ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS

Light robust location – routing

Leon van der Knaap
386320

Supervisor
dr. K. S. Postek

Co-reader
dr. S. Sharif Azadeh

Econometrics and Management Science
Operations Research and Quantitative Logistics
Erasmus School of Economics

October 28, 2018

Abstract

In this thesis we will address the location routing problem (LRP) with uncertainty in customer demand. The objective of the LRP is to minimize the sum of depot cost, vehicle cost and routing costs. The LRP is an integrated version of the facility location problem (FLP) and the vehicle routing problem (VRP), where the decisions on the opening of facilities and determining optimal routes are considered simultaneously. We will present a two-stage light robust problem formulation, where in the first stage the decisions regarding the locations of the open depots, the number of assigned vehicles and the *a priori* routing are considered. After observing the actual demand values, the *a posteriori* routing can be revised. We will show that we can provide non-conservative robust solutions using the light robust approach. We also introduce a set of valid inequalities specifically designed to tighten the light robust LRP formulation and evaluate the computational performance on benchmark instances. In addition, we consider a similar LRP with uncertainty in the customer set.

Contents

1 Introduction

The classical and widely studied vehicle routing problem (VRP) considers designing optimal vehicle routes along multiple customers. The main characteristics of this problem include that every customer can only be served by a single vehicle. However, a vehicle can serve multiple customers on any given route. Also, every vehicle must start and end their route at the same depot. The objective of the problem is to minimize the sum of the routing costs and the fixed vehicle deployment cost.

The location-routing problem (LRP) is derived from this problem and differs in the sense that the depot locations are not fixed in advance. Given a set of potential depot locations, a subset of locations needs to be opened in order to start the routes. This problem is similar to the commonly known facility location problem (FLP), which aims to open facilities in order to minimize the opening cost as well as the distances from customers to their nearest open facility. The LRP considers the decisions of opening facilities and determining optimal routes simultaneously. Hence, it can be seen as a combination and a generalization of both the VRP and the FLP. If the depot locations are fixed, the LRP becomes a VRP. Otherwise, if we require all routes to contain exactly one customer, the LRP reduces to a capacitated FLP. Since both the VRP and the capacitated FLP are NP-hard problems, the LRP is an NP-hard problem as well. Prior to the research on LRP, facility location and vehicle routing problems have generally been solved separately in a sequential manner. Here, the locations of open facilities are chosen first without taking into account the routing options, after which the optimal routes are determined. This approach ignores the interrelation between location and routing, which may result in non-optimal solutions. In addition, many different side constraints for all given problems have been studied, of which the most widely used deal with capacitated facilities and vehicles.

In this thesis, we study a robust approach to the stated LRP, where we consider uncertainty in customer demand. Similar to stochastic optimization (see e.g. Birge and Louveaux (2011)), robust optimization (see e.g. Ben-Tal and Nemirovski (1999)) deals with models in which a part of the information is unknown. In the field of robust optimization, the distribution of the unknown values is assumed to be unknown and only a bounded set of possible realizations is taken into account. The aim is to optimize the solution to the uncertain problem assuming the realizations attain a worst-case scenario. We introduce a two-stage method, where the actual demand is observed in the second stage. Before that, only a nominal demand value is known, alongside with a symmetric uncertainty interval around this nominal value. In the first stage, the decisions of opening facilities, deciding on the number of vehicles and determining optimal routes corresponding to the demand realizations are considered simultaneously. The locations of the open depots and the number of assigned vehicles become fixed in the second stage. Based on the known values of demand, the actual routes to perform can be revised. This also means that customers are not bound to a fixed depot nor a vehicle. Throughout this thesis, we will refer to the initial routes as *a priori* routes and the final executed routes as *a posteriori* routes.

A popular method within the field of robust optimization is a strict robust approach (Liebchen et al. (2009)). This approach is also known as classic robust optimization and min-max optimization. A solution is called *strictly robust* if it is feasible for all possible scenarios that compose the uncertainty set. When using a box uncertainty set, this becomes equivalent to solving the initial problem with all uncertain parameters simultaneously attaining their worst-case values. While being reasonably easy to solve, this approach is known to result in conservative solutions, yielding a large cost in order to be feasible for all possible scenarios. Since it is very unlikely that all uncertain parameters actually attain their worst-case value simultaneously, this solution might be too conservative. According to the central limit theorem, this effect becomes larger for multiple uncertain parameters (Rosenblatt (1956)). In order to avoid such conservative solutions, different types of uncertainty sets can be used by shrinking the original uncertainty set. A popular method is to use a budgeted uncertainty set (Bertsimas and Sim (2004)). This method defines a budget denoted by Γ and only takes into account the scenarios in which at most Γ uncertain parameters attain their worst-case value, per constraint. Therefore, the optimal solution might not be feasible in the case all uncertain parameters attain their worst-case values, but will hold for the more likely scenarios. Besides a budgeted uncertainty set, different alternative robust optimization concepts have been introduced (Goerigk and Schöbel (2016)), including the modeling framework of *light robustness*, introduced by Fischetti and Monaci (2009). Light robustness allows for a certain total cost and maximizes the robustness of the optimal solution, bounded by the given cost. By only allowing for a certain maximum cost for the optimal solution, the control parameter as input will be a concrete cost value. We can argue that this is more convenient from a managerial standpoint than the alternative of controlling an indistinct value of robustness, e.g. as in the budgeted uncertainty set. We will provide a more detailed explanation of the light robustness framework in the following sections.

The contribution of this thesis is threefold:

- we solve a two-index location-routing problem containing uncertainty in customer demand using light robust optimization. This is computationally easier than solving the problem using a budgeted uncertainty approach, which requires a three-index formulation to consider the sum of the uncertain parameters.

- we present a set of valid inequalities designed with the specific purpose of solving the light robust location-routing problem under demand uncertainty faster.
- we reformulate the light robust problem formulation at hand to account for the situation of uncertainty in the customer set

The outline of this thesis is as follows. In Section 2, we will present a literature review on the subjects of LRP and robust optimization. Section 3 contains an elaboration on the problem definition and the possible problem formulations. Then in Section 4 we introduce the setup of the experiments of which several numerical results are presented in Section 5. A conclusion of this thesis and suggested directions of future research are discussed in Section 6

Introductory example

To motivate the problem, we start with a small example. Consider a graph consisting of three customers $\{a, b, c\}$ on the vertices of an equilateral triangle in a Euclidean plane and two potential depot locations $\{A, B\}$, located on the geometric center and on the midpoint between a vertex and the center, see Figure 1. Furthermore, assume the following information:

- The demand of a customer is uncertain, but is known to belong to the set $U = \{d \in \mathbb{R}^3_+ \mid 0 \leq d_a, d_b, d_c \leq \frac{1}{2}Q\}$, where Q denotes the capacity of a vehicle. Hence, with uncertain demand located in uncertainty set U you need to deploy at least two vehicles to satisfy all demand scenarios.

- The cost of opening a depot and of assigning a vehicle is very large in relation to the routing costs. This assumption assures that there will be no more assigned vehicles and open depots in the optimal solution than minimally required.

- The routing costs corresponds to the Euclidean distances between the nodes for a triangle with sides of length 1:
  - $c(a, b) = c(a, c) = c(b, c) = 1$
  - $c(A, a) = c(A, b) = c(A, c) = 0.577$
  - $c(B, a) = 0.289$
  - $c(B, b) = c(B, c) = 0.736$

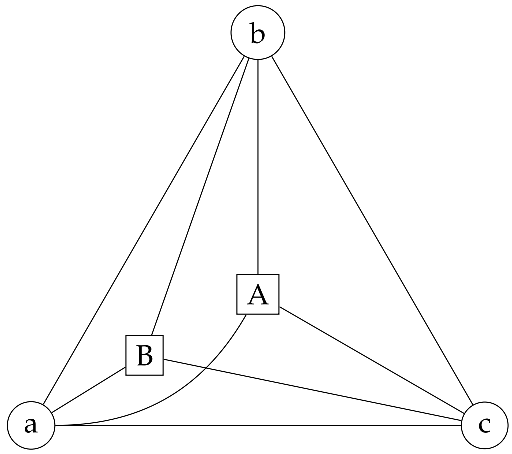

Figure 1: Topology design

We consider the following three different methods that all consider the worst-case scenarios for demand:

Heuristic approach The classical heuristic approach is a sequential method that first solves the FLP to optimality and later, when the customer demand becomes known, considers the VRP. The location of the open depot is based on minimzing the sum of distances to the customers. Hence, it ignores the routing costs. A VRP is solved when the actual demand values are observed.

Static approach The static approach is a single-stage robust LRP that determines all decisions simultaneously (i.e. the depots to open, the number of vehicles to assign and the final routing). Due to the static nature, no decisions can be revised once actual demand values are observed.

Two-stage approach The two-stage robust approach first considers the opening of depots, the number of vehicles and the a priori routes. Therefore, the first stage is equal to the approach described above. However, the final routing may be changed after the actual demand values are observed.

A property of an equilateral triangle is that depot A, located on the triangle's center, is the point that minimizes the sum of the distances to the vertices. Hence, this is the optimal place to open a depot according to the heuristic approach. The optimal vehicle routes from depot A have a total cost of 3.308. On the other hand, the optimal routes starting from depot B have a total cost of 3.05. The latter routes are optimal for all scenarios for which $d_a + d_b + d_c > Q$, which requires two vehicles. Due to the dynamic nature of the heuristic approach and the two-stage LRP, the final routes can be revised if this reduces the routing costs. In case the observed demands satisfy $d_a + d_b + d_c \leq Q$, a single route is sufficient and, according to the triangle inequality assumption, reduces the routing cost. Note that these revised routes are bounded by the decisions made in the first stage. The optimal solutions for the three approaches and the two different demand scenarios are visually shown in the Figures 2 and 3.

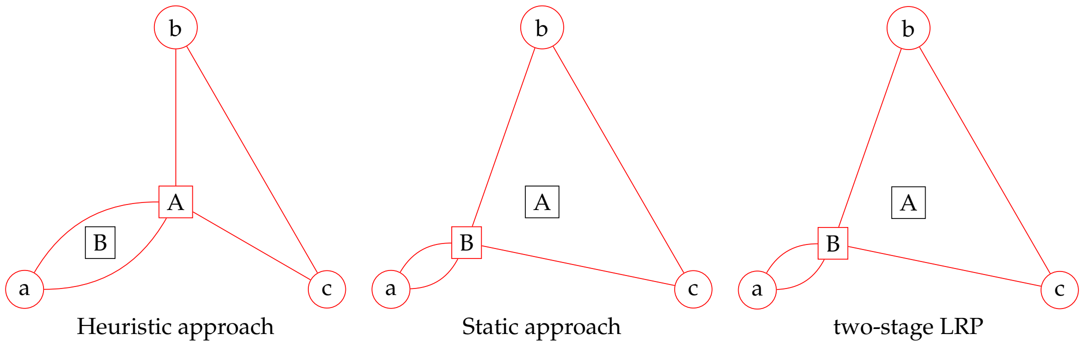

Figure 2: Optimal solutions for scenarios with observed demand $d_a + d_b + d_c > Q$.

