confirm that our new approach is useful as a decision support tool for dividing a territory into service districts for a fleet of vehicle when information about the demand distribution is limited. Moreover, it provides a data-driven model which can take better advantage of data when it is more available.

5.1 Introduction

One of the most complex factors that arises in formulating and solving robust travelling salesman problems (TSP) and vehicle routing problems (VRP) is the difficulty of describing one's ambiguity set in a way that is both useful and mathematically tractable. Recent works have seen many different approaches for describing these sets, such as taking convex combinations of observed demand vectors [63], general polyhedral constructions [121], and using mean and covariance information about the spatial distribution of destination points [65]. The choice of one's ambiguity set often yields qualitative insights into what demand patterns affect the outcome most significantly; for example, the worst-case spatial distribution for the Euclidean TSP is that which is as equitably distributed (uniform) as possible [136].

In this chapter, we consider a *distributionally robust* version of the Euclidean TSP: as input, we are given a compact, contiguous planar region $\mathcal{R}$ and a realization of sampled demand points in that region, and our objective is to construct a probability distribution on $\mathcal{R}$ that is sufficiently "close" to the empirical distribution consisting of the sampled points and is as "spread out" as possible, in the sense that the asymptotic length of a TSP tour of points drawn from that distribution should be as large as possible. In order to characterize our ambiguity set of distributions, we use a statistical metric called the *Wasserstein distance*, which is also known as the *earth mover's* or *Kantorovich* metric. Conceptually speaking, the Wasserstein distance is very simple and intuitive: if we visualize two probability distributions μ_1 and μ_2 as being two piles of equal amounts of sand, then the Wasserstein distance between them is simply the minimum amount of work needed to move one pile to take the shape of the other, as suggested in Figure 5.1a. A particularly attractive feature of the Wasserstein distance that is not present in many other statistical metrics is the ability to directly compare a discrete distribution and a continuous distribution, as illustrated in Figures 5.1e-5.1g. In addition, because the Wasserstein distance is a true metric, the set of all distributions within a certain distance of a reference distribution is a convex set that turns out to admit a simple representation.

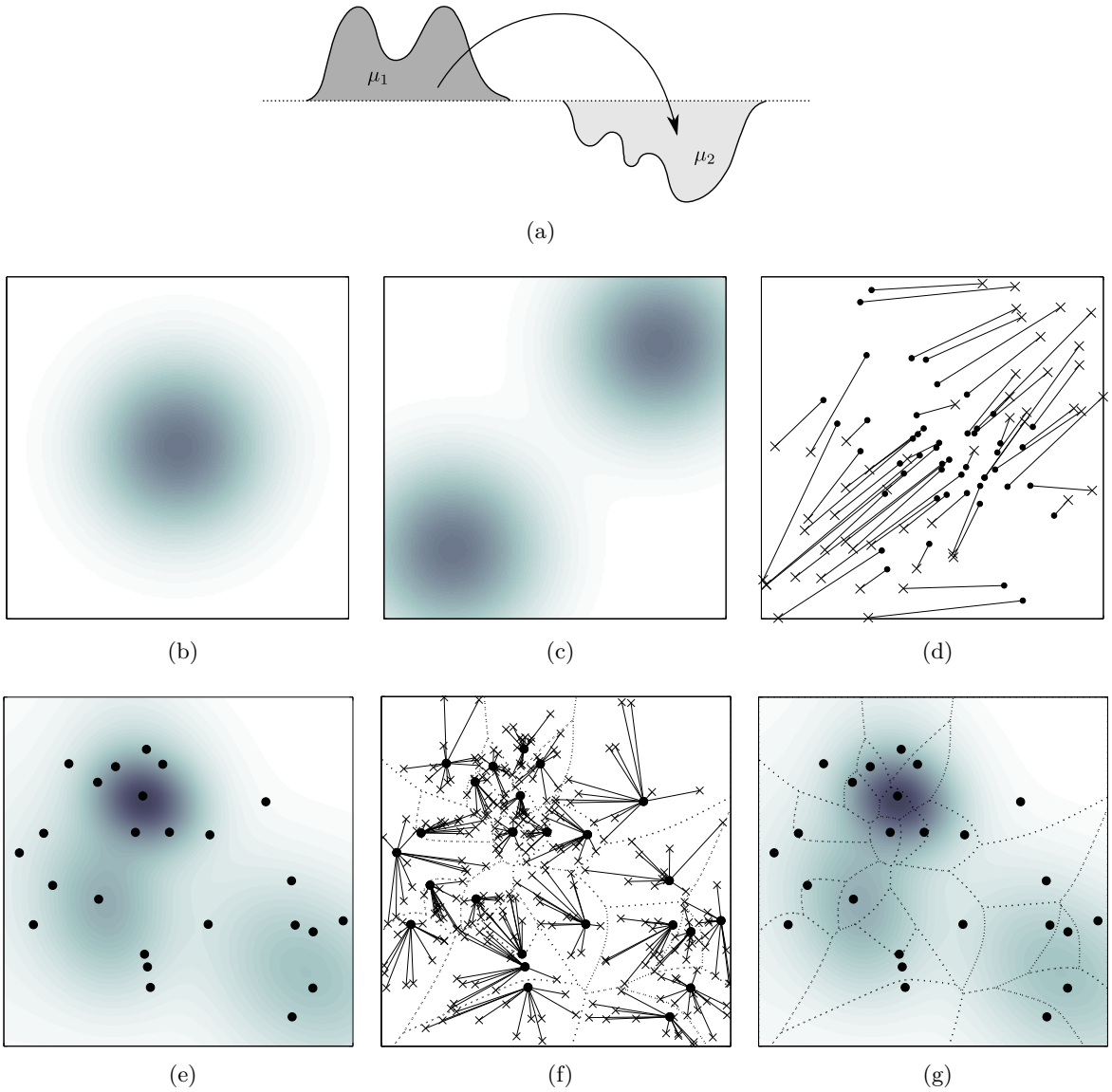


Figure 5.1: Figure 5.1a shows a Wasserstein distance problem between two univariate distributions μ_1 and μ_2. Figures 5.1b-5.1d show that a Wasserstein mapping can be thought of as an infinite-dimensional generalization of a bipartite matching; here μ_1 and μ_2 are shown in 5.1b and 5.1c, and 5.1d shows a bipartite matching between a large number of samples collected from μ_1 and μ_2. Figures 5.1e-5.1g show an interpretation of a Wasserstein distance problem when μ_1 is a smooth density and μ_2 is atomic. The two distributions are shown in 5.1e, and 5.1f shows the solution to an assignment problem between a large number of samples from μ_1 and the atomic distribution μ_2; a side consequence is that the Lagrange multipliers of this assignment induce a partition of the region $\mathcal{R}$, each of whose cells are associated with one of the elements of μ_2. Figure 5.1g shows this partition together with the atomic distribution μ_2; each cell contains $1/n$ of the mass of the density, which is to be transported to the point contained within it. By using previous results [2], these dashed curves are computationally easy to compute. If we let R_i denote the cell associated with each point x_i in the atomic distribution and f denote the density, then the Wasserstein distance is $\sum_i \iint_{R_i} f(x)\|x - x_i\| \, dA$.

This chapter is structured as follows: Section 5.2 describes the basic theoretical preliminaries that are needed for the analysis that we perform in Section 5.3, which describes the structure of the worst-case spatial distribution for the TSP under a Wasserstein distance constraint. Next, Section 5.4 describes a primal-dual algorithm that finds this worst-case distribution efficiently, and this algorithm is then implemented in two computational experiments involving both the single-vehicle and multi-vehicle TSP in Section 5.5.

5.1.1 Related work

This chapter describes a continuous approximation model that uses robust optimization to describe the worst-case demand distribution for the travelling salesman problem; this model is then applied to solve a districting problem that assigns vehicles to pre-specified zones in a region. As such, there are essentially three bodies of literature from which it stems.

Continuous approximation models

This chapter is concerned with a *continuous approximation* model for a transportation problem, and is therefore philosophically similar to (for example) [104], which analytically determines trade-offs between transportation and inventory costs, [105], which shows how to route emergency relief vehicles to beneficiaries in a time-sensitive manner, and [106], which describes a simple geometric model for determining the optimal mixture of a fleet of vehicles that perform distribution. The basic premise of the continuous approximation paradigm is that one replaces combinatorial quantities that are difficult to compute with simpler mathematical formulas, which (under certain conditions) provide accurate estimations of the desired quantity [107, 108]. Such approximations exist for many combinatorial problems, such as the travelling salesman problem [51, 111], facility location [112, 27, 11], and any *subadditive Euclidean functional* such as a minimum spanning tree, Steiner tree, or matching [113, 83, 82]. In our computational districting experiment, an approximation of this kind is used as the first level of an optimization problem in which we design service zones that are associated with different vehicles.

Districting problems in vehicle routing

The primary application of the theory derived in this chapter is in the design of *districts* for allocating a fleet of vehicles to visit a collection of customers when demand is uncertain. The problem of designing such districts is a foundational one in the continuous approximation literature, as can be seen in [148] or Chapter 4 of the seminal book [132], for example. The most common way that uncertainty is represented is by assuming that demand follows a

known probability density function (which is often further assumed to be uniform); this density then informs the districting decision in some way. To give a few examples, [149] uses a multiplicatively-weighted Voronoi partitioning scheme in which district sizes are determined by a set of scalar weights associated with the vehicles, [43] uses so-called "ham sandwich cuts" to recursively partition the region, and [110] uses a "disk model" that allows for explicit control of district sizes and implicit control of district shapes. Further additional examples from the robotics community include [150, 151, 152, 57], which all use various forms of the Voronoi paradigm (such as additive, quadratic, or logarithmic weighting schemes) to partition a geographic region, and place a particular emphasis on "decentralizing" the means by which partitions are constructed.

An alternative method to the preceding continuous models is to instead assume that demand is present on a known graph, and that vertices on the graph have probability weights. This is the approach taken by [54], which models the districting problem as a two-stage stochastic optimization program with recourse, by [153], which uses a three-phase procedure that aggregates data points into compact districts using a mixed-integer goal program, and by [154], which uses a steady-state spatial queueing model to simultaneously reason about the optimal locations of emergency response stations and the territories they serve. It is also possible to apply principles from continuous approximation theory to design districts in graph-based models, provided some basic geometric information is available; this is the case in [155, 156], which use the square-root approximation of [51] in conjunction with a graph-based model, and which show good performance when the inputs are known to be uniformly distributed over a geometric domain. Section 5.5.2 of this chapter also shows how to apply a continuous approximation scheme to a heterogeneous road network, namely, a map of Los Angeles County, to a set of inputs that are non-uniformly distributed.

In practice, one often does not have sufficient information (namely, a probability distribution defined on an entire geographic region) to apply the preceding models. For example, in recent years, numerous start-up companies have emerged that provide "last-minute" delivery of food and groceries such as Good Eggs, DoorDash, BiteSquad, and Caviar [126, 127, 128, 129], and such companies must make high-level strategic allocation decisions without an extensive set of historical data. Another example arises in threat detection and surveillance, in which case a set of vehicles begins with some *a priori* information about the distribution of targets and one seeks a policy for routing these vehicles that takes new information into account as it becomes available [157]. The problem of designing districts in such a data-driven fashion (i.e. when one has an ambiguous distribution setting) is considerably less understood, although the paper [65] describes one approach for doing so

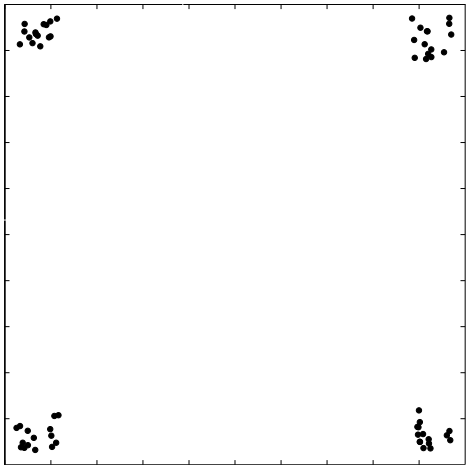


Figure 5.2: The above point sets in the unit square are extremely clustered and one would expect that their TSP tour should be short. However, because their sample mean and covariance matrix are the same as that of the uniform distribution, any robust methodology that uses only mean and covariance information will fail to recognize the clustering, thereby incurring significant overestimation.

when one knows the mean and covariance of the demand distribution. A major deficiency of this approach, which motivates our present work, is its inability to respond to *clustering* or even mere *multi-modality* in data points. For example, Figure 5.2 shows a data set that is very clustered, and whose TSP tour should therefore be short relative to (for example) a uniform distribution. However, this clustered data set actually has precisely the same mean and covariance matrix as the uniform distribution; such an approach therefore frequently leads to an over-conservative solution (or more generally, a solution that is not faithful to the true unknown demand distribution), even when a large number of samples is available. This over-conservatism is actually noted in Figure 10 of [65].

Robust optimization and vehicle routing

In most models of the robust VRP, one has a pre-defined ambiguity region and seeks a set of routes that is as good as possible with respect to all of the outcomes; this ambiguity region is usually described as a polyhedral set [119, 121, 63], although the recent paper [124] adopts a "robust mean-variance" approach that minimizes a weighted sum of the average cost and the variance of a route when sampled over many scenarios. In our problem, we are concerned with robustness in the *distributional* sense [125]: we seek the spatial distribution of demand for which the expected cost of a tour is as high as possible, while remaining consistent with some observed data samples or some parameters derived thereof.

By far, the most common parameters used in distributionally robust problems (in general domains, not just those arising in transportation) are the support and the first and second moments of the sample distribution [84, 158, 159], and the papers [160, 161] additionally make use of bounds on "directional deviation measures" that isolate stochastically independent components. We have previously seen the use of first and second moment information for the distributionally robust VRP in Chapter 4 and also the paper [65]; one major drawback of this method is an inability to detect clustering, as we have already noted in the preceding section. In order to remedy this, we propose the use of the Wasserstein distance as a means of defining the uncertainty region of demand distributions. The Wasserstein distance is very commonly used in machine learning and statistics [162, 163, 164, 165], and is also mentioned in the context of robust optimization in [166] for its relationship with the *Prokhorov metric*. To our knowledge, the first direct applications of the Wasserstein distance to optimization problems have occurred very recently in [167, 168]; the former uses Carathéodory-type results to reduce the support set of an infinite-dimensional optimization problem to a finite set and the latter uses the Wasserstein distance as one of several statistical metrics to define risk measures for portfolios. Even more recently, the paper [169] shows how to apply complementary slackness principles to solve a very large family of distributionally robust optimization problems subject to Wasserstein distance constraints; their use of convex duality theory is closely related to our own derivation in Section 5.3.

For general problems (i.e. not those related specifically to vehicle routing), a variety of other statistical metrics (or *pseudo-metrics*) have been used previously for solving distributionally robust optimization problems; such metrics include the Kullback-Leibler divergence, Hellinger distance, χ^2-distance, total variation distance, or Kolmogorov-Smirnov statistic. A few examples follow:

- The paper [170] solves a variety of robust linear programs using *ϕ-divergences*, a wide class of pseudo-metrics to which several of the aforementioned quantities belong.
- The paper [171] gives a highly flexible framework that builds ambiguity sets using classical statistical hypothesis tests, including the χ^2 test and the Kolmogorov-Smirnov test,
- The paper [172] computes robust financial portfolios using the Kullback-Leibler divergence.
- The paper [173] shows how to solve robust dynamic programs whose distributional ambiguity sets are defined using the Kullback-Leibler divergence.

- The paper [174] proposes a robust lot-sizing model whose distributional ambiguity set is defined via the χ^2 goodness-of-fit test.

There are three reasons why the Wasserstein distance is a particularly appropriate choice for our problem of interest: first, the Wasserstein distance allows one to directly make comparisons between a discrete distribution (such as the empirical distribution consisting of a collection of data points) and a continuous distribution, as we have previously noted in Figure 5.1; this it not possible in (for example) the Kullback-Leibler divergence, the Hellinger distance, or the total variation distance. Secondly, the Wasserstein distance is in a sense "inherited" from the Euclidean distance, inasmuch as the distance between two distributions is defined as an integral of Euclidean distances. Since we are concerned with obtaining a probability distribution whose induced TSP tour is as long as possible (in an asymptotic limit as many samples are taken), and a TSP tour is also measured using Euclidean distances, the Wasserstein distance is a particularly appropriate choice. The third reason is purely practical: it turns out that the ambiguity set of distributions characterized by a Wasserstein distance threshold gives a very concise, closed-form expression for the worst-case distribution for our problem. As we will later show in Section 5.3.3, a fourth *a posteriori* justification for the use of the Wasserstein metric is that the worst-case distribution that one obtains for this problem is closely related to that of the classical *geographical gravity model*, which arises in many models of spatial interaction.

5.1.2 Notational conventions

Our notational conventions throughout this chapter are as follows: integrals over regions in $\mathbb{R}^2$ are denoted with the double integral sign $\iint dA$. The diameter of a region $\mathcal{R}$, denoted $\text{diam}(\mathcal{R})$, is the largest possible distance between two points in $\mathcal{R}$, $\sup_{x,y\in\mathcal{R}} \|x-y\|$. The vector consisting of all 1's is written $\mathbf{e}$, whose dimension will always be clear from context, and the indicator function of a particular condition and the Dirac delta function are written as $\mathbb{1}(\cdot)$ and $\delta(\cdot)$ respectively. The Wasserstein distance between two distributions is written $\mathcal{D}(\cdot,\cdot)$ and is defined in the next section. We will commit a slight abuse of notation and use the expression $\mathsf{TSP}(x_1,\ldots,x_n)$ to represent both the shortest tour that goes through a set of points as well as the length of that shortest tour. Finally, for any univariate function $f(x)$, we say that $f(x) \in o(g(x))$ as $x \to \infty$ if $\lim_{x\to\infty} f(x)/g(x) = 0$.

5.2 Preliminaries

In order to retain mathematical rigor, we find the following results useful; the first two items are simplified from [175]:

Definition 30 (Wasserstein distance)**.** Let μ_1 and μ_2 denote two probability measures defined on a compact planar region $\mathcal{R}$. The *Wasserstein distance* between μ_1 and μ_2, written $\mathcal{D}(\mu_1, \mu_2)$, is defined as

$$\mathcal{D}(\mu_1, \mu_2) := \inf_{\pi \in \Pi(\mu_1, \mu_2)} \iiiint_{\mathcal{R} \times \mathcal{R}} \|x - y\| \, d\pi(x, y) \,, \tag{5.1}$$

where $\Pi(\mu_1, \mu_2)$ is defined as the set of all probability measures on $\mathcal{R} \times \mathcal{R}$ whose marginals are μ_1 and μ_2, that is, the set of all probability measures that satisfy $\pi(A \times \mathcal{R}) = \mu_1(A)$ and $\pi(\mathcal{R} \times B) = \mu_2(B)$ for all measurable subsets $A, B \subset \mathcal{R}$.

The Wasserstein distance can be thought of as a generalization of an *assignment problem*: for example, when μ_1 and μ_2 are discrete distributions consisting of n points each with equal mass, the Wasserstein distance between the two is simply computed as the cost of a bipartite matching (multiplied by a normalization term of $1/n$). This interpretation is suggested in Figure 5.1.

Our notion of distributional robustness relies on the following famous theorem, originally stated in [51] and further developed in [83, 82], which relates the length of a TSP tour of some points with the distribution from which they were sampled:

Theorem 31 (BHH Theorem)**.** *Suppose that $X = \{X_1, X_2, \dots\}$ is a sequence of random points i.i.d. according to a probability density function $f(\cdot)$ defined on a compact planar region $\mathcal{R}$. Then with probability one, the length $\mathsf{TSP}(X)$ of the optimal travelling salesman tour through X satisfies*

$$\lim_{N \to \infty} \frac{\mathsf{TSP}(X)}{\sqrt{N}} = \beta \iint_{\mathcal{R}} \sqrt{f_c(x)} \, dA$$

where β is a constant and $f_c(\cdot)$ represents the absolutely continuous part of $f(\cdot)$.

It is additionally known that $0.6250 \leq \beta \leq 0.9204$ and estimated that $\beta \approx 0.7124$; see [5, 51].

The following classical result from [139] will be useful in confirming the existence of an optimal solution of the problem that we will construct:

Theorem 32 (Lagrange Duality)**.** *Let ϕ be a real-valued convex functional defined on a convex subset Ω of a vector space $\mathcal{X}$, and let $\mathcal{G}$ be a convex mapping of $\mathcal{X}$ into a normed*

space $\mathcal{Z}$. Suppose there exists an χ_1 such that $\mathcal{G}(\chi_1) < \theta$, where θ is the zero element, and that $\inf\{\phi(\chi) : \chi \in \Omega,\ \mathcal{G}(\chi) \leq \theta\}$ *is finite. Then*

$$\inf_{\chi\in\Omega,\ \mathcal{G}(\chi)\leq\theta} \phi(\chi) \quad = \quad \max_{z^*\geq\theta} \inf_{\chi\in\Omega} \phi(\chi) + \langle \mathcal{G}(\chi), z^* \rangle$$

and the maximum on the right is achieved by some $z_0^* \in \mathcal{Z}^*$ such that $z_0^* \geq \theta$, where $\mathcal{Z}^*$ denotes the dual space of $\mathcal{Z}$ and $\langle \cdot, \cdot \rangle$ denotes the evaluation of a linear functional, i.e. $z^*(\mathcal{G}(\chi))$. If the infimum on the left is achieved by some $\chi_0 \in \Omega$, then $\langle \mathcal{G}(\chi_0), z_0^* \rangle = 0$, and χ_0 minimizes $\phi(\chi) + \langle \mathcal{G}(\chi), z_0^* \rangle$ over all $\chi \in \Omega$.

Finally, the Wasserstein distance between a discrete distribution consisting of points $\{x_1, \ldots, x_n\}$ with uniform probabilities $1/n$ and a continuous probability density function f defined on a compact planar region $\mathcal{R}$ can be obtained by solving the following infinite-dimensional optimization problem:

$$\begin{aligned}
\underset{I_1(\cdot),\ldots,I_n(\cdot)}{\text{minimize}} \sum_{i=1}^{n} \iint_{\mathcal{R}} \|x - x_i\| f(x) I_i(x)\, dA \qquad & s.t. \\
\iint_{\mathcal{R}} f(x) I_i(x)\, dA &= 1/n \quad \forall i \\
\sum_{i=1}^{n} I_i(x) &= 1 \quad \forall x \in \mathcal{R} \\
I_i(x) &\geq 0 \quad \forall i, \forall x \in \mathcal{R}\,;
\end{aligned}$$

here the value $I_i(x)$ simply describes the amount of the distribution at point $x \in \mathcal{S}$ that should be moved to point x_i. The lemma below summarizes some basic results on the Wasserstein distance between a probability density and an empirical distribution:

Lemma 33. *Let f denote a probability density function on a compact planar region $\mathcal{R}$ and let $\hat{f}$ denote an atomic distribution consisting of distinct points $x_1, \ldots, x_n \in \mathcal{R}$ each having probability mass $1/n$. Then the following statements are true:*

1. *The Wasserstein distance $\mathcal{D}(f, \hat{f})$ is the optimal objective value to the concave maximization problem*

$$\begin{aligned}
\underset{\boldsymbol{\lambda}\in\mathbb{R}^n}{\text{maximize}} \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i\}\, dA \qquad & s.t. \\
\mathbf{e}^T \boldsymbol{\lambda} &= 0\,,
\end{aligned} \tag{5.2}$$

   where $\mathbf{e} \in \mathbb{R}^n$ denotes a vector whose entries are all 1's.

