ROBUST FACILITY LOCATION UNDER DEMAND LOCATION UNCERTAINTY

by

Auyon Siddiq

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright 2013 by Auyon Siddiq
Abstract

Robust facility location under demand location uncertainty

Auyon Siddiq
Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto
2013

In this thesis, we generalize a set of facility location models within a two-stage robust optimization framework by assuming each demand is only known to lie within a continuous and bounded uncertainty region. Our approach involves discretizing each uncertainty region into a set of finite scenarios, each of which represents a potential location where the demand may be realized. We show that the gap between the optimal values of the theorized continuous uncertainty problem and our discretized model can be bounded by a function of the granularity of the discretization. We then propose a solution technique based on row-and-column generation, and compare its performance with existing solution methods. Lastly, we apply our robust location models to the problem of ambulance positioning using cardiac arrest location data from the City of Toronto, and show that hedging against demand location uncertainty may help decrease EMS response times to cardiac arrest emergencies.
Dedicated to my parents
Acknowledgements

I want to thank Professor Timothy Chan for being an exemplary thesis advisor. It is difficult for me to articulate the impact his mentorship has had on both my professional and personal development over the last two years. This thesis would have never been completed without Tim’s patience, encouragement and infectious passion for research. His ability to graciously manage the duties of being a supervisor while also respecting his students as research colleagues is truly remarkable. I would also like to thank (and apologize to) Laura for the times I kept Tim in his office well past dinner time.

A big thank you is also owed to my labmates – Taewoo, Sarina, Brendan, Derya, Houra, Daria, Yifang, Mark, Philip, Heyse and Velibor – for all of their support and friendship. Whether it was through helping me resolve technical issues, providing me with feedback on papers and presentations, partaking in imaginative discussions about research, or helping me decompress on a Friday evening, their support has been invaluable during my time at UofT. I would also like to thank my friends and UTORG colleagues outside of the AOL – Peter, Jane, Jenya, Jim, Kimia, Shefali and Curtiss – for injecting plenty of laughter and energy into the day-to-day grind of graduate school.

I also want to thank my committee members, Professor Chris Beck and Professor Roy Kwon, for the time and energy they spent providing invaluable feedback on this work.

Lastly, I want to thank my parents and older sister for their endless love and support from day 1.

This research was supported by a Canada Graduate Scholarship (Master’s) from the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship (Master’s) from the Government of Ontario.
# Contents

1 Introduction ........................................... 1  
   1.1 Related literature ................................. 2  
      1.1.1 Demand location uncertainty ..................... 2  
      1.1.2 Two-stage robust optimization .................... 3  
      1.1.3 Ambulance location models ....................... 4  
   1.2 Organization ...................................... 5  

2 Modeling demand location uncertainty .................. 6  
   2.1 Location models ................................... 6  
      2.1.1 P-median problem ............................... 6  
      2.1.2 P-center problem ............................... 10  
      2.1.3 Conditional value-at-risk ....................... 11  
   2.2 Relationship of p-median and p-center to CVaR ........... 13  
   2.3 Bounds on the discrete approximation error ............ 15  
   2.4 A continuous demand interpretation .................. 20  

3 Decomposition techniques ................................ 22  
   3.1 Primal methods .................................... 22  
      3.1.1 Row-generation only .............................. 22  
      3.1.2 Row and column generation ....................... 24  
   3.2 Dual methods ...................................... 25  
      3.2.1 Equivalent dual formulation ..................... 25  
      3.2.2 Dual master problem with row-and-column generation 26  
   3.3 Computational performance ........................... 26  

4 Computational study ................................... 28  
   4.1 Experimental design ................................ 29  
   4.2 Results ........................................... 31  
   4.3 Discussion .......................................... 36  

5 Conclusion ............................................. 38  

Bibliography ............................................. 40  

A Proofs ................................................. 44
List of Figures

2.1 Example of a three-move game with one demand (1 - facility placement by agent, 2 - demand realization by adversary, 3 - demand assignment by agent) ................. 8

2.2 Bound on continuous worst-case distance $Z_C$ (top); bound on discretization error, $\Delta_{med}$ (bottom) ................................................................. 18

4.1 12 demand regions with nominal and discrete scenario locations ................. 30

4.2 Total fraction of cardiac arrests within each demand region for training and testing data sets ................................................................. 31

4.3 Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance - p-median models .............................................. 32

4.4 Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance - CVaR models .............................................. 33

4.5 Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance - p-center models .............................................. 34

4.6 Locations of six facilities placed by robust p-median, robust CVaR and robust p-center location models ........................................... 36
List of Tables

3.1 Sensitivity of solution time to number of scenarios ($m = 50, n = 10, P = 5$) ............ 27
4.1 Performance of six location models on training dataset (506 cardiac arrests) ............ 35
4.2 Performance of six location models on test dataset (2023 cardiac arrests) .............. 35
4.3 Ranking of six models based on three distance metrics of test dataset .................... 35
Chapter 1

Introduction

In strategic facility location planning, uncertainty or changes in the operational environment – such as demand weights, facility capacities, travel time and distance – may lead to unforeseen costs or degrade system performance. As a result, facility location under uncertainty has received significant attention in the location science literature, particularly with respect to uncertainty in demand node weights and edge lengths (Snyder, 2006). However, to the best of our knowledge, relatively little attention has been given to problems where there is uncertainty in the location of the demands themselves.

The goal of this thesis is to propose a tractable model and solution method for facility location problems in which the location of each demand is subject to uncertainty. We are motivated by practical applications in which demand locations are uncertain at the time of facility siting, and where there exists a clear interest in hedging against the worst-case realization of the uncertainty. Facility location problems related to public health and emergency response are well aligned with such a goal. To illustrate, consider the problem of strategically placing ambulances in an urban area to minimize response time to future emergencies (Brotcorne et al., 2003). In this problem, demand locations might be approximated from demographic and historical call data, but the precise locations of future emergencies are impossible to know when ambulance positioning decisions are made. Other facility location problems where future demand locations may be uncertain include the siting of fire stations (Berman et al., 2013), treatment centers for medically evacuated soldiers (Bastian, 2010), vaccine clinics during an infectious outbreak (Lee et al., 2006), and placement of defibrillators in public areas (Chan et al., 2013; Siddiq et al., 2012). Additionally, knowledge of how public service systems might operate under a worst-case scenario can provide meaningful managerial insights that are likely to be obscured if only nominal or average cases are considered.

Further, in some applications there may be an interest in shaping the distribution of demand “costs”, where the cost associated with a demand might be the distance or travel time to its nearest facility. For example, a common standard in North American emergency medical service (EMS) systems is to respond to 90% of urban area calls within 9 minutes (Erkut et al., 2008). This response time target implies a clear interest in positioning ambulances in a way that minimizes the tail of the response time distribution. However, uncertainty in the location of future emergencies also introduces an uncertainty into the response time distribution, suggesting that these EMS standards may not be met if demand location uncertainty is unaccounted for during planning.

We make three contributions in this work:
1. First, we use a two-stage robust optimization framework to generalize a class of facility location problems where each demand is only known to lie within a continuous and bounded uncertainty region. While modeling the uncertainty regions as continuous may provide the most accurate measure of the true worst-case realization, solving large continuous location problems can be challenging. We present an alternate approach wherein each uncertainty region is approximated by a set of discrete locations – each representing a potential location where the demand could be realized – and provide a bound on the objective function error introduced by the discretization. We incorporate demand location uncertainty into three location models: the p-median problem, the p-center problem, and a conditional value-at-risk (CVaR) problem. Notably, we show that the p-median and p-center problems can be interpreted as special cases of the CVaR model, depending on the value of a tunable parameter. As a result, we show that combining our uncertainty framework with the concept of CVaR results in a general location model that can be tuned to optimize the mean (p-median), maximum (p-center) or tail-average of the distribution.

2. Second, we propose a solution technique based on row-and-column generation. The primary benefit of our method is that the computational performance of the algorithm is minimally impacted by the number of discrete scenarios enumerated for each uncertainty region. This allows each continuous uncertainty region to be finely discretized without compromising model tractability, leading to tight approximations of the corresponding continuous uncertainty problem. The solution algorithm we propose is applicable to all three of the models developed in Chapter 2.

3. Finally, we apply our robust optimization models to a case study on ambulance positioning. We use cardiac arrest location data from the City of Toronto. We examine the impact of accounting for demand location uncertainty by considering the distance metrics induced by the robust p-median, p-center and CVaR models (mean, maximum and tail average, respectively). We show that hedging against demand location uncertainty may improve the performance of EMS systems, particularly for those demands which are prone to be ill-served as a result of their location.

1.1 Related literature

In this section, we briefly provide an overview of relevant literature. We provide background on demand location uncertainty and two-stage robust optimization, as they are both present in our modeling framework. We also review the ambulance location literature to date.

1.1.1 Demand location uncertainty

Facility planning under uncertainty has received a significant amount of attention in the literature, with previous works taking both stochastic and robust optimization approaches to modeling uncertainty in demand node weights or edge lengths (Owen and Daskin, 1998; Snyder, 2006). Another source of uncertainty that has been considered is the risk of service disruptions at the facilities (see Lim et al. (2010), Shen et al. (2011) and Cui et al. (2010)).

With respect to demand location uncertainty, Cooper (1978) considers the problem of placing a single facility (the 1-median problem) in a network where each demand is only known to lie within an “uncertainty circle”. His main result is that the worst-case distance can be found by adding the sum of the circle radii to the optimal objective function value of the nominal problem, where the nominal
demand location is assumed to be at the center of the uncertainty circle. However, Cooper’s result does not extend to a multi-facility setting. For example, consider the simple case of four facilities at the corners of the unit square, with a single demand whose uncertainty circle is the largest possible circle inside the square – clearly, the worst-case demand location is also the nominal location. The models we present in this thesis can be viewed as a multi-facility extension to the problem in Cooper (1978), since we both consider local uncertainty in the location of each demand, and both assume only the boundaries of the uncertainty region are known. Averbakh and Bereg (2005) solve minimax regret 1-median and 1-center problems using rectilinear distances, where only interval estimates for each of the demand coordinates are known. For Euclidean distances, they only consider demand weight uncertainty. Drezner (1989) analyzes the 1-median on a sphere with random demand weights and locations, and shows that the difference between minimum and maximum possible objective value approaches zero as the number of demands approaches infinity.

We note here that demand location uncertainty can be interpreted as a special case of edge length uncertainty, since the practical consequence in both cases is uncertainty in travel time, distance or some other cost which is a function of edge length. However, no previous studies have modelled demand location uncertainty from the edge-length perspective. This is perhaps due to the difficulty in constructing tractable uncertainty sets for the edge lengths which accurately capture the change in each edge length that would result from a change in demand locations. Our approach can also be used to model edge length uncertainty, although a major difference from previous work is that our two-stage approach assumes that the assignment of demands to facilities occurs after the uncertainty has been realized.

1.1.2 Two-stage robust optimization

Single-stage robust optimization problems involve making all decisions prior to the realization or observation of any uncertainty. Two-stage robust optimization models (alternately, adjustable or adaptable robust optimization) have been proposed which involve recourse variables to represent decisions that are made after some or all of the uncertainty has been realized (Ben-Tal et al., 2004; Atamturk and Zhang, 2007). These recourse variables may represent true “wait-and-see” decisions, or may be auxiliary variables (slack or surplus variables) that do not necessarily correspond to material decisions. Unlike two-stage stochastic optimization, two-stage robust optimization does not require distributional information for the uncertain parameters. We refer the reader to Bertsimas et al. (2011) for a recent overview of the theory of two-stage robust optimization.

Two-stage robust optimization is a useful framework for modeling game theoretic problems, where the uncertain parameter is modeled as the decision of a real or fictitious (e.g. nature) adversary, and the planner has access to a recourse decision upon observation of the adversary’s move. Brown et al. (2006) and Brown et al. (2009) employ this game-theoretic framework when developing models for the defense of critical infrastructure and network interdiction. In both papers, a min-max-min mixed integer problem is obtained, and a Benders-based decomposition algorithm is used to solve the problem to either a desired tolerance or optimality.