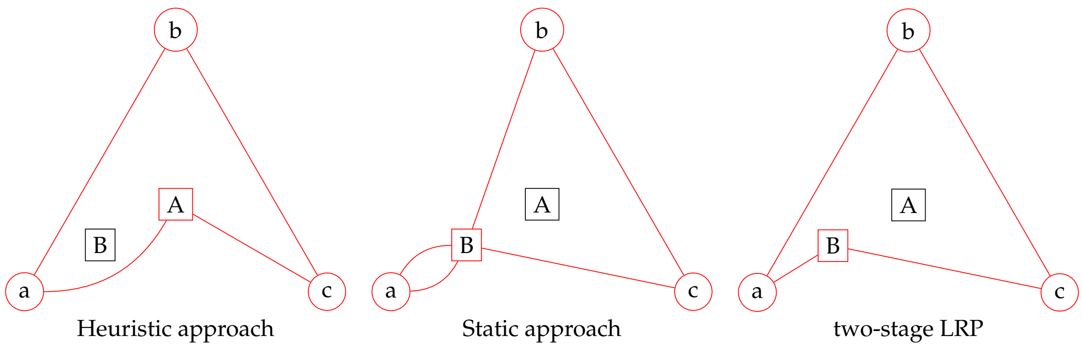

Figure 3: Optimal solutions for scenarios with observed demand $d_a + d_b + d_c \leq Q$.

Given the assumption of large cost for opening locations and assigning vehicles, all approaches will open a single depot with two vehicles assigned to it. This is sufficient to satisfy all possible scenarios. The location of the open depot and the determined routes differ. The total routing costs per solution, according to the Euclidean distances, are given in Table 1.

Table 1: Routing costs for different approaches and demand scenarios

Total demand	Heuristic approach	Static approach	Two-stage LRP
$> Q$	3.308	3.05	3.05
$\leq Q$	3.154	3.05	3.025

We conclude that the two-stage LRP outperforms the heuristic approach due to a better decision on the location of the open depot. The heuristic approach opens a different depot which results in larger routing costs due to the opportunity cost of ignoring vehicles routing costs while locating depots. Note that the difference in routing costs would be even larger if the location of depot B coincides with the location of one of the customers (Salhi and Rand (1989)). Furthermore, the two-stage LRP outperforms the static approach if the routing can be revised and optimized after observing the actual demand values. Since the two-stage LRP takes multiple demand scenarios into account, it is more difficult to formulate and to solve computationally. We will elaborate on this in Section 3.

2 Literature review

In this section we provide a literature review of the subjects of location-routing problems and (light) robust optimization.

2.1 Location-routing problem

The idea of combining location and routing problems dates from the late 1960s. However, the first paper to be credited for showing that an LRP can significantly outperform the discussed sequential heuristic approach was Salhi and Rand (1989). This study has instigated the research of integrating depot location and vehicle routing problems. Most of the early works consider a simplified problem with uncapacitated vehicles, as described in Min et al. (1998), for which a single vehicle per depot is sufficient if the cost obey the triangle inequality. Since then, similar to the research on VRP, many different problem variants and solution methods have been introduced. For elaborate surveys on the progress of LRPs, we refer to Nagy and Salhi (2007), Prodhon and Prins (2014) and Drexl and Schneider (2015).

Several different problem formulations have been introduced for the LRP. A three-index flow formulation has been introduced by Perl and Daskin (1985). Compared to a two-index formulation, the additional index denotes a specific vehicle. This notation allows for more general problem formulations including heterogeneous vehicles. Laporte et al. (1986) was the first to formulate a two-index formulation for the LRP with uncapacitated facilities. Karaoglan et al. (2011) have come up with a two-index flow formulation by introducing a set of constraints that prohibits illegal routes. This problem extends the original LRP by including simultaneous pickup and deliveries. They also consider several polynomial-size valid inequalities to strengthen the formulation. This formulation has been extended by Koç et al. (2016) to allow for heterogeneous fleet and time windows. For a computational comparison of different LRP formulations, including two- and three-index node-based and flow-based formulations, we refer to Contardo et al. (2013).

2.2 Robust optimization

Although the field of robust optimization stems back to Soyster (1973), it has experienced an increase in popularity since the late 90s. A series of papers (Ben-Tal and Nemirovski (1998), Ben-Tal and Nemirovski (1999), Ben-Tal and Nemirovski (2000)) provide a strong framework for strict robust optimization. Strict robustness using certain uncertainty sets has disadvantages in many situations due to the conservativeness of the solution. This is due to the very unlikely event that all uncertain parameters will attain their worst-case scenario simultaneously. Bertsimas and Sim (2004) introduce the budgeted uncertainty approach to address this issue. This approach shrinks the uncertainty set by limiting the maximum total deviation of uncertainties from their nominal values. This widely used concept has been generalized for multiple types of robust optimization in Bertsimas and Thiele (2006). For a survey about the different types of robust optimization and the development of the subject, we refer to Goerigk and Schöbel (2016).

2.2.1 Adjustable robustness

Ben-Tal et al. (2004) show that two-stage robust linear programming is computationally intractable and propose adjustable robust programming as an alternative approach. Specifically for two-stage problems, and allowing to extend for multiple stages (Bertsimas and Caramanis (2010)), the approach of adjustable robustness (Ben-Tal et al. (2004)) distinguishes variables that have to be decided on before and after some information becomes to be known. These are the so called *here-and-now variables* and *wait-and-see variables*. The first of these types of variables has to be determined when there still are uncertainty parameters, while the latter can be decided when the actual scenario becomes known. Adjustable robustness is a convenient tool to solve integrated problems. For example, Zeng and Zhao (2013) apply two-stage adjustable robustness to solve a location-transportation problem. A recent survey regarding adjustable robustness can be found in Yanıkoğlu et al. (2017).

2.2.2 Light robust optimization

Fischetti and Monaci (2009) introduce light robustness as a modeling framework to substitute the use of a budget of uncertainty. Instead of minimizing the objective cost, while satisfying a certain level of robustness, the aim is to maximize the level of robustness while satisfying a constraint on the allowed cost. The level of robustness is measured by a weighted sum of slack variables in the robust constraints. In this way, some of the constraints are allowed to be violated. The cost of a solution to the light robust problem is bounded by either: a) an absolute deviation in cost from the objective value of the nominal problem or b) a relative deviation in cost from the objective values of both the nominal problem and the strict robust problem. The optimal solution to the light robust problem has to satisfy the nominal problem and minimizes the weighted slack variables that allow to satisfy the strict robust problem. Note that, since the cost boundary is equal or larger than the obtained cost from the optimal nominal problem solution, the optimal solution to the nominal problem is always a feasible solution to the light robust problem. Fischetti and Monaci (2009) show for several problems that the problem at hand can relatively easy be reformulated as a light robust problem and that it will yield similar results as the budget of uncertainty.

The formulation of light robustness from Fischetti and Monaci (2009) has been generalized for multiple optimization problems and uncertainty sets in Schöbel (2014). The concept of light robustness is currently a popular tool in the field of timetable scheduling (e.g. Fischetti et al. (2009) and Goerigk et al. (2013)). It has also been shown to work well in several other problems, such as network slicing (Baumgartner et al. (2017)) and bi-objective robust problems, see Carrizosa et al. (2017) and Kuhn et al. (2016).

2.3 Uncertain routing problems

Laporte et al. (1989) was the first to introduce uncertainty in the LRP. They considered a chance-constrained stochastic two-stage problem containing uncertainty in customer demand. They required customers to be assigned to a specific vehicle and route. In case the vehicle capacity is exceeded by the total demand of the customers on the route, the vehicle needs to return to the depot in order to reload the capacity to continue the route. The objective is to minimize the cost of opening depots and performing the a priori routes. The solution is subject to either a limit on the probability of routes to fail or a limit on the expected penalty of failed routes. A similar formulation is applied by Albareda-Sambola et al. (2007), with the alternative objective of minimizing the cost of opening depots, expected routing costs and expected penalty cost. Penalty costs are incurred in case the a posteriori route needs to omit customers if the total demand on route exceeds the vehicle capacity. Furthermore, the authors only allow for one vehicle per depot. In addition, uncertainty in the customer set is included using independent Bernoulli distributions.

Research in LRP is limited in relation to the number of publications regarding location or routing problems. However, many problem variations and solution procedures can be generalized to the integrated LRP. Although we are not aware of any existing research on the subject of robust location-routing, several different robust routing problems have been presented in the field of VRP. A strict robust formulation with different types of demand uncertainty sets has been introduced by Sungur et al. (2008). Gounaris et al. (2013) derive a robust counterpart for several formulations. Using a single-stage problem the objective is to find optimal routes feasible for all scenarios of customer demand. Cao et al. (2014) study a robust open vehicle routing problem that allows for unsatisfied demand at a penalty cost. The open VRP deviates from the general problem by not restricting vehicles to end the route at the starting depot. They introduce different strategies such as minimizing total cost or minimizing unmet demand.

3 Model formulations

In this section we start with an introduction to the different problem formulations. Next, we will define the LRP and introduce the notation. We then present the mixed-integer programming (MIP) formulations to solve the LRP under demand uncertainty. These include the uncertain problem, the nominal and strict robust problem and the light robust problem formulations. In addition, we present a set of valid inequalities to strengthen the formulations. Finally, we also consider a variation of the LRP with deterministic demand for every customer, but with uncertainty in the customer set.