2. *For any* $\boldsymbol{\lambda}$*, a valid* supergradient *[176] for the objective function of (5.2) is the vector* $\mathbf{g} \in \mathbb{R}^n$ *defined by setting*

$$g_i = -\iint_{R_i} f(x)\, dA\,,$$

where each R_i *is a connected piecewise hyperbolic region characterized by*

$$R_i = \{x \in \mathcal{R} : \|x - x_i\| - \lambda_i \leq \|x - x_j\| - \lambda_j \quad \forall j \neq i\}\ ;$$

that is, for any other $\boldsymbol{\lambda}'$*, we have*

$$\iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i'\}\, dA \leq \iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA + \mathbf{g}^T(\boldsymbol{\lambda}' - \boldsymbol{\lambda})\,.$$

3. *If* $\boldsymbol{\lambda}^*$ *is a maximizer of (5.2), then an optimal Wasserstein mapping between* f *and* $\hat{f}$ *is obtained by defining*

$$R_i^* = \left\{x \in \mathcal{R} : \|x - x_i\| - \lambda_i^* \leq \|x - x_j\| - \lambda_j^* \quad \forall j \neq i\right\}$$

for each i *and transporting all of the mass of each* R_i^* *to its associated point* x_i*.*

4. *If* $f(x) > 0$ *for all* $x \in \mathcal{R}$*, then there exists a unique maximizer* $\boldsymbol{\lambda}^*$*.*

Proof. Statement 1 is a well-known special case of the *Kantorovich duality theorem*; see for example Theorem 1.3 of [175] or [45, 2, 177] for specific details. In addition, the economic interpretation of the regions R_i^* relative to the dual variables λ_i^* can be found in [67]; in a nutshell, the sub-regions R_i^* that characterize the mapping are equivalent to market regions induced by a *mill pricing* scheme at each of the points x_i. Proofs of statements 2-4 are routine and can be found in Section A.10 of the Online Supplement. □

5.3 Worst-case distributions with Wasserstein distance constraints

The input to our problem is a set of distinct demand points $x_1, \ldots, x_n$ in a compact planar region $\mathcal{R}$, which we assume are sampled from some (unknown) distribution function f. By rearranging the terms of Theorem 31, we can write

$$\mathsf{TSP}(x_1, \ldots, x_n) = \beta\sqrt{n} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA + o(\sqrt{n})$$

with probability one as $n \to \infty$. Since β is a constant and $\sqrt{n}$ is (presumably) not related to the distribution f, we therefore conclude that the "worst" distribution whose induced TSP workload is as large as possible (subject to whatever other constraints might be present) is precisely that distribution that maximizes $\iint_{\mathcal{R}} \sqrt{f(x)}\, dA$.

We now let $\hat{f}$ denote the empirical distribution on these n points x_i. We will search through all distributions f whose Wasserstein distance to $\hat{f}$ is sufficiently small, i.e. where $\mathcal{D}(f, \hat{f}) \leq t$; here $\mathcal{D}(\cdot, \cdot)$ is the Wasserstein distance from Definition 30 and t is a parameter that will be discussed in Section 5.4.2. The problem of finding the worst-case TSP distribution, subject to the Wasserstein distance constraint, is then written as the infinite-dimensional convex optimization problem

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \mathcal{D}(f, \hat{f}) &\leq t \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{5.3}$$

This is our problem of interest throughout this chapter. We will embed f in the Banach space $L^1(\mathcal{R})$, hereafter abbreviated simply to L^1, which consists of all functions that are absolutely Lebesgue integrable on $\mathcal{R}$.

5.3.1 Comparison with other approaches

The earlier paper [65] considers a problem closely related to (5.3) in which one has constraints on the mean and covariance of f instead of the constraint on $\mathcal{D}(f, \hat{f})$; that problem is written as

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} x f(x)\, dA &= \mu \\ \iint_{\mathcal{R}} x x^T f(x)\, dA &\preceq \Sigma + \mu\mu^T \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{5.4}$$

Since it is well-known (e.g. Section 2 of [178]) that the Wasserstein distance between the empirical distribution $\hat{f}$ and the true distribution f converges to zero with probability one

as $n \to \infty$, it is not surprising that our proposed formulation (5.3) is guaranteed to make better use of sample points as they become available, unlike the problem (5.4) written above:

Theorem 34. *Let $X = \{X_1, X_2, \dots\}$ be a sequence of random points i.i.d. according to an absolutely continuous probability density function $\bar{f}(\cdot)$ defined on a compact planar region $\mathcal{R}$. For any positive integer n, let $\hat{f}_n$ denote the empirical distribution on points $\{X_1, \dots, X_n\}$. Then with probability one there exists a sequence $\{t_1, t_2, \dots\}$, converging to 0, such that the optimal objective value of the problem*

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \mathcal{D}(f, \hat{f}_n) &\leq t_n \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R} \end{aligned} \tag{5.5}$$

approaches the ground truth (i.e. $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$) as $n \to \infty$.

Proof. See Section A.11 of the Online Supplement. □

5.3.2 Structure of the solution to (5.3)

To begin, we apply Lemma 33 to express the distance constraint $\mathcal{D}(f, \hat{f}) \leq t$ in (5.3) differently, obtaining the equivalent formulation

$$\begin{aligned} \underset{f \in L_1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i\}\, dA &\leq t \quad \forall \boldsymbol{\lambda} : \mathbf{e}^T \boldsymbol{\lambda} = 0 \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{5.6}$$

This is an infinite-dimensional problem with an infinite-dimensional constraint space and is therefore best addressed using Theorem 32; before doing so, we find the following result useful:

Lemma 35. *There exists a unique optimal solution f^* to problem (5.6), and $f^*(x) > 0$ for all $x \in \mathcal{R}$.*

Proof. The fact that the optimal solution is unique (provided one exists) is an immediate consequence of the fact that the square root function in the integrand of (5.6) is strictly

concave. To prove existence, let $\{f^j\}$ denote a sequence of feasible inputs to (5.6) whose objective values converge to a supremum. For each f^j, let $\boldsymbol{\lambda}^j$ denote a value of $\boldsymbol{\lambda}$ that induces the optimal Wasserstein mapping between f^j and $\hat{f}$ as described in Lemma 33, i.e. that solves problem (5.2). It is easy to verify that the iterates $\boldsymbol{\lambda}^j$ lie in the compact set Λ, defined by

$$\Lambda := \left\{ \boldsymbol{\lambda} \in \mathbb{R}^n : \mathbf{e}^T \boldsymbol{\lambda} = 0,\, \lambda_i \leq \operatorname{diam}(\mathcal{R})\, \forall i \right\},$$

because any $\boldsymbol{\lambda}$ lying outside Λ would force some sub-regions to be empty. Therefore, the sequence $\{\boldsymbol{\lambda}^j\}$ must have a convergent subsequence with a limit $\boldsymbol{\lambda}^*$, inducing a partition $R_1^*, \ldots, R_n^*$ as in statement 2 of Lemma 33 that satisfies $\iint_{R_i^*} f(x)\, dA = 1/n$ for all i. Standard arguments then show that the true worst-case distribution f^* is precisely the solution to the problem

$$\begin{aligned} \underset{f \in L_1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \sum_{i=1}^{n} \iint_{R_i^*} \|x - x_i\| f(x)\, dA &\leq t \\ \iint_{R_i^*} f(x)\, dA &= \frac{1}{n} \quad \forall i \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R} \end{aligned} \tag{5.7}$$

(this is an immediate consequence of the fact that the optimal objective cost to (5.7) varies continuously as the vector $\boldsymbol{\lambda}^*$, which defines the partition $R_1^*, \ldots, R_n^*$, is perturbed). This problem has a finite-dimensional constraint space, and it is routine to apply Theorem 32 to (5.7) to derive the dual problem

$$\begin{aligned} \underset{\boldsymbol{\nu} \geq \mathbf{0}}{\text{minimize}}\, \frac{1}{4} \sum_{i=1}^{n} \iint_{R_i^*} \frac{1}{\nu_0 \|x - x_i\| + \nu_i}\, dA + \nu_0 t + \frac{1}{n}(\nu_1 + \cdots + \nu_n) \qquad & s.t. \\ \nu_0 \|x - x_i\| + \nu_i \quad &\geq \quad 0 \quad \forall x \in R_i^* \quad \forall i \end{aligned} \tag{5.8}$$

whereby we conclude that the optimal solution f^* to (5.7) must take the form

$$f^*(x) = \frac{1}{4(\nu_0^* \|x - x_i\| + \nu_i^*)^2} \tag{5.9}$$

on each sub-region R_i^*. This satisfies $f(x) > 0$ for all $x \in \mathcal{R}$ and completes the proof. □

The functional form for the optimal f^* can in fact be simplified further:

Theorem 36. *The worst-case distribution that solves problem (5.6), and therefore (5.3), takes the form*

$$f^*(x) = \frac{1}{4(\nu_0^* \min_i\{\|x - x_i\| - \lambda_i^*\} + \nu_1^*)^2} \tag{5.10}$$

with $\nu_0^*, \nu_1^* \geq 0$ *and* $\mathbf{e}^T\boldsymbol{\lambda}^* = 0$.

Proof. The major difference between the form of f^* as written above and the form described in (5.9) is the fact that the expression in (5.9) is not guaranteed to vary continuously as we move from one region R_i^* to another; the expression (5.10) is continuous by inspection. We first note that the constraint

$$\iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA \leq t \quad \forall \boldsymbol{\lambda} : \mathbf{e}^T\boldsymbol{\lambda} = 0$$

can be restricted to merely the compact set

$$\Lambda := \left\{\boldsymbol{\lambda} \in \mathbb{R}^n : \mathbf{e}^T\boldsymbol{\lambda} = 0,\, \lambda_i \leq \operatorname{diam}(\mathcal{R})\, \forall i\right\}$$

because, if $\lambda_i > \operatorname{diam}(R)$ for some i, then $\|x - x_i\| - \lambda_i < 0$ for all x and the constraint is obviously satisfied. We will apply Theorem 32 where $\mathcal{X} = L^1$, Ω is the subset of the non-negative orthant in L^1 that integrates to 1, and $\mathcal{Z}$ consists of all continuous functions on Λ, i.e. $\mathcal{Z} = \mathscr{C}(\Lambda)$ (note that $\mathcal{Z}$ satisfies the interior point requirement of Theorem 32 because inequalities are simply taken elementwise in Λ). We define $\phi(\chi) : \mathcal{X} \to \mathbb{R}$ and $\mathcal{G}(\chi) : \mathcal{X} \to \mathcal{Z}$ as the maps sending

$$f \mapsto \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$$

and

$$f \mapsto \iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA - t$$

respectively, where the right-hand side of the second expression is regarded as a continuous function of $\boldsymbol{\lambda}$. The dual space $\mathcal{Z}^*$ consists of all *regular signed Borel measures* on Λ (this is the *Riesz representation theorem*; see e.g. [179]). However, Lemma 35 shows that $f^*(x) > 0$ on $\mathcal{R}$, and therefore the optimal $\boldsymbol{\lambda}^*$ that solves problem (5.2) is unique by statement 4 of Lemma 33. This implies that $\mathcal{G}(\chi) = \iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA - t < 0$ whenever $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}^*$, and therefore, since $\langle \mathcal{G}(\chi), z^* \rangle = 0$ at optimality, it must be the case that z^* is zero everywhere except for (possibly at) $\boldsymbol{\lambda}^*$. Thus, we conclude that z^* is an evaluation functional at $\boldsymbol{\lambda}^*$ (multiplied by a scalar), so that

$$\langle \mathcal{G}(\chi), z^* \rangle = q^* \left(\iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i^*\}\, dA - t \right)$$

for all feasible f, where $q^* \geq 0$ is some scalar. Theorem 32 then says that f^* must also be the solution to the problem

$$\begin{aligned} \underset{f \in L^1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA + q^* \left(\iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i^*\}\, dA - t \right) \qquad & s.t. \\ \iint_{\mathcal{R}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R} \end{aligned}$$

or equivalently, the problem

$$\begin{aligned} \underset{f \in L^1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i^*\}\, dA &\leq t \\ \iint_{\mathcal{R}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned}$$

It is routine to verify that the constraint $\iint_{\mathcal{R}} f(x)\, dA = 1$ can be replaced with an inequality (in a nutshell, this is because we are allowed to make $f(x)$ as large as we like when $\|x - x_i\| - \lambda_i^* \leq 0$ for some index i). Thus, we can apply Theorem 32 again to the problem

$$\begin{aligned} \underset{f \in L^1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i^*\}\, dA &\leq t \\ \iint_{\mathcal{R}} f(x)\, dA &\leq 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R} \end{aligned}$$

to derive the 2-dimensional dual problem

$$\begin{aligned} \underset{\nu_0, \nu_1}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4(\nu_0 \min_i \{\|x - x_i\| - \lambda_i^*\} + \nu_1)}\, dA + \nu_0 t + \nu_1 \qquad & s.t. \\ \nu_0 \min_i \{\|x - x_i\| - \lambda_i^*\} + \nu_1 &\geq 0 \quad \forall x \in \mathcal{R} \\ \nu_0, \nu_1 &\geq 0\,; \end{aligned} \tag{5.11}$$

the optimality conditions of (5.11) describe precisely the desired form of f^*, which completes the proof. □

Remark 37. Many problems in distributionally robust optimization have objective functions and constraints that are linear in terms of the unknown distribution f (for example, the expectation operator). For such problems, Carathéodory-type theorems imply that the worst-case distribution will consist of a finite number of points, even when one uses ambiguity sets defined by the Wasserstein distance; see for example [167]. As a consequence of this fact, it is sometimes the case that one can determine the worst-case distribution using only one iteration of a finite-dimensional optimization problem; this turns out to hold (for the Wasserstein metric) in [169, 168], for example. Because of the non-linearity in the objective, our worst-case distribution f^* is smooth and requires an iterative method to solve, which we will describe in Section 5.4.

5.3.3 Variations and extensions

Uneven data weights: By definition, the empirical distribution $\hat{f}$ of the points $x_1, \ldots, x_n$ consists of a collection of n atomic masses at each point, each having mass of $1/n$. It is easy to envision scenarios in which one desires uneven weights: for example, one might use an exponential weighting scheme to emphasize more recent measurements, or one might use different weights to distinguish between activity on weekends versus weekdays (or other seasonal effects). If we require that point x_i have a mass q_i associated with it, then we can find the worst-case distribution f^* by solving problem (5.6), with the one change that we replace the restriction that $\mathbf{e}^T\boldsymbol{\lambda} = 0$ with a restriction that $\mathbf{q}^T\boldsymbol{\lambda} = 0$ instead; the form of f^* is otherwise unchanged.

Capacitated vehicles: Our approach can also be adapted to solve problems when vehicles have capacities and originate from a central depot located at the origin. To do so, suppose that each vehicle can visit c destinations before returning to the depot. The following theorem from [118] provides useful upper and lower bounds for the cost of a capacitated vehicle routing tour:

Theorem 38. *Let $X = \{x_1, \ldots, x_n\}$ be a set of demand points in the plane serviced by a fleet of vehicles with capacity c that originate from a single depot located at the origin. The length of the optimal set of capacitated VRP tours of X, written $\mathsf{VRP}(X)$, satisfies*

$$\max\left\{\frac{2}{c}\sum_{i=1}^{n}\|x_i\|, \mathsf{TSP}(X)\right\} \leq \mathsf{VRP}(X) \leq 2\left\lceil\frac{|X|}{c}\right\rceil \cdot \frac{\sum_{i=1}^{n}\|x_i\|}{|X|} + (1 - 1/c)\,\mathsf{TSP}(X)\,. \tag{5.12}$$

The probabilistic version of this, as derived in Section B.1 of the online supplement,

uses the BHH Theorem (Theorem 31 of this chapter) to characterize the length of the TSP term:

$$\sqrt{n}\cdot\max\left\{\frac{2}{s}\iint_{\mathcal{R}}\|x\|f(x)\,dA,\,\beta\iint_{\mathcal{R}}\sqrt{f_c(x)}\,dA\right\}\lesssim\mathsf{VRP}(X)\lesssim\sqrt{n}\cdot\left(\frac{2}{s}\iint_{\mathcal{R}}\|x\|f(x)\,dA+\beta\iint_{\mathcal{R}}\sqrt{f_c(x)}\,dA\right), \tag{5.13}$$

where we set $s=c/\sqrt{n}$ and we have adopted the notation "$\lesssim$" to denote an "approximate" inequality, both of which are also explained in Section B.1. It is immediately obvious that the upper and lower bounds are within a factor of 2 of one another. Applying the same analysis as in Section 5.3, the worst-case distribution that maximizes the right-hand side of (5.13) subject to a Wasserstein distance constraint takes the form

$$f^*(x)=\frac{1}{4(\nu_0^*\min_i\{\|x-x_i\|-\lambda_i^*\}+\nu_1^*-\frac{2}{s}\|x\|)^2}\,;$$

its level sets, i.e. those curves for which $\nu_0^*(\|x-x_i\|-\lambda_i^*)+\nu_1^*-\frac{2}{s}\|x\|$ is constant, consist of piecewise components of so-called *Cartesian ovals* [180].

Higher dimensions: The BHH Theorem (Theorem 31) is also applicable in higher dimensions; the general form says that, when the service region $\mathcal{R}$ belongs to $\mathbb{R}^d$, we have

$$\lim_{N\to\infty}\frac{\mathsf{TSP}(X)}{N^{(d-1)/d}}=\beta_d\iiint_{\mathcal{R}}f_c(x)^{(d-1)/d}\,dV\,,$$

for dimension-dependent constants β_d. Applying the same analysis as in Section 5.3, the worst-case distribution that maximizes the right-hand side of the above, subject to a Wasserstein distance constraint, takes the form

$$f^*(x)=\frac{(d-1)^{d-1}}{d^d}\cdot\frac{1}{(\nu_0^*\min_i\{\|x-x_i\|-\lambda_i^*\}+\nu_1^*)^d}\,.$$

The gravity model One of the salient attributes of the worst-case distribution f^* as established in Theorem 36 is that the presence of the square root in the objective of (5.3) establishes an inverse proportionality between the optimal solution $f^*(x)$ and the *square* of the distance to one of the data points x_i (with some additional additive and multiplicative weights from the dual variables $\boldsymbol{\nu}^*$ and $\boldsymbol{\lambda}^*$). This same inverse proportionality is shared by the classical *geographic gravity model* [140, 141, 142], which is "the most common formulation of the spatial interaction method" [141] and has historically been used to model a wide variety of demographic phenomena such as population migration [143], spatial utility for retail stores [144], and trip distributions between cities [145]. This would appear to lend credibility to our solution f^*, inasmuch

as it takes a form that closely matches that of distributions for related problems.

5.4 Solving (5.3) efficiently

The preceding section established that the worst-case distribution that solves (5.3) can be expressed in terms of optimal vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^n$ and $\boldsymbol{\nu}^* \in \mathbb{R}^2$. This section describes a simple method for calculating $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ efficiently by way of an *analytic center cutting plane method* [181]. Recall that our problem of interest, as written in (5.6), is

$$\begin{aligned} \underset{f \in L_1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i\}\, dA &\leq t \quad \forall \boldsymbol{\lambda} : \mathbf{e}^T \boldsymbol{\lambda} = 0 \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,; \end{aligned}$$

thus, it is certainly true that if we fix any specific value $\bar{\boldsymbol{\lambda}}$ such that $\mathbf{e}^T \bar{\boldsymbol{\lambda}} = 0$, then the following problem is a relaxation of (5.6) and hence has an objective value that is at least as large as that of (5.6):

$$\begin{aligned} \underset{f \in L_1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \bar{\lambda}_i\}\, dA &\leq t \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned}$$

It is natural to consider the problem of selecting the particular value of $\bar{\boldsymbol{\lambda}}$ that makes the above relaxation as tight as possible. In fact, our proof of Theorem 36 says that there exists a particular value of $\bar{\boldsymbol{\lambda}}$, namely $\boldsymbol{\lambda}^*$, such that the above relaxation is actually tight; in other words, the optimal distribution f^* is the solution to the problem

$$\begin{aligned} \underset{f \in L^1}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i^*\}\, dA &\leq t \\ \iint_{\mathcal{R}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R} \end{aligned}$$

for an appropriately chosen vector $\boldsymbol{\lambda}^*$. Thus, the problem of finding $\boldsymbol{\lambda}^*$ actually reduces to the optimization problem

$$\begin{aligned} \underset{\boldsymbol{\lambda}\in\mathbb{R}^n}{\text{minimize}} \max_{f\in\Omega(\boldsymbol{\lambda})} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \mathbf{e}^T\boldsymbol{\lambda} &= 0 \\ \lambda_i &\leq \operatorname{diam}(\mathcal{R}) \quad \forall i \end{aligned} \tag{5.14}$$

where $\Omega(\boldsymbol{\lambda})$ is the subset of L^1 consisting of all functions f such that

$$\begin{aligned} \iint_{\mathcal{R}} f(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA &\leq t \\ \iint_{\mathcal{R}} f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned}$$

Of course, the inner problem of maximizing f given $\boldsymbol{\lambda}$ is easily solved because the gradient vector for the dual problem

$$\begin{aligned} \underset{\nu_0,\nu_1}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4(\nu_0 \min_i\{\|x - x_i\| - \lambda_i\} + \nu_1)}\, dA + \nu_0 t + \nu_1 \qquad & s.t. \\ \nu_0 \min_i\{\|x - x_i\| - \lambda_i\} + \nu_1 &\geq 0 \quad \forall x \in \mathcal{R} \\ \nu_0, \nu_1 &\geq 0\,, \end{aligned} \tag{5.15}$$

as derived in the proof of Theorem 36, can be computed explicitly. Thus, we simply require a better understanding of problem (5.14):

Lemma 39. *The (outer) objective function of problem (5.14) is quasiconvex, i.e. its sub-level sets are convex.*

Proof. For notational compactness, let $G(\boldsymbol{\lambda})$ denote the objective function of (5.14). Recall [182] that $G(\boldsymbol{\lambda})$ is quasiconvex if and only if, for any $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2$ and any $\theta \in [0, 1]$, we have

$$G(\theta\boldsymbol{\lambda}^1 + (1-\theta)\boldsymbol{\lambda}^2) \leq \max\{G(\boldsymbol{\lambda}^1), G(\boldsymbol{\lambda}^2)\}\,.$$

Let $\bar{\boldsymbol{\lambda}} = \theta\boldsymbol{\lambda}^1 + (1-\theta)\boldsymbol{\lambda}^2$ and let $\bar{f}$ denote the distribution that maximizes $\iint_{\mathcal{R}} \sqrt{f(x)}\, dA$ over all $f \in \Omega(\bar{\boldsymbol{\lambda}})$; it will suffice to prove that either $\bar{f} \in \Omega(\boldsymbol{\lambda}^1)$ or $\bar{f} \in \Omega(\boldsymbol{\lambda}^2)$. By definition, we have

$$\iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \bar{\lambda}_i\}\, dA \leq t\,,$$

and the left-hand side of the above inequality is a concave function in $\bar{\boldsymbol{\lambda}}$ (if we fix the

function $\bar{f}$). Thus, if we let $\mathcal{S}$ denote the line segment joining $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$ (which, of course, contains $\bar{\boldsymbol{\lambda}}$), we see that the problem

$$\underset{\boldsymbol{\lambda}\in\mathcal{S}}{\text{minimize}} \iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \lambda_i\}\, dA$$

must realize its minimizer on the boundary of $\mathcal{S}$ (since we are minimizing a *concave* function), i.e. the point $\boldsymbol{\lambda}^1$ or $\boldsymbol{\lambda}^2$. Therefore, it must be the case that $\iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \lambda_i^1\}\, dA \leq t$ or $\iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \lambda_i^2\}\, dA \leq t$, which completes the proof. □

The following theorem describes a cutting plane oracle for the outer problem (5.14):

Theorem 40. *Let $\bar{\boldsymbol{\lambda}}$ satisfy $\mathbf{e}^T\bar{\boldsymbol{\lambda}} = 0$ and let $\bar{f}$ be the solution to the inner problem of (5.14) (i.e. $\bar{f}$ maximizes $\iint_{\mathcal{R}} \sqrt{f(x)}\, dA$ over all $f \in \Omega(\bar{\boldsymbol{\lambda}})$). Then the vector $\bar{\mathbf{g}} \in \mathbb{R}^n$ defined by setting*

$$\bar{g}_i = -\iint_{\bar{R}_i} \bar{f}(x)\, dA$$

for all i, where $\bar{R}_i$ is defined as

$$\bar{R}_i = \left\{x \in \mathcal{R} : \|x - x_i\| - \bar{\lambda}_i \leq \|x - x_j\| - \bar{\lambda}_j \quad \forall j \neq i\right\},$$

defines a valid cutting plane for problem (5.14); that is, if $\bar{\mathbf{g}}^T(\boldsymbol{\lambda}' - \bar{\boldsymbol{\lambda}}) \leq 0$ for some $\boldsymbol{\lambda}'$ satisfying $\mathbf{e}^T\boldsymbol{\lambda}' = 0$, and f' is the solution to the inner problem of (5.14) associated with $\boldsymbol{\lambda}'$, then $\max_{f\in\Omega(\boldsymbol{\lambda}')} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \geq \max_{f\in\Omega(\bar{\boldsymbol{\lambda}})} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$.