Two-stage robust optimization can be a particularly useful framework for modeling facility location problems under uncertainty, due to the implicit two-stage structure of many location models: first, the placement of facilities in the network and second, the assignment of demands to facilities. In deterministic location models, these decisions are made simultaneously as a single stage optimization problem, since all demand information is known prior to facility placement. However, in an environment where the
location of future demands is uncertain, it may be unrepresentative of practical applications to make facility assignment decisions before the demand locations are observed (for example, the location of an emergency must be known before an ambulance can be assigned to it). Therefore, interpreting the assignment variable as a recourse decision and deferring it until after the uncertainty has been realized can be an accurate way of capturing the sequential nature of real world systems.

1.1.3 Ambulance location models

The strategic placement of ambulances or ambulance stations within urban areas is a well-studied problem in the facility location literature. Given the time critical nature of some emergencies such as cardiac arrest, healthcare planners have a clear interest in ensuring that ambulances are strategically located so they can respond to emergencies as quickly as possible. Earlier ambulance location models typically modelled all parameters as deterministic, while later formulations introduced probabilistic aspects into the models.

One of the first ambulance location models was the location set covering model (LSCM) (Toregas et al., 1971). The LSCM is a deterministic location model where the goal is to minimize the number of ambulance required to cover a set of demand nodes. An alternate coverage-based model for ambulance location is the maximal covering location problem (MCLP), where the goal is to cover as many demand points as possible for a fixed number of ambulances (Church and ReVelle, 1974). The LSCM and MCLP have served as the foundation for many other ambulance models that have been proposed over the last four decades. We provide a brief overview of some ambulance location models here, but refer the reader to Brotcorne et al. (2003) for a more detailed review of the ambulance location literature.

Schilling et al. (1979) propose a variation of the MCLP which uses two different types of ambulances to reflect the two tiers of service – advanced life support (ALS) and basic life support (BLS) – used by many EMS systems. Schilling’s model imposes the constraint that a “Type A” ambulance could only be placed at a candidate node if a “Type B” ambulance is also placed there. Daskin and Stern (1981) propose another modified MCLP model where each demand can be covered multiple times, with the objective of maximizing the number of demands covered more than once. A coverage constraint is included to ensure all demands are covered at least once. Gendreau et al. (1997a) proposes a more general form of the Daskin and Stern model by introducing two coverage radii, \( r_1 \leq r_2 \), where the objective is to maximize the number of demands covered twice within \( r_1 \) of an ambulance, with the constraint that all demands are at least within \( r_2 \) of an ambulance. Gendreau et al. (1997b) introduces a dynamic ambulance placement model where the ambulance locations are resolved at every period \( t \), with an associated cost for repositioning the ambulances.

Many ambulance models also incorporate probabilistic coverage to capture the reality that an ambulance may not always be available. Daskin (1983) proposes a model where each ambulance is independent and has a probability \( q \) of being unavailable, called the busy fraction. The model is based on the MCLP model, but employs an expected coverage objective function in lieu of a simple coverage objective. ReVelle and Hogan (1989) formulate a related problem with the added constraint that each demand must be covered with a probability of at least \( \alpha \). They also allow the busy probability \( q \) to vary with each candidate site. Goldberg et al. (1990) introduces an expected coverage model with stochastic travel times with the objective of maximizing the expected number of demands covered in under 8 minutes. Similarly, Repede and Bernardo (1994) consider variations in travel time throughout the day, with the same objective of maximizing expected coverage. With respect to probabilistic set covering models, Ball
and Lin (1993) propose a model based on the LSCM which seeks to minimize total ambulance cost such that all demands are covered with a probability of $\alpha$. Marianov and ReVelle (1993) formulate a queueing-based covering problem with busy fraction that vary with each ambulance site. This model focuses on the minimum number of ambulances needed to cover a demand such that the probability of all ambulances being busy at the same time is no greater than some specified threshold. Mandell (1998) proposes a two-tier system for ALS and BLS ambulances, and uses a queueing model to determine the busy fractions. In this model, the service level at a demand node depends on the number of ALS and BLS ambulances that are located within $r_1$ and $r_2$ of the node, respectively. More recently, Erkut et al. (2008) propose location models that use an expected survival objective, and argue that ambulance location should be viewed from the perspective of survival rather than the classical notion of coverage.

Another location problem in emergency medicine that is closely related to ambulance positioning is the deployment of automated external defibrillators (AEDs) in public locations. Publicly located defibrillators can enable bystanders to administer treatment to victims of sudden cardiac arrest prior to the arrival of EMS responders, and have been shown to improve chances of survival (Valenzuela et al., 2000). The AED location problem has primarily been studied using coverage-based models (Chan et al., 2013; Siddiq et al., 2012), although the methods developed in this thesis can easily be extended to the placement of AEDs as well.

While ambulance location has clearly been well investigated in the facility location literature, we note that much of the modeling focus has been on the “facility-side”, and aims to realistically model ambulance behaviour through busy probabilities, travel time uncertainty and multiple vehicle types. All of the previous models also aggregate ambulance demand into discrete points in space, and assume the location are known at the time of ambulance positioning. By contrast, the approach that we take in this thesis focuses on the “demand-side”, by modeling the demand for an ambulance as continuous and uncertain in space, making this work novel within the ambulance location literature.

### 1.2 Organization

The remainder of this thesis is organized as follows. In Chapter 2, we develop our uncertainty framework, discuss its application to three location models, and examine some theoretical relationships between the models. We provide a bound on the error introduced by the discretization of the uncertainty regions, and analyze the relationship between the granularity of the discretization and the tightness of the bound.

In Chapter 3, we present our row-and-column generation method. We also present a duality-based reformulation of the robust models, and show them to be amenable to our row-and-column generation as well. We conclude the chapter with computational results in which we benchmark our decomposition technique against alternative solution methods.

In Chapter 4, we demonstrate the applicability of our models through a computational study on ambulance positioning using historical cardiac arrest data from the City of Toronto. We solve the nominal and robust formulations of the models discussed in Chapter 2 and evaluate model performance using three different distance-based metrics.

In Chapter 5, we conclude and offer remarks on future directions for research.
Chapter 2
Modeling demand location uncertainty

In this chapter, we develop robust formulations for three location models: the p-median problem, the p-center problem, and a model we propose based on conditional value-at-risk. In Section 2.1, we introduce our uncertainty framework by generalizing the p-median problem, and show that the same framework can be applied to the other two models. In Section 2.2, we explore a theoretical relationship between the three models presented. In Section 2.3, we analyze the relationship between continuous and discrete uncertainty regions and provide a bound on the error introduced by the discretization. Lastly, in Section 2.4, we discuss an alternate interpretation of our models, motivated by the stochastic nature of demand arrivals in practical facility location problems.

2.1 Location models

The p-median problem seeks to place \( p \) facilities in a network such that the total weighted distance between all demands and their nearest facilities is minimized (Tansel et al., 1983). This classical location model has served as the foundation for many other location problems (Owen and Daskin, 1998). As a result of its significance in the facility location literature, we view it as an appropriate vehicle for introducing our framework for demand location uncertainty. We therefore focus most of this section on developing a robust analogue of the p-median problem, but will show that our approach extends to the p-center and CVaR models as well. Proofs for this section which do not appear in the body are located in the Appendix.

2.1.1 P-median problem

We begin by formulating the classical p-median problem. Let \( I \) denote a set of \( m \) candidate sites for the placement of \( P \) facilities. Let \( J \) denote a set of \( n \) demand locations, each of which has a demand weight of \( h_j \). Let the parameter \( d_{ij} \) be the distance between locations \( i \) and \( j \). Lastly, let \( y_i \) and \( z_{ij} \) be binary decision variables, where \( y_i \) is 1 if a facility is sited at location \( i \), and \( z_{ij} \) is 1 if demand \( j \) is assigned to
Chapter 2. Modeling demand location uncertainty

A facility at location $i$. The p-median problem can be formulated as the following integer program:

$$\begin{align*}
\text{minimize} & \quad \sum_{i \in I} \sum_{j \in J} h_j d_{ij} z_{ij} \\
\text{subject to} & \quad \sum_{i \in I} y_i = P, \\
& \quad \sum_{i \in I} z_{ij} = 1, \quad j \in J, \\
& \quad z_{ij} \leq y_i, \quad i \in I, j \in J, \\
& \quad y_i, z_{ij} \in \{0, 1\}, \quad i \in I, j \in J.
\end{align*}$$

(2.1a)

(2.1b)

(2.1c)

(2.1d)

(2.1e)

In the remainder of this thesis, we will refer to formulation (2.1) as the **nominal p-median model**. The nominal p-median problem can equivalently be interpreted as seeking to minimize the mean of the distribution of distances, which we will refer to as the **distance distribution**.

Suppose now that each demand is only known to lie within a continuous and bounded uncertainty region. We discretize each uncertainty region to overcome the difficulty in modeling them as continuous. In Section 2.3, we discuss in further detail how this discretization can result in an arbitrarily close approximation of the continuous uncertainty regions with minimal impact on model tractability.

We now formulate a two-stage robust optimization model as follows. Let $K_j$ represent the set of discrete points which approximate the uncertainty region for demand $j$. We let each point $k \in K_j$ represent a potential **scenario** for the realization of demand $j$. We note here that in most of the facility location literature, a “scenario” refers to a particular arrangement of demands throughout the entire network. In this thesis, we define a scenario as being local to each demand, so that the location $k \in K_j$ is a potential realization of only the demand $j$. Accordingly, we let $d_{ij}^k$ represent the distance between candidate site $i$ and demand $j$ when it is realized at the scenario location $k$.

Let $h_j$ and $y_i$ retain the same definitions as in the nominal p-median model shown in formulation (2.1). Let $z_{ij}^k$ be a binary assignment variable equal to 1 if the scenario $k$ of demand $j$ is assigned to location $i$. Lastly, let $x_{ij}^k$ be a binary decision variable equal to 1 if the demand $j$ is realized at location $k$. While from the perspective of the decision maker $x$ is an uncertain parameter, we will model it the decision variable of an adversary whose aim is to realize demands in the locations that maximize the objective. The two-stage problem can then be formulated as follows:

$$\begin{align*}
\text{min} & \quad \text{max} \quad \text{min} & \quad \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k \\
\text{subject to} & \quad \sum_{i \in I} y_i = P, \\
& \quad \sum_{i \in I} \sum_{k \in K_j} z_{ij}^k = 1, \quad j \in J, \\
& \quad z_{ij}^k \leq y_i x_{ij}^k, \quad i \in I, k \in K_j, j \in J, \\
& \quad \sum_{k \in K_j} x_{ij}^k = 1, \quad j \in J, \\
& \quad y_i, x_{ij}^k, z_{ij}^k \in \{0, 1\}, \quad i \in I, j \in J, k \in K_j.
\end{align*}$$

(2.2a)

(2.2b)

(2.2c)

(2.2d)

(2.2e)

(2.2f)

Formulation (2.2) can be understood as a two-player, three-move sequential game between an agent and
an adversary (nature). In the first move, an agent decides on a set of facility locations, based only on knowledge of the uncertainty regions of each demand and the corresponding demand weights. In the second move, the adversary realizes each demand at a location within its respective uncertainty region. In the third move, having now observed the realized demand locations, the agent then assigns the demands to facilities. The equilibrium that results from this game represents the optimal solution to (2.2). Figure 2.1 shows a simple example of one of these games.

This model belongs to a class of games known as Stackelberg games, wherein an agent must commit to a strategy prior to the adversary, allowing the adversary to observe the agent’s decisions and respond accordingly (Paruchuri et al., 2008; Brown et al., 2006, 2009). Within the optimization framework, these games typically take on a min-max-min structure, where each operator represents a move in the game.

Brown et al. (2006) simplify a min-max-min formulation by replacing the inner minimization problem with its dual equivalent, resulting in a simpler min-max optimization problem. Applying this approach to (2.2) yields a bilinear objective in the equivalent min-max problem as a result of constraint (2.2e). To avoid the bilinearity, we instead present a reformulation which allows us to obtain an equivalent min-max problem with a linear objective. This reformulated problem then lends itself to several solution approaches that we discuss in Chapter 3.