3.1 Introduction to strict and light robustness

Consider the following standard linear programming problem.

$$
\begin{aligned}
\max_{x} \quad & \sum_{j=1}^{n} c_j x_j, \\
\text{s.t} \quad & \sum_{j=1}^{n} a_{ij} x_j \le b_i \quad \forall i = 1, \dots, m, \\
& x_j \ge 0 \quad \forall j = 1, \dots, n.
\end{aligned}
$$

Suppose A is actually uncertain and lies in the uncertainty set $\mathcal{U} = \mathcal{U}_{11} \times \ldots \times \mathcal{U}_{mn}$, with $\mathcal{U}_{ij} = [\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$. In the nominal problem formulation, (Nom), we replace uncertain parameters a_{ij} by the known nominal values $\bar{a}_{ij}$.

$$
\begin{aligned}
\text{(Nom)} \quad \max_{x} \quad & \sum_{j=1}^{n} c_j x_j, \\
\text{s.t} \quad & \sum_{j=1}^{n} \bar{a}_{ij} x_j \le b_i \quad \forall i = 1, \dots, m \\
& x_j \ge 0 \quad \forall j = 1, \dots, n
\end{aligned}
$$

The problem solution of the robust counterpart should satisfy all possible scenarios. Note that the *worst-case scenario* corresponds to all uncertain parameters attaining their maximum values. We formulate the robust counterpart as follows.

$$
\begin{aligned}
\text{(SR)} \quad \max_{x} \quad & \sum_{j=1}^{n} c_j x_j, \\
\text{s.t} \quad & \sum_{j=1}^{n} (\bar{a}_{ij} + \hat{a}_{ij}) x_j \le b_i \quad \forall i = 1, \dots, m, \\
& x_j \ge 0 \quad \forall j = 1, \dots, n.
\end{aligned}
$$

This is called a strict robust formulation for the box uncertainty set which is known to often yield conservative solutions. The light robust approach aims to minimize the violation of the strict robust solution, while satisfying the nominal problem. The solution is subject to a limit in total cost that is linearly dependent on (a) the nominal problem solution cost and (b) the strict robust problem solution cost. Let $\overline{C}$ and $\widehat{C}$ denote the obtained objective values from (Nom) and (SR), respectively. We introduce parameter $\rho \in [0, 1]$ to denote a convex combination of the two obtained costs values. Let slack variables γ_i denote the violations of the strict robust constraints, which we aim to minimize. Finally, we present the light robust problem formulation.

$$
\begin{aligned}
\text{(LR)} \quad \min_{\gamma} \quad & \sum_{i=1}^{n} \gamma_i, \\
\text{s.t} \quad & \sum_{j=1}^{n} c_j x_j \geq (1-\rho)\overline{C} + \rho \widehat{C}, \\
& \sum_{j=1}^{n} \bar{a}_{ij} x_j \leq b_i \quad \forall i = 1, \dots, m, \\
& \sum_{j=1}^{n} (\bar{a}_{ij} + \hat{a}_{ij}) x_j \leq b_i + \gamma_i \quad \forall i = 1, \dots, m, \\
& x_j \geq 0 \quad \forall j = 1, \dots, n, \\
& \gamma_i \geq 0 \quad \forall i = 1, \dots, m.
\end{aligned}
$$

Note that the optimal solution of (LR) corresponds to the solution to (Nom) in case $\rho = 0$ and to the solution of (SR) if $\rho = 1$. The latter is equivalent to a formulation including a budget of uncertainty in case $\Gamma = n$ (Schöbel (2014)). We can interpret the optimal solution to the light robust formulation as follows. Out of all feasible solutions to the nominal problem that have a total cost lower than the limit, dependent on ρ, $\overline{C}$ and $\widehat{C}$, it is the solution that is most robust (i.e. closest to feasibility in case all customers attain their maximum demand).

3.2 Notation and problem definition

In this section we will consider an LRP with capacitated depots and vehicles. We will consider the two-index flow-based formulation introduced by Karaoglan et al. (2011). The formulation is adapted as a two-stage formulation and extended to allow for uncertainty in demand. Furthermore, we do not explicitly require that every customer needs to be served. Instead, it is allowed for demand to remain unsatisfied, which will yield a penalty cost. This allows for a more gradual transition towards the second stage problem formulation, where it might not be possible to serve every customer due to the dependence on previously made decisions on the opening of depots and assignment of vehicles.

Consider a complete directed graph $G = (\mathcal{N}, \mathcal{A})$, where the set of nodes is composed of both potential depot and customer locations $\mathcal{N} = \mathcal{N}_D \cup \mathcal{N}_C$ and arcs $\mathcal{A} = \{(i,j) \mid i,j \in \mathcal{N}\} \setminus \{(i,j) \mid i,j \in \mathcal{N}_D,\ i \neq j\}$. We introduce the following variables:

- $x_{ij} \in \{0,1\}$ equals 1 if node $i \in \mathcal{N}$ is followed by $j \in \mathcal{N}$ on a route, 0 otherwise.
- $y_k \in \{0,1\}$ equals 1 if depot $k \in \mathcal{N}_D$ is opened, 0 otherwise.
- $v_k \in \mathbb{Z}_+$ denotes the number of vehicles stationed at depot k.
- $z_{ik} \in \{0,1\}$ equals 1 if customer i is assigned to depot k, 0 otherwise.
- $u_{ij} \in \mathbb{R}_+$ denotes the current load of a vehicle when traveling from node i to j.

A nonnegative cost, denoted by c_{ij}, is incurred for traveling over arc $(i,j) \in \mathcal{A}$. It is assumed that the traveling cost is symmetric and satisfies the triangle inequality. Furthermore, fixed costs f_k and g_k are incurred for opening depot k and for every vehicle assigned to depot k, respectively. Let p_i denote the penalty cost per unit of unsatisfied demand to customer i, which can be interpreted in multiple ways e.g. loss in revenue or outsourcing cost.

The vehicles are homogeneous and all have a capacity Q. Potential depots are heterogeneous and have a capacity D_k for $k \in \mathcal{N}_D$. Every customer $i \in \mathcal{N}_C$ has a nonnegative uncertain demand $d_i \in \mathcal{D}_i$, where $\mathcal{D}_i$ is an uncertainty set. Assume that the actual demand is known to be bounded by a symmetric interval around the known nominal value. Let $\hat{d}_i$ denote the deviation in demand from the nominal value $\bar{d}_i$ for customer i. Hence it holds that $d_i \in [\bar{d}_i - \hat{d}_i, \bar{d}_i + \hat{d}_i]$. We define the demand uncertainty set as $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_n$, with

$$\mathcal{D}_i = \{d_i \in \mathbb{R}_+ \mid |d_i - \bar{d}_i| \leq \hat{d}_i\}$$

being the uncertainty set for customer i.

3.3 Integer programming formulations

We will now present the different problem formulations. We first consider the LRP formulation containing uncertainty in customer demand and we will clarify the objective and all the constraints. Afterwards, we will use the formulation containing uncertainty to derive the deterministic nominal and strict robust formulations. We will use the obtained optimal cost values from these two approaches in the derived light robust formulation. To test the performance of the optimal solutions from the light robust formulation, we introduce an artificial second stage formulation. Here, we assume that the actual demand is now observed and the *wait-and-see* variables have to be determined, subject to the previously decided *here-and-now* variables on the depot location and vehicle allocation. Finally, we present a set of valid inequalities to strengthen the various formulations. The robust LRP with uncertain demand formulation is defined as follows:

$$\text{(P1)} \quad \min_{y,v} \quad \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + \sup_{d\in\mathcal{D}} F(y,v,d), \tag{1}$$

$$\text{s.t.} \quad y_k \in \{0,1\} \quad \forall k \in \mathcal{N}_D, \tag{2}$$

$$v_k \geq 0 \quad \forall k \in \mathcal{N}_D, \tag{3}$$

where $F(y,v,d)$ is defined as the optimal value of the second stage problem:

$$F(y,v,d) := \min_{x,u,z} \quad \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \sum_{i\in\mathcal{N}_C} p_i d_i (1 - \sum_{j\in\mathcal{N}} x_{ij}), \tag{4}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{N}} x_{ij} = \sum_{k\in\mathcal{N}_D} z_{ik} \quad \forall i \in \mathcal{N}_C, \tag{5}$$

$$\sum_{j\in\mathcal{N}} x_{ij} = \sum_{i\in\mathcal{N}} x_{ji} \quad \forall j \in \mathcal{N}, \tag{6}$$

$$\sum_{i\in\mathcal{N}_C} x_{ki} \leq v_k \quad \forall k \in \mathcal{N}_D, \tag{7}$$

$$u_{ij} \leq Q x_{ij} \quad \forall i,j \in \mathcal{N}, i \neq j, \tag{8}$$

$$u_{ij} + (d_i - Q)x_{ij} \leq 0 \quad \forall i \in \mathcal{N}_C, j \in \mathcal{N}, \tag{9}$$

$$u_{ij} - d_j x_{ij} \geq 0 \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_C, \tag{10}$$

$$\sum_{i\in\mathcal{N}_C} u_{ik} = 0 \quad \forall k \in \mathcal{N}_D, \tag{11}$$

$$\sum_{j\in\mathcal{N}} u_{ji} - \sum_{j\in\mathcal{N}} u_{ij} \geq \sum_{j\in\mathcal{N}} d_i x_{ij} \quad \forall i \in \mathcal{N}_C, \tag{12}$$

$$\sum_{k\in\mathcal{N}_D} z_{ik} \leq 1 \quad \forall i \in \mathcal{N}_C, \tag{13}$$

$$\sum_{i\in\mathcal{N}_C} d_i z_{ik} \leq D_k y_k \quad \forall k \in \mathcal{N}_D, , \tag{14}$$

$$x_{ik} \leq z_{ik} \quad \forall i \in \mathcal{N}_C, k \in \mathcal{N}_D, \tag{15}$$

$$x_{ki} \leq z_{ik} \quad \forall i \in \mathcal{N}_C, k \in \mathcal{N}_D, \tag{16}$$

$$x_{ij} + z_{ik} + \sum_{m\in\mathcal{N}_D, m\neq k} z_{jm} \leq 2 \quad \forall i,j \in \mathcal{N}_C, i \neq j, k \in \mathcal{N}_D, \tag{17}$$

$$x_{ij} \in \{0,1\} \quad \forall i,j \in \mathcal{N}, \tag{18}$$

$$z_{ik} \in \{0,1\} \quad \forall i \in \mathcal{N}_C, k \in \mathcal{N}_D, \tag{19}$$

$$u_{ij} \geq 0 \quad \forall i,j \in \mathcal{N}. \tag{20}$$

In the above formulation: the objective function and subsequent constraints can be interpreted as follows:

- (1) the total objective cost consists of the first stage decisions considering the cost for opening depots and cost for deploying vehicles. The supremum over $F(y, v, d)$ denotes the optimal second stage cost dependent on the first stage decisions, for the most expensive scenarios of demand within the uncertainty set.