Proof. Statement 2 of Lemma 33 says that, for any other $\boldsymbol{\lambda}'$, the assumption that $\bar{\mathbf{g}}^T(\boldsymbol{\lambda}' - \bar{\boldsymbol{\lambda}}) \leq 0$ yields

$$\iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \lambda_i'\}\, dA \leq \iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \bar{\lambda}_i\}\, dA + \bar{\mathbf{g}}^T(\boldsymbol{\lambda}' - \bar{\boldsymbol{\lambda}}) \leq \iint_{\mathcal{R}} \bar{f}(x) \min_i\{\|x - x_i\| - \bar{\lambda}_i\}\, dA \leq t$$

which implies that $\bar{f} \in \Omega(\boldsymbol{\lambda}')$ and therefore that

$$\max_{f\in\Omega(\boldsymbol{\lambda}')} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \geq \max_{f\in\Omega(\bar{\boldsymbol{\lambda}})} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$$

as desired. □

We now have a fast method for generating cutting planes associated with problem (5.14) and thereby recovering the distribution f^* that solves problem (5.3); see Algorithm 7 for a formal description.

Input: A compact, planar region $\mathcal{R}$ containing a set of distinct points $x_1, \ldots, x_n$ which are interpreted as an empirical distribution $\hat{f}$, a distance parameter t, and a tolerance ϵ.

Output: An ϵ-approximation of the distribution f^* that maximizes $\iint_{\mathcal{R}} \sqrt{f(x)}\, dA$ subject to the constraint that $\mathcal{D}(f, \hat{f}) \leq t$.

`/* This is a standard analytic center cutting plane method applied to problem (5.14), which has an` n`-dimensional variable space. */`

Set $\texttt{UB} = \infty$ and $\texttt{LB} = -\infty$;

Set $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^n : \mathbf{e}^T\boldsymbol{\lambda} = 0,\ \lambda_i \leq \text{diam}(\mathcal{R})\ \forall i\}$;

while $\texttt{UB} - \texttt{LB} > \epsilon$ **do**

Let $\bar{\boldsymbol{\lambda}}$ be the analytic center of Λ;

`/* Build an upper bounding` $\bar{f}$ `for the original problem (5.3). */`

Let $\bar{\nu}_0, \bar{\nu}_1$ be the solution to problem (5.15) with $\bar{\boldsymbol{\lambda}}$ as an input;

Let $\bar{f}(x) = \frac{1}{4}(\bar{\nu}_0 \min_i\{\|x - x_i\| - \bar{\lambda}_i\} + \bar{\nu}_1)^{-2}$;

Let $\texttt{UB} = \iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$;

`/* Build a lower bounding` $\tilde{f}$ `that is feasible for (5.3) by construction. */`

Let $\bar{R}_i = \{x \in \mathcal{R} : \|x - x_i\| - \bar{\lambda}_i \leq \|x - x_j\| - \bar{\lambda}_j \quad \forall j \neq i\}$ for $i = \{1, \ldots, n\}$;

Let $\tilde{\boldsymbol{\nu}} \in \mathbb{R}^{n+1}$ be the solution to problem (5.8) with inputs $\bar{R}_1, \ldots, \bar{R}_n$;

Let $\tilde{f}$ be defined by setting $\tilde{f}(x) = \frac{1}{4}(\tilde{\nu}_0\|x - x_i\| + \tilde{\nu}_i)^{-2}$ on each $\bar{R}_i$;

Let $\texttt{LB} = \iint_{\mathcal{R}} \sqrt{\tilde{f}(x)}\, dA$;

Let $g_i = -\iint_{\bar{R}_i} \bar{f}(x)\, dA$ for $i = \{1, \ldots, n\}$;

Let $\mathcal{H} = \{\boldsymbol{\lambda} \in \mathbb{R}^n : \mathbf{g}^T\boldsymbol{\lambda} \geq \mathbf{g}^T\bar{\boldsymbol{\lambda}}\}$ and set $\Lambda = \Lambda \cap \mathcal{H}$;

end

return $\tilde{f}$;

Algorithm 7: Algorithm WorstTSPDensity takes as input a compact planar region containing a set of n distinct points, a distance threshold t, and a tolerance ϵ.

5.4.1 Districting Service Region

When one has multiple vehicles available to perform service, a natural strategy for allocating them – especially in the presence of uncertainty – is to use a *districting* strategy in which we divide the region $\mathcal{R}$ into sub-regions, then associate each vehicle with one of these sub-regions [43, 154, 54, 152]. In the context of this chapter, the most natural procedure would be to partition $\mathcal{R}$ into districts $D_1, \ldots, D_m$ and calculate the worst-case workloads associated with each district D_j by solving the problem

$$\begin{aligned} \underset{f \in L_1}{\text{maximize}} \iint_{D_j} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{\mathcal{R}} f(x) \min_i \{\|x - x_i\| - \lambda_i\}\, dA &\leq t \quad \forall \boldsymbol{\lambda} : \mathbf{e}^T \boldsymbol{\lambda} = 0 \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{5.16}$$

for each sub-region D_j (this is identical to (5.6) except for the domain of integration in the objective). The worst-case distribution associated with each district D_j is characterized as follows:

Theorem 41. *The worst-case distribution that solves problem (5.16) takes the form*

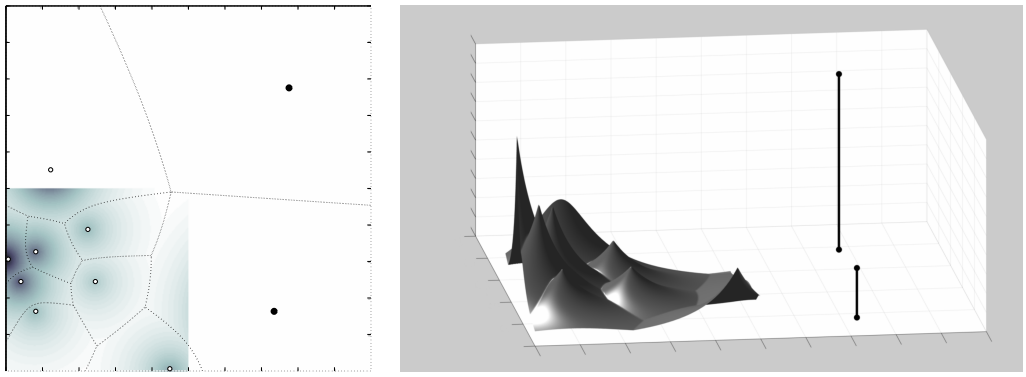
$$f^*(x) = \left[\frac{1}{4(\nu_0^* \min_i \{\|x - x_i\| - \lambda_i^*\} + \nu_1^*)^2} \right] \mathbb{1}(x \in D_j) + \sum_{i=1}^{n} p_i^* \delta(x - x_i)$$

with $\nu_0^*, \nu_1^* \geq 0$, $\mathbf{e}^T \boldsymbol{\lambda}^* = 0$, *and* $0 \leq p_i^* \leq 1$. *Moreover, we have* $p_i^* = 0$ *whenever* $x_i \in D_j$.

Proof. This is almost identical to the proof of Theorem 36 and we omit it here for brevity. The intuition behind the Dirac delta components is not difficult: if any mass of f^* is located outside district D_j, then it does not contribute to the objective and therefore should contribute as little as possible towards the Wasserstein distance constraint. See Figure 5.3 for an example of such a distribution.

□

In Section 5.5.2, we will apply the preceding result to a computational experiment in which we seek to divide a service region into districts in such a way as to minimize the maximum workload over any district.



(a) (b)

Figure 5.3: Two views of an example of $f^*(x)$ as described in Theorem 41, where there are $n = 10$ points and the sub-region D_j is the lower quarter of the unit square. At left, the shading represents f^* and the dashed lines indicate the optimal Wasserstein map between f^* and $\hat{f}$; the Dirac delta functions are indicated by the thick black circles in both images.

5.4.2 Selecting the distance parameter t

From the preceding discussion, it is clear that the parameter t in the Wasserstein distance constraint $\mathcal{D}(f, \hat{f}) \leq t$ from our original problem (5.3) has a significant impact on the problem solution. Of course, in practice, we do not have any way of *a priori* calculating an exact value of t. However, in order to estimate t in a data-driven fashion, the following result is helpful:

Theorem 42. *Let $\hat{f}_1$ and $\hat{f}_2$ denote empirical distributions associated with two sets of samples of n points from a distribution f. Then*

$$\frac{1}{2}\mathbf{E}\mathcal{D}(\hat{f}_1, \hat{f}_2) \leq \mathbf{E}\mathcal{D}(f, \hat{f}_1) \leq \mathbf{E}\mathcal{D}(\hat{f}_1, \hat{f}_2)\,.$$

Proof. This is due to [178], and follows from Jensen's inequality and the triangle inequality. □

The above result is useful because the distance between the two empirical distributions $\mathcal{D}(\hat{f}_1, \hat{f}_2)$ is simply the cost of a minimum-weight bipartite matching between the elements of $\hat{f}_1$ and $\hat{f}_2$, multiplied by a factor of $1/n$. Thus, one simple, "back-of-the-envelope" procedure to select the distance parameter t would be to sample two sets of n points each, let c be equal to the cost of the minimum-weight bipartite matching between them, and set $t = \alpha c$ with $\alpha \in [1/2, 1]$.

If we desire rigorous probabilistic bounds on t, more sophisticated machinery is required. Theorem 6.15 of [183] gives a useful bound on the Wasserstein distance between two probability density functions f_1 and f_2 by

$$\mathcal{D}(f_1, f_2) \leq \iint_{\mathcal{R}} \|x_0 - x\| \cdot |f_1(x) - f_2(x)| \, dA$$

for any $x_0 \in \mathcal{R}$. Theorem 1(i) of [184] relates the right-hand side of the above to the *relative entropy* $H(f_1|f_2)$ between f_1 and f_2 by the expression

$$\iint_{\mathcal{R}} \|x_0 - x\| \cdot |f_1(x) - f_2(x)| \, dA \leq \left(\frac{3}{2} + \log \iint_{\mathcal{R}} e^{2\|x-x_0\|} f_2(x) \, dA\right) \left(\sqrt{H(f_1|f_2)} + \frac{1}{2} H(f_1|f_2)\right),$$

where we define

$$H(f_1|f_2) = \iint_{\mathcal{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} \, dA \,.$$

Let $r = \min_{x_0 \in \mathcal{R}} \max_{x \in \mathcal{R}} \|x - x_0\|$ denote the "radius" of $\mathcal{R}$, whence $\log \iint_{\mathcal{R}} e^{2\|x-x_0\|} f_2(x) \, dA \leq \log e^{2r} = 2r$. Thus, if $\mathcal{D}(f_1, f_2) \geq t$, we have

$$\begin{aligned} t \leq \mathcal{D}(f_1, f_2) &\leq \iint_{\mathcal{R}} \|x_0 - x\| \cdot |f_1(x) - f_2(x)| \, dA \\ &\leq \left(\frac{3}{2} + 2r\right) \left(\sqrt{H(f_1|f_2)} + \frac{1}{2} H(f_1|f_2)\right) \\ \implies H(f_1|f_2) &\geq \frac{8r - 2\sqrt{16r^2 + 16rt + 24r + 12t + 9} + 4t + 6}{3 + 4r} \,. \end{aligned} \tag{5.17}$$

Next, the paper [185] shows that, for any distribution f with empirical distribution $\hat{f}$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \Pr(\mathcal{D}(f, \hat{f}) \geq t) \leq -\alpha(t) \,,$$

where the function $\alpha(t)$ is defined as

$$\alpha(t) = \inf_{g : \mathcal{D}(f,g) \geq t} H(f|g) \,.$$

the result (5.17) establishes that $\alpha(t) \geq (8r - 2\sqrt{16r^2 + 16rt + 24r + 12t + 9} + 4t + 6)/(3 + 4r)$, and therefore we find that

$$\begin{aligned} \limsup_{n \to \infty} \frac{1}{n} \log \Pr(\mathcal{D}(f, \hat{f}) \geq t) &\leq -\frac{8r - 2\sqrt{16r^2 + 16rt + 24r + 12t + 9} + 4t + 6}{3 + 4r} \\ \implies \Pr(\mathcal{D}(f, \hat{f}) \geq t) &\lesssim \exp\left(-n \frac{8r - 2\sqrt{16r^2 + 16rt + 24r + 12t + 9} + 4t + 6}{3 + 4r}\right), \end{aligned} \tag{5.18}$$

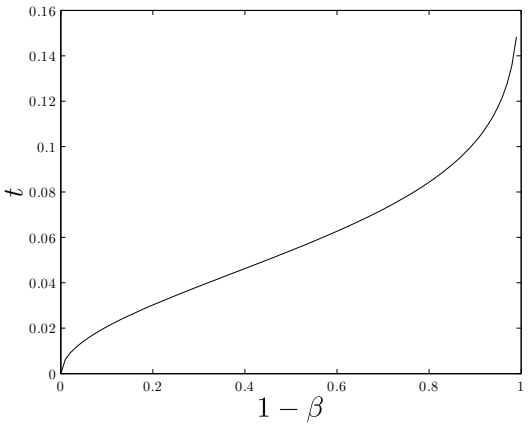


Figure 5.4: Threshold values of t for significance levels between 0 and 1, where $n = 100$ points are sampled in the unit square. For example, at $1 - \beta = 0.9$, we have $t = 0.102$; this means that, if 100 samples are drawn from any distribution f in the unit square, then there is at least a 90% probability that $\mathcal{D}(f, \hat{f}) \leq 0.102$.

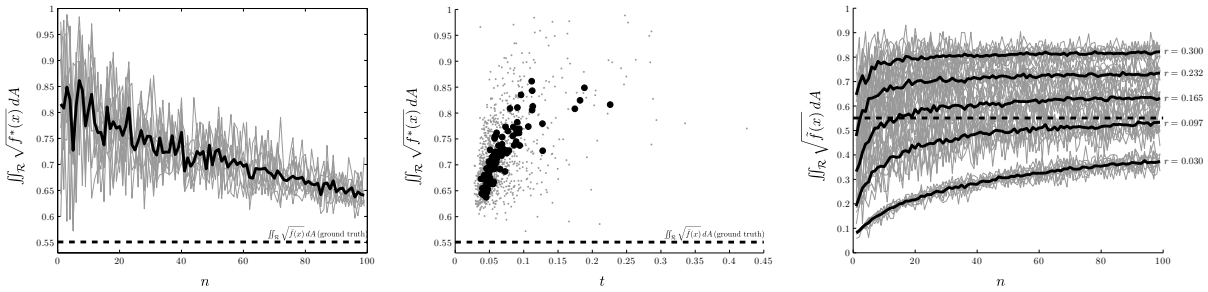
where the notation "$\lesssim$" reflects the approximate inequality that results from dropping the "lim sup" term. Thus, given a desired significance level $1 - \beta$, we can construct a threshold distance t by equating the right-hand side of (5.18) to $1 - \beta$ and solving for t. Figure 5.4 shows a plot of these threshold values of t as a function of β, for $n = 100$ samples in the unit square.

5.5 Computational experiments

In this section, we apply our theoretical results to two computational experiments: the first experiment shows the impact of increasing the number of samples n, and the second is a districting strategy in which we divide a map of Los Angeles County into pieces so as to to minimize the worst-case workload of any vehicle.

5.5.1 Varying values of n

In our first experiment, we let $\mathcal{R}$ be the unit square, and as a ground truth distribution $\bar{f}$ we use an even mixture of two truncated Gaussian distributions with means $\mu_1, \mu_2 = (0.400, 0.187), (0.795, 0.490)$ and covariance matrices $\Sigma_1 = \Sigma_2 = \left(\begin{smallmatrix} 0.070 & 0 \\ 0 & 0.070 \end{smallmatrix}\right)$. This mixture was chosen because it satisfies $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA = 0.55$ and therefore represents a compromise between extreme clustering (which would have $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$ close to zero) and a perfect uniform distribution (which would have $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$ equal to one). For $n \in \{2, \ldots, 100\}$,



(a) (b) (c)

Figure 5.5: Figure 5.5a shows the worst-case costs that are computed during the 99×10 executions of our algorithm; the gray plots indicate the results obtained from individual samples and the thick line indicates the sample averages of the 10 trials for fixed n. Figure 5.5b shows the same data set, only we plot the worst-case costs as a function of the Wasserstein distance threshold t rather than a function of n; the gray points indicate individual experiments and the dark points again indicate the sample averages of the 10 trials for fixed n. For purposes of comparison, Figure 5.5c shows the estimates of $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\,dA$ obtained when one uses a uniform kernel density estimator; that is, if we draw n samples $x_1, \ldots, x_n$ from $\bar{f}$, then we define an estimator $\tilde{f}$ by setting $\tilde{f}(x) = \frac{1}{C}\sum_i \mathbb{1}(\|x - x_i\| \leq r)$, where r is a "bandwidth" parameter and C is a normalization constant. As in the preceding two figures, the gray plots indicate the results from individual samples, the thick lines indicate sample averages of 10 trials for fixed n, and the dashed line indicates the true value of $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\,dA$; furthermore, as indicated, we used 5 different values of r between 0.03 and 0.3; note that the estimate $\iint_{\mathcal{R}} \sqrt{\tilde{f}(x)}\,dA$ is highly sensitive to the choice of r.

we performed 10 independent experiments where we drew n samples from $\bar{f}$ and then obtained the worst-case TSP distribution f^* by solving problem (5.3) via Algorithm 7 (hence, 99×10 experiments in total). For each experiment, we defined our distance constraint using Theorem 42 by setting t to be the cost of a minimum-weight bipartite matching between two independent collections of samples of size n from $\bar{f}$ (multiplied by a factor of $1/n$). Figure 5.5a shows a plot of the worst-case TSP costs $\iint_{\mathcal{R}} \sqrt{f^*(x)}\,dA$ as n varies, and Figure 5.5b shows the same data, only using the Wasserstein distance threshold t as the independent variable. Not surprisingly, it is clear that the worst-case cost decreases as n increases and as t decreases. Figure 5.5b suggests that the worst-case cost, measured as a function of t, decreases in a *concave* fashion as $t \to 0$. For purposes of comparison, Figure 5.5c shows the estimates of $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\,dA$ obtained when one uses a uniform kernel density estimator.

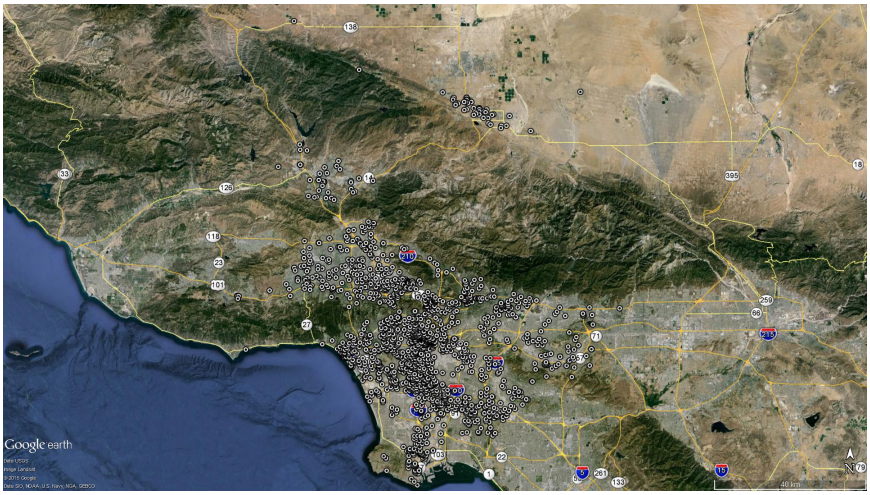


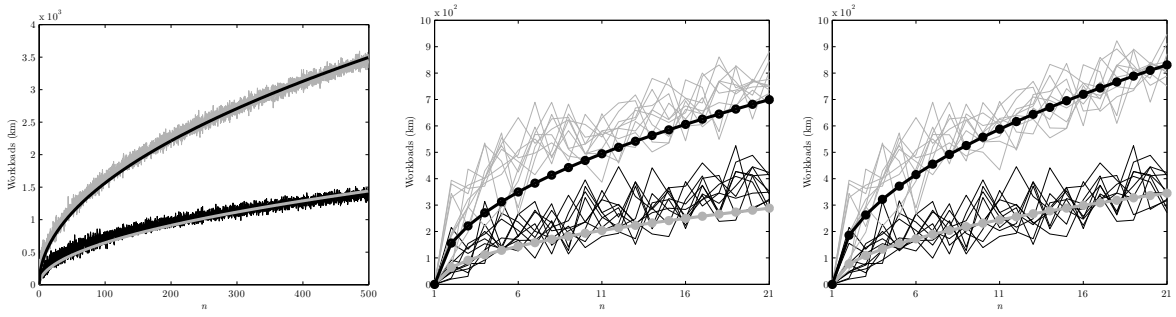
Figure 5.6: Locations of 1704 crime reports filed in Los Angeles County during the first week of July, obtained from the website [3].

5.5.2 A districting experiment with road network data

In this section, we describe an experiment in which we divide a service region $\mathcal{R}$ into 4 pieces so as to allocate the workloads of a fleet of vehicles. This experiment is much more elaborate than that of the preceding section because we compute our TSP tours using data from an actual road network, rather than simply make an assumption that distances are Euclidean. Specifically, our service region $\mathcal{R}$ is a map of Los Angeles County, and distances between points are measured according to the driving distance, as obtained via the Google Distance Matrix API [6]. Our sampled points $x_1, \ldots, x_n$ are the locations of crime reports filed in the first week of July, which were extracted from the "Detailed Report" tool at the CrimeMapping.com website developed by the Omega Group [3], and are shown in Figure 5.6. The purpose of this experiment is two-fold: first, we must demonstrate that the proposed continuous approximation techniques are actually useful for solving practical problems, and second, we must then show that our proposed methodology is superior to that of existing approaches.

Validation of continuous approximation methods

In order to apply our results to solve a practical problem, it is necessary to first confirm that the continuous approximation method remains valid and useful even when point-to-point distances are not Euclidean. As an initial sanity check, we first sample between $n = 2$ and $n = 500$ points uniformly at random within the map $\mathcal{R}$ of Los Angeles County, and solve the TSP for these points where distances are given by the driving distance as obtained from the



(a) (b) (c)

Figure 5.7: The lengths of the TSP tours of n points sampled in Los Angeles County, where point-to-point distances are induced by a road network. In all three diagrams, the upper plot corresponds to uniformly distributed points and the lower plot corresponds to points that are sampled from the CrimeMapping.com website [3]. For each value of $n \in \{1, \ldots, 500\}$ and for each of the two sampling strategies (uniform or non-uniform), we perform 10 independent experiments in which n points are sampled and their TSP tour is calculated using Concorde [4]. Figure 5.7a shows the tour lengths (the thin lines) together with the best fit of these tour lengths to a curve of the form $C\sqrt{n}$, where we have $C \approx 159$ for uniformly sampled points and $C \approx 64$ for non-uniformly distributed points. Figure 5.7b shows a close-up of this plot for $n \in \{1, \ldots, 21\}$, where we can see that the fitted curve underestimates the true tour lengths when n is small. A better fit for these small values of n, as shown in Figure 5.7c, is to set $C \approx 185$ for uniformly sampled points and $C \approx 77$ for non-uniformly distributed points (as an aside, for small values of n, it is clearly common sense to fit a curve of the form $C\sqrt{n-1}$, since the TSP tour of a single point has length 0). This establishes that, provided one has an approximate estimate of the number of points n, the square root approximation is indeed a reasonable one.

Google Distance Matrix API [6]. We also sample between $n = 2$ and $n = 500$ points from the non-uniform distribution consisting of the locations of crime reports obtained from [3], as shown in Figure 5.6. The lengths of these tours, computed via the Concorde TSP solver [4], are shown in Figure 5.7, and are consistent with the findings of Table 16.7 of [133] for the Euclidean TSP in the unit square: specifically, we see that the square-root approximation tends to slightly *underestimate* the tour length for small values of n. For example, [133] says that a TSP tour of $n \approx 100$ points in the unit square (with Euclidean distances) is approximately $0.78\sqrt{n}$, whereas a TSP tour of $n \approx 1000$ points is approximately $0.73\sqrt{n}$. Figure 5.7 suggests that, provided one has some vague estimate of the number of destination points n, the square root approximation is indeed a valid one, even when distances are not Euclidean and the spatial distribution is not uniform.