First, we modify constraint (2.2c) by removing the inner sum over all \( k \in K_j \) and enforcing the constraint over all \( k \in K_j \) and all \( j \in J \). Intuitively, this modified constraint forces every scenario of every demand to be assigned to a facility, instead of only assigning the scenarios which host a realized demand. The \( x^h_j \) term is then dropped from constraint set (2d) to maintain feasibility. To ensure that each demand-facility pairwise distance only contributes to the objective function once, we introduce \( x^h_j \) to the objective. This ensures that only the distance between the realized demand location and its
nearest facility is counted. These changes yield the following problem:

\[
\min_{y} \max_{x} \min_{z} \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k x_{ij}^k
\]  \hspace{1cm} (2.3a)

subject to

\[
\sum_{i \in I} y_i = P, \hspace{1cm} (2.3b)
\]

\[
\sum_{i \in I} z_{ij}^k = 1, \hspace{0.5cm} k \in K_j, j \in J, \hspace{1cm} (2.3c)
\]

\[
z_{ij}^k \leq y_i, \hspace{0.5cm} i \in I, k \in K_j, j \in J, \hspace{1cm} (2.3d)
\]

\[
\sum_{k \in K_j} x_{ij}^k = 1, \hspace{0.5cm} j \in J, \hspace{1cm} (2.3e)
\]

\[
y_i, x_{ij}^k, z_{ij}^k \in \{0, 1\}, \hspace{0.5cm} i \in I, k \in K_j, j \in J. \hspace{1cm} (2.3f)
\]

**Lemma 1** The optimal values of (2.2) and (2.3) are equal.

We now make two additional observations which allow us to further simplify (2.3).

**Lemma 2** For a feasible \( \hat{y} \), the inner max-min problem in (2.3) contains a saddle point.

First, we note that as a result of the reformulation, the two-move game represented by the inner max-min problem in (2.3) now contains a saddle point. This allows the order of the \( \max \) and inner \( \min \) operators to be exchanged without affecting the optima. In a game-theoretic context, this implies that agent now gains no useful information from the observation of the demand locations \( x \), since each scenario of each demand can be optimally “pre-assigned” to its nearest facility. This reduces the Stackelberg game to two-moves, as the facility placement and demand assignment decisions can now both be made in the first move.

**Lemma 3** For a feasible \( \hat{y} \), the optimal value of (2.3) remains unchanged when the integrality constraints on \( x \) and \( z \) are relaxed.

Second, we observe that the integrality constraints on \( x \) and \( z \) can both be relaxed without changing the optima. As will be shown in Section 3.2, relaxing \( x \) to be a continuous variable allows us to exploit linear programming duality to obtain an exact duality-based reformulation. This relaxation also greatly reduces the number of binary variables in the model.
These two observations allows us to obtain our final formulation for the robust p-median model:

\[
\begin{align*}
\min \quad & \max \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k x_{ij}^k z_{ij}^k \\
\text{subject to} \quad & \sum_{i \in I} y_i = P \quad (2.4b) \\
& \sum_{i \in I} x_{ij}^k = 1, \quad k \in K_j, j \in J, \quad (2.4c) \\
& z_{ij}^k \leq y_i, \quad i \in I, k \in K_j, j \in J, \quad (2.4d) \\
& \sum_{k \in K_j} x_{ij}^k = 1, \quad j \in J, \quad (2.4e) \\
& y_i \in \{0,1\}, \quad i \in I, \quad (2.4f) \\
& x_{ij}^k, z_{ij}^k \geq 0, \quad i \in I, k \in K_j, j \in J. \quad (2.4g)
\end{align*}
\]

For conciseness, in the remainder of the thesis we let

\[
X = \left\{ x_{ij}^k \geq 0 \mid \sum_{k \in K_j} x_{ij}^k = 1, \forall j \in J \right\},
\]
\[
Y = \left\{ y_i \in \{0,1\} \mid \sum_{i \in I} y_i = P \right\},
\]
\[
Z(y) = \left\{ z_{ij}^k \geq 0 \mid \sum_{i \in I} \sum_{k \in K_j} z_{ij}^k = 1, \forall j \in J; z_{ij}^k \leq y_i, \forall i \in I, k \in K_j, j \in J \right\}.
\]

Clearly, the uncertainty in demand locations also introduces uncertainty in the distance between each demand and its nearest facility. As a consequence, we no longer obtain a single distance distribution for a given set of facilities, but are instead presented with a family of distributions. Similarly, the mean (weighted) distance, which can be calculated if all demand locations are known, is now replaced with an interval for the true mean. Thus, just as the nominal p-median minimizes the mean of the distance distribution, the robust p-median model seeks to minimize the worst-case mean, i.e. the upper endpoint of the interval in which the true mean lies.

### 2.1.2 P-center problem

The classical p-center problem seeks to locate \( p \) facilities in a network such that some maximum cost over a set of demands is minimized, typically the weighted distance of the demand to its nearest facility (Tansel et al., 1983). Using the same notation and constraint set as the nominal p-median formulation (2.1), the nominal p-center problem can be formulated as follows:

\[
\begin{align*}
\min_{y,z} \quad & \max_{j \in J} \left\{ \sum_{i \in I} h_j d_{ij} z_{ij} \right\} \\
\text{subject to} \quad & \sum_{i \in I} \sum_{j \in J} h_{ij} z_{ij} \quad (2.1b) - (2.1e).
\end{align*}
\]
By introducing discrete uncertainty regions for each demand using the same framework as in formulation (2.4), we obtain the following robust optimization model:

\[
\min_{y, z, x} \max_{x} \left\{ \max_{j \in J} \left\{ \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k x_j \right\} \right\}
\]

(2.6a)

\[
x \in X, \ z \in Z(y), \ y \in Y.
\]

(2.6b)

We note that the feasible set \( X \) can be separated in \( j \) such that \( X = \prod_{j \in J} X_j \), where

\[
X_j = \left\{ x_j \ \bigg| \ \sum_{k \in K_j} x_j^k = 1 \right\}.
\]

This implies that for a feasible \( \hat{z} \),

\[
\max_{x} \left\{ \max_{j \in J} \left\{ \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k x_j \right\} \right\} = \max_{j \in J} \left\{ \max_{x} \left\{ \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k x_j \right\} \right\}
\]

Using this result and a dummy variable \( t \), we can reformulate (2.6) to obtain our final robust p-center problem:

\[
\text{minimize} \quad t
\]

(2.7a)

\[
\text{subject to} \quad t \geq \max_{x} \sum_{j \in J} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k x_j, \quad j \in J,
\]

(2.7b)

\[
x \in X, \ z \in Z(y), \ y \in Y.
\]

(2.7c)

Whereas the nominal p-center model minimizes the maximum value in the distance distribution, the robust p-center model similarly seeks to minimize the maximum value within the worst-case distance distribution.

### 2.1.3 Conditional value-at-risk

**Conditional value-at-risk (CVaR)** is a coherent risk measure with origins in the mathematical finance literature and is based on the related value-at-risk (VaR) measure (Rockafellar and Uryasev, 2000). Briefly, suppose \( f(y, \xi) \) is a loss (or cost) function that depends on a decision variable \( y \) and a random vector \( \xi \), with density \( p(\xi) \) assumed to exist. Given a risk tolerance \( 1 - \beta \) (where \( \beta \) is commonly 0.95 or 0.99), the value-at-risk is the lowest threshold \( \alpha \) such that the probability of \( f(y, \xi) \) exceeding \( \alpha \) is exactly \( 1 - \beta \). Then, \( \beta\text{-CVaR} \) is the expected loss conditional on the loss exceeding \( \alpha \). In other words, it is the average value of the outcomes at or above the \( \beta \)th percentile of the distribution. CVaR’s popularity in the optimization literature stems from its tractability. Additionally, since by definition CVaR is always greater than VaR, minimizing CVaR offers a tractable way of limiting VaR as well. In general, CVaR can be minimized by (Rockafellar and Uryasev, 2000):

\[
\text{minimize} \quad \alpha + \frac{1}{1 - \beta} \int_{f(y, \xi) \geq \alpha} (f(y, \xi) - \alpha) p(\xi) d\xi
\]

(2.8)
Chapter 2. Modeling demand location uncertainty

The corresponding value of VaR is simply the optimal $\alpha$ in (2.8). Further, if the continuous density $p(\xi)$ is replaced with a discrete distribution, the problem reduces to a linear program. Suppose $\Xi = \{\xi_1, \xi_2, ..., \xi_{|S|}\}$ is the set of possible outcomes of a discrete random variable, which occur with probability $p(\xi_s) = p_s$. The minimization of CVaR can then be modelled as the following linear program:

$$\min_{y, \alpha} \alpha + \frac{1}{1 - \beta} \sum_{s \in S} p_s \max \left\{ f(y, \xi_s) - \alpha, 0 \right\}$$  \hspace{1cm} (2.9)

We now consider a CVaR-based location problem with deterministic demand locations. Let us use the same variable and parameter definitions and constraint set as in (2.1). Let the loss function be $f_j(z_j) = \sum_{i \in I} d_{ij} z_{ij}$, which is the distance between a demand and its assigned facility. In this model, we let each demand $j$ be analogous to an event $s$ from (2.9). Similarly, in place of the event probability $p_s$, we use the demand weight $h_j$, which can be normalized so that $\sum_{j \in J} h_j = 1$. Then, we can minimize the average distance at the tail of the distance distribution – where the tail is defined as all distances at or above $100\beta^{th}$ percentile – by solving the following nominal CVaR problem:

$$\min_{y, z} \alpha + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \left\{ f_j(z_j) - \alpha, 0 \right\}$$  \hspace{1cm} (2.10a)

subject to (2.1b) – (2.1e).

We include demand weight parameter $h_j$ to generalize the model, so that some demands can be emphasized more than others. Suppose now that the demand locations are uncertain, where $x_j$ represents the true location of demand $j$. Let the loss function for a single demand $j$ again be the distance to its assigned facility. However, the distance to its assigned facility is now also a function of $x_j$:

$$f_j(x_j, z_j) = \sum_{i \in I} \sum_{k \in K_j} d_{ijk} z_{ikj} x_{kj}$$

The uncertainty in demand locations means that CVaR cannot be directly evaluated. However, given a set of facilities and assignment decisions, we can evaluate the worst-case CVaR as follows:

$$\max_{x} \min_{\alpha} \alpha + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \left\{ f_j(x_j, z_j) - \alpha, 0 \right\}$$  \hspace{1cm} (2.11)

We can minimize the worst-case CVaR with the following model:

$$\min_{y, z} \max_{x} \min_{\alpha} \alpha + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \left\{ f_j(x_j, z_j) - \alpha, 0 \right\}$$  \hspace{1cm} (2.12a)

subject to $x \in X$, $z \in Z(y)$, $y \in Y$.  \hspace{1cm} (2.12b)

Finally, we provide a reformulation of (2.12) which simplifies it from a two-stage model to a min-max

Chapter 2. Modeling demand location uncertainty

formulation:

\[
\begin{align*}
\text{minimize} & \quad \alpha + \frac{1}{1 - \beta} \sum_{j \in J} h_j \gamma_j \\
\text{subject to} & \quad \gamma_j \geq \max_{x} f_j(x_j, z_j) - \alpha, \quad \forall j \in J, \\
x \in X, \quad z \in Z(y), \quad y \in Y.
\end{align*}
\] (2.13)

Lemma 4 The optimal values of (2.12) and (2.13) are equal.

We envision coherent risk measures such as CVaR as being useful metrics in healthcare operations, where there is an interest in hedging against risk. Additionally, these measures may be useful in guiding decision making in contexts where fairness and equity play a significant role. For example, in ambulance location problems, an expected distance minimization scheme might not be appropriate if a low-weighted area receives extremely poor service. Further, as mentioned earlier, risk-based targets are already present in emergency medicine. CVaR-based location models may therefore be useful in assisting health service planners in meeting these standards.

We note here that Chen et al. (2006) also develop a scenario-based facility location model with a CVaR objective, which they refer to as the mean-excess regret model. Their model takes a stochastic approach by assuming the probability of each potential demand realization is known. Our model differs in that our uncertainty is in the location of each individual demand, and we use no distributional information.

2.2 Relationship of p-median and p-center to CVaR

The CVaR models (both nominal and robust) can be viewed as a compromise between the p-median and p-center models, in that they focus on optimizing the tail of the distribution, as opposed to the mean or the absolute maximum. In fact, we can show that both the p-median and p-center problems are special cases of the CVaR model, depending on how \( \beta \) is chosen. This result implies that, on its own, the robust CVaR model provides a highly general framework for shaping the distance distribution, and allows the decision maker to optimize the mean, maximum, or tail average by simply choosing an appropriate value for \( \beta \).