- (4) the second stage cost consists of the total routing cost and a penalty cost for unsatisfied demand. Note that the second term allows for a feasible solution of $F(y, v, d)$ for every first stage solution, since $x_{ij} = 0 \; \forall i, j \in \mathcal{N}$ is feasible for all *here-and-now* variables $y_k, v_k \; \forall k \in \mathcal{N}_D$;

- (5) every customer can be served by only one vehicle and only if they are assigned to a depot;

- (6) the number of arcs entering each node is equal to the number of arcs leaving the node;

- (7) the number of vehicles departing from a depot is bounded by the number of vehicles assigned to the depot;

- (8) the load of a vehicle never exceeds the vehicle capacity;

- (9) the current load of a vehicle is at most equal to the vehicle capacity minus the demand of the most recently visited customer;

- (10) a vehicle can only visit a customer if the current load is sufficient to satisfy all the customer's demand;

- (11) the load of all vehicles is empty when traveling back to the depots;

- (12) the load of a vehicle is updated correctly after a delivery. This implies that both pickups and partial deliveries are prohibited;

- (13) each customer can only be assigned to a single depot;

- (14) customers can only be assigned to open depots and the total demand of customers assigned to a depot cannot exceed the depot capacity;

- (15)-(17) together prevent illegal vehicle routes that start and end at different depots. Constraint (17) states that each adjacent pair of customers in a route needs to be assigned to the same depot. Constraints (15) and (16) only allow vehicles to travel between a depot and a customer if the customer is assigned to that depot.

Together with (13), this ensures that all customers on a route need to be assigned to the same depot and each vehicle has to start and end at the same depot.

Problem (P1) is a large MIP problem that contains a supremum to consider uncertain parameters. In order to solve this problem, we will derive: 1) a nominal formulation, 2) a strict robust formulation and 3) a light robust formulation, in the following subsections.

3.3.1 Nominal and strict robust problem formulation

We observe uncertainty in customer demand in several constraints in the above formulation. The nominal approach replaces the uncertain parameters d_i by the known nominal values $\bar{d}_i$, to allow for a deterministic problem. By setting $F(y, v, d) = F(y, v, \bar{d})$, we can compactly formulate the nominal LRP. Similarly, replacing all uncertain demand d_i by their maximum attainable value $\bar{d}_i + \hat{d}_i$ and defining $F(y, v, d) = F(y, v, \bar{d} + \hat{d})$ coincides with the strict robust formulation, where each demand attains its maximum value. We can now define the nominal (P1N) and the strict robust model (P1SR):

$$\text{(P1N)} \min_{y,v} \quad \sum_{k \in \mathcal{N}_D} f_k y_k + \sum_{k \in \mathcal{N}_D} g_k v_k + F(y, v, \bar{d}), \tag{1'}$$
$$\text{s.t.} \quad (2) - (20).$$

$$\text{(P1SR)} \min_{y,v} \quad \sum_{k \in \mathcal{N}_D} f_k y_k + \sum_{k \in \mathcal{N}_D} g_k v_k + F(y, v, \bar{d} + \hat{d}), \tag{1''}$$
$$\text{s.t.} \quad (2) - (20).$$

3.3.2 Light robust problem formulation

In the light robust formulation, we aim to maximize the robustness of the solution. This is equivalent to the minimization of the constraint violations in the strict robust approach. At the same time, the solution is required to:

- satisfy all the constraints under the nominal demand scenario
- have a total cost that is bounded by a predetermined threshold relative to the nominal problem objective value.

The constraints in the light robust (LR) problem formulation are related to those of a strict robust formulation. Once again, we consider for every uncertain parameter its worst-case value within the uncertainty set. The LR problem differs by adding a slack variable γ_{ij} to all constraints that contain the uncertain parameter d_i. This allows the solution to not satisfy all the strict robust problem. The objective is to minimize the sum of the slack variables.

Let $\overline{C}$ and $\widehat{C}$ denote the optimal objective values from (P1N) and (P1SR), respectively. The light robust cost value upper bound will be a convex combination of these objective values, defined as $C \in \{(1-\rho)\overline{C} + \rho\widehat{C} \mid \rho \in [0,1]\}$. Our formulation is analogous to that of Schöbel (2014), Section 4.3.

$$
\begin{aligned}
\text{(P1LR)}\ \min_{\gamma}\quad & \sum_{(i,j)\in\mathcal{A}} \gamma_{ij} && (21)\\
\text{s.t.}\quad & \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \\
& \qquad \sum_{i\in\mathcal{N}_C} p_i(\bar{d}_i + \hat{d}_i)(1 - \sum_{j\in\mathcal{N}} x_{ij}) \le (1-\rho)\overline{C} + \rho\widehat{C}, && (22)\\
& u_{ij} + (\bar{d}_i - Q)x_{ij} \le 0 \quad \forall i \in \mathcal{N}_C, j \in \mathcal{N}, && (9')\\
& u_{ij} - \bar{d}_j x_{ij} \ge 0 \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_C, && (10')\\
& \sum_{j\in\mathcal{N}} u_{ji} - \sum_{j\in\mathcal{N}} u_{ij} \ge \sum_{j\in\mathcal{N}} \bar{d}_i x_{ij} \quad \forall i \in \mathcal{N}_C, && (12')\\
& \sum_{i\in\mathcal{N}_C} \bar{d}_i z_{ik} \le D_k y_k \quad \forall k \in \mathcal{N}_D, && (14')\\
& u_{ij} + (\bar{d}_i + \hat{d}_i - Q)x_{ij} \le \gamma_{ij} \quad \forall i \in \mathcal{N}_C, j \in \mathcal{N}, && (9'')\\
& u_{ij} - (\bar{d}_j + \hat{d}_j)x_{ij} \ge -\gamma_{ij} \quad \forall i \in \mathcal{N}, j \in \mathcal{N}_C, && (10'')\\
& \sum_{j\in\mathcal{N}} u_{ji} - \sum_{j\in\mathcal{N}} u_{ij} \ge \sum_{j\in\mathcal{N}} (\bar{d}_i + \hat{d}_i)x_{ij} - \sum_{j\in\mathcal{N}} \gamma_{ji} \quad \forall i \in \mathcal{N}_C, && (12'')\\
& \sum_{i\in\mathcal{N}_C} (\bar{d}_i + \hat{d}_i) z_{ik} \le D_k y_k + \sum_{i\in\mathcal{N}_C} \gamma_{ki} \quad \forall k \in \mathcal{N}_D, && (14'')\\
& \gamma_{ij} \ge 0 \quad \forall i, j \in \mathcal{N}, && (23)\\
& (2), (3), (5)-(8), (11), (13), (15)-(20).
\end{aligned}
$$

The objective function and constraints can be interpreted as follows:

- (21) the objective is to minimize the total sum of the slack variables;
- (22) the relative increase in total cost compared to the cost obtained from the nominal formulation is limited;

- (9′), (10′), (12′) and (14′) are nominal counterparts of the original uncertain constraints. Hence, the nominal demand $\bar{d}_i$ of every customer i must either be satisfied or stay unsatisfied at a penalty cost that has remained unchanged;
- (9″), (10″), (12″) and (14″) are the robust counterparts of the original uncertain constraints with the addition of the slack variables γ_{ij} to allow for feasible solutions in case a strict robust solution is unattainable.

Note that the optimal solution of (P1LR) with $\rho = 0$ is equivalent to the nominal problem solution. Furthermore, the optimal solution in case $\rho = 1$ is equal to the strict robust solution, in which case $\gamma_{ij} = 0$ for all $i, j \in \mathcal{N}$. Throughout this thesis, we will refer to the solutions of (P1N) and (P1SR) as the solutions of (P1LR) with $\rho = 0$ and $\rho = 1$, respectively.

3.3.3 Formulation of the solution evaluation

We now provide a formulation that we can use to test the performance of the light robust problem solutions. The location of the open depots and the number of assigned vehicles are decided on in the first stage. These decisions now remain unchanged. In the second stage the actual demand values d_i are known and we determine the optimal a posteriori routes. Hence, we solve $\min_{x,u,z}\{F(y, v, d)\}$. This problem is equivalent to solving a multi-depot capacitated vehicle routing problem (MDCVRP). To test the performance of the first stage decisions dependent on the different formulations, the second stage is solved T times for random sampled customer demand scenarios. Every time, actual demand d_{it} of all customers at time t is observed. Since there are no linking constraints over index t, every instance of time is independent. Since there are no linking constraints over index $t \in \{1, \dots, T\}$, every instance of time is independent. Therefore we do not actually need to include index t. We present the second stage formulation as follows:

$$\text{(P2)} \min_{x,u,z} \quad \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \sum_{i\in\mathcal{N}_C} p_i d_i (1 - \sum_{j\in\mathcal{N}} x_{ij}), \tag{24}$$

$$\text{s.t} \quad (5) - (20) \tag{25}$$

3.3.4 Valid inequalities

In the research on LRP, valid inequalities (also known an valid cuts) are a popular tool to tighten the LP-relaxation of the problem. Valid inequalities are not needed to obtain an optimal solution, but can decrease the solving time by restricting the solution space of a problem. We first discuss a few valid inequalities that can be included in all described problems. These inequalities are derived from Karaoglan et al. (2011), who show a positive effect in terms of solving times.

Two-customer subtour elimination constraints

The first inequalities are subtour elimination constraints, classically used for the Traveling Salesman Problem (Miller et al. (1960)).

$$x_{ij} + x_{ji} \leq 1 \quad \forall i, j \in \mathcal{N}_C. \tag{26}$$

Constraint (26) eliminates subtours involving only two customers. This type of elimination constraint can be included for subtours involving a larger number of customers, but adding the constraint for all subsets results in an exponential number of constraints, thereby increasing the problem's difficulty.

Only assign a customer to an open depot

The following valid inequality imposes that a customer can only be assigned to an open depot.

$$z_{ik} \leq y_k \quad \forall\, i \in \mathcal{N}_C, k \in \mathcal{N}_D. \tag{27}$$

Lower bound on the number of routes

This inequality assigns a lower bound to the minimum number of vehicles to use, based on the total demand and capacity of the vehicles.

$$\sum_{i \in \mathcal{N}_C} \sum_{k \in \mathcal{N}_D} x_{ij} \geq \left\lceil \frac{\sum_{i \in \mathcal{N}_C} d_i}{Q} \right\rceil. \tag{28}$$

This inequality is only valid if all demand has to be satisfied. This only holds if the penalty cost p_i per unit of unsatisfied demand is sufficiently large, on which we will elaborate in the next section. If this is the case, the optimal solution to the light robust problem for any value of ρ will satisfy at least all nominal demand values due to the nominal problem constraints. Therefore, constraint (28) can be included with $d_i = \bar{d}_i$. In addition, the optimal solution to the strict robust problem will satisfy the maximum attainable demand for every customer. Therefore, if $\rho = 1$, constraint (28) can be included with $d_i = \bar{d}_i + \hat{d}_i$.

We will now consider some new valid inequalities.