We have thus far established that the length of a TSP tour of n points in $\mathcal{R}$ scales proportionally to $\sqrt{n}$. However, it is necessary to take into account the heterogeneity of

the road network as well: for example, point-to-point distances in the middle of downtown Los Angeles are likely to be close to the ℓ_1 metric because the streets form a regular grid, whereas point-to-point distances elsewhere will likely be longer due to sparser road coverage. In order to take this factor into account, we should first note that Theorem 31 actually holds under much more general conditions than the Euclidean TSP, and remains valid when one considers the TSP under many "natural" norms or even other combinatorial structures such as the minimum spanning tree or Steiner tree (more precisely, Theorem 31 holds whenever the underlying structure is a *subadditive Euclidean functional*; see [83, 82] for an extensive study thereof). Obviously, the coefficient β depends on the choice of structure. The following example is useful in designing an appropriate framework to handle this disparateness:

Example 43 (Varying metrics in a region)**.** Consider a set of n points sampled according to a distribution f in the unit square, with distances $d(x_1, x_2)$ between pairs of points $x_1 = (x_1^1, x_1^2)$ and $x_2 = (x_2^1, x_2^2)$ defined as follows:

- If x_1 and x_2 are in the lower left quadrant, then $d(x_1, x_2)$ is the Euclidean distance between x_1 and x_2.
- If x_1 and x_2 are in the lower right quadrant, then $d(x_1, x_2)$ is the ℓ_1 distance between x_1 and x_2.
- If x_1 and x_2 are in the upper left quadrant, then $d(x_1, x_2)$ is the ℓ_∞ distance between x_1 and x_2.
- If x_1 and x_2 are in the upper right quadrant, then $d(x_1, x_2) = \sqrt{(x_1 - x_2)^T A (x_1 - x_2)}$, where A is a symmetric positive definite matrix.
- If x_1 and x_2 are in different quadrants, then $d(x_1, x_2)$ is determined by a tie-breaking rule of some sort (the details of which are not relevant).

The TSP tour of a set of points under these assumptions is shown in Figure 5.8. If we let $Q_1, \ldots, Q_4$ denote the four quadrants of the square, then it is routine to verify that we in fact have

$$\lim_{N\to\infty} \frac{\mathsf{TSP}(X_1, \ldots, X_N)}{\sqrt{N}} = \sum_{i=1}^{4} \beta_i \iint_{Q_i} \sqrt{f_c(x)}\, dA$$

where each β_i is associated with the metric on quadrant Q_i (e.g. β_1 is the Euclidean TSP coefficient); one can verify this by proceeding through the proof of the BHH theorem in (for example) Chapter 2.4 of [82].

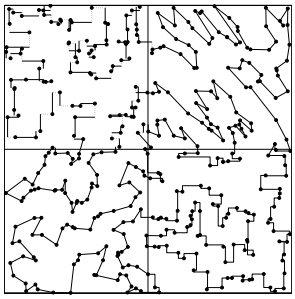


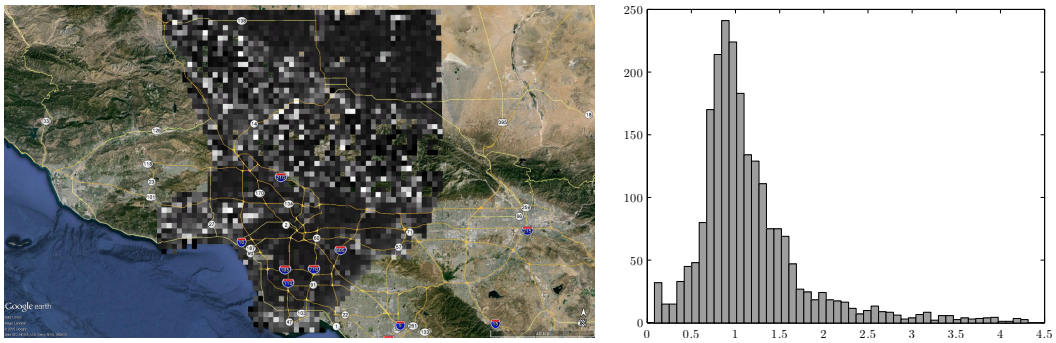
Figure 5.8: The TSP tour of a collection of points uniformly sampled in the unit square, with varying metrics depending on quadrants. The dashed lines in the paths in the upper left quadrant correspond to the smaller of the two directions (horizontal or vertical) between points, which is relevant because the ℓ_∞ distance is used.

Example 43 suggests a general approach that is useful when the service region $\mathcal{R}$ has a heterogeneous road network: if we decompose $\mathcal{R}$ into a collection of "patches" $Q_1, \ldots, Q_K$, then we adopt the approximation

$$\mathsf{TSP}(X_1, \ldots, X_n) \approx \sqrt{n} \cdot \sum_{i=1}^{K} \beta_i \iint_{Q_i} \sqrt{f_c(x)}\, dA\,.$$

One can estimate the values β_i as follows: if we sample a set of k points *uniformly* in Q_i and compute the length of their TSP tour ℓ_i (using road network distances), we would expect to see that $\ell_i \approx \beta_i \sqrt{\mathrm{Area}(Q_i) \cdot k}$; this is simply the uniform case of the BHH theorem applied to points constrained to lie in Q_i. Thus, a sensible estimate of β_i is given by setting $\beta_i = \ell_i / \sqrt{\mathrm{Area}(Q_i) \cdot k}$. We discretized the region $\mathcal{R}$ into a collection of patches Q_i of size $2\,\mathrm{km} \times 2\,\mathrm{km}$, and estimated each coefficient β_i using $k = 10$ samples (a larger number would be preferable, but the Google Maps API [6] imposes a limit of at most $100,000$ queries per day); Figure 5.9 shows the resulting values. We found a total of $K = 2564$ patches in which road coverage was adequate for distances to be estimated, whence $\mathrm{Area}(\mathcal{R}) = (2\,\mathrm{km} \times 2\,\mathrm{km}) \cdot K = 10256\,\mathrm{km}^2$.

In order to validate these estimates β_i, we revisit the earlier experiment in which we sampled up to 500 points in $\mathcal{R}$, where we found (as suggested in Figure 5.7a) that the length of the TSP tour of n points sampled uniformly in $\mathcal{R}$ is approximately $159\sqrt{n}$ kilometers. It follows that, for uniformly distributed points X_i in $\mathcal{R}$, the fact that $f(x) = 1/\,\mathrm{Area}(\mathcal{R})$



(a) (b)

Figure 5.9: The shading in Figure 5.9a indicates the values of β_i associated with each of the square patches (lighter patches correspond to higher values of β_i). Figure 5.9b is a histogram of these same values; note that a handful of these values are actually *lower* than the current estimate [5] of the Euclidean TSP coefficient $\beta \approx 0.7124$; this appears to be a combination of statistical noise and an artifact of the Google Maps API [6] that was used to compute driving distances; for example, two locations belonging to the same venue (e.g., two stores on opposite ends of a large shopping mall) report a driving distance of 0.

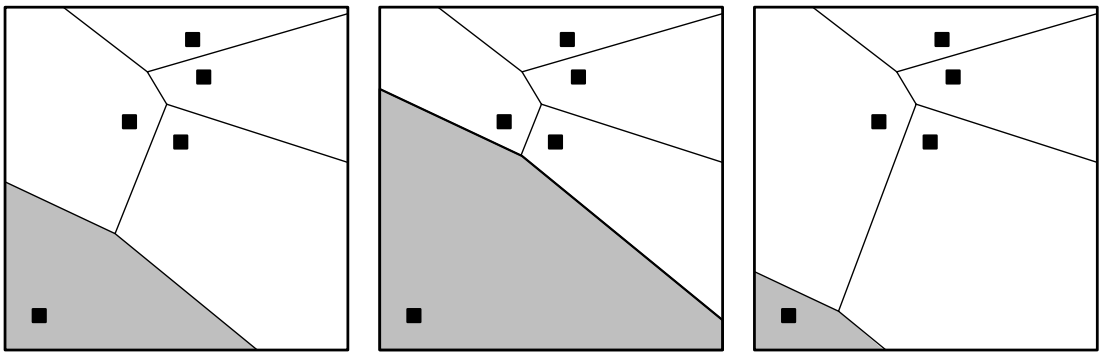
implies that

$$\begin{aligned} 159\sqrt{n} \approx \mathsf{TSP}(X_1, \ldots, X_n) &\approx \sqrt{n} \cdot \sum_{i=1}^{K} \beta_i \iint_{Q_i} \sqrt{f_c(x)}\, dA \\ &= \sqrt{n} \cdot \sum_{i=1}^{K} \beta_i \cdot (4\,\mathrm{km}^2) \cdot \sqrt{1/\operatorname{Area}(\mathcal{R})}\, dA \\ &= \sqrt{n} \cdot 0.0395 \sum_{i=1}^{K} \beta_i\,, \end{aligned}$$

and so we expect to find that $0.0395 \sum_{i=1}^{K} \beta_i$ should sum to approximately 159. Indeed, we find that $0.0395 \sum_{i=1}^{K} \beta_i \approx 142$, so that we introduce a relative error of approximately 10%. In order to compensate for this error, we re-scale all terms $\beta_i \mapsto \frac{159}{142}\beta_i$ for all i.

Districting criteria

In this section, we apply the theory developed in this chapter to solve a problem in which we seek to partition $\mathcal{R}$ (a map of Los Angeles County) into service districts so as to divide the workloads of a fleet of vehicles in a balanced way. In order to divide $\mathcal{R}$ into districts, we use a computational geometric structure called a *power diagram* [147], which has frequently been applied to districting problems in existing literature on vehicle routing [186, 65, 150, 187, 152]. Given a set of "depot points" $p_1, \ldots, p_m$ in $\mathcal{R}$ and any vector $\mathbf{w} \in \mathbb{R}^m$, the power



(a) (b) (c)

Figure 5.10: Figure 5.10a is a Voronoi partition, that is, a power diagram with all weights w_i equal (each district consists of those points that are closer to their associated landmark point than the others). Figures 5.10b and 5.10c present the two power diagrams obtained by increasing and decreasing the weight associated with the shaded cell.

diagram of $\mathcal{R}$ with respect to $p_1, \ldots, p_m$ and $\mathbf{w}$ is a partition of $\mathcal{R}$ into districts $D_1, \ldots, D_m$ defined by

$$D_i = \left\{ x \in \mathcal{R} : \|x - p_i\|^2 - w_i \leq \|x - p_j\|^2 - w_j \; \forall j \neq i \right\} . \tag{5.19}$$

An example of a power diagram is shown in Figure 5.10. It is straightforward to verify that the pieces D_i are always convex. In our experiment, we let $p_1, \ldots, p_{m=4}$ be the locations of the 4 major police stations associated with the 4 largest cities in Los Angeles County, namely Los Angeles, Long Beach, Glendale, and Santa Clarita, which are shown in Figure 5.11. Given these locations p_i, our goal is to select a weight vector $\mathbf{w}$ that determines an "optimal" partition $D_1, \ldots, D_4$ with respect to some cost function that approximates the workloads in these regions; selected cost functions will be described later in this section.

It would of course be desirable to replace the Euclidean distance terms in (5.19) with the driving distance (which would result in boundaries between sub-regions that are characterized, in some sense, by the underlying road network). Unfortunately, for practical reasons, we are prevented from doing this because of the limit of $100,000$ queries per day of the Google Maps API, which we used to calculate driving distances. When one does not have such a limit (e.g. when driving distances are computed "in-house"), then obviously, such a method is indeed feasible. We next describe three objective functions or attributes associated with the districts D_i that we seek to optimize by selecting the weight vector $\mathbf{w}$.

Equal n_i: A common-sense criterion in designing districts would be to require that the number of customers in each district D_i, which we denote as n_i, be the same. This is eminently sensible when the spatial distribution of customers is uniform, and has been

Figure 5.11: Locations of 4 police stations associated with the 4 largest cities in Los Angeles County.

applied previously to large-scale routing problems in [188, 189], for example. Given a collection of points $x_1, \ldots, x_n$ in $\mathcal{R}$, it turns out that it is computationally extremely simple to compute a weight vector $\mathbf{w}^*$ such that each district D_i contains the same number n/m of customers (with the possibility of being off by one if m does not divide n evenly): the desired weight vector is the Lagrange multiplier vector associated with the first set of constraints in the assignment problem

$$\begin{aligned} \underset{z_{ij}}{\text{minimize}} \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} z_{ij} \qquad & s.t. \\ \sum_{i=1}^{n} z_{ij} &= \frac{n}{m} \; \forall j \\ \sum_{j=1}^{m} z_{ij} &= 1 \; \forall i \\ z_{ij} &\geq 0 \; \forall i, j \, , \end{aligned}$$

where we set $c_{ij} = \|x_i - p_j\|^2$. This is solvable as a linear program.

Equal $\sqrt{A_i n_i}$: By far the most popular approximation in designing districts for vehicle routing problems is the estimation

$$\mathsf{TSP}(D_i) \approx \beta \sqrt{A_i n_i} \, ,$$

where $A_i = \text{Area}(D_i)$, n_i is the number of customers in D_i as before, and the notation

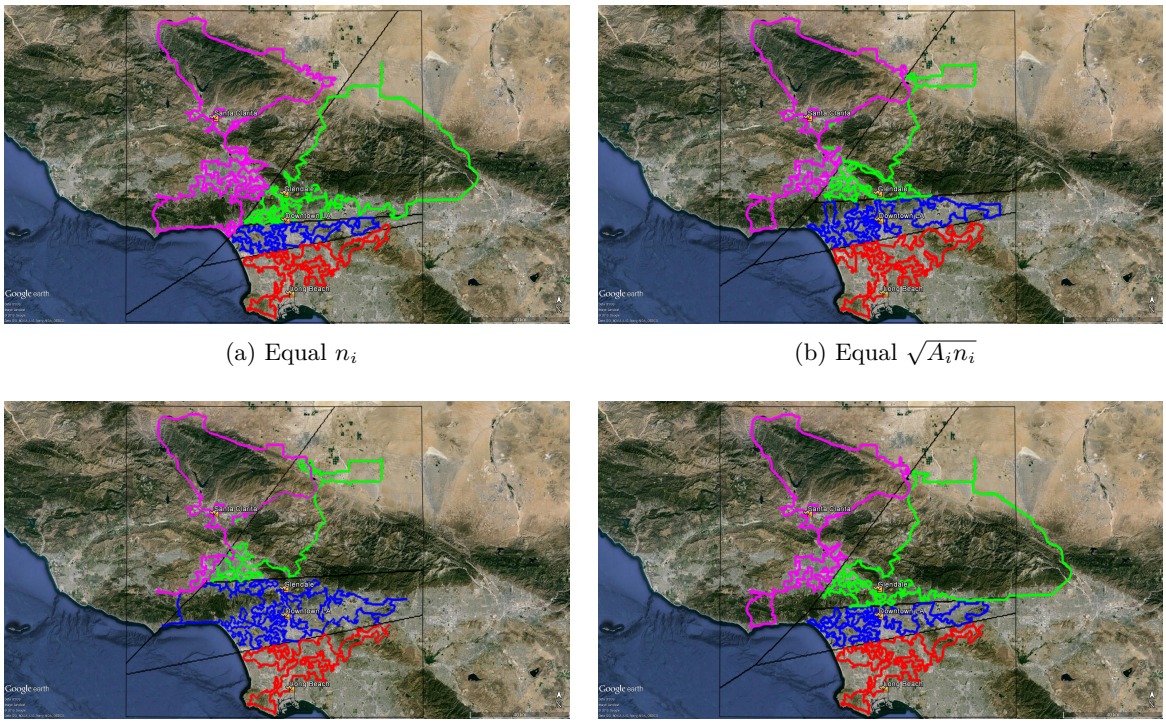
$\mathsf{TSP}(D_i)$ denotes the length of the TSP tour through the n_i points in D_i. This has been used previously in [104, 149, 155, 156, 190], for example, and is predicated on the assumption that the points are uniformly distributed *within* each district D_i (in other words, the distribution may vary over $\mathcal{R}$, but the distribution is assumed to be more or less uniform within each of the districts). This is nothing more than the BHH theorem, applied to a set of points that are uniformly distributed within each district D_i. It is possible to compute a weight vector $\mathbf{w}^*$ such that $\sqrt{A_i n_i}$ is equal for all districts D_i using a simple gradient descent scheme similar to that used in [152] (with the same caveat that an "off by one" error may exist as in the previous example).

Mean-covariance robust partitioning: Section 5 of the paper [65] describes a branch-and-bound method for partitioning a region $\mathcal{R}$ into a power diagram partition in which the worst-case workloads for all districts D_i are equal. Here, the "worst-case workloads" are defined via robust optimization, specifically the solution to (5.4). In other words, we construct a power diagram partition against all distributions whose mean and covariance matrix are equal to the fixed values obtained from the sampled points $x_1, \ldots, x_n$.

Wasserstein robust partitioning Problem (5.16) in Section 5.4.1 of this chapter describes the structure of the worst-case distribution that maximizes the asymptotic workload in a particular district D_i, subject to a Wasserstein distance constraint. Thus, it is sensible to seek a weight vector $\mathbf{w}^*$ that results in districts $D_1, \ldots, D_m$ such that the solution to (5.16) is equal for each district (in other words, the worst-case workloads are the same for all districts). The branch-and-bound scheme from Section 5 of [65] is based on a simple set of monotonicity properties and can be applied to find these districts as well.

Results

In order to demonstrate the practicality of our proposed approach, we compare the districts that result from enforcing the four criteria from the preceding section, where the data points are the locations of crime reports filed in the first week of July as previously shown in Figure 5.6. We give the first three partitioning criteria (i.e. the non-Wasserstein criteria) an advantage by building their districts using knowledge of all 1704 sample points (the mean-covariance partitioning scheme uses exact knowledge of the mean and covariance of the data points). By comparison, the Wasserstein partitions are computed using a sample of only 50 points drawn from the full dataset, and with a threshold distance t calculated



(a) Equal n_i

(b) Equal $\sqrt{A_i n_i}$

(c) Robust mean-covariance

(d) Wasserstein

Figure 5.12: The power diagram districts obtained according to the four partitioning criteria.

according to equation (5.18) with a 90% confidence level. The four sets of districts, together with the resulting workloads, are shown in Figures 5.12 and 5.13. Not surprisingly, we see that the non-uniformity of the samples leads to districts whose workloads are substantially unbalanced. Even more surprising is the fact that the mean-covariance robust partitioning method is by far the *worst* of the three; we attribute this to the fact that the crime locations are distributed in a highly multi-modal, non-uniform fashion, and that the worst-case distribution with given first and second moments as derived in [65] always has a unimodal shape (together with a mixture of Dirac delta components).

This experiment establishes that our approach is useful for balancing workloads of vehicles when one has limited sample information and when the underlying distribution is highly non-uniform. We also conducted another set of experiments in which we sampled 1000 points *uniformly* in $\mathcal{R}$, as shown in Figure 5.14, and applied the four districting criteria. As Figure 5.15 shows, the four criteria are roughly indistinguishable in this case.

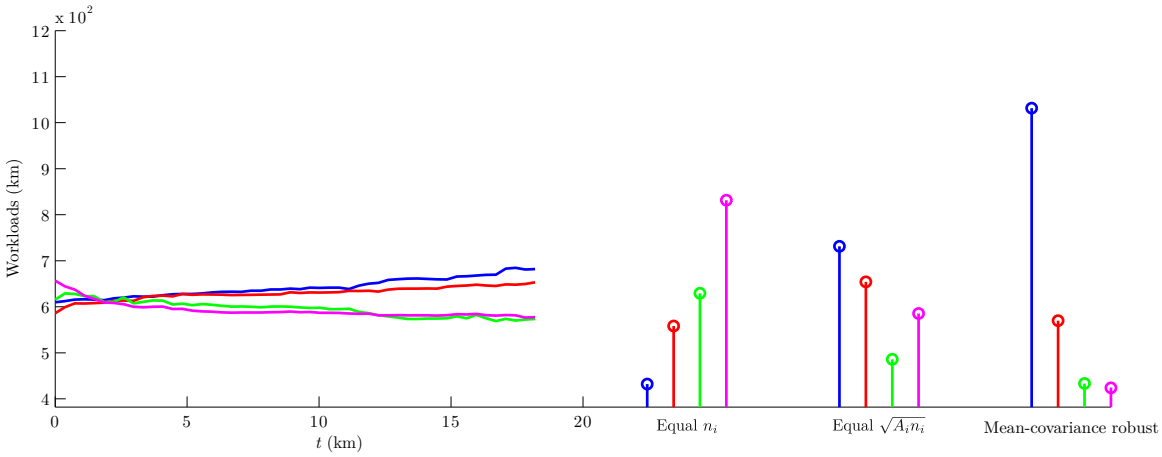


Figure 5.13: The workloads associated with the four districts and the four partitioning criteria from Figure 5.12, interpreted as follows: the colors of the various plots correspond to the workloads in the district of the same color (e.g. the magenta plots correspond to the magenta district, which belongs to Santa Clarita). The left set of plots corresponds to the workloads that result when we design districts according to the Wasserstein partitioning criterion: more precisely, for each value of t in the range shown, we construct a weight vector $\mathbf{w}^*$ such that the worst-case workloads in problem (5.16) are all equal. Thus, different values of t correspond to different values of $\mathbf{w}^*$, and thereby different partitions. The left-hand plot shows the true workloads for each of the districts as t varies; the maximum value of t of 18.3 km is calculated according to equation (5.18) with a 90% confidence level with $n = 50$ samples. The three sets of stem plots on the right show the workloads in each of the four districts that are obtained when one partitions according to the first three criteria of Section 5.5.2. The Wasserstein partitioning criterion consistently produces districts whose workloads are more balanced than those of the other three criteria, even though the Wasserstein partitions are constructed using a small number of samples, whereas the other three methods are actually permitted to make complete use of all 1704 sample points (the mean-covariance partitioning scheme uses exact knowledge of the mean and covariance of the data points). Surprisingly, we found that the the mean-covariance robust partitioning method is by far the *worst* of the three.

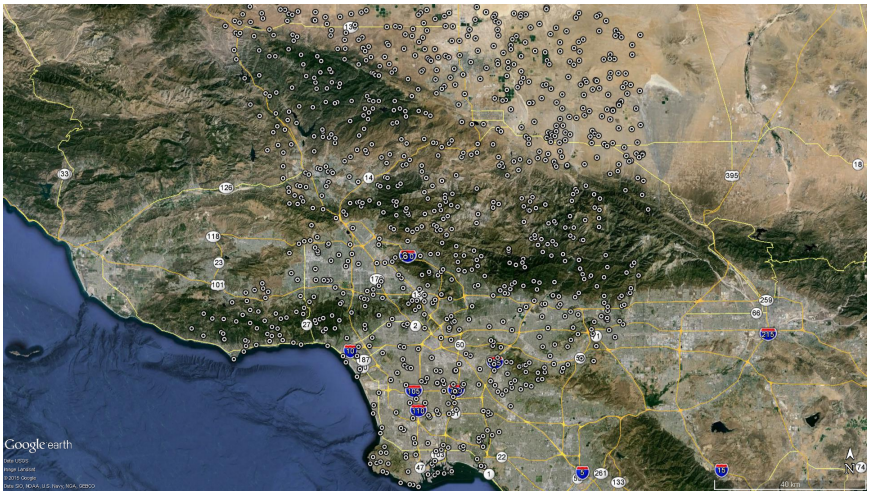
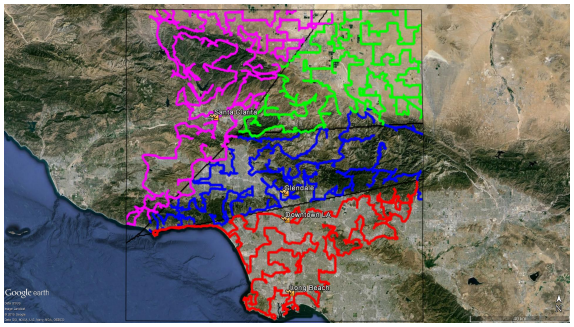
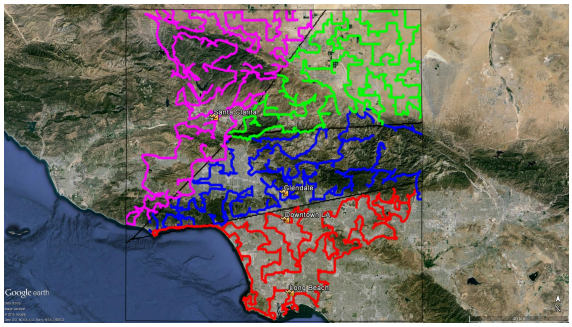


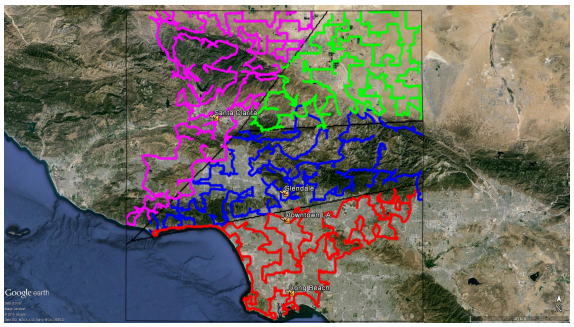
Figure 5.14: 1000 uniformly sampled points in Los Angeles County.