Specifically, we show that for a \( \beta \) value of 0, the CVaR model specializes to the p-median, and for \( \beta \) values that are sufficiently close to 1, the CVaR model specializes to the p-center. As a result, by adjusting the \( \beta \) parameter from 0 to 1, we can “sweep out” a spectrum of models between the p-median and the p-center problem.

Proposition 1 Let \( Z_V \) and \( Z_M \) be the optimal values of the robust CVaR and the robust p-median models, respectively. If \( \beta = 0 \), then \( Z_V = Z_M \).

Proof: Let \((\alpha^*, x^*, y^*, z^*)\) be an optimal solution to the robust CVaR model as shown in formulation (2.12). The optimal value is then given by

\[
Z_V = \alpha^* + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \left\{ f_j(x_j^*, z_j^*) - \alpha^*, 0 \right\}
\]
We observe that with $\beta = 0$, if $f_j(x_j, z_j) - \alpha^* \geq 0$ for all $j$, then the CVaR objective reduces to the p-median objective:

$$Z_V = \alpha^* + \frac{1}{1-\beta} \sum_{j \in J} h_j \left[ f_j(x_j^*, z_j^*) - \alpha^* \right]$$

$$= \alpha^* + \sum_{j \in J} h_j f_j(x_j^*, z_j^*) - \sum_{j \in J} h_j \alpha^*$$

$$= \alpha^* \left(1 - \sum_{j \in J} h_j \right) + \sum_{j \in J} h_j f_j(x_j^*, z_j^*)$$

$$= \sum_{j \in J} h_j f_j(x_j^*, z_j^*)$$

$$= \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} h_{ij} d_{ij}^k s_{ij} k^* x_j^*$$

Formulations (2.4) and (2.12) show that the feasible regions of $y$, $x$ and $z$ are identical in both the p-median and CVaR models. Since the objective functions are also the same if $f_j(x_j^*, z_j^*) - \alpha^* \geq 0$, $\forall j \in J$, then $(x^*, y^*, z^*)$ must also be an optimal solution to the robust p-median model, which gives us $Z_V = Z_M$.

We are now required to show that $f_j(x_j^*, z_j^*) - \alpha^* \geq 0$ in all cases where $\beta = 0$. This is clearly true in the case where $\alpha^* = 0$, since $f_j(x_j^*, z_j^*)$ is always positive for all $j \in J$.

For $\alpha^* > 0$, suppose by contradiction that there exists some $j$ such that $f_j(x_j^*, z_j^*) - \alpha^* < 0$. Now suppose we replace $\alpha^*$ with $(\alpha^* - \delta) \geq 0$. The new objective value is given by:

$$Z_{V}^{new} = (\alpha^* - \delta) + \sum_{j \in J} h_j \max \left\{ f_j(x_j^*, z_j^*) - (\alpha^* - \delta), 0 \right\}$$

$$< (\alpha^* - \delta) + \sum_{j \in J} h_j \max \left\{ f_j(x_j^*, z_j^*) - \alpha^*, 0 \right\} + \sum_{j \in J} h_j \delta$$

$$< \alpha^* + \sum_{j \in J} h_j \max \left\{ f_j(x_j^*, z_j^*) - \alpha^*, 0 \right\} + \left( \sum_{j \in J} h_j - 1 \right) \delta$$

$$< Z_V + \left( \sum_{j \in J} h_j - 1 \right) \delta$$

$$< Z_V$$

But $Z_{V}^{new} < Z_V$ contradicts the optimality of $Z_V$, so $f_j(x_j^*, z_j^*) - \alpha^* \geq 0$ for all $j \in J$. ■

**Proposition 2** Let $Z_V$ and $Z_T$ be the optimal values of the robust CVaR and the robust p-center models, respectively. If $\beta$ is arbitrarily close to 1, then $Z_V = Z_T$.

**Proof:** Let $(\alpha^*, x^*, y^*, z^*)$ be an optimal solution to the robust CVaR model. Note that if $\alpha^* =
max \{f_j(x^*_j, z^*_j)\}, then the robust CVaR objective reduces to the robust p-center objective:

\[
Z_V = \alpha^* + \frac{1}{1-\beta} \sum_{j \in J} h_j \max \{f_j(x^*_j, z^*_j) - \alpha^*, 0\}
\]

\[
= \max_{j \in J} \{f_j(x^*_j, z^*_j)\} + \frac{1}{1-\beta} \sum_{j \in J} h_j \max \left\{f_j(x^*_j, z^*_j) - \max_{j \in J} \{f_j(x^*_j, z^*_j)\}, 0 \right\}
\]

\[
= \max_{j \in J} \{f_j(x^*_j, z^*_j)\}
\]

Since the feasible regions in the robust p-center and robust CVaR models are the same, this implies that (x^*, y^*, z^*) is also an optimal solution to the p-center problem, which gives us Z_V = Z_T. It remains to show that \(\alpha^* = \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\) in all cases where \(\beta\) is arbitrarily close to 1.

We now show that neither \(\alpha^* > \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\) nor \(\alpha^* < \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\) can hold at optimality. For the first inequality, we can find a \(\delta > 0\) such that \(\max_{j \in J} \{f_j(x^*_j, z^*_j) - (\alpha^* - \delta), 0\} = 0\) for all \(j \in J\). This suggests that we could obtain a new objective value:

\[
Z_V^{\text{new}} = \alpha^* - \delta + \frac{1}{1-\beta} \sum_{j \in J} h_j \max \{f_j(x^*_j, z^*_j) - (\alpha^* - \delta), 0\}
\]

\[
= Z_V - \delta < Z_V
\]

This contradicts the optimality of \(Z_V\), so \(\alpha^* = \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\) cannot hold. For the second inequality, suppose \(\alpha^* < \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\). This implies there exists some \(j \in J\) such that \(\max_{j \in J} \{f_j(x^*_j, z^*_j) - \alpha^*, 0\} > 0\). But then we obtain

\[
\alpha^* + \lim_{\beta \to 1} \frac{1}{1-\beta} \sum_{j \in J} h_j \max \{f_j(x^*_j, z^*_j) - \alpha^*, 0\} = \infty > \max_{j \in J} \{f_j(x^*_j, z^*_j)\}
\]

This implies that as \(\beta\) gets close to 1, any solution where \(\alpha^* < \max_{j \in J} \{f_j(x^*_j, z^*_j)\}\) will clearly be sub-optimal. This completes the proof.

The relationship between CVaR and the p-median and p-center models is intuitive when we recall that \(\beta\)-CVaR is defined as the mean loss of all demands at or above the \(100\beta^{th}\) percentile of the distribution. Thus when \(\beta\) is zero, the \(\beta\)-CVaR is the mean of the distribution itself. Similarly, when \(\beta\) is arbitrarily close to 1, the \(\beta\)-CVaR is the loss of the demand at the extreme upper-tail of the distance distribution, which is the demand that is furthest from any facility.

### 2.3 Bounds on the discrete approximation error

The discretization of the uncertainty regions introduces an error in the evaluation of the worst-case total distance for a set of facilities, since the possible locations where demands may be realized are constrained to a finite subset of points within the continuous uncertainty region. In this section, we present a bound on the error introduced by the discretization, which then allows us to bound the true worst-case distance to the nearest facility for a given demand. This then allows us to construct bounds on the theoretical
worst-case distance objective of the corresponding continuous problem for each of the three location models.

Consider an arbitrary demand \( j \). Suppose a set of facilities is fixed. Let \( U_j \) be the continuous uncertainty region around \( j \), and let \( Q_j = \{ q^1, q^2, ..., q^{|K_j|} \} \) be the set of points representing the discrete scenarios for the realization of demand \( j \). Let \( R \) be the set of facility locations and let \( d_j(\cdot,\cdot) \) measure the Euclidean distance between any two points in demand region \( j \). Next, we define the parameter \( \ell_j = \min_{\mathbf{p} \in U_j, \mathbf{q} \in Q_j} d_j(\mathbf{p}, \mathbf{q}^k) \), which is the maximum distance between any point \( \mathbf{p} \in U_j \) and its closest discrete scenario location \( \mathbf{q}^k \). Let \( \mathbf{q}^d \) be the location which represents the worst-case realization of demand \( j \) within the discrete uncertainty region, and let \( r_d \) be the nearest facility to \( \mathbf{q}^d \). Similarly, let \( \mathbf{p}^c \) be the location of the worst-case realization assuming a continuous uncertainty region, and let \( r_c \) the facility nearest to \( \mathbf{p}^c \).

**Lemma 5** For an arbitrary demand \( j \), \( 0 \leq d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^d, r_d) \leq \ell_j \).

**Proof:** Consider the first inequality. First, the worst-case distance assuming a continuous uncertainty region must be at least as large as the worst-case distance when the demand is constrained to be realized at one of the \( |K_j| \) scenarios. This yields \( d_j(\mathbf{q}^d, r_d) \leq d_j(\mathbf{p}^c, r_c) \) and thus \( 0 \leq d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^d, r_d) \).

For the second inequality, let \( \mathbf{q}^l \in U_j \) be the nearest discrete demand scenario to \( \mathbf{p}^c \). We first show that \( d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^l, r_f) \leq \ell_j \). The triangle inequality implies \( d_j(\mathbf{p}^c, r_f) - d_j(\mathbf{q}^l, r_f) \leq d_j(\mathbf{p}^c, \mathbf{q}^l) \). By definition of \( \ell_j \), \( d_j(\mathbf{p}^c, \mathbf{q}^l) \leq \ell_j \), so \( d_j(\mathbf{p}^c, r_f) - d_j(\mathbf{q}^l, r_f) \leq \ell_j \). Also by definition, since \( r_c \) is the closest facility to \( \mathbf{p}^c \), \( d_j(\mathbf{p}^c, r_c) \leq d_j(\mathbf{p}^c, r_f) \), which implies \( d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^l, r_f) \leq \ell_j \). Lastly, since \( \mathbf{q}^d \) is the worst-case discrete location, the distance to its nearest facility must be at least as large as \( d_j(\mathbf{q}^l, r_f) \). This gives us \( d_j(\mathbf{q}^l, r_f) \leq d_j(\mathbf{q}^d, r_d) \), which implies \( d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^d, r_d) \leq d_j(\mathbf{p}^c, r_c) - d_j(\mathbf{q}^l, r_f) \leq \ell_j \).

We note that the problem of determining \( \ell_j \) for the discrete uncertainty set \( Q_j \) is related to the largest empty circle problem (Toussaint, 1983). Given a finite set of points \( P \) in a plane, the largest empty circle problem asks how large a circle can be constructed such that its center lies within the convex hull of \( P \) and no point in \( P \) lies within the circle. Similarly, the calculation of \( \ell_j \) is equivalent to the radius of the largest circle possible such that its center lies within \( U_j \) and no point from \( Q_j \) lies within the circle’s interior. However, calculating \( \ell_j \) can be challenging, especially if each of the uncertainty regions assumes a different form. Alternatively, we make two assumptions which allow us to quantify an upper bound on \( \ell_j \).

**Assumption 1** The discrete set \( Q_j \) takes the form of a square lattice over the region \( U_j \) with a grid spacing of length \( s \).

**Assumption 2** Suppose \( S \) is a set of discrete points which creates a square lattice over the entire plane such that \( Q_j \subset S \), \( \forall j \in J \). The shape of the uncertainty region \( U_j \) is such that for any point \( \mathbf{p} \in U_j \), at least one of the four points in \( S \) which define the smallest square around \( \mathbf{p} \) is inside \( U_j \).

Assumption 1 is a modeling decision that can be satisfied easily. Assumption 2 excludes pathologically shaped uncertainty regions which make it difficult to bound the discretization error. With this, we present a bound on the error introduced by the discretization of the uncertainty regions. Let the discretization error, \( \Delta_{med} \), be defined as follows:

\[
\Delta_{med} = \frac{Z_C - Z_D}{Z_C}
\]
where $Z_D$ represents the optimal value from solving (2.4) and $Z_C$ represents the theoretical optimal value from solving the same problem assuming continuous uncertainty regions. Similarly, let $Z_N$ represent the optimal value of the associated nominal problem (where the uncertainty region is a singleton for each demand).