Three-customer subtour elimination constraints

The first set of constraints is related to the *two-customer subtour elimination constraints* (26). The total number of constraints increases exponentially when considering subtours involving an increasing number of customers. We find that the solution time effectively increases when including these subtour elimination constraints for three or more customers using this formulation. The first observation to make is that for all $i, j \in \mathcal{N}_C$, constraint (26) considers every distinct pair of customers twice, as well as all cases where $i = j$. Since the formulation for two customers already includes both arc directions, we can instead formulate given constraints for $i, j \in \mathcal{N}_C, i > j$ and we still consider all unique distinct customer pairs. Hereby, we reduce the number of constraints by more than half. We can extend this formulation for subtours containing more than two customers. Extending (26) to three customers can be formulated as follows

$$x_{ij} + x_{jl} + x_{li} \leq 2 \quad \forall i, j, l \in \mathcal{N}_C.$$

If we only consider the customers $i, j, l \in \mathcal{N}_C, i > j > l$, we have to take into account both directions, i.e. $x_{ij} + x_{jl} + x_{li} \leq 2$ and $x_{il} + x_{lj} + x_{ji} \leq 2$. However, since subtours of two customers are also illegal, we can combine both directions in the following set of constraints.

$$x_{ij} + x_{ji} + x_{jl} + x_{lj} + x_{li} + x_{il} \leq 2 \quad \forall i, j, l \in \mathcal{N}_C, i > j > l. \tag{29}$$

By only allowing to use a maximum of two out of the six possible edges between each unique set of three distinct customers, all subtours involving three customers are forbidden. The addition of constraint (26) separately for subtours involving both two and three customers results in an addition of $|\mathcal{N}_C|^3 + |\mathcal{N}_C|^2$ constraints. Constraint (29) yields $\frac{|\mathcal{N}_C|(|\mathcal{N}_C|-1)(|\mathcal{N}_C|-2)}{6}$ constraints, and an additional $\frac{|\mathcal{N}_C|(|\mathcal{N}_C|-1)}{2}$ constraints for the subtours involving two customers.

Symmetry breaking constraints

Next, we consider a set of symmetry breaking constraints. Due to the absence of time windows and the presence of symmetry in routing cost, the direction of an optimal route is irrelevant, resulting in 2^N non-unique optimal solutions, where N denotes the number of distinct routes in an optimal solution. Ultimately, we want to eliminate all but one of these non-unique optimal solutions. However, due to the absence of an index denoting individual vehicles or routes, this is not possible. Consider the following symmetry breaking constraint.

$$\sum_{i \in \mathcal{N}_C} \sqrt{i} \cdot x_{ki} \leq \sum_{i \in \mathcal{N}_C} \sqrt{i} \cdot x_{ik} \quad \forall k \in \mathcal{N}_D. \tag{30}$$

Constraint (30) removes symmetry using a lexicographical order of the routes. The sum of the square root of the indices of the last customer in every route to depot k must be larger than the sum of the square root of the indices of the first customer in every route from depot k. We can prove this by contradiction: consider an optimal set of routes that violates (30). For every depot k that does not satisfy the symmetry breaking constraint, change the directions of all routes starting at k. Due to the symmetry in traveling cost, the routing costs will remain unchanged and constraint (30) now holds. First note that this set of constraints eliminates all but one non-unique solutions if there is a single route per depot k. With multiple routes per depot, this effect is reduced. Furthermore, the square root reduces the number of instances in which an equality in constraint (30) holds. Hence, it tightens the inequality compared to the index i as the constant term. Consequently, the constraint can be formulated even tighter (e.g. using higher power roots), although the improvements will be close to non-existent.

Slack variables bounding constraints

We now introduce some valid inequalities that relate to the slack variables γ_{ij} in (P1LR). Consider the constraints from the light robust formulation that contain slack variables γ_{ij}: (9″), (10″), (12″). The bounding constraints already ensured that $u_{ij} > 0$ if and only if $x_{ij} = 1$ for $i, j \in \mathcal{N}_C$. Due to the referred constraints, we can extend this with $\gamma_{ij} > 0$ if and only if $x_{ij} = 1$ for $i, j \in \mathcal{N}_C$. Constraint (31) offers a tighter bound based on the the maximum relative deviation in demand.

$$\gamma_{ij} \leq \frac{\hat{d}_j}{\bar{d}_j} u_{ij} \quad \forall i, j \in \mathcal{N}_C. \tag{31}$$

Satisfied demand shortage constraints

The following valid inequalities provide a lower bound to the required total sum of the slack variables. Recall that, due to the bounding constraints in combination with constraint (11), $\sum_{k \in \mathcal{N}_D} \sum_{i \in \mathcal{N}_C} u_{ki}$ denotes the total satisfied demand to all customers. Inequality (32) requires the total sum of slack variables to be at least as large as the difference between the maximum attainable demand and the total satisfied demand.

$$\sum_{i \in \mathcal{N}_C} \Big(\sum_{k \in \mathcal{N}_D} u_{ki} + \sum_{j \in \mathcal{N}} \gamma_{ji} \Big) \geq \sum_{i \in \mathcal{N}_C} \bar{d}_i + \hat{d}_i. \tag{32}$$

Capacity shortage constraints

Finally, the slack variables are needed to denote the demand that cannot be satisfied, in case of maximum demand, due to the lack of sufficient capacity. This lack of capacity equals either the the total depot capacity or the total vehicle capacity. Therefore we can formulate

$$\sum_{j\in\mathcal{N}}\sum_{i\in\mathcal{N}_C}\gamma_{ji}+\min\{\sum_{k\in\mathcal{N}_D}D_k y_k,\sum_{k\in\mathcal{N}_D}Qv_k\}\geq\sum_{i\in\mathcal{N}_C}\bar{d}_i+\hat{d}_i,$$

which can be rewritten as linear constraints as follows:

$$\sum_{j\in\mathcal{N}}\sum_{i\in\mathcal{N}_C}\gamma_{ji}+\sum_{k\in\mathcal{N}_D}D_k y_k\geq\sum_{i\in\mathcal{N}_C}\bar{d}_i+\hat{d}_i, \tag{33}$$

$$\sum_{j\in\mathcal{N}}\sum_{i\in\mathcal{N}_C}\gamma_{ji}+\sum_{k\in\mathcal{N}_D}Qv_k\geq\sum_{i\in\mathcal{N}_C}\bar{d}_i+\hat{d}_i. \tag{34}$$

3.4 Uncertainty in the customer set

So far we have considered the location-routing problem with uncertainty in customer demand. In this section we consider an alternative problem in which we face uncertainty in the set of customers to consider. In other words, we have a set of *potential* customers, but it is uncertain which customers have a positive demand. For simplicity, we assume demand per customer (if positive) is known. A combination of both uncertainties can readily be included.

We consider that every customer $i\in\mathcal{N}_C$ will either have a positive demand or a demand equal to zero. We first have to define the subset of potential customers to be included in the problem. Let the uncertain parameter $a_i\in\{0,1\}$ denote whether a customer has a positive demand, with $a_i=1$ denoting that customer i is included in the customer set, and 0 otherwise. We define an uncertain realization as $a_1\times\ldots\times a_{|\mathcal{N}_C|}$ and the customer uncertainty set as $\mathcal{A}=\{0,1\}^{|\mathcal{N}_C|}$. Later on, values a_i will be randomly sampled for every $i\in\mathcal{N}_C$, in a similar fashion as in Albareda-Sambola et al. (2007). This will be explained in further detail in Section 4.

We rewrite the integer programming formulation containing uncertainty in the customer set as follows:

$$
\begin{aligned}
\text{(P1c)} \quad \min_{y,v} \quad & \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + \sup_{a\in\mathcal{A}} F(y,v,a), && (35)\\
\text{s.t.} \quad & (2),(3),
\end{aligned}
$$

where $F(y, v, a)$ is defined as the optimal value of the second stage problem:

$$
\begin{aligned}
F(y,v,a) = \min_{x,u,z} \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \sum_{i\in\mathcal{N}_C} p_i d_i a_i (1 - \sum_{j\in\mathcal{N}} x_{ij}), && (36)\\
\text{s.t.} \quad & \sum_{j\in\mathcal{N}} x_{ij} = a_i \sum_{k\in\mathcal{N}_D} z_{ik} \quad \forall i\in\mathcal{N}_C, && (37)\\
& (6) - (20).
\end{aligned}
$$

The demand values d_i are now deterministic, while a_i is uncertain. The objective function (36) only takes a penalty cost into account for a priori unsatisfied demand for the potential customers included in the actual customer set. This can be justified by stating that the demand of customers not included in given set is equal to zero. Furthermore, we have included uncertain parameter a_i in constraint (37). We could have rewritten the entire formulation by including a_i in all relevant constraints or by introducing a subset $\overline{\mathcal{N}}_C \subseteq \mathcal{N}_C$ consisting of the customers included in the customer set for the nominal problem. However, this is not necessary: the combination of constraints (5) and (13) states that a customer has to be included in exactly one route if and only if it is assigned to exactly one depot. If a customer is not assigned to a depot; it is not visited on a route, which results in a cost due to the second term in objective (4). By including a_i in this cost term in objective (36), there will be no cost incurred for customers not included in a route if $a_i = 0$. This effect ensures that, if $a_i = 0$, then $x_{ij} = x_{ji} = 0, \forall j \in \mathcal{N}$ will be optimal due to the absence of routing costs. Hence, customer i will not be included in any route and subsequently, u_{ij} and z_{ik} will also be equal to zero $\forall j, k \in \mathcal{N}$. Concluding, we do not need to redefine all constraints by including parameter a_i. We did, however, rewrite constraint (37) such that we can introduce the formulation of the light robust problem in a more gradual manner.