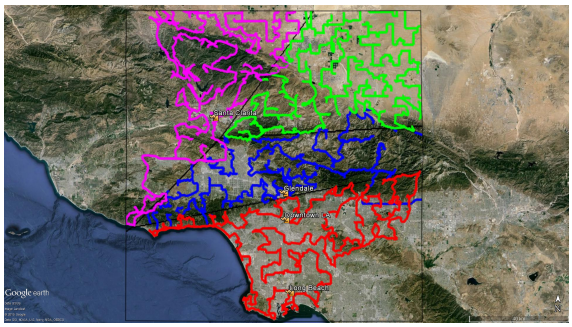
(a) Equal n_i



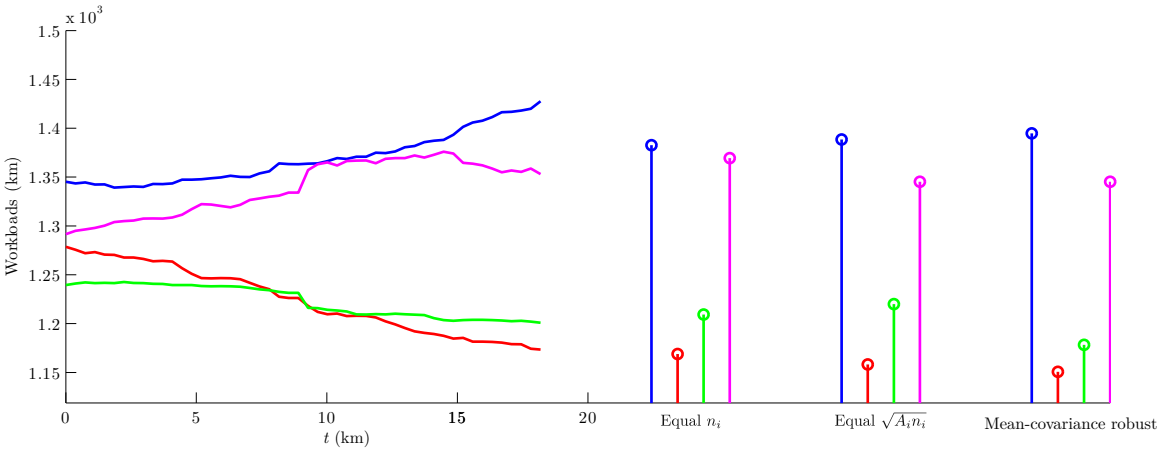
(b) Equal $\sqrt{A_i n_i}$



(c) Robust mean-covariance



(d) Wasserstein



(e) Workloads

Figure 5.15: The power diagram districts obtained according to the four partitioning criteria, and the workloads that result therein, for the case where all samples are uniformly distributed. In this case, all four partitioning criteria perform more or less the same.

Chapter 6

Conclusion

In this research we provided robust solutions for some of the geographic resource allocation problems and conversely provide new tools for robust optimization using the techniques of geometric partitioning. We have accomplished to finish four research problems, namely designing robust networks, approximation algorithms for k-centers problem, finding worst-case demand distribution in vehicle routing, and developing a data-driven model for the distributionally robust travelling salesman problem.

In the robust network design problem, presented in Chapter 2, we have presented three approximation algorithms for partitioning a convex region C into sub-regions so as to minimize the connectivity radius of any set of points inside those sub-regions. Our analysis further led us to a fourth algorithm, based on similar principles, that solves the continuous k-centers problem in a convex polygon with approximation factor 1.99, presented in Chapter 3. To the best of our knowledge there is no other approximation algorithm for this problem with approximation factor less than 2.

In Chapter 4, we have considered the asymptotic behavior of the vehicle routing problem with time windows and the capacitated vehicle routing problem in the Euclidean plane under limited information about the demand distribution. In studying the VRP with time windows, we find that the worst possible temporal distribution of demand is either to have a large amount of demand concentrated in a single time period (when total demand is small), or uniform demand throughout all time periods (when total demand is large). Thus, for upstart businesses with a small customer base, the worst possible outcome is to have a single period of concentrated activity, and for established business with a large customer base, the worst possible outcome is to have all demand spread uniformly throughout the service period.

For the capacitated VRP, we initially found that the worst-case distribution follows an

"inverse square" law that closely resembles the classical gravity model of spatial interaction. In order to extend this result to a decision support system, we next computed a closed-form expression for the worst-case demand distribution when one knows the first and second moments and as well as the support of the distribution, which we used in a computational districting experiment. Such districting strategies are useful when one has limited information about the locations of demand points and seeks an assignment scheme that can be performed without full knowledge of the demand distribution.

In order to overcome the intrinsic issues accompanied with methods that merely rely on the information about the moments of demand distribution, we proposed a new model in Chapter 5. By using the Wasserstein distance to define our region of ambiguity, we have developed a new data-driven tool for estimating the worst-case workload that one might face in visiting a sequence of points that is not affected by problems that would arise if we used only mean and covariance information as has been previously attempted. To the best of our knowledge, our use of the Wasserstein distance in such an application, that first appeared in our article [90], is the first of its kind.

6.1 Future Work

There are so many possible extensions to the above works and it is hard to enlist all of them, but here are a few that are more interesting.

6.1.1 Extensions for Robust Network Design

Imposing Different Shape Constraints One potential direction for future research regarding the design of robust networks would be the imposition of other (weaker) shape constraints on the sub-regions (as opposed to equal area or convexity), such as requiring star convexity or simple connectivity (i.e. no "holes"); structures of such shapes are prevalent in the computational geometry literature, see e.g. [191, 192], and we suspect that efficient approximation algorithms exist for such scenarios as well.

k-Medians Problem We can also use the algorithm used for k-centers problem to solve k-medians problem and our conjecture is that it improves the existing results including [13, 14].

6.1.2 Extensions for Distributionally Robust VRP

One might seek to understand the impact of *first* moment information in a back-of-the-envelope analysis similar to that in Section 4.5.1; we expect that the location of the center of mass of a distribution should impact the worst-case workloads in a distinctly different manner from that of the covariance matrix and we intend to study it in the near future. Another natural direction for further research would be to tighten the initial bounds in (5.13). This appears tractable because they are based on the earlier bounds in (5.12). Since (5.12) holds for any Euclidean point set, it is stronger than an asymptotic result, and it would appear that this property may be exploitable in improving (5.13). Moreover, it is worth to investigate applying alternate distributional ambiguity sets (as opposed to support and first and second moment information) which could be useful for the distributionally robust VRP; in Chapter 5, we have presented one possibility which is the use of Wasserstein distance. Other possibilities would be the use of other statistical metrics such as the *Kullback-Leibler divergence* or *Kolmogorov-Smirnov statistic*, which have recently been applied to distributionally robust optimization problems in [170, 171] and will likely yield additional insights into the costs of providing service in VRP.

6.1.3 Extensions for Distributionally Robust Optimization

Our use of the square root functional $\iint_{\mathcal{R}} \sqrt{f(x)}\, dA$ to approximate lengths of TSP tours is just one possibility; we can extend this analysis to handle more sophisticated objective functionals in the field of transportation and network optimization. This may also have further applications outside the transportation domain, e.g. in information theory, threat detection, and finance, where we will have different objectives such as entropy maximization, highest posterior density (HPD) minimization, and conditional value at risk (CVaR) minimization. Moreover, we can develop a general framework for distributionally robust optimization with uncertainty set being defined by Wasserstein distance.

6.1.4 Robust Optimization for Other Transportation Network Models

Almost all the network structures presented in Figure 1.1 have enormous applications in facility location and transportation models. In this dissertation, we have presented robust solutions for two most common structures, namely TSP tours and VRP tours, but it is worth to further investigate similar approaches for other structures such as minimum spanning tree (MST) or Steiner tree.

One might also want to apply the methods of this research for the generalized version

of these networks. For example in our article [193], we have investigated the asymptotic analysis of the behavior of a *generalized travelling salesman problem* (GTSP) under the assumption that all relevant locations are independently and identically distributed uniformly in a region. Finding the worst-case distribution for the location of destination points in GTSP is an interesting topic for further studies.

Bibliography

[1] J. B. M. Melissen and P. C. Schuur. Covering a rectangle with six and seven circles. *Discrete Applied Mathematics*, 99(1):149–156, 2000.

[2] J. G. Carlsson and R. Devulapalli. Dividing a territory among several facilities. *INFORMS Journal on Computing*, 25(4):730–742, 2012.

[3] The Omega Group. Crime Mapping – Building Safer Communities! `http://www.crimemapping.com/map.aspx?aid=3db8cf99-a73b-46d2-b218-bd24cf491577`, 2015.

[4] D. Applegate, R. Bixby, V. Chvatal, and W. Cook. Concorde TSP solver, 2006.

[5] D. Applegate, W. Cook, D. S. Johnson, and N. J. A. Sloane. Using large-scale computation to estimate the Beardwood-Halton-Hammersley TSP constant. Presentation at 42 SBPO, 2010.

[6] Google Developers. The Google Distance Matrix API. `https://developers.google.com/maps/documentation/distancematrix/intro`, 2015.

[7] Paul J Nahin. *When least is best: how mathematicians discovered many clever ways to make things as small (or as large) as possible*. Princeton University Press, 2007.

[8] LaDawn Haws and Terry Kiser. Exploring the brachistochrone problem. *American Mathematical Monthly*, pages 328–336, 1995.

[9] Nimrod Megiddo. Linear programming in linear time when the dimension is fixed. *Journal of the ACM (JACM)*, 31(1):114–127, 1984.

[10] Alfred Weber, Carl Joachim Friedrich, et al. Alfred weber's theory of the location of industries. 1929.

[11] C. H. Papadimitriou. Worst-case and probabilistic analysis of a geometric location problem. *SIAM Journal on Computing*, 10:542, 1981.

[12] Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem. *J. Comput. Syst. Sci.*, 65:129–149, August 2002.

[13] John Gunnar Carlsson, Fan Jia, and Ying Li. An approximation algorithm for the continuous k-medians problem in a convex polygon. *INFORMS Journal on Computing*, 26(2):280–289, 2013.

[14] Raghuveer Devulapalli and John Gunnar Carlsson. A factor 2.02 approximation algorithm for the continuous k-medians problem. *INFORMS Journal on Computing*, Under Review.

[15] Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for euclidean k-medians and related problems. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 106–113. ACM, 1998.

[16] Joseph SB Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric tsp, k-mst, and related problems. *SIAM Journal on Computing*, 28(4):1298–1309, 1999.

[17] Sándor P. Fekete, Joseph S. B. Mitchell, and Karin Beurer. On the continuous Fermat-Weber problem. *Oper. Res.*, 53:61–76, January 2005.

[18] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[19] Sanjoy Dasgupta. The hardness of k-means clustering. Technical Report CS2008-0916, University of California, San Diego, 2008.

[20] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar k-means problem is np-hard. In *WALCOM: Algorithms and Computation*, pages 274–285. Springer, 2009.

[21] Stuart P Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.

[22] Rafail Ostrovsky, Yuval Rabani, Leonard J Schulman, and Chaitanya Swamy. The effectiveness of lloyd-type methods for the k-means problem. *Journal of the ACM (JACM)*, 59(6):28, 2012.

[23] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[24] W Fernandez de la Vega, Marek Karpinski, Claire Kenyon, and Yuval Rabani. Approximation schemes for clustering problems. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 50–58. ACM, 2003.

[25] Amit Kumar, Yogish Sabharwal, and Sandeep Sen. A simple linear time (1+ ε)-approximation algorithm for geometric k-means clustering in any dimensions. In *Proceedings-Annual Symposium on Foundations of Computer Science*, pages 454–462. IEEE, 2004.

[26] S Louis Hakimi. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459, 1964.

[27] D. S. Hochbaum. When are NP-hard location problems easy? *Annals of Operations Research*, 1:201–214, 1984.

[28] Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

[29] D. S. Hochbaum. *Approximation algorithms for NP-hard problems*, volume 20. PWS publishing company Boston, 1997.

[30] Teofilo F Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[31] Dorit S Hochbaum and David B Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM (JACM)*, 33(3):533–550, 1986.

[32] DR Shier. A min-max theorem for p-center problems on a tree. *Transportation Science*, 11(3):243–252, 1977.

[33] E Erkut and S Neuman. Comparison of four models for dispersing facilities. *Infor*, 29(2):68–86, 1991.

[34] Barun Chandra and Magnús M Halldórsson. Approximation algorithms for dispersion problems. *Journal of algorithms*, 38(2):438–465, 2001.

[35] Kevin M Curtin and Richard L Church. A family of location models for multiple-type discrete dispersion. *Geographical Analysis*, 38(3):248–270, 2006.

[36] DW Wang and Y-S Kuo. A study on two geometric location problems. *Information processing letters*, 28(6):281–286, 1988.

[37] Erhan Erkut. The discrete p-dispersion problem. *European Journal of Operational Research*, 46(1):48–60, 1990.

[38] Sekharipuram S Ravi, Daniel J Rosenkrantz, and Giri K Tayi. Heuristic and special case algorithms for dispersion problems. *Operations Research*, 42(2):299–310, 1994.

[39] Refael Hassin, Shlomi Rubinstein, and Arie Tamir. Approximation algorithms for maximum dispersion. *Operations research letters*, 21(3):133–137, 1997.

[40] B. Aronov, P. Carmi, and M.J. Katz. Minimum-cost load-balancing partitions. *Algorithmica*, 54(3):318–336, July 2009.

[41] J.G. Carlsson, B. Armbruster, and Y. Ye. Finding equitable convex partitions of points in a polygon efficiently. *ACM Transactions on Algorithms*, To appear, 2010.

[42] J.G. Carlsson and Y. Ye. On equitably partitioning a convex polygon and a point set. *ACM Transactions on Algorithms*, Submitted for review, 2009.

[43] J. G. Carlsson. Dividing a territory among several vehicles. *INFORMS Journal on Computing*, 24(4):565–577, 2012.

[44] Marco Pavone, Alessandro Arsie, Emilio Frazzoli, and Francesco Bullo. Equitable partitioning policies for robotic networks. In *ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pages 3979–3984, Piscataway, NJ, USA, 2009. IEEE Press.

[45] F. Aurenhammer, F. Hoffmann, and B. Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1):61–76, 1998.

[46] Marco Pavone, Alessandro Arsie, Emilio Frazzoli, and Francesco Bullo. Distributed policies for equitable partitioning: theory and applications. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 4191–4197, Piscataway, NJ, USA, 2008. IEEE Press.

[47] Yossi Azar. On-line load balancing. In *Online Algorithms*, pages 178–195. Springer, 1998.

[48] Hans Kellerer, Vladimir Kotov, Maria Grazia Speranza, and Zsolt Tuza. Semi on-line algorithms for the partition problem. *Operations Research Letters*, 21(5):235–242, 1997.

[49] Opher Baron, Oded Berman, Dmitry Krass, and Qian Wang. The equitable location problem on the plane. *European Journal of Operational Research*, 183(2):578–590, 2007.

[50] O. Berman, Z. Drezner, A. Tamir, and G. O. Wesolowsky. Optimal location with equitable loads. *Annals of Operations Research*, 167(1):307–325, 2009.

[51] J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327, 1959.

[52] Marco Pavone, Nabhendra Bisnik, Emilio Frazzoli, and Volkan Isler. Decentralized vehicle routing in a stochastic and dynamic environment with customer impatience. In *Proceedings of the 1st international conference on Robot communication and coordination*, page 24. IEEE Press, 2007.

[53] J.G. Carlsson, D. Ge, A. Subramaniam, and Y. Ye. Solving the min-max multi-depot vehicle routing problem. In *Proceedings of the FIELDS Workshop on Global Optimization*, 2007.

[54] D. Haugland, S. C. Ho, and G. Laporte. Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, 180(3):997 – 1010, 2007.

[55] Z. Drezner and A. Suzuki. Covering continuous demand in the plane. *Journal of the Operational Research Society*, 61(5):878–881, 2010.

[56] Markus Jäger and Bernhard Nebel. Dynamic decentralized area partitioning for cooperating cleaning robots. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, volume 4, pages 3577–3582. IEEE, 2002.

[57] L. C. A. Pimenta, V. Kumar, R. C. Mesquita, and G. A. S. Pereira. Sensing and coverage for a network of heterogeneous robots. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 3947–3952. IEEE, 2008.

[58] Sam L Savage. *The flaw of averages: Why we underestimate risk in the face of uncertainty*. John Wiley & Sons, 2009.

[59] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[60] Moses Charikar, Samir Khuller, David M Mount, and Giri Narasimhan. Algorithms for facility location problems with outliers. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 642–651. Society for Industrial and Applied Mathematics, 2001.

[61] Lawrence V Snyder. Facility location under uncertainty: a review. *IIE Transactions*, 38(7):547–564, 2006.

[62] Roberto Montemanni, János Barta, Monaldo Mastrolilli, and Luca Maria Gambardella. The robust traveling salesman problem with interval data. *Transportation Science*, 41(3):366–381, 2007.

[63] I. Sungur, F. Ordónez, and M. Dessouky. A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, 40(5):509–523, 2008.

[64] O. Berman, P. Jaillet, and D. Simchi-Levi. Location-routing problems with uncertainty. In Z. Drezner, editor, *Facility Location: A Survey of Applications and Methods*, pages 427–452. Springer, 1995.

[65] J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):727–744, 2013.

[66] Satish V Ukkusuri, Tom V Mathew, and S Travis Waller. Robust transportation network design under demand uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, 22(1):6–18, 2007.

[67] J.G. Carlsson, E. Carlsson, and R. Devulapalli. Shadow prices in territory division. *Networks and Spatial Economics*, 2015.

[68] P Kouvelis and G Yu. Robust discrete optimization and its applications. 1997. *Kluwer Academic Publishers, Boston*.

[69] A Altın, Edoardo Amaldi, Pietro Belotti, and MÇ Pınar. Provisioning virtual private networks under traffic uncertainty. *Networks*, 49(1):100–115, 2007. "[An earlier version presented at 18th Int. Symp. Math. Programming, Copenhagen, 2003]".

[70] Neil Kelvin Olver. *Robust network design*. PhD thesis, Department of Mathematics and Statistics, McGill University, Montréal Québec, Canada, 2010.

[71] Arie MCA Koster, Manuel Kutschka, and Christian Raack. Towards robust network design using integer linear programming techniques. In *Next Generation Internet (NGI), 2010 6th EURO-NF Conference on*, pages 1–8. IEEE, 2010.

[72] Chandra Chekuri, F Bruce Shepherd, Gianpaolo Oriolo, and Maria Grazia Scutellà. Hardness of robust network design. *Networks*, 50(1):50–54, 2007.

[73] Alper Atamtürk and Muhong Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.

[74] Arie MCA Koster, Manuel Kutschka, and Christian Raack. Robust network design: Formulations, valid inequalities, and computations. *Networks*, 61(2):128–149, 2013.

[75] Supakom Mudchanatongsuk, Fernando Ordóñez, and Jie Liu. Robust solutions for network design under transportation cost and demand uncertainty. *Journal of the Operational Research Society*, 59(5):652–662, 2008.

[76] John Gunnar Carlsson, Mehdi Behroozi, and Xiang Li. Geometric partitioning and robust ad-hoc network design. *Annals of Operations Research*, 238(1-2):41–68, 2016.

[77] Atsuo Suzuki and Zvi Drezner. The p-center location problem in an area. *Location science*, 4(1):69–82, 1996.

[78] Sung Kwon Kim and Chan-Su Shin. Efficient algorithms for two-center problems for a convex polygon. In *Computing and Combinatorics*, pages 299–309. Springer, 2000.

[79] Dan HALPERIN, Micha SHARIR, and Ken GOLDBERG. The 2-center problem with obstacles. *Journal of algorithms*, 42(1):109–134, 2002.

[80] Hai Du and Yinfeng Xu. An approximation algorithm for k-center problem on a convex polygon. *Journal of Combinatorial Optimization*, 27(3):504–518, 2014.

[81] Jürgen Eckhoff et al. Helly, radon, and carathéodory type theorems. *Handbook of convex geometry*, pages 389–448, 1993.

[82] J.M. Steele. *Probability Theory and Combinatorial Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[83] J. M. Steele. Subadditive euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, 9(3):pp. 365–376, 1981.

[84] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[85] L. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81, 2007.

[86] Dimitris Bertsimas, Xuan Vinh Doan, Karthik Natarajan, and Chung-Piaw Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[87] John Gunnar Carlsson and Mehdi Behroozi. Worst-case demand distributions in vehicle routing. *European Journal of Operational Research*, 2016.

[88] Leonid Vaseršteĭn. Markov processes with countable state space describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–73, 1969. "(In Russian)".

[89] Roland L Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[90] John Gunnar Carlsson and Mehdi Behroozi. Wasserstein distance and the distributionally robust tsp. 2015.

[91] J. Brimberg and S. Salhi. A continuous location-allocation problem with zone-dependent fixed cost. *Annals of Operations Research*, 136(1):99–115, 2005.

[92] K. M. Alzoubi, P.-J. Wan, and O. Frieder. Message-optimal connected dominating sets in mobile ad hoc networks. In *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*, pages 157–164. ACM, 2002.

[93] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser. An analysis of the optimum node density for ad hoc mobile networks. In *Communications, 2001. ICC 2001. IEEE International Conference on*, volume 3, pages 857–861. IEEE, 2001.

[94] I. Stojmenovic. Position-based routing in ad hoc networks. *Communications Magazine, IEEE*, 40(7):128–134, 2002.

[95] R. Wattenhofer, L. Li, P. Bahl, and Y.-M. Wang. Distributed topology control for power efficient operation in multihop wireless ad hoc networks. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1388–1397. IEEE, 2001.

[96] J. Wu and H. Li. On calculating connected dominating set for efficient routing in ad hoc wireless networks. In *Proceedings of the 3rd international workshop on Discrete algorithms and methods for mobile computing and communications*, pages 7–14. ACM, 1999.

[97] G. Valiente. *Algorithms on Trees and Graphs*. Springer Berlin Heidelberg, 2013.

[98] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1: Theory. *Medical Imaging, IEEE Transactions on*, 1(2):81–94, October 1982.

[99] Franco P. Preparata and Michael I. Shamos. *Computational geometry: an introduction*. Springer-Verlag New York, Inc., New York, NY, USA, 1985.

[100] G. B. Dantzig and J. H Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

[101] C. S. Orloff. Route constrained fleet scheduling. *Transportation Science*, 10(2):149–168, 1976.

[102] D. J. Bertsimas and D. Simchi-Levi. A new generation of vehicle routing research: robust algorithms, addressing uncertainty. *Operations Research*, 44(2):286–304, 1996.

[103] C. F. Daganzo. Modeling distribution problems with time windows: Part i. *Transportation Science*, 21(3):171–179, 1987.

[104] L. D. Burns, R. W. Hall, D. E. Blumenfeld, and C. F. Daganzo. Distribution strategies that minimize transportation and inventory costs. *Operations Research*, 33(3):469–490, 1985.

[105] M. Huang, K. R. Smilowitz, and B. Balcik. A continuous approximation approach for assessment routing in disaster relief. *Transportation Research Part B: Methodological*, 50:20–41, 2013.

[106] O. Jabali, M. Gendreau, and G. Laporte. A continuous approximation model for the fleet composition problem. *Transportation Research Part B: Methodological*, 46(10):1591–1606, 2012.

[107] J. F. Campbell. Location and allocation for distribution systems with transshipments and transportion economies of scale. *Annals of Operations Research*, 40(1):77–99, 1992.