**Theorem 1** For any $\varepsilon > 0$, if $s \leq \frac{\sqrt{2}Z_N}{\sum_j h_j}$, then $\Delta_{med} \leq \varepsilon$.

**Proof:** Assumptions 1 and 2 imply $\ell_j \leq \sqrt{2}s$. Thus, $d_j(p^c, r_c) - d_j(q^d, r_d) \leq \sqrt{2}s$, and

$$Z_C - Z_D = \sum_{j \in J} h_j d_j(p^c, r_c) - \sum_{j \in J} h_j d_j(q^d, r_d) \leq \sqrt{2}s \sum_{j \in J} h_j$$

Dividing both sides by $Z_C$,

$$\Delta_{CD} = \frac{Z_C - Z_D}{Z_C} \leq \frac{\sqrt{2}s \sum_j h_j}{Z_C}$$

Lastly, since $Z_N \leq Z_D \leq Z_C$, we obtain

$$\Delta_{CD} \leq \frac{\sqrt{2}s \sum_j h_j}{Z_C} \leq \frac{\sqrt{2}s \sum_j h_j}{Z_D} \leq \frac{\sqrt{2}s \sum_j h_j}{Z_N} = \varepsilon. \quad \blacksquare$$

If $Z_N$ is known, then by Theorem 1 an upper bound on the discretization error can be expressed in closed form as a function of the discrete spacing $s$. As a result, this allows us to determine how finely the uncertainty regions must be discretized such that the discretization error $\Delta_{med}$ is no greater than some target value, without having to first solve the robust problem. Alternatively, since $Z_D \geq Z_N$, the robust problem itself could be solved to obtain $Z_D$, which leads to a tighter bound through the substitution of $Z_N$ with $Z_D$ in Theorem 1.

The bound can be further tightened by enforcing a slightly stronger condition on the shape of the uncertain regions. If we modify Assumption 2 so that *three* of the four points in a square lattice $S$ which define the smallest square around any point $p \in U_j$ are within $U_j$, then $\ell_j$ can be bounded by $s$ instead of $\sqrt{2}s$. For example, this stronger assumption is satisfied by circular uncertainty regions.

To examine how the bound improves as a function of $s$, we construct a test problem with 50 candidate sites, 10 equally-weighted demands and 5 facilities. Each uncertainty region is modeled as a circle with a radius of 50, and discretized with a square lattice. The problem is solved for values of $s$ ranging from 50 to 0.1, corresponding to a range of 5 to 820,000 scenarios per demand, respectively (we use an efficient solution method developed in Chapter 3 to solve all problem instances). The results are shown in Figure 2.2. The two upper bounds on $Z_C$ shown in Figure 2.2 are calculated using Theorem 1 as follows:

$$Z_{CN} = \frac{Z_D}{1 - \varepsilon_N}, \quad \text{where} \quad \varepsilon_N = \frac{\sqrt{2}s \sum_j h_j}{Z_N}$$

$$Z_{CD} = \frac{Z_D}{1 - \varepsilon_D}, \quad \text{where} \quad \varepsilon_D = \frac{\sqrt{2}s \sum_j h_j}{Z_D}$$

Figure 2.2 shows how the solution to the robust formulation (2.4) serves as a lower bound on $Z_C$. 


The results show that the error bound may be weak if the uncertainty regions are discretized with only a small number of scenarios. However, in Section 3.3 we show that the uncertainty regions can be discretized finely to obtain tight bounds on $Z_C$ without sacrificing model tractability.

We can also develop a similar bound on the discretization error for the robust p-center problem, $\Delta_{cen}$. We redefine $Z_C$ and $Z_D$ to be the optimal value of the robust p-center problem assuming continuous and discrete uncertainty regions, respectively, and $Z_N$ to be the optimal value of the nominal problem. We assume that Assumptions 1 and 2 hold here as well.

**Theorem 2** For any $\varepsilon > 0$, if $s \leq \frac{\varepsilon Z_N}{\sqrt{\max_{j \in J} \{h_j\}}}$, then $\Delta_{cen} \leq \varepsilon$.

**Proof:** For conciseness we let

$$f_j(x_j, z_j) = \sum_{i \in I} \sum_{k \in K_j} d_{ij}^k z_{ij}^k x_{ij}^k$$

Now let $(y^*, z^*, x^*)$ be the optimal solution to (2.6). By Lemma 5, an upper bound on the true worst-case distance for a demand $j$ for the given solution is

$$d_j(q^d, r_d) + \sqrt{2}s = f_j(x_j^*, z_j^*) + \sqrt{2}s.$$
The absolute gap between the optimal values of (2.6) and the corresponding continuous problem can thus be bounded as follows:

\[
Z_C - Z_D \leq \max_{j \in J} \left\{ h_j \left( f_j(x_j, z_j) + \sqrt{2}s \right) \right\} - Z_D \\
\leq \max_{j \in J} \{ h_j f_j(x_j, z_j) \} + \max_{j \in J} \{ h_j \sqrt{2}s \} - Z_D \\
= \sqrt{2}s \max_{j \in J} \{ h_j \}.
\]

Then the relative gap in the robust p-center model can be bounded by

\[
\Delta_{cen} := \frac{Z_C - Z_D}{Z_C} \leq \frac{\sqrt{2}s \max_{j \in J} \{ h_j \}}{Z_N} = \varepsilon.
\]

Lastly, we can again use Lemma 5 to provide a bound for the robust CVaR model. As before, let the optimal value of the robust CVaR model shown in (2.13) be \(Z_D\), and let the optimal value of the associated continuous uncertainty problem be \(Z_C\).

**Theorem 3** For any \(\varepsilon > 0\), if \(\sqrt{2}s \leq \frac{(1 - \beta)Z_N}{\sum_{j \in J} h_j}\), then \(\Delta_{CVaR} \leq \varepsilon\).

**Proof:** Suppose \((y^*, z^*, x^*, \alpha^*)\) is the optimal solution to (2.13). Let \(g_j(x_j, z_j)\) be the loss function for a demand \(j\) assuming a continuous uncertainty region. Lemma 5 implies \(g_j(x_j^*, z_j^*) \leq f_j(x_j^*, z_j^*) + s\). A bound on the absolute gap between \(Z_C\) and \(Z_D\) can therefore be given by:

\[
Z_C - Z_D \leq \alpha^* + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \{ g_j(x_j^*, z_j^*) - \alpha^*, 0 \} - Z_D \\
\leq \alpha^* + \frac{1}{1 - \beta} \sum_{j \in J} h_j \max \{ f_j(x_j^*, z_j^*) + \sqrt{2}s - \alpha^*, 0 \} - Z_D \\
\leq \alpha + \frac{1}{1 - \beta} \sum_{j \in J} h_j \left[ \max \{ f_j(x_j^*, z_j^*) - \alpha^*, 0 \} + \sqrt{2}s \right] - Z_D \\
= \frac{1}{1 - \beta} \sum_{j \in J} h_j \sqrt{2}s.
\]

Letting \(Z_N\) be the optimal value of the nominal problem, we can bound the discretization error by

\[
\Delta_{CVaR} := \frac{Z_C - Z_D}{Z_C} \leq \frac{1}{(1 - \beta)Z_N} \sum_{j \in J} h_j \sqrt{2}s = \varepsilon.
\]

As in the previous two models, we do not need to actually solve (2.13) in order to develop a bound on the relative error \(\Delta_{CVaR}\), since only \(Z_N\) is required.

These bounds show that while a continuous uncertainty region may be more representative of the actual uncertainty in demand locations, our discrete approach can still provide a good approximation if the discrete grid size is selected appropriately. In particular, Figure 2.2 shows the grid size necessary to reach a certain discretization error. As a result, instead of solving a robust optimization problem with continuous uncertainty regions, which can be difficult, we can solve the discretized formulations and then provide an upper bound on the true worst-case objective function, \(Z_C\).
2.4 A continuous demand interpretation

In practice, demands may not be concentrated at a finite set of nodes, but are instead likely to exist continuously in the plane. For example, the demand for an ambulance may realistically emanate from any location in an urban area. As a result, it may be beneficial to view demand as a continuous distribution rather than existing at a set of nodes. In this section we show how an alternate interpretation of our robust optimization models allows us to model continuous demand in settings where there is uncertainty in the spatial distribution.

One possible approach to modeling continuous demand is to uniformly discretize the continuous space with a very large number of discrete nodes, and to assign a demand weight to each of these nodes. However, there are limitations to this approach. First, this may lead to an intractable formulation due to the extremely large number of demand nodes that would be required to reasonably discretize the plane. Second, even in the absence of computational challenges, it may be difficult to accurately estimate an appropriate weight for each of these demand nodes. To illustrate, consider again the problem of siting emergency services facilities in an urban area. Suppose we discretize the continuous demand in space into an extremely large number of discrete points $S$, such that each point $s$ represents an individual building in the city, and the node weight $p_s$ represents the probability that a random future emergency will occur in building $s$ (alternatively, each $s$ could represent an arbitrarily sized area in the city, such as a single city block, depending on how finely we discretize the continuous demand). We can then minimize the expected distance between a future emergency and its nearest facility with a stochastic interpretation of the p-median model:

$$\min_{y, z} \sum_{i \in I} \sum_{s \in S} p_s d_{is} z_{is}$$

In practice, however, it is extremely difficult to predict the future demand for an ambulance at the individual building or even city block level, meaning any estimate of $p_s$ is likely to be highly uncertain. A more viable approach is to use historical data to forecast the total demand emanating from a geographic subregion, which would be represented as group of multiple discrete nodes. Letting $S$ again be the set of all demand points in the plane, suppose we now divide it into subregions indexed by $j$, so that $S = \bigcup_{j \in J} K_j$. We let $K_j$ be indexed by $k$, so that each index pair $(j,k)$ represents a unique demand node $s$.

Suppose now that the aggregate future demand emanating from region $j$ can be accurately predicted, $h_j$, but that the distribution of demand within region $j$ is uncertain. We can then formulate a problem which minimizes the worst-case expected distance to a facility:

$$\min_{y, z} \max_{k} \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} p_j^k d_{ij} z_{ij}^{k}$$

subject to

$$\sum_{k \in K_j} p_j^k = h_j, \quad \forall j \in J,$$  \hspace{1cm} (2.14b)

$$z \in Z(y), \quad y \in Y.$$  \hspace{1cm} (2.14c)

Constraint (2.14b) forces the total demand over subregion $j$ to be equal to $h_j$. Now, since the uncertainty
Chapter 2. Modeling demand location uncertainty

is in the distribution within each subregion \( j \), we can write

\[
\tilde{p}_j^k = h_j \tilde{x}_j^k
\]

where \( \tilde{x}_j^k \) is the uncertain parameter representing the fraction of \( h_j \) that lies at point \( k \in K_j \). This yields the following problem, which is clearly equivalent to (2.4):

\[
\begin{align*}
\min_{y,z} \max_{x} & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} h_j x_j^k d_{ij} z_{ij}^k \\
\text{subject to} & \sum_{k \in K_j} x_j^k = 1, \quad \forall j \in J, \\
& z \in Z(y), \ y \in Y.
\end{align*}
\] (2.15a, 2.15b, 2.15c)

This interpretation allows us to minimize the worst-case expected distance in problems where the demand is continuous in space but is subject to distributional uncertainty. In this context, \( x_j \) represents the unknown discrete demand distribution over the demand nodes \( K_j \) which constitute subregion \( j \). Note that since we impose no constraints on the demand distribution other than forcing it to sum to 1, the worst-case distribution over each demand region will always take the form of a delta function at the point furthest from any facility.

If this uncertainty set for the demand distribution is too conservative, we can add an additional constraint that forces the share of the probability mass to be no greater than \( \delta \) for any one location \( k \). Alternatively, if the nominal location of each demand is included as a scenario (e.g. \( k = 1 \)), then we can enforce linear constraints which limit the total number of demands that are allowed to deviate from their nominal locations.
Chapter 3

Decomposition techniques

In this chapter, we propose a row-and-column generation algorithm that can be used to solve the formulations developed in Chapter 2. Our technique relies on reducing the size of the optimization problem through the elimination of redundant variables and their associated constraints. For brevity, all solution approaches discussed in this section are in the context of the robust p-median formulation shown in (2.4). Our methodology can be extended to the robust p-center and robust CVaR formulations shown in (2.7) and (2.13), as a result of their similar structure.