Similar to the original problem (P1), problem (P1c) contains a supremum to consider uncertain values which is difficult to solve to optimality. Therefore, our solution approach remains unchanged. We will introduce the formulations of the nominal problem and the strict robust problem. The obtained objective values are stored in the values $\overline{C}$ and $\widehat{C}$. These values are used in the maximum cost constraint in the light robust problem formulation.

$$\text{(P1Nc)} \min_{y,v} \quad \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + F(y,v,\bar{a}), \tag{35$'$}$$
$$\text{s.t.} \quad (2),(3),(6)-(20),(36),(37).$$

$$\text{(P1SRc)} \min_{y,v} \quad \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + F(y,v,\hat{a}), \tag{35$''$}$$
$$\text{s.t.} \quad (2),(3),(6)-(20),(36),(37).$$

These formulations are similar to the case of uncertainty in demand, with the exception of using $F(y,v,a)$ instead of $F(y,v,d)$. Finally, we introduce the light robust problem:

$$\begin{aligned}
\text{(P1LRc)} \min_{\gamma} \quad & \sum_{i\in\mathcal{N}} d_i \gamma_i && (38)\\
\text{s.t.} \quad & \sum_{k\in\mathcal{N}_D} f_k y_k + \sum_{k\in\mathcal{N}_D} g_k v_k + \sum_{(i,j)\in\mathcal{A}} c_{ij} x_{ij} + \\
& \qquad \sum_{i\in\mathcal{N}_C} p_i d_i \bar{a}_i (1 - \sum_{j\in\mathcal{N}} x_{ij}) \le (1-\rho)\overline{C} + \rho\widehat{C}, && (39)\\
& \sum_{j\in\mathcal{N}} x_{ij} = \bar{a}_i \sum_{k\in\mathcal{N}_D} z_{ik} \quad \forall i \in \mathcal{N}_C : \ \bar{a}_i = 1, && (40)\\
& \sum_{j\in\mathcal{N}} x_{ij} = 1 - \gamma_i \quad \forall i \in \mathcal{N}_C, && (41)\\
& \gamma_i \ge 0 \quad \forall i \in \mathcal{N}_C, && (42)\\
& (2),(3),(6)-(20).
\end{aligned}$$

Once again, γ_i denotes a slack variable needed in case a strict robust solution is too costly and violates constraint (39). Customer demand value d_i is used as a weight in the objective function. Hereby we ensure that customers with small demand values are not prioritized due to their lower requirement for capacity. Constraint (40) ensures that the optimal solution will be feasible for the scenario of nominal demand, while constraint (41) implies that the optimal solution should be as robust as possible.

The term $\sum_{k\in\mathcal{N}_D} z_{ik}$ is absent in constraint (41), as well as $\hat{a}_i$. First, note that the *worst-case scenario* consists of $a_i = 1, \forall i \in \mathcal{N}_C$, therefore $\hat{a}_i = 1$. Furthermore, in case $a_i = 0$, customer i is still included if $\sum_{j\in\mathcal{N}} x_{ij} = 1$. If customer i can not be included (due to a violation of constraint (22)), γ_i will be equal to one. Thus, $\sum_{i\in\mathcal{N}_i} \gamma_i$ is equal to the number of customers not included in the problem. For both cases, the presence of constraints (13), (15), (16) and (17), results that

$$\sum_{j\in N} x_{ij} = \sum_{k\in\mathcal{N}_D} z_{ik},$$

will hold by default.

Similar to the case of demand uncertainty, the value a_i will become known in the second stage. We can redefine the second stage problem without rewriting the formulation by only including customer i in the customer set $\mathcal{N}_C$ in case $a_i = 1$. Hence (P2c) is equal to (P2).

In addition, we can use the valid inequalities (26), (27), (29) from the previous section. Furthermore, valid inequality (28), considering the minimum number of required vehicles to satisfy all demand, can be rewritten to

$$\sum_{i\in\mathcal{N}_C}\sum_{k\in\mathcal{N}_D} x_{ij} \geq \left\lceil \frac{\sum_{i\in\mathcal{N}_C} \bar{a}_i d_i}{Q} \right\rceil.$$

As explained in Section 3.3.4, this valid inequality provides a lower bound on the number of routes if all demand has to be satisfied and should therefore only be used if the penalty cost per unit of unsatisfied demand is sufficiently large.

4 Experimental setup

In this section we will introduce the setup of the experiments that will be presented in Section 5. We will first introduce the test environment in Section 4.1, followed by the implemented solution procedure in Section 4.2. Afterwards we present the performance measures used to discuss the obtained results in Section 4.3.

4.1 Test environment

We first introduce the test environment of the problem formulations. Unless stated otherwise, all parameters will be equal to these given values. In further subsections, we will perform a sensitivity analysis by varying one or a few values, while keeping the remaining parameters unchanged. To the best of our knowledge, no benchmark values exist for the LRP containing any type of uncertainty. However, multiple deterministic benchmark instances exist that are traditionally used to compare the performance of solution methods for the deterministic LRP. For fairness, our test environment will be closely related to these instances with the addition of data uncertainty.

4.1.1 Base parameter values

All node locations are uniformly sampled from the rectangle of size $[0, 100]^2$. The routing costs c_{ij} are equal to the Euclidean distance between node i and j. Each testing instance consists of three distinct possible depots, each having a uniformly sampled depot capacity $D_k \sim U(100, 400)$ and opening cost $f_k = 5D_k$. Furthermore, the base number of customers is $|\mathcal{N}_C| = 10$, where each customer has a nominal demand sampled from $[10, 30]$. The maximum deviation in demand for customer i from its nominal value is equal to $0.5\bar{d}_i$. Each vehicle has a capacity of $Q = 70$ and the cost of assigning a vehicle to a depot $g_k = 2Q$ for all $k \in \mathcal{N}_D$. The penalty cost for each unit of unsatisfied demand p_i equals $10,000$ for each customer $i \in \mathcal{N}_C$. Note that, if $p_i > M$ for a sufficiently large value of M, all demand will be satisfied, with $M = \max_i\{\frac{\min_k\{f_k+g_k+2c_{ik}\}}{d_i}\}$. This holds provided that the total capacity of the available depots is larger than the maximum attainable total demand. Unless mentioned otherwise, we simulate #Sims = 100 instances which are solved for $\rho \in P = \{0, 0.1, \dots, 0.9, 1\}$. Furthermore, to evaluate the performance of the solution for every instance, we sample $T = 20$ demand scenarios uniformly from $\mathcal{D}_i$ and solve (P2) for every scenario. Thus, the light robust LRP (P1LR) is solved #Sims $\times|P|$ times and (P2) is solved $T\times$ #Sims $\times|P|$ times. All base parameter values are summarized in Table 2.

Table 2: Base parameter values

#Sims	100
T	20
$\vert\mathcal{N}_C\vert$	10
$\vert\mathcal{N}_D\vert$	3
Q	70
D_k	$\sim U(100, 400)$
$\bar{d}_i$	$\sim U(10, 30)$
$\hat{d}_i$	0.5 $\bar{d}_i$
f_k	$5D_k$
g	$2Q$
k	10,000

For the problem with uncertainty in customer set, $|\mathcal{N}_C|$ now denotes the number of potential customers, of which $p \times |\mathcal{N}_C|$ customers have a positive demand, with $0 \leq p \leq 1$. For every problem instance, we simulate a nominal scenario of the customers with a positive demand using a Bernouilli distribution with probability p to be included. That is, $\bar{a}_i \sim B(1, p),\ \forall i \in \mathcal{N}_C$. The worst-case scenario corresponds to $\hat{a}_i = 1,\ \forall i \in \mathcal{N}_C$. Customers still have heterogeneous nominal demands, but $\hat{d} = 0$, removing the uncertainty in demand. Furthermore, the experiment remains unchanged.

4.1.2 Classical instances from the literature

In addition to simulated instances, we make use of publicly available benchmark instances, introduced by Prins et al. (2006) and available on Prodhon (2006). We refer to individual benchmark instances by their instance names as follows $|\mathcal{N}_C|$ - $|\mathcal{N}_D|$ - #clusters- $\{a, b\}$, with a and b denoting low and high vehicle capacity Q. For several instances, we selected a subset of the first $n \in \mathcal{N}_C$ of customers, to create more unique smaller instances. In addition, we include a certain level of uncertainty in the given instances by changing the given deterministic demand to nominal demand and we hold on to the maximum deviation $\hat{d}_i = 0.5\bar{d}_i$ for all customers i. Therefore the optimal solution to the nominal problem in case all customers are included (and only then) is equal to the optimal solution of the benchmark instance.

4.2 Solution and testing procedure

In this section we will elaborate on the method of solving and testing the solutions. For any given instance we first solve the nominal problem (P1N) and the strict robust problem (P1SR), with the single purpose of obtaining the optimal cost values $\overline{C}$ and $\widehat{C}$. We then solve (P1LR) for all values of $\rho \in P$. Finally, we sample T scenarios of customer demand from the given demand uncertainty set and solve (P2) for every demand sample subject to every distinct solution to (P1LR) based on ρ. We will report the running times for (P1LR) for all values $\rho \in P$. The optimal solution to the light robust problem for a certain value of ρ is always a feasible solution for all larger values of ρ. The consecutive problems will therefore be solved in increasing order of ρ where the yielded optimal solutions are used as a starting point for the next problem. A pseudocode of the solution method can be found in Algorithm 1. The same procedure holds for solving the problem with uncertainty in the customer set by replacing formulations (P1) for (P1c).

All experiments have been performed on an HP laptop with intel core i7 CPU (16GB RAM) using Gurobi 7.5.2 with its default settings as the optimizer to solve the MIP formulations.

Algorithm 1 Solution Procedure

γ, τ, S1Sol *are arrays of size $|P|$ that store the total sum of slack variables, the solving time and the optimal strategic decisions, respectively for every solution based on ρ.*

$n \leftarrow |P|$
for Sim $\leftarrow 1$ to #Sims **do**
    locations $i \leftarrow$ randomly sampled $\forall i \in \mathcal{N}$
    $d_{ij} \leftarrow$ Euclidean distances based on distances $\forall i, j \in \mathcal{N}$
    $D_k \leftarrow$ randomly sampled $\forall k \in \mathcal{N}_D$
    $\bar{d}_i \leftarrow$ randomly sampled $\forall i \in \mathcal{N}_C$
    solve (P1SR) **Stage 1**
    $\widehat{C} \leftarrow z^*;\ \tau[n-1] \leftarrow time;\ \hat{x} \leftarrow x$
    solve (P1N)
    $\overline{C} \leftarrow z^*;\ \tau[0] \leftarrow time;\ \bar{x} \leftarrow x$
    for $r \leftarrow 0$ to $n-1$ **do**
        $\rho \leftarrow P[r]$
        if $\rho = 0$ **then**
            Solve (P1LR) with $\bar{x}$ as start solution
            $\gamma[0] \leftarrow z^*$
            S1Sol$[0] \leftarrow$ optimal strategic decisions
            StartSol $\leftarrow$ optimal solution
        else if $\rho > 0$ and $\rho < 1$ **then**
            Solve (P1LR), s.t. $\overline{C}$, $\widehat{C}$ with StartSol as start solution
            $\gamma[r] \leftarrow z^*;\ \tau[r] \leftarrow time$
            S1Sol$[r] \leftarrow$ optimal strategic decisions
            StartSol $\leftarrow$ optimal solution
        else
            Solve (P1LR) with $\hat{x}$ as start solution
            $\gamma[n-1] \leftarrow z^*$
            S1Sol$[n-1] \leftarrow$ optimal strategic decisions
        end if
    end for
    for $t \leftarrow 1$ to T **do** **Stage 2**
        $d_i \leftarrow$ randomly sampled from $\mathcal{D}_i,\ \forall i \in \mathcal{N}$
        for $r \leftarrow 0$ to $n-1$ **do**
            solve (P2) s.t. S1Sol$[r]$
        end for
    end for
end for

4.3 Performance measures

In this section we introduce the different performance measures that we will use throughout this thesis. First, we will focus on the average optimal solution values from (P1LR) for different values of ρ. This optimal value, being the total sum of the slack variables, will be referred to as the Tot. slack = $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}}\gamma_{ij}$. We also consider the average total cost of these optimal solutions. This total cost, denoted by Tot. cost in all tables, is equal to the LHS of constraint (22), consisting of the cost of opening depots, assigning vehicles and the a priori route and unsatisfied demand cost. To denote the increasing cost for solutions with larger values of ρ, we use a percentage increase in cost relative to the optimal cost of the nominal solution, where %Wors. = $100 \times \frac{C_\rho - \overline{C}}{\overline{C}}$, with C_ρ denoting the cost value from (L1RP) for ρ.