[108] J. Geunes, Z.-J. M. Shen, and A. Emir. Planning and approximation models for delivery route based services with price-sensitive demands. *European journal of operational research*, 183(1):460–471, 2007.

[109] A. G. N. Novaes, J. E. S. de Cursi, and O. D. Graciolli. A continuous approach to the design of physical distribution systems. *Computers & Operations Research*, 27(9):877–893, 2000.

[110] Y. Ouyang. Design of vehicle routing zones for large-scale distribution systems. *Transportation Research Part B: Methodological*, 41(10):1079–1093, 2007.

[111] L. Few. The shortest path and the shortest road through n points. *Mathematika*, 2:141–144, 1955.

[112] M. Haimovich and T. L. Magnanti. Extremum properties of hexagonal partitioning and the uniform distribution in Euclidean location. *SIAM J. Discrete Math.*, 1:50–64, 1988.

[113] C. Redmond and J. E. Yukich. Limit theorems and rates of convergence for euclidean functionals. *The Annals of Applied Probability*, 4(4):pp. 1057–1073, 1994.

[114] C. F. Daganzo. Modeling distribution problems with time windows. part ii: Two customer types. *Transportation Science*, 21(3):180–187, 1987.

[115] M. A. Figliozzi. Planning approximations to the average length of vehicle routing problems with time window constraints. *Transportation Research Part B: Methodological*, 43(4):438–447, 2009.

[116] D. J. Bertsimas and G. van Ryzin. Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles. *Operations Research*, 41(1):60–76, 1993.

[117] C. F. Daganzo. The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. *Transportation Science*, 18(4):331–350, 1984.

[118] M. Haimovich and A. H. G. Rinnooy Kan. Bounds and heuristics for capacitated routing problems. *Mathematics of Operations Research*, 10(4):527–542, 1985.

[119] A. Agra, M. Christiansen, R. Figueiredo, L. M. Hvattum, M. Poss, and C. Requejo. The robust vehicle routing problem with time windows. *Computers & Operations Research*, 40(3):856–866, 2013.

[120] M. Barkaoui and M. Gendreau. An adaptive evolutionary approach for real-time vehicle routing and dispatching. *Computers & Operations Research*, 2013.

[121] C. E. Gounaris, W. Wiesemann, and C. A. Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 2013.

[122] C. Lee, K. Lee, and S. Park. Robust vehicle routing problem with deadlines and travel time/demand uncertainty. *Journal of the Operational Research Society*, 63(9):1294–1306, 2011.

[123] O. Solyalı, J.-F. Cordeau, and G. Laporte. Robust inventory routing under demand uncertainty. *Transportation Science*, 46(3):327–340, 2012.

[124] M. Allahviranloo, J. Y. J. Chow, and W. W. Recker. Selective vehicle routing problems under uncertainty without recourse. *Transportation Research Part E: Logistics and Transportation Review*, 62:68–88, 2014.

[125] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[126] Caviar. `http://www.trycaviar.com/`. Accessed: 2014-09-27.

[127] DoorDash Food Delivery. `http://www.doordash.com`. Accessed: 2014-10-27.

[128] Food Delivery & Restaurants Delivery - Order Food Online - BiteSquad.com. `http://www.bitesquad.com`. Accessed: 2014-10-27.

[129] Good Eggs. `http://www.goodeggs.com`. Accessed: 2014-10-27.

[130] GrubHub Inc. Form S-1: Registration Statement under the Securities Act of 1933, February 2014. `https://www.sec.gov/Archives/edgar/data/1594109/000119312514075544/d647121ds1.htm`. [Online; posted 28-Feb-2014].

[131] Kalev Leetaru. The $200 Uber ride and the realtime data-driven sharing economy, January 2016. `http://www.forbes.com/sites/kalevleetaru/2016/01/02/the-200-uber-ride-and-the-realtime-data-driven-sharing-economy/`. [Online; posted 2-Jan-2016].

[132] C. Daganzo. *Logistics Systems Analysis*. Springer, 2005.

[133] D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2011.

[134] L. A. Goddyn. Quantizers and the worst-case euclidean traveling salesman problem. *Journal of Combinatorial Theory, Series B*, 50(1):65–81, 1990.

[135] H. J. Karloff. How long can a euclidean traveling salesman tour be? *SIAM Journal on Discrete Mathematics*, 2(1):91–99, 1989.

[136] T. L. Snyder and J. M. Steele. Equidistribution in all dimensions of worst-case point sets for the traveling salesman problem. *SIAM Journal on Discrete Mathematics*, 8(4):678–683, 1995.

[137] Herbert Robbins. A remark on stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.

[138] William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Ed.* Wiley, 1968.

[139] D. G. Luenberger. *Optimization by vector space methods.* John Wiley & Sons, 1968.

[140] J. E. Anderson. The gravity model. Technical report, National Bureau of Economic Research, 2010.

[141] J.P. Rodrigue, C. Comtois, and B. Slack. *The Geography of Transport Systems.* Routledge, 2009.

[142] J. Tinbergen. Shaping the world economy; suggestions for an international economic policy. *Books (Jan Tinbergen)*, 1962.

[143] J. Q. Stewart. Demographic gravitation: evidence and applications. *Sociometry*, pages 31–58, 1948.

[144] W. J. Reilly. *The law of retail gravitation.* WJ Reilly, 1931.

[145] A. M. Voorhees. A general theory of traffic movement. In *Proceedings of the Institute of Traffic Engineers*, pages 46–50, 1955.

[146] United States Census Bureau. American FactFinder. `http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml`. Accessed: 2014-05-20.

[147] F. Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96, 1987.

[148] G. F. Newell and C. F. Daganzo. Design of multiple-vehicle delivery tours - I a ring-radial network. *Transportation Research Part B: Methodological*, 20(5):345–363, 1986.

[149] L. C. Galvão, A. G. N. Novaes, J. E. Souza De Cursi, and J. C. Souza. A multiplicatively-weighted voronoi diagram approach to logistics districting. *Computers & Operations Research*, 33(1):93–114, 2006.

[150] J. Cortés. Coverage optimization and spatial load balancing by robotic sensor networks. *Automatic Control, IEEE Transactions on*, 55(3):749–754, 2010.

[151] J. W. Durham, R. Carli, P. Frasca, and F. Bullo. Discrete partitioning and coverage control for gossiping robots. *Robotics, IEEE Transactions on*, 28(2):364–378, 2012.

[152] M. Pavone, A. Arsie, E. Frazzoli, and F. Bullo. Distributed algorithms for environment partitioning in mobile robotic networks. *Automatic Control, IEEE Transactions on*, 56(8):1834–1848, 2011.

[153] J. F. Bard and A. I. Jarrah. Large-scale constrained clustering for rationalizing pickup and delivery operations. *Transportation Research Part B: Methodological*, 43(5):542–561, 2009.

[154] N. Geroliminis, M. G. Karlaftis, and A. Skabardonis. A spatial queuing model for the emergency vehicle districting and location problem. *Transportation research part B: methodological*, 43(7):798–811, 2009.

[155] H. Lei, G. Laporte, and B. Guo. Districting for routing with stochastic customers. *EURO Journal on Transportation and Logistics*, 1(1-2):67–85, 2012.

[156] H. Lei, G. Laporte, Y. Liu, and T. Zhang. Dynamic design of sales territories. *Computers & Operations Research*, 56:84–92, 2015.

[157] M. Flint, M. Polycarpou, and E. Fernandez-Gaucherand. Cooperative control for multiple autonomous uav's searching for targets. In *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, volume 3, pages 2823–2828. IEEE, 2002.

[158] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[159] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

[160] X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

[161] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[162] C. Caillerie, F. Chazal, J. Dedecker, and B. Michel. Deconvolution for the Wasserstein metric and geometric inference. In *Geometric Science of Information*, pages 561–568. Springer, 2013.

[163] A. Irpino and R. Verde. A new wasserstein based distance for the hierarchical clustering of histogram symbolic data. In *Data science and classification*, pages 185–192. Springer, 2006.

[164] E. Pampalk, A. Flexer, G. Widmer, et al. Improvements of audio-based music similarity and genre classificaton. In *ISMIR*, volume 5, pages 634–637. London, UK, 2005.

[165] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[166] E Erdoğan and G Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61, 2006.

[167] D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47, 2012.

[168] D. Wozabal. Robustifying convex risk measures for linear portfolios: a nonparametric approach. *Operations Research*, 62(6):1302–1315, 2014.

[169] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116*, 2015.

[170] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[171] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *arXiv preprint arXiv:1401.0212*, 2013.

[172] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

[173] G. N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[174] D. Klabjan, D. Simchi-Levi, and M. Song. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management*, 22(3):691–710, 2013.

[175] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[176] S. Boyd. Subgradients. `http://stanford.edu/class/ee364b/lectures/subgradients_slides.pdf`, 2014.

[177] M. J. Todd. Note-solving the generalized market area problem. *Management Science*, 24(14):1549–1554, 1978.

[178] G. Canas and L. Rosasco. Learning probability measures with respect to optimal transport metrics. In *Advances in Neural Information Processing Systems*, pages 2492–2500, 2012.

[179] H. L. Royden and P. Fitzpatrick. *Real analysis*, volume 32. Macmillan New York, 1988.

[180] E. H. Lockwood. *A book of curves*. Cambridge University Press, 1967.

[181] S. Boyd. Localization and cutting-plane methods. `http://stanford.edu/class/ee364b/lectures/localization_methods_slides.pdf`, 2014.

[182] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[183] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[184] F. Bolley and C. Villani. Weighted csiszár-kullback-pinsker inequalities and applications to transportation inequalities. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 14, pages 331–352, 2005.

[185] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.

[186] A. Caiti, V. Calabro, F. Di Corato, D. Meucci, and A. Munafo. Cooperative distributed algorithm for AUV teams: a minimum entropy approach. In *OCEANS-Bergen, 2013 MTS/IEEE*, pages 1–6. IEEE, 2013.

[187] J. Le Ny and G. J. Pappas. Adaptive deployment of mobile robotic networks. *Automatic Control, IEEE Transactions on*, 58(3):654–666, 2013.

[188] J. Kytöjoki, T. Nuortio, O. Bräysy, and M. Gendreau. An efficient variable neighborhood search heuristic for very large scale vehicle routing problems. *Computers & Operations Research*, 34(9):2743–2757, 2007.

[189] A. Wade and S. Salhi. An ant system algorithm for the mixed vehicle routing problem with backhauls. In *Metaheuristics: computer decision-making*, pages 699–719. Springer, 2004.

[190] A. G. N. Novaes and O. D. Graciolli. Designing multi-vehicle delivery tours in a grid-cell format. *European Journal of Operational Research*, 119(3):613–634, 1999.

[191] D. Avis and G. T. Toussaint. An efficient algorithm for decomposing a polygon into star-shaped polygons. *Pattern Recognition*, 13(6):395–398, 1981.

[192] S. Bereg, P. Bose, and D. Kirkpatrick. Equitable subdivisions within polygonal regions. *Computational Geometry*, 34(1):20–27, 2006.

[193] John Gunnar Carlsson, Mehdi Behroozi, Raghuveer Devulapalli, and Xiangfei Meng. Household-level economies of scale in transportation. *Operations Research*, 0(0):null, 0, http://dx.doi.org/10.1287/opre.2016.1533.

[194] S.G. Krantz and H.R. Parks. *Geometric Integration Theory*. Cornerstones. Birkhäuser, 2008.

[195] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

Appendix A

Additional Proofs

A.1 Proof of Theorem 12

It remains to consider the special case where $q_0 = 2$; we will decompose this into two further sub-cases in which we will either divide Q into $\lfloor (n-1)/2 \rfloor \times 2$ sub-rectangles or $\lfloor (n-1)/3 \rfloor \times 3$ sub-rectangles. Recall from (2.3) that we must have

$$\frac{\sqrt{2(n-1)}}{3} < w \leq \frac{\sqrt{2(n-1)}}{2},$$

so that we must have either $\sqrt{2(n-1)}/3 < w \leq \sqrt{2(n-1)}/2.6$ or $\sqrt{2(n-1)}/2.6 < w \leq \sqrt{2(n-1)}/3$:

- If $\sqrt{2(n-1)}/3 < w \leq \sqrt{2(n-1)}/2.6$, then we will decompose Q into $\lfloor (n-1)/3 \rfloor \times 3$ sub-rectangles. By computing vertical and horizontal differences between rectangle centers as before, the approximation ratio is at most

$$\frac{\max\left\{\frac{2}{3w}, \frac{3w}{(n-1)-3}\right\}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}}.$$

  The first term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as small as possible, which occurs when $w = \sqrt{2(n-1)}/3$; the above ratio is then equal to

$$\begin{aligned}\frac{\frac{2}{3w}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{2}{\sqrt{2(n-1)}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\ &= \frac{\sqrt{(9\sqrt{3}+6\pi)(n-1)+18\pi}}{3\sqrt{n-1}} < 2.77 \text{ for all } n \geq 21\end{aligned}$$

and the second term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as large as possible, which occurs when $w = \sqrt{2(n-1)}/2.6$; the above ratio is then equal to

$$\begin{aligned}\frac{\frac{3w}{(n-1)-3}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{3(\sqrt{2(n-1)}/2.6)}{(n-1)-3}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\ &< \frac{\sqrt{5.10(n-1)^2+8.37(n-1)}}{n-4} < 2.77 \text{ for all } n \geq 21\end{aligned}$$

as desired.

- If $\sqrt{2(n-1)}/2.6 < w \leq \sqrt{2(n-1)}/2$, then we will decompose Q into $\lfloor (n-1)/2 \rfloor \times 2$ sub-rectangles. By computing vertical and horizontal differences between rectangle centers as before, the approximation ratio is at most

$$\frac{\max\left\{\frac{1}{w}\,,\,\frac{2w}{(n-1)-2}\right\}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}}\,.$$

The first term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as small as possible, which occurs when $w = \sqrt{2(n-1)}/2.6$; the above ratio is then equal to

$$\begin{aligned}\frac{\frac{1}{w}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{1}{\sqrt{2(n-1)}/2.6}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\ &< \frac{\sqrt{6.47(n-1)+10.62}}{\sqrt{n-1}} < 2.77 \text{ for all } n \geq 21\end{aligned}$$

and the second term of the $\max\{\cdot,\cdot\}$ expression is largest when w is as large as possible, which occurs when $w = \sqrt{2(n-1)}/2$; the above ratio is then equal to

$$\begin{aligned}\frac{\frac{2w}{(n-1)-2}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} &= \frac{\frac{\sqrt{2(n-1)}}{(n-1)-2}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} \\ &= \frac{\sqrt{(9\sqrt{3}+6\pi)(n-1)^2+18\pi(n-1)}}{3(n-3)} < 2.77 \text{ for all } n \geq 21\,,\end{aligned}$$

which completes the proof.

A.2 Proof of Theorem 16

If $q_0 = 1$, then we have $1 \leq \sqrt{2k}/w < 2$. We have already seen that our ratio of 1.99 is valid whenever $w \geq \sqrt{k}$, and thus we may restrict ourselves to the domain $\sqrt{2} < \sqrt{2k}/w < 2$, or equivalently $\sqrt{2} < t < 2$ with $t = \sqrt{2k}/w$ as before. The parity of k is now relevant; if k is even, then we can divide Q into a $k/2 \times 2$ grid, wherein each sub-rectangle has dimensions $\frac{w}{k/2} \times h/2$. The distance from any point $x \in C \subseteq Q$ is at most half of the diagonal of such a rectangle, which is

$$\frac{1}{2}\sqrt{\left(\frac{w}{k/2}\right)^2 + \left(\frac{h}{2}\right)^2} = \sqrt{\frac{1}{k}} \cdot \sqrt{\left(\frac{w^2}{k} + \frac{k}{4w^2}\right)} = \sqrt{\frac{1}{k}} \cdot \sqrt{t^2/8 + 2/t^2} < \frac{1}{2}\sqrt{\frac{5}{k}}$$

since $\sqrt{2} < t < 2$. Our approximation ratio is met because

$$\mathsf{Rat} \leq \frac{\frac{1}{2}\sqrt{5/k}}{1/\sqrt{\pi k}} < 1.99\,.$$

It remains to consider the case where k is odd, which will complete the proof.

Lemma 44. *Let R be a rectangle with dimensions $a \times b$, where $a \geq b$. If 5 points $x_1, \ldots, x_5$ are placed inside R according to Figure 3.1, then the distance between any point $x \in R$ and its nearest neighbor x_i is at most*

$$\min_i \|x - x_i\| \leq \frac{a}{\pi^2} + \frac{b}{2\varphi}\,,$$

where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio.

Proof. The configurations in Figure 3.1 are due to [1], which also gives a precise closed form expression for the maximum possible nearest-neighbor distance over all $x \in R$, i.e. $\max_{x \in R} \min_i \|x - x_i\|$. The above inequality is merely a crude upper bound thereof. □

We will next divide our rectangle Q with a vertical line into two rectangles Q_1 and Q_2 that have dimensions $\ell \times h$ and $(w - \ell) \times h$ respectively, where we let $\ell = 11.08/\sqrt{k} - 6.10/w$, and we will place 5 points in Q_1 and $k - 5$ points in Q_2. By Lemma 44, the maximum distance between a point $x \in Q_1$ and its nearest neighbor $x_i \in Q_1$ (which is placed according to Figure 3.1) is

$$\frac{\ell}{\pi^2} + \frac{h}{2\varphi} \approx \frac{1.122638714}{\sqrt{k}} - \frac{0.00003}{w} < 1.1227/\sqrt{k}$$

and thus our desired approximation ratio is met because

$$\frac{1.1227/\sqrt{k}}{1/\sqrt{\pi k}} < 1.99\,.$$

We can verify that the approximation ratio is met for Q_2 as well, which will complete our entire proof. If we divide Q_2 into a $(k-5)/2 \times 2$ grid, then each sub-rectangle will have dimensions $\frac{w-\ell}{(k-5)/2} \times \frac{h}{2}$, and thus using half of the diagonal of such a rectangle is

$$\begin{aligned}
\frac{1}{2}\sqrt{\left[\frac{w-\ell}{(k-5)/2}\right]^2 + \left(\frac{h}{2}\right)^2} &= \frac{\sqrt{\ell^2 - 2w\ell + w^2 + \frac{1}{w^2}(k^2/4 - 5k/2 + 25/4)}}{k-5} \\
&= \frac{\sqrt{\frac{122.7664}{k} - \frac{135.176}{w\sqrt{k}} - \frac{22.16w}{\sqrt{k}} + w^2 + \frac{1}{w^2}(k^2/4 - 5k/2 + 43.46) + 12.2}}{k-5} \\
&= \frac{\sqrt{\frac{122.7664}{k} - \frac{67.588\sqrt{2}t}{k} - \frac{22.16\sqrt{2}}{t} + \frac{2k}{t^2} + \frac{kt^2}{8} - \frac{5t^2}{4} + \frac{21.73t^2}{k} + 12.2}}{k-5}
\end{aligned}$$

where we have again substituted $t = \sqrt{2k}w$. The inner term of the square root is convex in t for $k \geq 31$ (by routine algebra) and thus, for fixed k, the above quantity is maximized at $t = \sqrt{2}$ or $t = 2$, and the above expression simply reduces to

$$\frac{1}{k-5}\max\left\{\sqrt{\frac{3.10504}{k} + 1.25k - 12.46}\,,\ \sqrt{\frac{209.6864 - 135.176\sqrt{2}}{k} + k - 11.08\sqrt{2} + 7.2}\right\}.$$

We therefore merely need to prove that the approximation ratio holds for all $k \geq 31$, i.e. that the ratio of the above expression to the lower bound $1/\sqrt{\pi k}$, which is

$$\begin{aligned}
&\frac{\frac{1}{k-5}\max\left\{\sqrt{\frac{3.10504}{k} + 1.25k - 12.46}\,,\ \sqrt{\frac{209.6864 - 135.176\sqrt{2}}{k} + k - 11.08\sqrt{2} + 7.2}\right\}}{1/\sqrt{\pi k}} \\
&= \frac{1}{k-5}\max\left\{\sqrt{\pi(3.10504 + 1.25k^2 - 12.46k)}\,,\ \sqrt{\pi(209.6864 - 135.176\sqrt{2} + k^2 - (11.08\sqrt{2} - 7.2)k}\right\} \\
&< \frac{1}{k-5}\max\left\{\sqrt{3.93k^2 - 39.14k + 9.76}\,,\ \sqrt{3.15k^2 - 26.60k + 58.19}\right\}
\end{aligned}$$

is bounded above by 1.99 for all $k \geq 31$. This is a simple univariate function in k and it is routine to verify that the desired result holds, which completes the proof.

A.3 Proof of Theorem 13

Suppose that $\sqrt{2n}/4 \leq w < \sqrt{n/3}$ and we divide Q into a grid of dimensions $p \times q = \lfloor n/3 \rfloor \times 3$. Each grid cell has dimensions $w/p \times h/q = \frac{w}{\lfloor n/3 \rfloor} \times \frac{h}{3}$ and by Lemma 14 the connectivity radius r is at most

$$\begin{aligned} r &\leq \max\left\{\sqrt{\left(\frac{2w}{\lfloor n/3 \rfloor}\right)^2 + \left(\frac{h}{3}\right)^2}, \sqrt{\left(\frac{w}{\lfloor n/3 \rfloor}\right)^2 + \left(\frac{2h}{3}\right)^2}\right\} \\ &\leq \max\left\{\sqrt{\left[\frac{2w}{(n-2)/3}\right]^2 + \left(\frac{h}{3}\right)^2}, \sqrt{\left[\frac{w}{(n-2)/3}\right]^2 + \left(\frac{2h}{3}\right)^2}\right\} \\ &= \sqrt{\max\left\{\left(\frac{6w}{n-2}\right)^2 + \left(\frac{2}{3w}\right)^2, \left(\frac{3w}{n-2}\right)^2 + \left(\frac{4}{3w}\right)^2\right\}} \end{aligned}$$

For fixed values of n, the inner term of the square root is the maximum of two convex functions in w and is therefore convex in w, and is therefore maximized at when w is as large or as small as possible, i.e. at $w = \sqrt{2n}/4$ or at $w = \sqrt{n/3}$. At $w = \sqrt{2n}/4$ we have

$$\sqrt{\max\left\{\left(\frac{6w}{n-2}\right)^2 + \left(\frac{2}{3w}\right)^2, \left(\frac{3w}{n-2}\right)^2 + \left(\frac{4}{3w}\right)^2\right\}}\Bigg|_{w=\sqrt{2n}/4} = \sqrt{\frac{9n}{8(n-2)^2} + \frac{128}{9n}}$$

which gives an approximation ratio bounded by the univariate function

$$\mathsf{Rat} \leq \frac{\sqrt{\frac{9n}{8(n-2)^2} + \frac{128}{9n}}}{\sqrt{\frac{1}{\pi + (n-1)(\pi/3 + \sqrt{3}/2)}}} < 5.94 \text{ for all } n \geq 33$$

and at $w = \sqrt{n/3}$ we have

$$\sqrt{\max\left\{\left(\frac{6w}{n-2}\right)^2 + \left(\frac{2}{3w}\right)^2, \left(\frac{3w}{n-2}\right)^2 + \left(\frac{4}{3w}\right)^2\right\}}\Bigg|_{w=\sqrt{n/3}} = \sqrt{\frac{12n}{(n-2)^2} + \frac{4}{3n}}$$

whence

$$\mathsf{Rat} \leq \frac{\sqrt{\frac{12n}{(n-2)^2} + \frac{4}{3n}}}{\sqrt{\frac{1}{\pi + (n-1)(\pi/3 + \sqrt{3}/2)}}} < 5.94 \text{ for all } n \geq 33\,.$$

The area of each sub-region is

$$\frac{2}{3\lfloor n/3 \rfloor} \leq \frac{2}{3(n/3 - 2)} = \frac{2}{n-6} \leq \frac{22}{9n}$$

since $n \geq 33$. Finally, if $\sqrt{n/3} \leq w < \frac{3}{5}\sqrt{n}$, then (as we have already seen in the very beginning of this proof) the connectivity radius of a grid of dimensions $n \times 1$ is at most $\sqrt{(2w/n)^2 + h^2} = \sqrt{(2w/n)^2 + (2/w)^2}$. For fixed n, this is once again maximized for extreme values of w. At $w = \sqrt{n/3}$ we have

$$\mathsf{Rat} \leq \frac{\sqrt{(2w/n)^2 + (2/w)^2}\Big|_{w=\sqrt{n/3}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} = \frac{\sqrt{\frac{40}{3n}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} < 5.94 \text{ for all } n \geq 33$$

and at $w = \frac{3}{5}\sqrt{n}$ we have

$$\mathsf{Rat} \leq \frac{\sqrt{(2w/n)^2 + (2/w)^2}\Big|_{w=\frac{3}{5}\sqrt{n}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} = \frac{\sqrt{\frac{2824}{225n}}}{\sqrt{\frac{1}{\pi+(n-1)(\pi/3+\sqrt{3}/2)}}} < 5.94 \text{ for all } n \geq 33$$

as desired, which completes the proof (clearly, each sub-region has area equal to $2/n$).