In Section 3.1 we review a classic row generation algorithm used for decomposing min-max problems and compare it with our row-and-column generation algorithm. We will refer to these as the “primal methods” because the adversary problem retains its primal form in the min-max formulation that is decomposed. In Section 3.2 we consider an exact duality-based reformulation of (2.4), and then apply a similar decomposition technique based on row-and-column generation. Similarly, we refer to these approaches as the “dual methods”, since we take the dual of the adversary problem in both methods. In Section 3.3 we present results from a short study on the computational performance of these solution methods.

3.1 Primal methods

Brown et al. (2006, 2009) use a row generation algorithm for solving min-max problems, which they refer to as Benders-based decomposition. We briefly review this approach and then present our row-and-column generation algorithm.

3.1.1 Row-generation only

Let the objective function of the robust p-median model be \( f(x, z) \). Formulation (2.4) can be reformulated with the use of a dummy variable as follows:

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq f(\bar{x}, z), \quad \bar{x} \in X \tag{3.1a} \\
& \quad z \in Z(y), \quad y \in Y. \tag{3.1c}
\end{align*}
\]
Since we allow $x$ to be continuous, this problem cannot be solved as a MIP due to the infinite constraint set (3.1b). Even in the case where $x$ remains binary, this would still result in a constraint set which grows exponentially with the number of scenarios per demand (assuming all demands have the same number of scenarios).

Alternatively, (3.1) can be solved using the following row generation algorithm. First, we solve a relaxed version of (3.1) by only enforcing the constraint for the subset $\hat{X} = \{x^1\}$, where $x^1$ represents the nominal demand pattern. After obtaining the optimal solution $(y^*, z^*, t^*)$, we identify the corresponding worst-case demand locations $x^*$ by solving the inner maximization problem of (2.4). If we obtain a violating constraint such that $t^* < f(x^*, z^*)$, then the current solution $(y^*, z^*, t^*)$ cannot be feasible for all $\bar{x} \in X$, and is therefore cannot be an optimal solution to (3.1). Then, we add the constraint $t \leq f(x^*, z^*)$ to (3.1) and re-solve it to obtain a new optimal set of facility locations. When no more violating constraints can be identified, the current solution is optimal.

Let us describe the approach formally. Let $S$ be an index set for the iterations of the algorithm, and let the optimal solution of the subproblem on the $s^{th}$ iteration of the algorithm be $\bar{x}^s$. Thus $\hat{X}$ is updated with each iteration with $\hat{X} \leftarrow \hat{X} \cup \bar{x}^s$. The set $S$ also serves to index all of the “demand patterns” in $\hat{X}$.

The master and subproblem for the row generation algorithm can be written as follows:

(Master problem)

\[
\begin{align*}
\text{minimize} \quad & t \\
\text{subject to} \quad & t \geq \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} z^k_{ij} d_{ij} h_j k x_{jk}^s, \quad s \in S, \\
& z \in Z(y), \quad y \in Y.
\end{align*}
\]

(Sub-problem)

\[
\begin{align*}
\text{maximize} \quad & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K_j} z^k_{ij} d_{ij} h_j k x_{jk}^k \\
\text{subject to} \quad & x \in X.
\end{align*}
\]

An alternate stopping criterion is to terminate the algorithm once the optimal values of the master and subproblem converge to within some desired tolerance. This is the approach used in Brown et al. (2006).

As a result of the structure of $X$, subproblem (3.3) can be solved in closed form as follows. Let $I_1 \in I$ be the set where $I_1 = \{i|y_i = 1\}$. Then for each demand $j$, we set $x_{jk}^* = 1$ where

\[
k^* = \arg \max_{k \in K_j} \left\{ \min_{i \in I_1} \{d_{ij}^k h_j\} \right\}
\]

Instead of solving the subproblem as a linear program, we can use a simple sorting algorithm based on the rule above. The computational performance of the algorithm is therefore determined primarily by the solution of the master problem, because the sorting algorithm is extremely fast even for large lists.
3.1.2 Row and column generation

Our method improves upon the approach in Section 3.1.1 by eliminating variables and constraints from the master problem that we identify as redundant. We consider a constraint redundant if removing the constraint and then resolving the master problem results in no change in the optimal solution. Similarly, we consider a variable redundant if the objective function of the master problem is independent of the variable’s value. Letting each demand contain \( K \) scenarios, the total number of variables in the master problem of the row-generation approach is given by \( m(1 + nK) \). This remains unchanged with each iteration, since only new constraints are generated throughout the algorithm. However, many of these variables are redundant and also lead to redundant (non-binding) constraints in the master problem.

Consider the cut generated by the subproblem in the first iteration \((s = 1)\). This constraint is:

\[
t \geq \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} h_{ij} d_{ij} x_{ij}^{k1}
\]

This constraint can also be written as

\[
t \geq \sum_{i \in I} \sum_{j \in J} \left( d_{ij}^{1} z_{ij}^{1} x_{ij}^{1} + h_{ij} d_{ij}^{2} z_{ij}^{2} x_{ij}^{2} + \ldots + h_{ij} d_{ij}^{K} z_{ij}^{K} x_{ij}^{K1} \right)
\] (3.4)

Now consider an arbitrary demand \( j \). If there exists some \( k \in K \) such that \( x_{ij}^{k1} = 0 \), then the contribution of the term \( h_{ij} d_{ij}^{k} z_{ij}^{k} x_{ij}^{k1} \) to the right hand side of the constraint is zero. As a result, the value of the assignment variable \( z_{ij}^{k} \) in that term is always multiplied by zero and thus has no impact on the objective function.

Further, since we know \( x_{ij}^{k} \) is 1 for exactly one of the \( K \) scenarios, this implies that for each \((i, j)\), all but one of the terms on the right hand side of (3.4) are zero, meaning \( K - 1 \) decision variables within \( z \) are redundant. Thus in the first iteration of the row generation algorithm, for every demand \( j \) there are \((K - 1)m\) extraneous variables in the master problem, each of which incurs a redundant constraint as well \((z_{ij}^{k} \leq y_{i})\). In the first iteration, there are also an additional \((K - 1)n\) redundant constraints, as a result of the constraint \( \sum_{i} z_{ij}^{k} = 1 \).

From an intuitive perspective, the extraneous variables are the components of \( z \) that are used to assign the scenarios where the demand is never realized during the algorithm. The master problem in our row-and-column generation algorithm is able to remain limited in size by creating the \( z_{ij}^{k} \) assignment variables and associated \( d_{ij}^{k} \) parameters in the master problem only when they have been identified as corresponding to the worst-case demand realization (for that iteration). Since our method relies on introducing decision variables on an as-needed basis, this incorporates an aspect of column generation that is not present in the algorithm from Section 3.1.1.

Let us describe the modified master problem formally. Let \( K_{j}^{s} = \{k | x_{j}^{ks} = 1\} \), so that \( K_{j}^{s} \) represents the scenario \( k \) for demand \( j \) that is realized at iteration \( s \) of the algorithm. The master problem for the
row-column generation algorithm can then be written as

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \sum_{i \in I} \sum_{j \in J} \sum_{k \in K^s_j} h_{ij} d^{k}_{ij} z_{ij}^{k} k_{ij}, & \quad s \in S, \\
& \sum_{i \in I} z_{ij}^{k} = 1, & \quad k \in K^s_j, j \in J, s \in S, \\
& z_{ij}^{k} \leq y_{i}, & \quad i \in I, k \in K^s_j, j \in J, s \in S, \\
& y \in Y.
\end{align*}
\]

The subproblem in this algorithm is the same as in the row-generation algorithm. Now, if at the \(s^{th}\) iteration we let \(d_{ij}^s = d_{ij}^{\hat{k}}, \forall i \in I, j \in J,\) where \(\hat{k} = \{k | x_{ij}^{k} = 1\},\) the master problem can be reformulated as:

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \sum_{i \in I} \sum_{j \in J} z_{ij}^{s} h_{ij}, & \quad s \in S, \\
& \sum_{i \in I} z_{ij}^{s} = 1, & \quad j \in J, s \in S, \\
& z_{ij}^{s} \leq y_{i}, & \quad i \in I, j \in J, s \in S, \\
& y \in Y.
\end{align*}
\]

This equivalent formulation shows that the complexity of the master problem is independent of the number of scenarios enumerated for each demand.

### 3.2 Dual methods

In this section we briefly consider a duality based reformulation of (2.4) and show how a row-and-column generation algorithm can be applied to it as well using a similar argument as in Section 3.1.2.

#### 3.2.1 Equivalent dual formulation

The relaxation of the binary constraint on \(x\) transforms the inner maximization problem of (2.4) into a linear program. This allows us to exploit strong duality to obtain an equivalent problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{j \in J} w_{j} \\
\text{subject to} & \quad \sum_{i \in I} h_{ij} d^{k}_{ij} z_{ij}^{k} - w_{j} \leq 0, & \quad j \in J, k \in K_j, \\
& z \in Z(y), y \in Y, \\
& w \text{ free.}
\end{align*}
\]
This formulation can then be solved as a single-stage mixed-integer program. A drawback of this approach is that adding an additional scenario to each demand increases the number of variables and constraints by \(mn\) and \((m + 1)n\), respectively. As a result, (3.7) becomes intractable if a large number of scenarios are used to discretize the uncertainty regions. However, formulation (3.7) may be an appropriate way of solving smaller instances of (2.4).

### 3.2.2 Dual master problem with row-and-column generation

Formulation (3.7) can be decomposed and solved using a row-and-column generation technique similar to the approach discussed in Section 3.1.2. For a fixed \(\hat{y}\), it is clear that the optimal \(w^*\) in (3.7) is given by:

\[
w_j^* = \max_{k \in K_j} \left\{ \min_{i \in I} \{ h_j d_{ij}^k \} \right\}
\]

where \(\tilde{I} = \{ i \mid \hat{y}_i = 1 \}\). Thus only the constraints that correspond to the worst-case location of each demand will be binding at optimality. Similarly, the variable \(z_{ij}^k\) for the non-binding constraints will have no impact on the objective function, and can be removed. This allows us to eliminate extraneous variables and redundant constraints from the dual problem to obtain a smaller master problem. Then for a given master problem solution \((y^*, z^*)\), the subproblem from (3.3) can be solved to identify the worst-case locations, which are used to generate the relevant variables and constraints. The master problem for the duality-based decomposition is:

\[
\begin{align*}
\text{minimize} \quad & \sum_{j \in J} w_j \\
\text{subject to} \quad & \sum_{i \in I} h_j d_{ij}^s z_{ij}^s - w_j \leq 0, \quad j \in J, s \in S, \\
& z \in Z(y), \quad y \in Y, \\
& w \text{ free}.
\end{align*}
\]

Comparing the duality-based master problem in formulation (3.8) with the original master problem in (3.6), we see that the duality-based method we add \(n\) cuts at each iteration, one for each \(w_j\) (constraint set 3.8b). In the primal method however, we only add one cut on \(t\) at each iteration (constraint set 3.6b).

### 3.3 Computational performance

We use a random data set to test the computational performance of the solution techniques. We randomly generate a set of candidate sites and nominal demand locations on a plane \(x, y \in [0, 1000]\) according to a uniform distribution. We let each nominal demand location be the center of a circular uncertainty region with a radius of 50 that we discretize with a square lattice. Since we use circular uncertainty regions, the number of scenarios in each demand region can be expressed as follows:

\[
|K| = \frac{\pi}{4} \left( \frac{2R}{s} \right)^2
\]
All problem instances were solved on a computing cluster using a 2.9 GHz quad-core CPU. Models were implemented in MATLAB R2011a and solved using CPLEX 12.1 with default parameter settings. The subproblems in all three decomposition methods were solved using sorting rule described in Section 3.1. Table 3.1 show solution times in CPU seconds for each method with varying problem sizes.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Primal</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row</td>
<td>Row-Col</td>
</tr>
<tr>
<td>4</td>
<td>6.2</td>
<td>10.1</td>
</tr>
<tr>
<td>12</td>
<td>4.6</td>
<td>8.8</td>
</tr>
<tr>
<td>48</td>
<td>28.3</td>
<td>9.4</td>
</tr>
<tr>
<td>196</td>
<td>281.4</td>
<td>7.1</td>
</tr>
<tr>
<td>796</td>
<td>34.7</td>
<td>9.7</td>
</tr>
<tr>
<td>3,208</td>
<td>105,276.7</td>
<td>10.1</td>
</tr>
<tr>
<td>12,852</td>
<td>-</td>
<td>11.4</td>
</tr>
<tr>
<td>51,432</td>
<td>-</td>
<td>19.1</td>
</tr>
<tr>
<td>205,860</td>
<td>-</td>
<td>47.6</td>
</tr>
</tbody>
</table>

Table 3.1: Sensitivity of solution time to number of scenarios ($m = 50$, $n = 10$, $P = 5$)

The row-and-column generation algorithm scales well with the number of scenarios per demand. This is because the size of the master problem in the row-and-column generation algorithms grows only with the number of iterations, and not with the number of scenarios. The computational results indicate that the continuous uncertainty regions can be discretized very finely without adversely affecting the performance of the solution algorithm.