Value θ indicates the instance percent tightness ratio, introduced by Sungur et al. (2008). It can be calculated as follows:

$$\theta = \frac{\sum_{i\in\mathcal{N}_C}\bar{d}_i}{\min\{\sum_k Qv_k, \sum_k D_k y_k\}}.$$

This ratio can be interpreted as the ratio of the total nominal customer demand and the minimum of the total available vehicle capacity and available depot capacity. If the minimum of the total depot capacity and the total vehicle capacity is equal to the total nominal customer demand, then $\theta = 100\%$. This implies that it is necessary to fully utilize the maximum available capacity to satisfy the nominal demand. Therefore, every customer sample with a larger total demand than the nominal demand will result in unsatisfied demand. Thus, lower values of θ imply robust solutions. In case the available capacity is smaller than the nominal demand, θ can be larger than 100%.

Finally, we report two computational measures of robustness for any solution to the light robust problem, based on the demand scenario sampling. The first measure, referred to as % Infeas., indicates the percentage of demand scenarios where not all customer demand can be satisfied. This is the case if the minimum of the available depot capacity and total vehicle capacity is smaller than the total sampled demand. The second measure, % Uns., equals the average percentage of demand that remains unsatisfied in the optimal solutions. Hence, given an instance where the optimal second stage solution yields a positive total unsatisfied demand, we refer to the instance as *infeasible* with a certain percentage of unsatisfied demand. Note that feasible instances can also result in unsatisfied demand, although this only occurs if the penalty cost for unsatisfied demand is very low.

5 Computational experiments

In this section we present results of the light robust LRP as formulated in the previous sections, that are obtained from random scenario sampling and solving public instances from literature. We present and discuss the results obtained using different parameters in Sections 5.1 to 5.3. Finally, in Section 5.4 we will present results from the LRP with uncertainty in the customer set.

5.1 Homogeneous customers and varying vehicle capacities

In this section we focus on the results obtained from randomly sampled instances with different vehicle capacities Q. For simplicity we only consider homogeneous customers with a nominal demand $\bar{d}_i = 20$ for all $i \in \mathcal{N}_C$ to show the relevance of the different performance measures.

Table 3 presents the results obtained from sampling instances with homogeneous customers. Here, every vehicle has a capacity $Q = 70$. We present the values and measures introduced in the Section 4.3. The average computation times of solving the light robust problem (P1LR) and the solution evaluation (P2) are given in seconds.

Table 3: Results obtained from instances with homogeneous customers and vehicle capacity $Q = 70$.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	2349.100	0.000	82.62	76.453	5.144	8.30	0.678	2.622
0.1	2403.784	2.328	69.82	76.237	23.148	7.90	0.652	2.748
0.2	2468.192	5.070	63.54	75.301	31.273	6.30	0.511	2.633
0.3	2539.437	8.103	55.22	74.778	35.795	5.30	0.440	2.635
0.4	2597.531	10.576	49.52	73.303	28.179	4.10	0.346	2.629
0.5	2664.158	13.412	43.88	71.989	37.758	2.80	0.227	2.726
0.6	2729.956	16.213	36.86	69.979	29.108	1.10	0.089	2.632
0.7	2773.411	18.063	33.28	68.292	31.931	0.80	0.065	2.691
0.8	2871.344	22.232	24.78	62.563	27.971	0.10	0.010	2.855
0.9	2958.754	25.953	22.38	61.709	62.645	0.10	0.010	2.814
1.0	3069.667	30.674	0.00	56.915	1.085	0.00	0.000	2.941

We first observe that the formulation of the light robust approach in case $\rho = 1$ coincides with the strict robust formulation. This is confirmed by the total slack being equal to zero, implying that no slack variables are needed to satisfy the worst-case demand scenarios. In addition, all demand is satisfied in all sampled demand scenarios. Furthermore, the sum of slack variables as well as both robustness measures are monotonously decreasing for increasing values of ρ, while the total cost increases continuously. The same holds for the tightness ratio θ. This confirms that allowing for higher total cost results in more robust solutions. The next observation is that strict robust solutions can lead to conservative solutions, yielding unnecessarily high cost for common scenarios. The percentage of unsatisfied demand quickly increases for low values of ρ, while the improvements are relatively small for larger values. Note that the increase in total cost is gradual due to the dependency of the maximum allowed increase in cost. Thus, an average cost increase of 30.674 % compared to the nominal problem solution guarantees a solution that is always feasible, while an increase of 17 - 22 % can provide a solution that will be close to always feasible, assuming uniform demand sampling. Due to the additional complexity of the light robust problem for values of $\rho \in (0, 1)$ compared to the deterministic nominal and strict robust problems, the running times are significantly larger due to the addition of the slack variables. We observe fairly balanced running times for solving the sampled scenarios (P2) with slightly larger times for lower values of ρ. This result is due to the increase in difficulty in solving the optimal *a posteriori* routes for customer demand samples where not all demand can be satisfied.

Despite these observations, we can conclude that even for the solutions for lower values of ρ, the percentage of unsatisfied demand is still reasonably small. We will explain how these findings are caused by the tightness of both location and routing problems. We discuss two examples with $Q = 70$ and $Q = 100$ that yield very different performances. First consider the following example: for simplicity, we ignore the depot capacity. Given $|\mathcal{N}_C| = 10, \bar{d} = 20$ and $Q = 70$, as in the scenarios used in Table 3, four vehicles are required to satisfy all nominal demand (recall: split deliveries are not allowed). Hence, an excess vehicle capacity of 80 units remains to accommodate sampled additional demand. Considering that the worst-case scenario coincides with $\bar{d} + \hat{d} = 30$, which requires a total of five vehicles, we observe that the presence of a large excess capacity in the nominal solution results in tight instances. This is confirmed by the low values for θ and can explain the reasonably adequate performance in terms of satisfied demand. Now consider the same example with vehicle capacity $Q = 100$. In this scenario, assigning two vehicles is sufficient to satisfy all nominal demand, leaving an excess vehicle capacity of zero. This implies that every scenario with a total demand larger than the total nominal demand, results in unsatisfied demand. This is shown in Table 4, containing the results from the same sampled instances as before with vehicle capacity $Q = 100$.

Table 4: Results obtained from instances with homogeneous customers and vehicle capacity $Q = 100$.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	2109.926	0.000	106.815	100.000	0.847	48.85	4.695	7.175
0.1	2164.524	2.587	106.464	100.000	10.435	48.85	4.695	7.008
0.2	2254.804	6.879	96.440	86.565	12.149	38.40	3.724	5.814
0.3	2333.849	10.632	72.006	76.791	8.368	22.45	2.008	3.645
0.4	2396.179	13.587	63.445	74.026	9.625	17.05	1.502	3.034
0.5	2475.032	17.343	45.523	72.542	12.187	4.35	0.333	2.184
0.6	2567.818	21.751	36.814	70.811	15.641	2.50	0.174	2.287
0.7	2648.278	25.556	31.489	68.139	18.638	1.50	0.118	2.545
0.8	2777.168	31.699	23.822	63.643	25.036	1.00	0.079	2.875
0.9	2897.092	37.356	17.219	61.390	28.950	0.00	0.000	2.681
1.0	3026.733	43.516	0.000	56.878	32.043	0.00	0.000	2.576

We can clearly observe the differences in terms of performance resulting from the robustness measures. First consider the value of $\theta = 100$ for $\rho = 0$. This indicates that the minimum of the capacity of the opened depots and the deployed vehicles is equal to the sum of the nominal demands. This coincides with the assigning of two vehicles in every scenario due to the homogeneous nominal demand $\bar{d} = 20$. As argued in the example above, this will result in infeasible instances for every demand sample larger than the nominal demand. This can be deduced from the percentage of infeasible instances being close to 50% and larger percentages of unsatisfied demand.

Furthermore, we observe that both the average value of θ as well as the performance measures are equal for $\rho = 0$ and $\rho = 0.1$. This implies that the additional allowed cost is in no instance sufficient to assign a third vehicle. The total of slack variables, however, is slightly lower, since larger depot capacity has become available. For the running times, we observe longer times if the strategic decisions are less obvious. We also see longer running times for the solution evaluation, due to the presence of more scenarios resulting in unsatisfied demand. Finally, compared to Table 3, both performance measures decrease faster as ρ increases due to the clear connection with θ. Note that the percentage of unsatisfied demand is still small. In case of a scenario where not all demand can be satisfied, it is often possible to serve all but one customers, where the excluded customer might have the lowest demand value. Thus the large majority of the total demand can still be satisfied.

5.2 Varying cost ratios

The main purpose of the previous section was to clarify the realization of the performance measures. In this section we consider the base case including heterogeneous customers. Hence, all parameter values are equal to those, or distributed accordingly, as stated in Table 2. Afterwards we will consider similar problem settings, but with different cost values associated with the strategic decisions, i.e. f_k and g_k, denoting the cost for opening depots and assigning vehicles, respectively. The relative difference between the strategic cost and the operational cost is of high relevance for the location-routing problem. If we consider the strategic cost to be very high compared to the routing cost, additional depots will only be opened for the sake of sufficient capacity. Otherwise, if the strategic cost is very low, additional depots will be opened to reduce the length of the needed routes. The latter scenario resembles more an uncapacitated FLP where additional depots are opened if and only if it reduces the total shipping cost. Table 5 reports the average results of the LRP over multiple instances and customer demand samples for different values of ρ.