A.4 Proof of Lemma 21

By the union bound, we see that

$$\Pr(\text{length}(\mathsf{TSP}(X_1, \ldots, X_n)) \leq \ell) \leq n! \underbrace{\Pr(\|X_1 - X_2\| + \cdots + \|X_{n-1} - X_n\| \leq \ell)}_{=:p}$$

since there are $n!$ permutations of the point set. We can think of the n-tuple $(X_1, \ldots, X_n)$ as a vector belonging to $[0,1]^{2n}$, in which case p is simply equal to the volume of the domain $[0,1]^{2n} \cap \mathcal{V}$, where $\mathcal{V}$ is defined by

$$\mathcal{V} = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^{2n} : x_1 \in [0,1]^2 \text{ and } \sum_{i=2}^{n} \|x_i - x_{i-1}\| \leq \ell \right\}.$$

Consider the mapping $\Phi : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ defined by

$$\Phi(x_1, \ldots, x_n) = \begin{pmatrix} x_1 \\ x_2 - x_1 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix}$$

for which the image of $\mathcal{V}$ is

$$\Phi(\mathcal{V}) = \left\{ (u_1, \ldots, u_n) \in \mathbb{R}^{2n} : u_1 \in [0,1]^2 \text{ and } \sum_{i=2}^{n} \|u_i\| \leq \ell \right\}.$$

By building its Jacobian matrix, it is straightforward to verify that $\Phi(\cdot)$ is volume-preserving, whence

$$\begin{aligned} p = \mathrm{vol}([0,1]^{2n} \cap \mathcal{V}) &\leq \mathrm{vol}(\mathcal{V}) \\ &= \mathrm{vol}(\Phi(\mathcal{V}))\,. \end{aligned}$$

By the *coarea formula* [194], using the map $\xi_i \mapsto \|u_i\|$, we see that

$$\mathrm{vol}(\Phi(\mathcal{V})) = (2\pi)^{n-1} \iiint_{\Delta} \xi_2 \xi_3 \cdots \xi_n \, d\boldsymbol{\xi} = \frac{(2\pi)^{n-1}\ell^{2(n-1)}}{(2n-2)!}\,,$$

where $\Delta \subset \mathbb{R}^{n-1}$ denotes the simplex where $\xi_i \geq 0$ for all i and $\sum_{i=2}^{n-1} \xi_i \leq \ell$. As $n \to \infty$, we have

$$\begin{aligned} \Pr(\mathrm{length}(\mathsf{TSP}(X_1,\ldots,X_n)) \leq \ell) &\leq n!\Pr(\|X_1 - X_2\| + \cdots + \|X_{n-1} - X_n\| \leq \ell) \\ &\leq n!\,\mathrm{vol}(\Phi(\mathcal{V})) = \frac{n!}{(2n-2)!} \cdot (2\pi\ell^2)^{n-1} \end{aligned}$$

as desired.

A.5 Proof of Theorem 24

Assume that $\mathcal{R}$ is polygonal[1] with unit area and let $r_0 = \max_{x\in\mathcal{R}} ||x||$ denote the longest possible distance from the depot at the origin to a point in $\mathcal{R}$. It is then clear that the denominator of the left-hand side of (4.7) is simply $\max\{2r_0/t, \beta\}$ (which we could obtain by letting $f(\cdot)$ be equal either to an atomic distribution located a distance r_0 away from the origin or a uniform distribution on $\mathcal{R}$). We will first prove that the "worst" possible shape that $\mathcal{R}$ could take is a circular sector having radius r_0 centered at the origin:

Lemma 45. *Let $r_0 > 0$ be a fixed radius. Among all convex planar shapes $\mathcal{R}$ of unit area that contain the origin and have $\max_{x\in\mathcal{R}} ||x|| = r_0$, the ratio in (4.7) is maximized when $\mathcal{R}$ is a circular sector of radius r_0 centered at the origin that subtends an angle of $\theta = 2/r_0^2$.*

Proof. We will prove the above result by constructing a mapping from an arbitrary region $\mathcal{R}$ that increases the numerator of (4.7) while holding the denominator constant. Let $\mathcal{T} = \{T_1,\ldots,T_m\}$ be a triangulation of $\mathcal{R}$ based at the origin, as shown in Figure A.1a. In order to show that a circular sector is the worst possible shape that $\mathcal{R}$ can take, it will suffice to

[1]This assumption is made without loss of generality since any convex region can be approximated to arbitrary precision by a polygonal region.

give a mapping from any of the triangles T_i to a circular sector S_i of radius r_0 with area equal to that of T_i that increases the numerator of (4.7), since we could then piece all of those sectors S_i back together, as shown in Figure A.1b. In other words, we can assume without loss of generality that $\mathcal{R}$ is *itself* a single triangle, one of whose vertices is the origin (and whose longest side therefore has length equal to r_0). We can also further assume that $\mathcal{R}$ is isosceles because we can shear the point that is opposite its longest side until it is also at a distance r_0 from the origin, as shown in Figure A.1c; this transformation increases $\|x\|$ for every $x \in \mathcal{R}$ and therefore increases the numerator of (4.7) without affecting the denominator. Thus, we now assume that $\mathcal{R}$ is an isosceles triangle with unit area having vertices $(0,0)$, (a,b), and $(a,-b)$. Consider the mapping $M(x_1,x_2) : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$M(x_1,x_2) = x_1\sqrt{\frac{b^2}{a^2}+1}\begin{pmatrix} \cos\frac{ax_2}{(a^2b+b^3)x_1} \\ \sin\frac{ax_2}{(a^2b+b^3)x_1} \end{pmatrix};$$

the image of $\mathcal{R}$ under $M(\cdot,\cdot)$ is precisely a circular sector of radius $r_0 = \sqrt{a^2+b^2}$ with unit area; see Figure A.1d. It is easy to see that $\|M(x_1,x_2)\| \geq \|(x_1,x_2)\|$ for all $(x_1,x_2) \in \mathcal{R}$ because we must have $-\frac{b}{a}x_1 \leq x_2 \leq \frac{b}{a}x_1$, whence

$$\|(x_1,x_2)\| = \sqrt{x_1^2+x_2^2} \leq \sqrt{x_1^2+\frac{b^2}{a^2}x_1^2} = x_1\sqrt{1+b^2/a^2} = \|M(x_1,x_2)\|$$

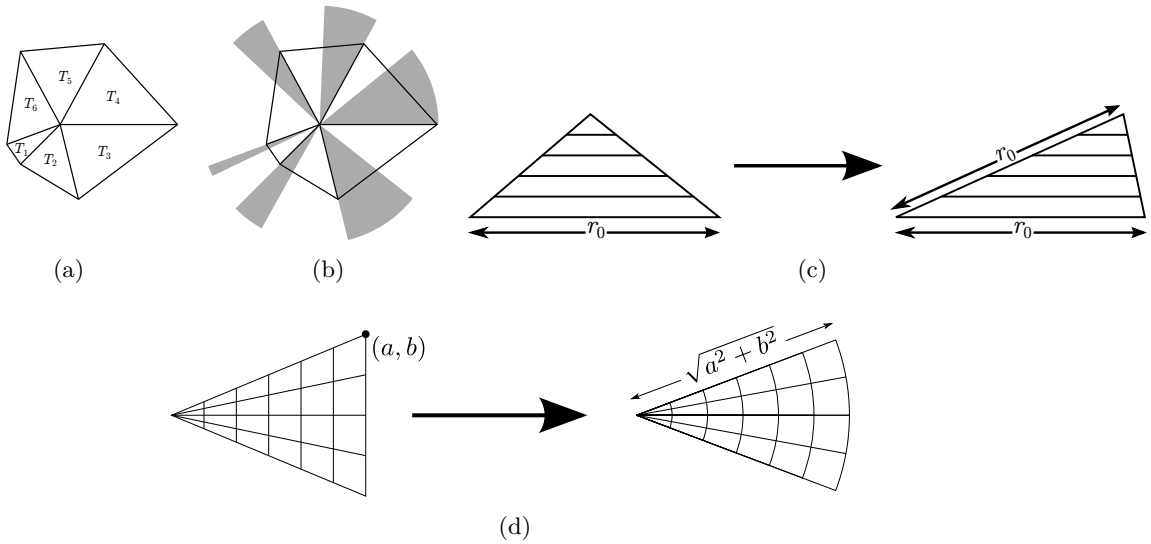
which can be verified algebraically. To complete the proof of this lemma, we must show that $M(\cdot,\cdot)$ is area-preserving, which we merely accomplish by verifying algebraically that its Jacobian matrix has a determinant equal to 1.

□

Using Lemma 45 and the duality result of equation (4.6), we can then assume that $\mathcal{R}$ is a circular sector and therefore evaluate the left-hand side of (4.7) in polar coordinates as

$$\begin{aligned}\frac{\max_{f(\cdot)} \frac{2}{t}\iint_{\mathcal{R}} \|x\| f(x)\,dA + \beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA}{\max_{f(\cdot)}\max\left\{\frac{2}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA, \beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right\}} &= \frac{\min_{\nu\geq 2r_0/t}\iint_{\mathcal{R}}\frac{1}{4}\cdot\frac{\beta^2}{\nu-\frac{2}{t}\|x\|}\,dA+\nu}{\max\{2r_0/t,\beta\}} \\ &= \frac{\min_{\nu\geq 2r_0/t}\int_0^{\theta_0}\int_0^{r_0}\frac{1}{4}\cdot\frac{\beta^2}{\nu-\frac{2}{t}r}\,r\,dr\,d\theta+\nu}{\max\{2r_0/t,\beta\}} \\ &= \frac{\min_{\nu\geq 2r_0/t}\frac{\beta^2\theta_0}{16}\left[\nu t\log\left(\frac{\nu t}{\nu t-2r_0}\right)-2tr_0\right]+\nu}{\max\{2r_0/t,\beta\}},\end{aligned}$$

where we set $\theta_0 = 2/r_0^2$ to ensure that $\mathcal{R}$ has area 1. We can now assume that $2r_0/t \geq \beta$,



(a) (b) (c) (d)

Figure A.1: Figure A.1a shows a triangulation of a convex polygon; the origin is common to all triangles. Figure A.1b shows a collection of circular sectors, each having radius $r_0 = \max_{x \in \mathcal{R}} \|x\|$ and area equal to the triangle that they intersect. Figure A.1c shows a simple shearing operation that makes a triangle isosceles, and Figure A.1d shows the action of the map $M(\cdot, \cdot)$.

i.e. that $t \leq 2r_0/\beta$, because otherwise the denominator is a constant and the numerator is strictly decreasing in t by definition (since it represents the worst-case cost with capacity t); thus, the maximum value of the above ratio must be attained for $t \in (0, 2r_0/\beta]$. Under this assumption, we then have

$$\begin{aligned} \frac{\min_{\nu \geq 2r_0/t} \frac{\beta^2\theta_0}{16}\left[\nu t \log\left(\frac{\nu t}{\nu t - 2r_0}\right) - 2tr_0\right] + \nu}{2r_0/t} &\geq \frac{\frac{\beta^2\theta_0}{16}\left[\nu t \log\left(\frac{\nu t}{\nu t - 2r_0}\right) - 2tr_0\right] + \nu\Big|_{\nu = 2.4 \cdot r_0/t}}{2r_0/t} \\ &= \frac{\frac{\beta^2\theta_0 t}{16}\left(2.4 \cdot \log 3 - 2\right) + 2.4}{2} =: \left(1.2 \cdot \log 3 - 1\right) z + 1.2 \end{aligned}$$

where we introduce $z = \frac{\beta^2 t}{8r_0^2}$ which is between 0 and 1/2 by our earlier assumption. The above function of z is clearly bounded above by 3/2 as desired, thus completing the proof of Theorem 24.

A.6 Proof of Proposition 25

In order to prove Proposition 25, we find it convenient to begin with the problem (4.9) which we will show has the original problem (4.8) as its Lagrangian dual. First, it is necessary to verify that (4.9) can be restricted to a compact set, and therefore must attain its minimizer:

Lemma 46. *Problem (4.9) attains its minimizer for finite values ν^*, λ^*, and Q^*.*

Proof. Assume without loss of generality that $\mathcal{R}$ is oriented in such a way that Σ is diagonal, and assume that μ is the origin and that the depot point is p_0, giving the problem

$$\begin{aligned}\underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^2,Q\in\mathbb{R}^{2\times 2}}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2}{\nu+\lambda^T x+x^TQx-\frac{2}{t}\|x-p_0\|}\,dA+\nu+s_{11}q_{11}+s_{22}q_{22} \qquad s.t. \qquad (A.1)\\ \nu+\lambda^T x+x^TQx \;\geq\; \frac{2}{t}\|x\| \quad \forall x\in\mathcal{R}\\ Q \;\succeq\; \mathbf{0}\,.\end{aligned}$$

Let M denote the objective value of (A.1) when $\nu = 1+\frac{2}{t}\max_{x\in\mathcal{R}}\|x-p_0\|$ and λ and Q are both equal to the zero vector and matrix respectively. Since the origin (i.e. μ) is contained in $\mathcal{R}$, it is then immediately clear that we can impose the restriction that $0\leq\nu\leq M$ without affecting the problem (since ν appears in the linear term of the objective function), and furthermore, since we must have $\Sigma\bullet Q\leq M$ with $s_{11},s_{22}>0$, it is not hard to derive the bound $\|Q\|_1\leq 3M/\min\{s_{11},s_{22}\}$. This merely leaves us with the task of bounding λ; to do so, let $\mathcal{B}$ be a ball about the origin (i.e. μ) of radius ϵ that is contained in $\mathcal{R}$. At the point $x=-\epsilon\lambda/\|\lambda\|$ on the boundary of $\mathcal{B}$ we see that $\lambda^Tx=-\epsilon\|\lambda\|$. Thus, it must be true that $\nu+\lambda^Tx+x^TQx-\frac{2}{t}\|x-p_0\|\geq 0$, i.e.

$$\begin{aligned}\underbrace{\nu}_{\leq M}+\underbrace{\lambda^Tx}_{=-\epsilon\|\lambda\|}+\underbrace{x^TQx}_{\leq\|Q\|\epsilon^2}-\underbrace{\frac{2}{t}\|x-p_0\|}_{\geq 0} &\geq 0\\ M-\epsilon\|\lambda\|+\underbrace{\|Q\|\epsilon^2}_{\leq 3\epsilon^2M/\min\{s_{11},s_{22}\}} &\geq 0\\ \|\lambda\| &\leq M\left(\frac{1}{\epsilon}+\frac{3\epsilon}{\min\{s_{11},s_{22}\}}\right),\end{aligned}$$

and thus we conclude that (4.9) can be restricted to a compact set and thus attains its minimizer. □

We next note that (4.9) can also be written as

$$\begin{aligned}\underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^2,Q\in\mathbb{R}^{2\times 2},\sigma(\cdot)\in\mathcal{L}^2}{\text{minimize}} \iint_{\mathcal{R}}\frac{1}{4}\cdot\frac{\beta^2}{\sigma(x)}\,dA+\nu+\lambda^T\mu+(\Sigma+\mu\mu^T)\bullet Q \qquad s.t.\\ \nu+\lambda^Tx+x^TQx-\frac{2}{t}\|x\|-\sigma(x) \;\leq\; 0 \quad \forall x\in\mathcal{R}\\ \frac{2}{t}\|x\|-\nu-\lambda^Tx-x^TQx \;\leq\; 0 \quad \forall x\in\mathcal{R}\\ -Q \;\preceq\; \mathbf{0}\,;\end{aligned}$$

note that a disadvantage of the above formulation is that its feasible region does not have an interior point and thus we cannot directly apply standard results of strong duality. It is, however, straightforward to verify that the dual of the above problem is precisely problem (4.8):

Definition 47. (Lagrange Duality; see Section 8.6 of [139]) Let $\mathfrak{f}$ be a real-valued convex functional defined on a convex subset Ω of a vector space $\mathfrak{X}$, and let $\mathfrak{G}$ be a convex mapping of $\mathfrak{X}$ into a normed space $\mathfrak{Z}$, where $\mathfrak{Z}$ contains a zero element θ. The *Lagrangian dual* of the optimization problem

$$\begin{array}{rcll} \underset{\mathfrak{x}}{\text{minimize}}\, \mathfrak{f}(\mathfrak{x}) & & & s.t. \\ \mathfrak{G}(\mathfrak{x}) & \leq & \theta & \\ \mathfrak{x} & \in & \Omega & \end{array}$$

is defined by

$$\begin{array}{rcll} \underset{\mathfrak{z}^\star}{\text{maximize}}\, \varphi(\mathfrak{z}^\star) & & & s.t. \\ \mathfrak{z}^* & \geq & \theta & \end{array}$$

where

$$\varphi(\mathfrak{z}^\star) := \inf_{\mathfrak{x}\in\Omega} \mathfrak{f}(\mathfrak{x}) + \langle \mathfrak{G}(\mathfrak{x}), \mathfrak{z}^\star \rangle$$

and satisfies $\mathfrak{f}(\mathfrak{x}) \geq \varphi(\mathfrak{z}^\star)$ for all feasible pairs $\mathfrak{x}$ and $\mathfrak{z}^\star$.

In order to apply this definition to our problem, we let $\mathfrak{X} = \mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^{2\times 2} \oplus \mathcal{L}^2$ and let $\Omega = \mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{S}^2 \oplus \mathcal{L}^2$, where $\mathbb{S}^2$ denotes the set of symmetric positive semidefinite 2×2 matrices. Let $\mathfrak{Z} = \mathcal{L}^2 \oplus \mathcal{L}^2 \oplus (-\mathbb{S}^2)$ and let an element of $\mathfrak{Z}^\star = \mathcal{L}^2 \oplus \mathcal{L}^2 \oplus (-\mathbb{S}^2)$ be denoted by the triple $(g(\cdot), h(\cdot), P)$, with $\mathfrak{G}(\cdot) : \mathfrak{X} \to \mathfrak{Z}$ defined by

$$\mathfrak{G}(\nu, \lambda, Q, \sigma(\cdot)) = \begin{pmatrix} \nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\| - \sigma(x) \\ \frac{2}{t}\|x\| - \nu - \lambda^T x - x^T Q x \\ -Q \end{pmatrix}.$$

We then find that

$$\begin{aligned}
\varphi(\mathfrak{z}^\star) &= \inf_{\mathfrak{x}\in\Omega} \mathfrak{f}(\mathfrak{x}) + \langle \mathfrak{G}(\mathfrak{x}), \mathfrak{z}^\star \rangle \\
&= \inf_{(\nu,\lambda,Q,\sigma(\cdot))\in\Omega} \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2}{\sigma(x)}\, dA + \nu + \lambda^T\mu + (\Sigma + \mu\mu^T) \bullet Q \\
&+ \iint_{\mathcal{R}} g(x)\left(\nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\| - \sigma(x)\right) dA + \iint_{\mathcal{R}} h(x)\left(\frac{2}{t}\|x\| - \nu - \lambda^T x - x^T Q x\right) dA - Q \bullet P \\
&= \inf_{(\nu,\lambda,Q,\sigma(\cdot))\in\Omega} \iint_{\mathcal{R}} \frac{1}{4}\cdot\frac{\beta^2}{\sigma(x)} - g(x)\sigma(x)\, dA + \nu\left(1 + \iint_{\mathcal{R}} g(x) - h(x)\, dA\right) + \lambda^T\left(\mu + \iint_{\mathcal{R}} x\,(g(x) - h(x))\, dA\right) \\
&+ Q \bullet \left(\Sigma + \mu\mu^T - P + \iint_{\mathcal{R}} xx^T\,(g(x) - h(x))\, dA\right) - \iint_{\mathcal{R}} \frac{2}{t}\|x\|(g(x) - h(x))\, dA\,.
\end{aligned}$$

In order for the above expression to be bounded, it must be the case that

$$\begin{aligned}
1 + \iint_{\mathcal{R}} g(x) - h(x)\, dA &= 0 \\
\mu + \iint_{\mathcal{R}} x\,(g(x) - h(x))\, dA &= 0 \\
\Sigma + \mu\mu^T - P + \iint_{\mathcal{R}} xx^T\,(g(x) - h(x))\, dA &\succeq \mathbf{0}\,.
\end{aligned}$$

Furthermore, we can also assume that $\sigma(x) = \frac{\beta}{2\sqrt{-g(x)}}$, and make the substitution $g(\cdot) \mapsto -g(\cdot)$, so that the problem of maximizing $\varphi(\mathfrak{z}^\star)$ is equivalent to

$$\begin{aligned}
\underset{g(\cdot),h(\cdot),P}{\text{maximize}}\ \beta \iint_{\mathcal{R}} \sqrt{g(x)}\, dA + \frac{2}{t}\|x\|(g(x) + h(x))\, dA \qquad & s.t. \\
1 - \iint_{\mathcal{R}} g(x) + h(x)\, dA &= 0 \\
\mu - \iint_{\mathcal{R}} x\,(g(x) + h(x))\, dA &= 0 \\
P + \iint_{\mathcal{R}} xx^T\,(g(x) + h(x))\, dA &\preceq \Sigma + \mu\mu^T\,.
\end{aligned}$$

It is easy to verify that we can assume that $h(x) = 0$ for all $x \in \mathcal{R}$ and that $P = \mathbf{0}$, which then verifies that (4.8) is indeed the dual problem of (4.9). In order to complete the proof of Proposition 25, it will suffice to check that the primal and dual problems both have the same objective value when $f^*(\cdot)$ is defined according to equation (4.10). Recall that at

optimality for problem (4.8), we have

$$\begin{aligned}
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA &= 1 \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA &= \mu \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x x^T}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\, dA + P^* &= \Sigma + \mu\mu^T \\
P^* \bullet Q^* &= 0 \\
P^*, Q^* &\succeq \mathbf{0}\,,
\end{aligned}$$

where we have included a Lagrange multiplier matrix P associated with the constraint that $Q \succeq \mathbf{0}$ (see Section 5.9 of [182]). Setting

$$f^*(x) = \frac{1}{4} \cdot \frac{\beta^2}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2}\,,$$

we then see that indeed

$$\begin{aligned}
&\iint_{\mathcal{R}} \frac{2}{t}\|x\| f^*(x) + \beta\sqrt{f^*(x)}\, dA \\
&= \iint_{\mathcal{R}} \frac{2}{t}\|x\| \left(\frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2} \right) + \beta \sqrt{\frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}}\, dA \\
&= \iint_{\mathcal{R}} \frac{2}{t}\|x\| \left(\frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2} \right) + \frac{1}{2} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA \\
&+ \iint_{\mathcal{R}} \frac{2}{t}\|x\| \left(\frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2} \right) + \frac{1}{4} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA + \frac{\beta^2}{4} \iint_{\mathcal{R}} \frac{\nu^* + (\lambda^*)^T x + x^T Q^* x}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}\, dA \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA + \nu^* \underbrace{\frac{\beta^2}{4} \iint_{\mathcal{R}} \frac{1}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}\, dA}_{=1} \\
&+ (\lambda^*)^T \underbrace{\frac{\beta^2}{4} \iint_{\mathcal{R}} \frac{x}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}\, dA}_{=\mu} + Q^* \bullet \underbrace{\left(\frac{\beta^2}{4} \iint_{\mathcal{R}} \frac{x x^T}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2}\, dA \right)}_{=\Sigma + \mu\mu^T - P^*} \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|}\, dA + \nu^* + (\lambda^*)^T \mu + (\Sigma + \mu\mu^T) \bullet Q^*
\end{aligned}$$

as desired, which completes the proof.