In Theorem 1, we demonstrated that the error introduced by the discretization can be made very small by selecting a small value for the discrete spacing $s$, thus enumerating a large number of scenarios for each demand region. The results shown here suggests that a bound on the true worst-case distance objective can be tractably computed.
Chapter 4

Computational study

Cardiac arrest is a leading health concern in North America, and is responsible for over 300,000 deaths each year (Nichol et al., 2008). Successful treatment of cardiac arrest is extremely time sensitive, with chances of survival decreasing by up to 10 percent with each minute of delay in treatment (Larsen et al., 1993), which comes in the form of cardiopulmonary resuscitation (CPR) and defibrillation. Currently, only 5-10 percent of out-of-hospital cardiac arrest victims survive to hospital discharge (Weisfeldt et al., 2010). In the vast majority of cases, emergency medical service (EMS) personnel are exclusively relied upon to treat cardiac arrest, making survival highly dependent on ambulance response times.

The time sensitive nature of cardiac arrest typically results in the nearest available ambulance being dispatched to the site of the emergency in an effort to minimize time to treatment. Shorter ambulance response times, which allow the patient to be treated sooner, have been shown to improve survivability from out-of-hospital cardiac arrest (Valenzuela et al., 2000; Pell et al., 2001). As a result, the strategic positioning of “standby” ambulances within an urban area can play a crucial role in strengthening the response to cardiac arrest, as well as other time sensitive emergencies. The problem of strategic ambulance positioning has received significant attention in the operations research literature over the last four decades (Brotcorne et al., 2003).

As we discussed in 1.1.3, much of the ambulance location literature has focused on realistic modeling of ambulance behaviour, by including busy probabilities and multiple vehicle types, for example. However, the modeling of ambulance demand has received relatively little attention in the literature, with most studies modeling demand as a small number of nodes with deterministic locations. To that end, our primary goal in this study is to propose a novel approach to demand modeling within ambulance location problems. We are motivated by two real world considerations.

First, as mentioned in Section 2.4, we note that an emergency might reasonably be expected to occur in any building or public location in an urban area, suggesting that the demand for ambulances is continuous in space rather than being restricted to a small set of demand nodes. Second, the exact locations of future emergencies are impossible to know in advance of positioning the ambulances. While historical call data may be useful in approximating the continuous demand, any forecast of future emergency locations from historical data alone is likely to be uncertain.

We also note that satisfying EMS response time standards (i.e. 90 percent of calls within 9 minutes) may be challenging if the positioning model does not account for uncertainty in emergency locations or the spatially continuous nature of the demand. Our robust CVaR model can assist EMS planners in
positioning ambulances in a manner that is explicitly focused on adherence to the VaR-based response
time standards. We also posit that a robust optimization approach to modeling demand can equip health
service planners with useful information by quantifying how the EMS system might perform under the
worst-case realization of demand.

4.1 Experimental design

Since our focus in this study is to model demand location uncertainty, we take a simple static location
approach to ambulance positioning. Our objective is to identify a set of fixed “standby” locations which
minimize various distance based objectives in an environment where the spatial distribution of demand
is uncertain. These standby locations may represent actual EMS stations, or simply parked vehicles
which are ready to respond to a call. Our study implicitly assumes that each of the standby locations
will always have an available ambulance, and that each emergency will be responded to by an ambulance
from the nearest standby location.

We focus on a densely populated area of Toronto that covers approximately 25 km$^2$. We partition
this area into a total of 12 demand regions guided by the census boundaries. Given that the incidence of
cardiac arrest is driven by underlying demographic factors which are relatively stable over time (Sasson
et al., 2010), we assume that historical cardiac arrest data provides a reasonably accurate estimate of
the expected share of future cardiac arrest emergencies that will occur in each of the 12 demand regions.
We let the demand weight $h_j$ represents the known probability that a random future cardiac arrest will
occur in region $j$, and we interpret $x_j$ as the unknown distribution of future cardiac arrests within region
$j$.

**Data**

Our dataset includes all cardiac arrests that occurred in the City of Toronto from January 2006 to April
2013. There were a total of 2529 cardiac arrests within the 12 demand regions during this time period.
Approximately 15% of these cardiac arrests occurred in public places, with the remaining 85% occurring
in private dwellings. For the purposes of this study, we do not differentiate between public and private
cardiac arrests, as they are both treated as high priority emergencies by EMS responders.

While cardiac arrest may not be an exact proxy for all high priority emergencies, using cardiac arrest
location data for ambulance positioning is appropriate for several reasons (Erkut et al., 2008). First,
cardiac arrest belongs to the highest priority class of emergencies (due to the high risk of a fatal outcome)
and also accounts for a large share of all high priority calls. Further, the development of EMS response
time standards have primarily been motivated by studies on cardiac arrest. Lastly, there exists a clear
consensus within the medical community that reducing time to treatment for cardiac arrest can increase
the probability of survival, indicating that guiding ambulance positioning using cardiac arrest location
data can save lives.

**Model setup**

We solved six models in our study: the nominal and robust formulations of the p-median, p-center
and CVaR location models. We set $\beta$ to 0.9 in the CVaR models. Ambulance location models typically
aggregate the demand over a given region into a single demand point. Thus for our nominal formulations,
we designated the centroid of each demand region as its nominal location (although we recognize there might exist other approaches to selecting nominal demand locations). For the robust models, a square lattice with a spacing of 200 metres was constructed over the entire study region, with each discrete point being assigned to the demand region in which it lay. All models were solved to optimality using the dualized row-and-column generation algorithm described in Section 3.2.2.

We performed a simple random sample without replacement of 20% of the 2529 cardiac arrests to obtain a “training data set” (506 data points), and designated the remaining 80% of cardiac arrests (2023 data points) as the testing data set. The training set is used to determine \( h_j \) for each demand region and is therefore used directly in the optimization. The training data set can be interpreted as representing the location of all historical cardiac arrests prior to the ambulance positioning. The testing set is used post-optimization to measure the goodness of the facility locations, and can be viewed as representing future demand locations that were unknown at the time of optimization. This allows us to measure how each of the models might behave in the face of true uncertainty regarding demand locations.

We counted the fraction of training data points which fell within each of the 12 demand regions, and used this value as \( h_j \) for the p-median and CVaR models. For the p-center model, all demands were weighted equally. This has the effect of focusing the model on minimizing the distance to the point on the map that is furthest from any ambulance location.

Figure 4.2 shows the fraction of cardiac arrests that fall within each of the 12 demand regions for both the training and testing datasets, and shows that estimating the aggregate future demand in each region using historical data can be fairly accurate.

We obtained a list of over 700 major intersections in the City of Toronto to use as a pool for selecting candidate sites for ambulance placement. We selected 80 of these locations at random for use as the candidate sites in all models, ensuring that each site fell within the study area. The same 80 sites were

Figure 4.1: 12 demand regions with nominal and discrete scenario locations
Although the critical factor in treating cardiac arrest is response time, for demonstrative purposes we use Euclidean distance as the primary parameter in each model, and implicitly assume time to be proportional to distance. We note that response time could also be used directly by estimating the travel time associated with every $d_{ij}$ parameter value in the model. We fixed $P$ to six facilities for all runs.

### 4.2 Results

For each of the six models, we produce cumulative distributions for the distance between each cardiac arrest and its nearest ambulance, using both the training and testing datasets. The cumulative distributions in Figures 4.3-4.5 show the fraction of cardiac arrests that fall within some distance of a placed ambulance for both the nominal and robust models, for every distance between 0 and 2 km.
Figure 4.3: Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance - p-median models
Figure 4.4: Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance
- CVaR models
Figure 4.5: Empirical cumulative distribution of distances between cardiac arrests and nearest ambulance - p-center models
We also use three distance metrics to evaluate the performance of each model: mean distance from a cardiac arrest location to its nearest ambulance, the mean distance to the nearest ambulance of the 10% of “worst-served” cardiac arrests, and the maximum distance between any cardiac arrest and its the nearest ambulance. These three metrics correspond to the objectives of the p-median, CVaR and p-center location models, respectively. These results are shown for both the training set and the testing set in Tables 4.1 and 4.2.

<table>
<thead>
<tr>
<th>Distance metric (m)</th>
<th>p-median</th>
<th>CVaR ($\beta = 90%$)</th>
<th>p-center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Nominal</td>
<td>Robust</td>
<td>Nominal</td>
</tr>
<tr>
<td>Mean</td>
<td>788.1</td>
<td>686.4</td>
<td>954.9</td>
</tr>
<tr>
<td>CVaR ($\beta = 90%$)</td>
<td>1523.4</td>
<td>1422.7</td>
<td>1575.2</td>
</tr>
<tr>
<td>Max</td>
<td>2823.2</td>
<td>2469.9</td>
<td>2046.0</td>
</tr>
</tbody>
</table>

Table 4.1: Performance of six location models on training dataset (506 cardiac arrests)

<table>
<thead>
<tr>
<th>Distance metric (m)</th>
<th>p-median</th>
<th>CVaR ($\beta = 90%$)</th>
<th>p-center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Nominal</td>
<td>Robust</td>
<td>Nominal</td>
</tr>
<tr>
<td>Mean</td>
<td>771.9</td>
<td>690.5</td>
<td>919.4</td>
</tr>
<tr>
<td>CVaR ($\beta = 90%$)</td>
<td>1471.1</td>
<td>1434.9</td>
<td>1572.2</td>
</tr>
<tr>
<td>Max</td>
<td>2688.1</td>
<td>2334.6</td>
<td>2046.0</td>
</tr>
</tbody>
</table>

Table 4.2: Performance of six location models on test dataset (2023 cardiac arrests)

Table 4.3 shows a ranking of the six models based on each of the three distance metrics, where a ranking of 1 is given to the model which achieves the best (lowest) distance metric. The rankings are the same for both the testing and training data set.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Mean</th>
<th>90%-CVaR</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robust p-median</td>
<td>Robust CVaR</td>
<td>Robust p-center</td>
</tr>
<tr>
<td>2</td>
<td>Nominal p-median</td>
<td>Robust p-median</td>
<td>Robust CVaR</td>
</tr>
<tr>
<td>3</td>
<td>Robust CVaR</td>
<td>Nominal p-median</td>
<td>Nominal p-center</td>
</tr>
<tr>
<td>4</td>
<td>Nominal CVaR</td>
<td>Robust p-center</td>
<td>Nominal CVaR</td>
</tr>
<tr>
<td>5</td>
<td>Nominal p-center</td>
<td>Nominal CVaR</td>
<td>Robust p-median</td>
</tr>
<tr>
<td>6</td>
<td>Robust p-center</td>
<td>Nominal p-center</td>
<td>Nominal p-median</td>
</tr>
</tbody>
</table>

Table 4.3: Ranking of six models based on three distance metrics of test dataset
Lastly, in 4.6 we show the six optimal ambulance locations as determined by each of the three robust location models.

![Map of Cardiac Arrest Locations](image)

Figure 4.6: Locations of six facilities placed by robust p-median, robust CVaR and robust p-center location models

### 4.3 Discussion

Although we used straight-line distances as a proxy for ambulance accessibility in our study, previous work in the medical literature suggests a high correlation between straight-line distance and travel time (Phibbs and Luft, 1996; Haynes et al., 2006). This is particularly true for ambulances which, in North America, are generally unimpeded by traffic congestion. We therefore posit that the improvements in straight-line distance shown in our results are likely to correspond with real improvements in EMS response time. However, it may still be beneficial to compare these results with an identical study which uses road distance or estimated travel time.