A.7 Proof of Proposition 26

The dual problem of (4.11) is

$$\begin{aligned} \underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^2,Q\in\mathbb{R}^{2\times 2}}{\text{minimize}}\ \nu + \lambda^T\mu + s(q_{11}+q_{22}) \qquad & s.t. \\ \nu + \lambda^T x + x^T Q x &\geq \frac{2}{t}\|x\| \quad \forall x \in \mathcal{D} \\ Q &\succeq \mathbf{0}\,. \end{aligned} \tag{A.2}$$

By symmetry of the region $\mathcal{D}$ and concavity of the objective function, we know that the first-moment constraint $\iint_{\mathcal{R}} x f(x)\, dA = \mu$ is redundant and we can therefore assume that $\lambda = 0$ in the dual problem. By similar reasoning, we can also assume that Q is also a diagonal matrix whose entries are both a scalar q. Thus, an equivalent form of (A.2) is

$$\begin{aligned} \underset{\nu,q\in\mathbb{R}}{\text{minimize}}\ \nu + 2sq \qquad & s.t. \\ \nu + 2q\|x\|^2 &\geq \frac{2}{t}\|x\| \quad \forall x \in \mathcal{D} \\ q &\geq 0\,. \end{aligned} \tag{A.3}$$

The above problem can be solved explicitly: if $s \leq A/\pi$, then the optimal solution has $q^* = \frac{1}{2t\sqrt{s}}$ and $\nu^* = \sqrt{s}/t$ and an objective function value of $2\sqrt{s}/t$. If $s > A/\pi$, the optimal solution has $q^* = 0$ and $\nu^* = 2\sqrt{A/\pi}/t$ and an objective function value of $2\sqrt{A/\pi}/t$. This implies that the optimal solution to the primal problem (4.11) is to let $f(\cdot)$ be a uniform mixture of two atomic masses at the points $(\pm r, 0)^T$, where $r = \min\{\sqrt{s}, 2\sqrt{A/\pi}\}$, as desired.

A.8 Proof of Theorem 27

The dual problem of interest is

$$\begin{aligned} \underset{\nu,q\in\mathbb{R}}{\text{minimize}} \iint_{\mathcal{D}} \frac{1}{4}\cdot\frac{\beta^2}{\nu + q\|x\|^2}\, dA + \nu + 2sq \qquad & s.t. \\ \nu, q &\geq 0\,, \end{aligned} \tag{A.4}$$

which can be written explicitly as

$$\begin{aligned} \underset{\nu,q\in\mathbb{R}}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log(\nu\pi+qA)-\log\nu-\log\pi}{q}\right)+\nu+2sq \qquad & s.t. \\ \nu, q \;\geq\; 0\,. & \end{aligned}$$

We can differentiate the above objective function to solve explicitly for ν, giving

$$\nu = \frac{\sqrt{q^2A^2+\pi^2\beta^2A}-qA}{2\pi}\,,$$

which we then substitute into the following simpler lower bound of (A.4):

$$\begin{aligned} & \underset{\nu,q\geq 0}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log A+\log q-\log v-\log\pi}{q}\right)+2sq \\ \equiv\; & \underset{q\geq 0}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log A+\log q-\log\left(\frac{\sqrt{q^2A^2+\pi^2\beta^2A}-qA}{2\pi}\right)-\log\pi}{q}\right)+2sq \\ \geqq\; & \underset{q\geq 0}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log A+\log q-\log\left(\sqrt{q^2A^2+\pi^2\beta^2A}-qA\right)}{q}\right)+2sq \\ \equiv\; & \underset{q\geq 0}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log A+\log\left(\frac{q}{\sqrt{q^2A^2+\pi^2\beta^2A}-qA}\right)}{q}\right)+2sq \\ \geqq\; & \underset{q\geq 0}{\text{minimize}}\, \frac{\beta^2\pi}{4}\left(\frac{\log A+\log\frac{2q^2}{\pi\beta^2}}{q}\right)+2sq \geq \frac{\beta^2\pi\log A}{4q}+2sq\,, \end{aligned}$$

which has an optimal objective function value of $\beta\sqrt{2\pi s\log A}$. In the above analysis, we have assumed that $q\to\infty$ as $A\to\infty$, which justifies the assumption that $\log\frac{2q^2}{\pi\beta^2}\geq 0$ in the last line. This is must be the case because if we were to bound q from above as $A\to\infty$, then the resulting objective value would be proportional to $\log A$. In order to complete the proof of Theorem 27, we must also show that there exists an *upper bound* of problem (A.4) that approaches $\beta\sqrt{2\pi s\log A}$ as $A\to\infty$. Setting

$$\begin{aligned} q &= \frac{\beta}{4}\cdot\sqrt{\frac{2\pi\log A}{s}} \\ \nu &= \frac{\beta^2}{4}\cdot\sqrt{\frac{2\pi s}{\log A}}\,, \end{aligned}$$

we see that

$$\begin{aligned} & \frac{\beta^2\pi}{4}\left(\frac{\log(\nu\pi+qA)-\log\nu-\log\pi}{q}\right)+\nu+2sq \\ = \; & \frac{\beta\sqrt{2\pi s}}{4}\cdot\frac{2\log(\beta\pi s+A\log A)+\beta+2\log A-2\log\beta-2\log\pi-2\log s}{\sqrt{\log A}} \\ \sim \; & \beta\sqrt{2\pi s\log A} \end{aligned}$$

as $A\to\infty$, holding s constant. This completes the proof.

A.9 Proof of Theorem 29

Suppose that $f(\cdot)$ happens to be a finite mixture of atomic distributions, i.e. $f(\cdot) := \sum_{i=1}^{N} s_i\delta(x-\bar{x}_i)$, with objective function value OBJ, mean μ, and covariance matrix $\bar{\Sigma} \preceq \Sigma+\mu\mu^T$. Consider the polyhedron $\mathcal{P}$ defined by

$$\mathcal{P}=\left\{(s_1,\ldots,s_N):\sum_{i=1}^{N}\frac{2}{t}\|\bar{x}_i-p_i\|s_i=\mathsf{OBJ},\quad \sum_{i=1}^{N}\bar{x}_is_i=\mu,\quad \sum_{i=1}^{N}\bar{x}_i\bar{x}_i^Ts_i=\bar{\Sigma},\quad \sum_{i=1}^{N}s_i=1,\quad s_i\geq 0\,\forall i\right\}.$$

Clearly, $\mathcal{P}$ can be represented in the form $As=b$ with $s\geq 0$, where $A\in\mathbb{R}^{7\times N}$, $s\in\mathbb{R}^N$, and $b\in\mathbb{R}^7$, and therefore its corner points have at most 7 non-zero entries. Thus, there exists an atomic distribution with at most 7 terms whose objective value is the same as that of the original density $f(\cdot)$. If $f(\cdot)$ is not atomic, we can always construct a mixture of atomic distributions $\hat{f}(\cdot) := \sum_{i=1}^{N} s_i\delta(x-\bar{x}_i)$ whose objective value is within ϵ of $\iint_{R_i}\frac{2}{t}\|x-p_i\|f(x)\,dA$, where $\hat{f}(\cdot)$ has the same mean μ and covariance as $f(\cdot)$.

A.10 Proof of Lemma 33

Proofs of statements 2 through 4 follow below:

Proof of statement 2 We seek to show that

$$\iint_{\mathcal{R}} f(x)\min_i\{\|x-x_i\|-\lambda_i'\}\,dA \leq \iint_{\mathcal{R}} f(x)\min_i\{\|x-x_i\|-\lambda_i\}\,dA+\mathbf{g}^T(\boldsymbol{\lambda}'-\boldsymbol{\lambda}),$$

which is equivalent to showing that

$$\iint_{\mathcal{R}} f(x)\min_i\{\|x-x_i\|-\lambda_i'\}\,dA \leq \sum_{i=1}^{n}\iint_{R_i} f(x)(\|x-x_i\|-\lambda_i)\,dA+g_i(\lambda_i'-\lambda_i).$$

Consider the right-hand side of the above; for each i, we have

$$\begin{aligned}\iint_{R_i} f(x)(\|x-x_i\|-\lambda_i)\,dA + g_i(\lambda_i^{'}-\lambda_i) &= \iint_{R_i} f(x)(\|x-x_i\|-\lambda_i)\,dA - (\lambda_i^{'}-\lambda_i)\iint_{R_i} f(x)\,dA \\ &= \iint_{R_i} f(x)(\|x-x_i\|-\lambda_i^{'})\,dA\end{aligned}$$

and therefore, if we define regions $R_1^{'},\ldots,R_n^{'}$ in the obvious way by

$$R_i^{'} = \left\{x \in \mathcal{R} : \|x-x_i\|-\lambda_i^{'} \leq \|x-x_j\|-\lambda_j^{'}\ \forall j \neq i\right\},$$

we see that

$$\iint_{\mathcal{R}} f(x)\min_i\{\|x-x_i\|-\lambda_i^{'}\}\,dA = \sum_{i=1}^n \iint_{R_i^{'}} f(x)(\|x-x_i\|-\lambda_i^{'})\,dA \leq \sum_{i=1}^n \iint_{R_i} f(x)(\|x-x_i\|-\lambda_i^{'})\,dA$$

is obvious because the partition $R_1^{'},\ldots,R_n^{'}$ is obtained by taking the *minimal* value of $\|x-x_i\|-\lambda_i^{'}$, and is therefore minimal over all partitions of $\mathcal{R}$. This completes the proof.

Proof of statement 3 We observe that the vector $-\frac{1}{n}\mathbf{e} \in \mathbb{R}^n$ must be a supergradient at $\boldsymbol{\lambda}^*$; this simply follows from the KKT conditions of (5.2), which is a finite-dimensional problem. Therefore, it follows that $\iint_{R_i^*} f(x)\,dA = 1/n$ for all i, and therefore the objective value of problem (5.2) is

$$\begin{aligned}\iint_{\mathcal{R}} f(x)\min_i\{\|x-x_i\|-\lambda_i^*\}\,dA &= \sum_{i=1}^n \iint_{R_i^*} f(x)(\|x-x_i\|-\lambda_i^*)\,dA \\ &= \sum_{i=1}^n \iint_{R_i^*} f(x)\|x-x_i\|\,dA - \lambda_i^* \iint_{R_i^*} f(x)\,dA \\ &= \sum_{i=1}^n \iint_{R_i^*} f(x)\|x-x_i\|\,dA - \frac{1}{n}\underbrace{\mathbf{e}^T\boldsymbol{\lambda}^*}_{=0} = \sum_{i=1}^n \iint_{R_i^*} f(x)\|x-x_i\|\,dA\end{aligned}$$

and therefore the Wasserstein distance between f and $\hat{f}$ as induced by the partition $R_1^*,\ldots,R_n^*$ is the same as that of the optimal objective value of (5.2), which completes the proof.

Proof of statement 4 We simply note that if $f(x) > 0$ then the supergradient inequality in the proof of statement 2 is actually strict:

$$\sum_{i=1}^n \iint_{R_i^{'}} f(x)(\|x-x_i\|-\lambda_i^{'})\,dA < \sum_{i=1}^n \iint_{R_i} f(x)(\|x-x_i\|-\lambda_i^{'})\,dA\,.$$

The objective function of problem (5.2) is therefore strictly concave, thus guaranteeing uniqueness of $\boldsymbol{\lambda}^*$. The fact that $\boldsymbol{\lambda}^*$ exists follows from the boundedness of $\mathcal{R}$, because if we were ever to have $\lambda_i - \lambda_j > \text{diam}(\mathcal{R})$, it would imply that $\|x - x_i\| - \lambda_i < \|x - x_j\| - \lambda_j$ for all $x \in \mathcal{R}$, thus rendering R_j to be empty.

A.11 Proof of Theorem 34

Purely for ease of exposition, we assume that $\mathcal{R}$ is the unit square. Section 2.1 of [178] says that $\mathcal{D}(\hat{f}_n, \bar{f}) \to 0$ with probability one because the Wasserstein distance metrizes weak convergence whenever $\mathcal{R}$ is compact. Thus, setting $t_n = \mathcal{D}(\hat{f}_n, \bar{f})$ for all $n \geq 1$ gives us a sequence that converges to 0 with probability one, with the added feature that $\bar{f}$ is feasible for problem (5.5) by construction. Next, for each n, the triangle inequality says that the set of distributions f on $\mathcal{R}$ such that $\mathcal{D}(f, \bar{f}) \leq 2t_n$ must contain the set of distributions where $\mathcal{D}(f, \hat{f}_n) \leq t_n$. Thus, an upper bound for problem (5.5) – which is itself always an upper bound for the ground truth cost $\iint_{\mathcal{R}} \sqrt{\bar{f}(x)}\, dA$ by our construction of t_n – is given by the problem

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \mathcal{D}(f, \bar{f}) &\leq 2t_n \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,; \end{aligned} \tag{A.5}$$

it will therefore suffice to verify that the optimal objective value to this problem approaches the ground truth cost as $t_n \to 0$.

We will relax problem (A.5) one step further by using an alternate metric to the Wasserstein distance, namely the *Prokhorov metric* $\mathcal{D}_P(\cdot, \cdot)$, defined by

$$\mathcal{D}_P(\mu_1, \mu_2) = \inf\{\epsilon > 0 : \mu_1(B) \leq \mu_2(B^\epsilon) + \epsilon \text{ for all Borel sets } B \text{ on } \mathcal{R}\}$$

where $B^\epsilon = \{x : \inf_{y \in B} d(x, y) \leq \epsilon\}$. Theorem 2 of [195] says that for any two distributions f and g on $\mathcal{R}$, we have $(\mathcal{D}_P(f, g))^2 \leq \mathcal{D}(f, g)$, and therefore we can study the relaxation of

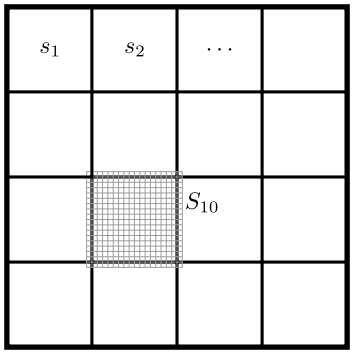


Figure A.2: A division of the unit square $\mathcal{R}$ into $N^2 = 16$ grid cells. The larger square S_{10} has side length $1/N + 2/N^3$and contains s_{10}.

(A.5) given by

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \mathcal{D}_P(f, \bar{f}) &\leq \sqrt{2t_n} \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{A.6}$$

as $n \to \infty$, whence $t_n \to 0$ with probability one. For ease of notation, we will define $\epsilon = \sqrt{2t_n}$.

Let N be a positive integer and suppose that $\epsilon = 1/N^3$. We then divide $\mathcal{R}$ into N^2 square grid cells s_i with side length $1/N$. The distance constraint $\mathcal{D}_P(f, \bar{f}) \leq \epsilon$ implies that for each $B = s_i$, we have $\iint_{s_i} f(x)\, dA \leq \iint_{S_i} \bar{f}(x)\, dA + \epsilon$, where S_i is the square of side length $1/N + 2/N^3$ that contains s_i (see Figure A.2). Define $m_i = N^2 \iint_{S_i} \bar{f}(x)\, dA$ for each m_i and consider the relaxation of (A.6) given by

$$\begin{aligned} \underset{f}{\text{maximize}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA \qquad & s.t. \\ \iint_{s_i} f(x)\, dA &\leq \frac{m_i}{N^2} + \epsilon \quad \forall i \\ \iint_R f(x)\, dA &= 1 \\ f(x) &\geq 0 \quad \forall x \in \mathcal{R}\,. \end{aligned} \tag{A.7}$$

If we ignore the constraint that $\iint_R f(x)\, dA = 1$, then clearly, our optimal solution f^* would simply have $\iint_{s_i} f^*(x)\, dA = m_i/N^2 + \epsilon$ for each i. This problem has a finite-dimensional constraint space and it is straightforward to see that its optimal solution f^* must be piecewise

constant on each piece s_i, so that $f^* = q_i^*$ on each s_i, defined by

$$\frac{q_i^*}{N^2} = \frac{m_i}{N^2} + \epsilon$$

or equivalently

$$q_i^* = m_i + 1/N\,.$$

Thus, the optimal objective value of (A.7) is at most

$$\frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{q_i^*} = \frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{m_i + 1/N} \leq \frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{m_i} + \frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{1/N} = \frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{m_i} + \sqrt{1/N}\,;$$

it is routine to verify that $\frac{1}{N^2}\sum_{i=1}^{N^2}\sqrt{m_i} \to \iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA$ (the only reason that this is not simply the definition of an integral is because the squares S_i that characterize the m_i's have an area of $(1/N + 2/N^3)^2$ rather than $1/N^2$), which thereby completes the proof.

Appendix B

Additional Analyses

B.1 Probabilistic analysis of the capacitated VRP

We first note that, if n samples are drawn from a distribution f, then $\mathbf{E}(\sum_{i=1}^{n} \|x_i\|) = n \iint_{\mathcal{R}} \|x\| f(x)\, dA$. The representation of capacity constraints via the substitution $c = s\sqrt{n}$ is a standard and useful technique that can be seen in Section 4.2 of [132] or the paper [117]. By exchanging the expectation and $\max\{\cdot,\cdot\}$ operators, we can express the bound (5.12) as

$$\begin{aligned} & \max\left\{\frac{2\sqrt{n}}{s}\iint_{\mathcal{R}} \|x\| f(x)\, dA,\, \beta\sqrt{n}\iint_{\mathcal{R}} \sqrt{f_c(x)}\, dA\right\} + o(\sqrt{n}) \\ \leq\ & \mathbf{E}\,\mathsf{VRP}(X) \\ \leq\ & 2\left\lceil\frac{\sqrt{n}}{s}\right\rceil \iint_{\mathcal{R}} \|x\| f(x)\, dA + \left(1 - \frac{1}{s\sqrt{n}}\right)\beta\sqrt{n}\iint_{\mathcal{R}} \sqrt{f_c(x)}\, dA + o(\sqrt{n})\,. \end{aligned}$$

Note that $\lceil\sqrt{n}/s\rceil$ is simply the number of vehicles needed to provide service. Since we are interested in the limiting behavior as $n \to \infty$, we have $\lceil\sqrt{n}/s\rceil \sim \sqrt{n}/s$ and $1/(s\sqrt{n}) \to 0$, so that we can write

$$\sqrt{n}\cdot\max\left\{\frac{2}{s}\iint_{\mathcal{R}} \|x\| f(x)\, dA,\, \beta\iint_{\mathcal{R}} \sqrt{f_c(x)}\, dA\right\} \lesssim \mathsf{VRP}(X) \lesssim \sqrt{n}\cdot\left(\frac{2}{s}\iint_{\mathcal{R}} \|x\| f(x)\, dA + \beta\iint_{\mathcal{R}} \sqrt{f_c(x)}\, dA\right)$$

as desired, where the approximate inequality implied by the "$\lesssim$" terms simply reflects the fact that we have disregarded the $o(\sqrt{n})$ terms.

B.2 Analysis of Algorithm 8

We find it convenient to emphasize the dependency of the sub-regions (i.e. the power diagram cells) R_i on the weight vector $\mathbf{w}$ and we will therefore use the notation $R_i(\mathbf{w})$ when appropriate. Let $d = \max_i \max_{x \in \mathcal{R}} \|x - p_i\|$. It is immediately clear that we can restrict ourselves to vectors $\mathbf{w} \in \mathbb{R}^n$ such that $|w_i - w_j| \leq d^2$ for all i and j because, if $w_i - w_j > d^2$, then $R_j(\mathbf{w}) = \emptyset$. We further observe that the cells defined by a power diagram are clearly invariant under addition by a constant and thereby restrict ourselves to vectors $\mathbf{w} \in \mathbb{R}^n$ such that $\sum_i w_i = 0$. Let $\mathcal{W}$ denote this feasible region for $\mathbf{w}$, i.e.

$$\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^n : |w_i - w_j| \leq d^2 \; \forall i, j \text{ and } \sum_i w_i = 0\} \,. \tag{B.1}$$

We are interested in determining the weight vector $\mathbf{w}^*$ such that

$$\Phi(\mathbf{w}) := \max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i(\mathbf{w}), \mathcal{R}) \tag{B.2}$$

is minimal. For any sub-domain $\mathcal{A} \subseteq \mathcal{W}$, we can determine a lower bound of $\min_{\mathbf{w} \in \mathcal{A}} \Phi(\mathbf{w})$ by setting

$$\Phi_{\mathsf{LB}}(\mathcal{A}) := \max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i(\mathbf{w}^{(i)}), \mathcal{R}) \,, \tag{B.3}$$

where $\mathbf{w}^{(i)}$ is defined by

$$\begin{aligned} w_i^{(i)} &= \min_{\mathbf{w} \in \mathcal{A}} w_i \\ w_j^{(i)} &= \max_{\mathbf{w} \in \mathcal{A}} w_j \quad \forall j \neq i \,. \end{aligned}$$

As an upper bounding function we may use $\Phi_{\mathsf{UB}}(\mathcal{A}) = \Phi(\mathbf{w})$, where $\mathbf{w}$ is any point in $\mathcal{A}$. It is entirely straightforward to verify that $\Phi_{\mathsf{LB}}(\cdot)$ and $\Phi_{\mathsf{UB}}(\cdot)$ satisfy the usual conditions for convergence of a branch-and-bound process and we will refrain from doing so here in the interest of brevity. The full procedure is shown in Algorithm 8.

Input: A convex region $\mathcal{R}$, a point $\mu \in \mathcal{R}$, a covariance matrix $\Sigma \succ \mathbf{0}$, a capacity coefficient t, n landmark points $p_i \in \mathcal{R}$, and a threshold ϵ.

Output: A partition of $\mathcal{R}$ into n power diagram cells R_i such that $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ is within ϵ of the minimum over all possible power diagram partitions.

Set $\mathbf{w}$ to be the analytic center of $\mathcal{W}$ as defined in (B.1);
Define $\Phi(\cdot)$ and $\Phi_{\mathsf{LB}}(\cdot)$ as in equations (B.2) and (B.3);
Set `BB_nodes` := $\{\mathcal{W}\}$;
Set `Upper_bounds` := $\{\Phi(\mathbf{w})\}$;
Set `Lower_bounds` := $\{\Phi_{\mathsf{LB}}(\mathcal{W})\}$;
Set `best_obj` := $\Phi(\mathbf{w})$;
Set `lowest_LB` := $\min\{$`Lower_bounds`$\}$;
while `best_obj` $-$ `lowest_LB` $\geq \epsilon$ **do**
    Set `node` := `BB_nodes`[0];
    Set `BB_nodes` := `BB_nodes`[1 : end];
    Set `Upper_bounds` := `Upper_bounds`[1 : end];
    Set `Lower_bounds` := `Lower_bounds`[1 : end];
    Split `node` into two nodes, `node_1` and `node_2`, along its longest dimension,
    $\arg\max_i \{\max_{\mathbf{w} \in \text{node}} w_i - \min_{\mathbf{w} \in \text{node}} w_i\}$;
    Let $\mathbf{w}'$ and $\mathbf{w}''$ be the analytic centers of `node_1` and `node_2`;
    Set `UB1` := $\Phi(\mathbf{w}')$ and `UB2` := $\Phi(\mathbf{w}'')$;
    Set `LB1` := Φ_{LB}(`node_1`) and `LB2` := Φ_{LB}(`node_2`);
    Set `best_obj` := $\min\{$`best_obj`, `UB1`, `UB2`$\}$;
    Set `BB_nodes` := `BB_nodes` $\cup$ `node_1` $\cup$ `node_2`;
    Set `Upper_bounds` := `Upper_bounds` $\cup$ `UB1` $\cup$ `UB2`;
    Set `Lower_bounds` := `Lower_bounds` $\cup$ `LB1` $\cup$ `LB2`;
    Set `lowest_LB` := $\min\{$`Lower_bounds`$\}$;
    /* Pruning unnecessary nodes */
    Let `I` be the indices `i` such that `Lower_bounds`[i] $\leq$ `best_obj`;
    Set `BB_nodes` := `BB_nodes`[I];
    Set `Upper_bounds` := `Upper_bounds`[I];
    Set `Lower_bounds` := `Lower_bounds`[I];
end
Let `i` denote the index of `Upper_bounds` such that `Upper_bounds`[i] = `best_obj`, and let $\mathbf{w}^*$ denote the analytic center of `BB_nodes`[i];
Set $\{R_1, \ldots, R_n\}$ to be the power diagram partition of $\mathcal{R}$, as defined via (4.17);
return $\{R_1, \ldots, R_n\}$;

Algorithm 8: Algorithm PowerDiagram gives a locally optimal solution to the robust partitioning problem.