Figure 4.3 shows that the distribution of the robust p-median tends to dominate the nominal p-median model at all distances. Similarly, the robust model performs better (produces lower distance values) than the nominal model when measured by average distance, the CVaR distance or maximum distance. Figure 4.4 shows that the robust CVaR does not necessarily perform better than the nominal model, except for cardiac arrests above the 65th percentile of the distribution, beyond which the robust model visibly dominates. However, the robust CVaR model still outperforms the nominal CVaR model on all three distance metrics. Lastly, the robust p-center model outperforms the nominal model beyond the 85th percentile, but significantly underperforms at lower distances. This is an intuitive result given that our implementation of the robust p-center model focuses improving service for the single point in
the entire study region that is furthest from any facility. This also results in the robust p-center model producing a higher mean distance that the nominal formulation.

Unsurprisingly, the model that achieves the best performance when measured by the mean, CVaR and maximum distance metrics are the robust p-median, robust CVaR and robust p-center models, respectively. Despite performing the best when measured by mean distance, the two p-median models rank last when measured by the maximum distance metric. This is an expected outcome, given that the goal of the p-median models is to minimize expected distance, whereas the CVaR and p-center models focus on improving performance at the tail of the distribution.

The rankings of the six models shown in 4.3 reiterate that the robust models all outperform their corresponding nominal models on each of their respective distance metrics. Further, comparing robust and nominal formulations for all three models across each of the three distance metrics provides a total of nine possible “robust vs nominal” comparisons. In eight of these, the robust model achieves a higher ranking than its nominal counterpart. Only the nominal p-center performs better than the robust formulation on the mean distance metric, suggesting the robust p-center might be overly conservative. From the rankings, we also observe that the robust CVaR model is in the top 3 for each of the distance metrics. This suggests that minimizing the tail-average serves as a good compromise between minimizing the average of the entire distribution and minimizing the maximum value. As a consequence, if there is uncertainty regarding which of the the three metric is most important is when optimizing the facility locations, the robust CVaR might be an appropriate choice as a result of its strong performance in all cases.

With respect to the facility locations themselves, Figure 4.6 shows that the facility that is furthest North, South, East or West within the study region are all placed by the robust p-center model (although the p-median model also places a facility at the same site as the northern most p-center facility). This reflects the fact that the p-center model focuses on minimizing distance for the absolute “worst-served” point on the map, which causes the optimal locations to be relatively more spread out. By contrast, the robust p-median model places two facilities in the region with the highest number of cardiac arrests, whereas the other two models place no facilities in this region.

While the exact risk of cardiac arrest throughout the study region is unknown, the empirical distributions can provide an estimate of the probability that a random cardiac arrest will occur within $x$ metres of one of the ambulance locations. The similarities between the training and testing distributions suggest that, in aggregate, historical data can provide a reasonably accurate estimate of the future distribution of cardiac arrest locations.
Chapter 5

Conclusion

In this thesis, we presented a two-stage robust optimization framework for modeling demand location uncertainty in multi-facility location problems. We have shown our methods to be generalizable by applying it to the classical p-median and p-center problems — two of the most studied models in the facility location literature. We also presented a novel location model based on conditional value-at-risk (CVaR), which we also modify to account for demand location uncertainty. We also discussed how the p-median and p-center models can be viewed as special cases of the CVaR model, which shows that the robust CVaR model itself represents a family of models that works within our uncertainty framework. This grants a decision maker the flexibility to tailor the optimization to different parts of the distance distribution by selecting appropriate parameter values in the CVaR formulation. While we only consider ambulance positioning in this thesis, modeling demand location uncertainty is likely to be beneficial in other real world problems, such as the location of fire stations, public access defibrillators or treatment centers for medical evacuation.

To solve our robust optimization models, we proposed a decomposition algorithm based on row-and-column generation. Notably, the efficiency of algorithm allows us to closely approximate continuous uncertainty regions by using a scenario-based approach to modeling the uncertainty. Further, our discretization approach is general and allows uncertainty regions to take on any shape.

In our final chapter, we performed a computational study on strategic ambulance positioning. Our study was motivated by the uncertainty in the location of future emergencies when ambulance positioning decisions are made. Additionally, our data shows that it may be more accurate to model ambulance demand as continuous in space rather than concentrated at a finite number of demand nodes. Our results suggest that our approach may lead to improvements in response time, particularly for those emergencies which occur the furthest from any ambulance location. Given the well established link between ambulance response time and the chances of survival from cardiac arrest, hedging against uncertainty in this instance may have a positive and tangible impact on health and human life.

With respect to future work, it may be beneficial to incorporate some measure of control over the conservatism of our robust optimization formulations. The models developed in this thesis assume that demands will always be realized in the locations that are furthest from existing facilities. While this allows us to hedge against worst-case realizations of demand, it may cause the system to perform poorly for typical cases. Allowing for some control over the conservatism of the uncertainty regions may further improve our robust optimization models. This may be done by enforcing additional constraints which
limit the demand distribution from taking on overly conservative forms.

Additionally, it may be valuable to examine whether our uncertainty framework can be incorporated into other location models. Although we focused primarily on classical formulations, it is conceivable that our approach might also extend to location models which consider facility reliability or capacity. This would further broaden the scope of this work.

Lastly, given the focus on modeling demand, our computational study takes a fairly straightforward approach to modeling ambulance behaviour. The existing ambulance literature takes a more nuanced approach to modeling the ambulances, by incorporating uncertain travel times, ambulance busy probabilities, multiple vehicle types and dynamic relocation into the models. However, it is our hope that this study serves as a reminder of why the demands themselves should also be modelled in a realistic manner. To the best of our knowledge, our incorporation of uncertainty in demand locations is novel within the general facility location literature, and therefore is also novel among existing ambulance location models. A logical next step would be to combine advanced modeling of ambulance behaviour with our model for demand location uncertainty. In our view, improving the response to cardiac arrest through ambulance location optimization requires that we continue to account for uncertainty in the real world. This includes uncertainty in the behaviour of the facilities as well as the demands themselves.
Bibliography


y G Sopko, J Powell, G Nichole, and LJ Morrison. Survival after 289 application of automatic
defibrillators before arrival of the emergency medical system: 290 evaluation in the resusci-
tation outcomes consortium population of 21 million. Journal of the American College of Cardiology,
Appendix A

Proofs

Lemma 1 The optimal values of (2.2) and (2.3) are equal.

Proof: The feasible sets $Y$ and $X$ are the same in both formulations, so we only require that for a feasible $(\hat{y}, \hat{x})$

$$\min_{x \in Z_2} \sum_{i \in I} \sum_{k \in K_j} \sum_{j \in J} h_j d_{ij}^k z_{ij}^k = \min_{x \in Z_3} \sum_{i \in I} \sum_{k \in K_j} \sum_{j \in J} h_j d_{ij}^k z_{ij}^k$$

where $Z_2$ and $Z_3$ are the feasible regions for $z$ in (2.2) and (2.3) given $\hat{y}$, respectively. Since $Z_2$ is separable in $j$, we have

$$\min_{x \in Z_2} \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k = \sum_{j \in J} \min_{x \in Z_2, i \in I, k \in K_j} h_j d_{ij}^k z_{ij}^k$$

The same holds for $Z_3$. Now we show that for any $j$

$$\min_{x \in Z_1} \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k = \min_{x \in Z_1, i \in I, k \in K_j} h_j d_{ij}^k z_{ij}^k$$

Let $\hat{I} = \{i \mid \hat{y}_i = 1\}$ and $\hat{K}_j = \{k \mid \hat{x}_j^k = 1\}$. First, we consider (2.2). Since $\sum_{i \in I} \sum_{k \in K_j} z_{ij}^k = 1, \forall j \in J$ and $x_j \in \{0, 1\}$, we have

$$\min_{x \in Z_2} \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k = \min_{i \in I, k \in K_j} \left\{ h_j d_{ij}^k \right\}$$

Similarly, for (2.3), since $\sum_{i \in I} z_{ij}^k = 1$, and $\sum_{k \in K_j} x_j^k = 1$, and $x_j \in \{0, 1\}$, we obtain

$$\min_{z \in Z_1} \sum_{i \in I} \sum_{k \in K_j} h_j d_{ij}^k z_{ij}^k = \min_{i \in I, k \in K_j} \left\{ h_j d_{ij}^k \right\} \quad \blacksquare$$

Lemma 2 For a feasible $\hat{y}$, the inner max-min problem in (2.3) contains a saddle point.

Proof: We are required to show that $f(x, z^*) \leq f(x^*, z^*) \leq f(x^*, z)$. Since $f(x, z)$ is concave (linear) in $x$, we have $f(x, z^*) \leq \max_{x \in X} f(x, z^*) = f(x^*, z^*)$, which proves the first inequality. Similarly, since $f(x, z)$ is convex (linear) in $z$, we have $\min_{z \in Z} f(x^*, z) = f(x^*, z^*) \leq f(x^*, z)$, proving the second inequality. $\blacksquare$
**Lemma 3** For a fixed $\hat{y}$, the optimal value of (2.3) remains unchanged when the integrality constraints on $x$ and $z$ are relaxed.

**Proof:** We are required to show that there exists an optimal solution in the relaxed formulation where $x^*$ and $z^*$ are binary. Due to separability, we can consider an arbitrary demand $j$ in both cases. For $x$, we note that relaxation of the binary constraint yields $X_j = \{ x_j^k | \sum_k x_j^k = 1, x_j^k \geq 0 \}$ where $X = \prod_j X_j$.

Since, for an arbitrary $j$, $X_j$ defines the $|K| - 1$ dimensional simplex, all corner points of $X_j$ are binary, and so any optimal $x^*$ must be binary.

Now we consider the relaxation on $z$. With the binary constraint relaxed and replaced with non-negativity constraints, $Z_j$ is given by $\{ z_{ij}^k | \sum_i z_{ij}^k = 1, \forall k \in K_j, z_{ij}^k \leq y_i, \forall i,j,k \}$. For there to exist an optimal $z$ that is non-integer, one of the binding constraint on $z$ must have a non-integer right hand side. However, all constraints on $z$ have a RHS value of 0 or 1, meaning $z$ must be binary at every corner point of $Z$.

**Lemma 4** The optimal values of (2.12) and (2.13) are equal.

**Proof:** Let $g(\alpha, x, z) = \alpha + \frac{1}{1-\beta} \sum_{j \in J} h_j \max \{ f_j(x_j, z_j) - \alpha, 0 \}$. Let $x_j = \arg\max_{x_j} f_j(x_j, z_j)$, and $x = \prod_{j \in J} x_j$. The feasible sets for $Y$ and $Z(y)$ are the same in both formulations, so we are required to show that for a feasible $z$,

$$\max_x \min_\alpha g(\alpha, x, z) = \min_\alpha g(\alpha, x, z)$$

Since the feasible set for $x$ is the same in (2.12) and (2.13), we obtain the following inequality

$$\max_x \min_\alpha g(\alpha, x, z) \geq \min_\alpha g(\alpha, x, z)$$

Now we consider the reverse direction. Since for any solution $(\hat{\alpha}, \hat{x})$, we obtain $f(\hat{x}_j, z_j) \leq f(\bar{x}_j, z_j)$, this implies

$$\max \{ f_j(\hat{x}_j, z_j) - \hat{\alpha}, 0 \} \leq \max \{ f_j(\bar{x}_j, z_j) - \hat{\alpha}, 0 \}$$

Which gives us $g(\hat{\alpha}, \bar{x}, z) \leq g(\hat{\alpha}, \hat{x}, z)$ for any $(\hat{\alpha}, \hat{x})$. Now, let $(\alpha_1, x_1)$ and $(\alpha_2, x_2)$ be the optimal solutions to (2.12) and (2.13), respectively. Since $x_2 = \bar{x}$, we finally obtain

$$\max_x \min_\alpha g(\alpha, x, z) = g(\alpha_1, x_1, z) \leq g(\alpha_2, x_1, z) \leq g(\alpha_2, \bar{x}, z) = \min_\alpha g(\alpha, \bar{x}, z).$$