Table 5: Results obtained from instances with base values.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	2246.570	0.000	93.356	86.059	2.392	19.20	1.686	3.434
0.1	2306.171	2.653	75.774	83.513	11.892	16.00	1.354	3.429
0.2	2382.798	6.064	64.104	75.880	14.031	10.50	0.838	3.102
0.3	2453.952	9.231	53.813	73.509	11.408	7.50	0.620	2.648
0.4	2531.682	12.691	43.187	73.025	8.300	4.30	0.346	2.641
0.5	2601.979	15.820	37.670	71.317	8.635	3.10	0.246	2.541
0.6	2668.891	18.798	32.157	68.761	7.970	2.20	0.178	2.684
0.7	2726.066	21.343	26.042	66.168	11.195	1.00	0.087	2.755
0.8	2822.812	25.650	17.522	63.559	18.903	0.10	0.011	2.766
0.9	2897.567	28.977	10.190	61.701	26.124	0.00	0.000	2.808
1.0	3033.944	35.048	0.000	57.513	1.582	0.00	0.000	2.611

The heterogeneity of the nominal customer demand removes the consistency in the number of customers contained in any route, compared to the previous results. As a result, the excess capacity of a vehicle will be smaller if this reduces the required number of vehicles. Hence, both the total sum of slack variables and θ are larger compared to Table 4. This lower level of robustness also negatively affects the performance for small values of ρ. This heterogeneity also implies that less additional vehicles are required to provide more robust solutions. Hence, the average performance of larger values of ρ is slightly better than before. We can conclude that a cost increase of 35.048% compared to the nominal problem solution guarantees a solution that is always feasible, while an increase of 20 - 25 % can provide a solution that will be close to always feasible. The percentage decrease in the total sum of slack variables and the performance measures against the increase in total cost are shown in Figure 4. It shows that the theoretical increase in robustness is close to linear for more costly solutions, while the actual robustness performance measures show a much steeper curve based on the random sampling.

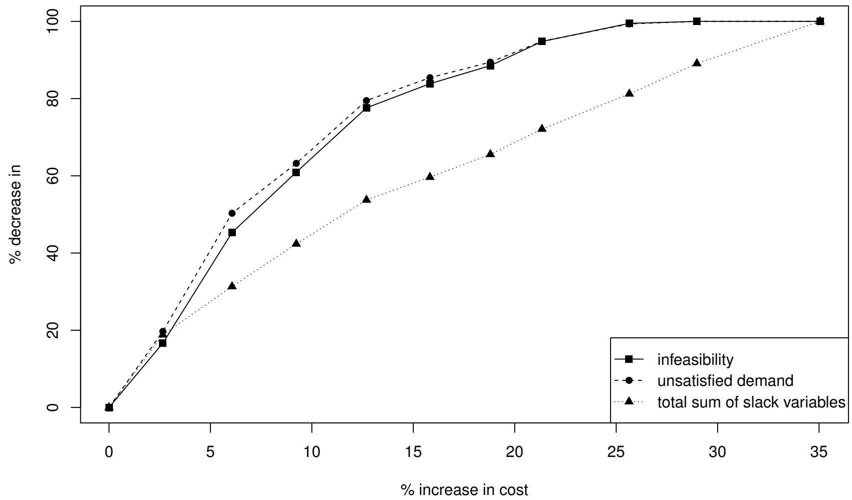

Figure 4: The effect of percentage increase in stage 1 cost, compared to $\overline{C}$ on percentage decrease in light robust objective value and the robustness measures

Tables 6 and 7 report the results in which the strategic cost values are relatively lower and higher, respectively. Specifically, in case of low strategic cost f_k and g_k are reduced by half to values $2.5D_k$ and $1.0Q$ for every depot $k \in \mathcal{N}_D$. For the high strategic cost these values are $20.0D_k$ and $8.0Q$, respectively

Table 6: Results obtained from instances with $f_k = 2.5D_k, g_k = Q$.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	1392.157	0.000	87.827	84.513	2.240	15.65	1.330	3.200
0.1	1422.277	2.164	70.853	81.113	6.202	11.95	0.980	3.018
0.2	1470.116	5.600	54.878	71.825	7.776	5.45	0.423	2.557
0.3	1514.071	8.757	44.912	69.668	7.773	4.05	0.329	2.554
0.4	1562.292	12.221	36.684	68.135	8.863	2.10	0.168	2.589
0.5	1603.907	15.210	30.907	66.172	9.035	1.30	0.104	2.618
0.6	1629.256	17.031	23.286	65.089	10.791	0.85	0.075	2.717
0.7	1676.723	20.441	18.145	62.454	10.614	0.30	0.029	2.651
0.8	1720.320	23.572	12.732	59.876	15.797	0.15	0.016	2.781
0.9	1765.053	26.785	6.711	58.321	25.429	0.00	0.000	2.788
1.0	1826.120	31.172	0.000	57.607	1.710	0.00	0.000	2.665

Table 7: Results obtained from instances with $f_k = 20D_k, g_k = 8Q$.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	7556.330	0.000	93.801	86.587	3.867	17.40	1.483	2.881
0.1	7720.611	2.174	79.027	86.059	30.538	15.75	1.329	3.034
0.2	7954.420	5.268	66.646	75.880	29.422	10.20	0.840	2.753
0.3	8228.666	8.898	55.871	73.536	50.399	6.60	0.517	2.664
0.4	8429.744	11.559	49.245	72.729	28.877	4.75	0.364	2.516
0.5	8668.333	14.716	38.201	72.542	12.427	2.00	0.154	2.280
0.6	8826.506	16.809	33.758	71.731	12.590	1.65	0.136	2.373
0.7	9062.553	19.933	31.295	69.731	25.975	1.45	0.122	2.387
0.8	9309.628	23.203	23.504	66.551	41.603	0.45	0.046	2.556
0.9	9609.433	27.171	12.243	62.541	46.409	0.00	0.000	2.522
1.0	10275.146	35.981	0.000	57.752	2.300	0.00	0.000	2.376

We can observe a few differences in terms of performance. First, a better performance is obtained for higher values of ρ for the instances with lower strategic cost. This is because the maximum available increase in cost allows for the opening and assigning of relatively more vehicles. This effect is also apparent in the relatively low improvements for lower values of ρ in Table 7. An explanation is that in many instances, the small additional increase in allowed cost does not enable an additional depot to be opened or a vehicle to be assigned. This is confirmed by the more gradual decrease in θ for lower strategic cost values.

Figure 5 shows the relative increase in the minimum capacity of opened depots and assigned vehicles, compared to the minimum capacity for the solution to the nominal problem with normal strategic cost. Similar to θ, the minimum capacity is defined as $\min\{\sum_k D_k y_k, \sum_k Q v_k\}$. It shows that lower strategic cost result in larger minimum capacities. Since the same demand values are taken into consideration, this implies that relatively more depots and vehicles are used to reduce the total routing cost.

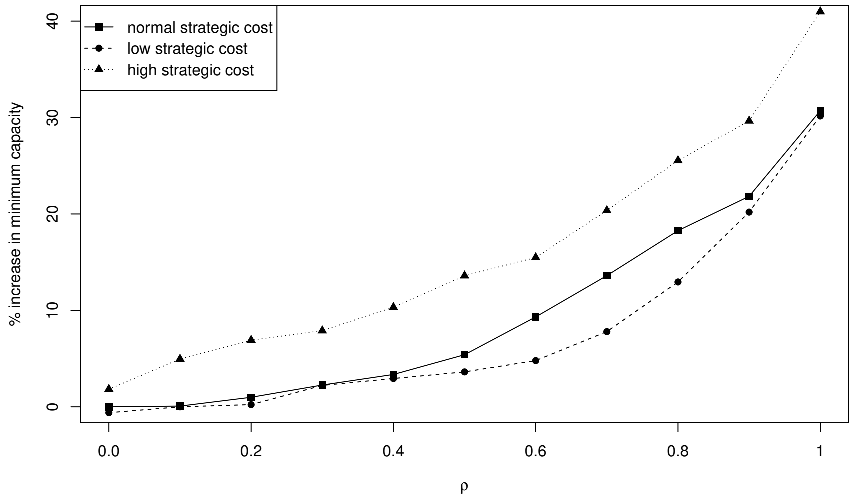

Figure 5: The relative increase in minimum capacity compared to the minimum capacity for $\rho = 0$ with normal strategic cost, for different values of strategic cost.

We can now compare the improvement in performances for more costly solutions for the different problems from this and the previous section, see Figure 6. For the variety of problems we can make the following important observation: both the percentage of unsatisfied demand and infeasible scenarios reduce quickly when using more expensive cost. However, this improvement in performance diminishes when solutions get more expensive, in the figure around 20% for all problems. From this we can conclude that, for all different problems, we seem to be able to find non-conservative robust solutions.

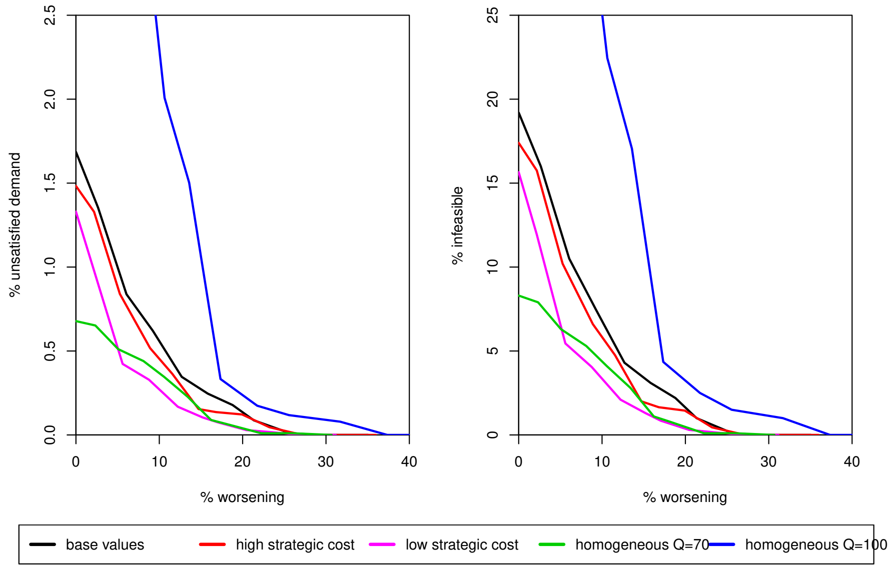

Figure 6: Problem comparison of the effect increase in cost compared to the nominal solution on percentage unsatisfied demand and infeasible scenarios.

5.3 Results from classical instances from literature

Table 8 reports on the performance of several (subsets) of benchmark instances consisting of $\{16, 20, 25\}$ customers and five potential depot locations. We compare the performance using the valid inequalities introduced in Section 3.3.4 with the performance in absence of these valid inequalities. The valid inequalities adopted from previous papers (i.e. constraints (26)-(28)) are contained in both formulations, since they have already been proven to yield significant improvements in running time (Karaoglan et al. (2011)).

For every instance we report the number of customers and depot locations. In terms of performance, we report the computation times (in seconds) and the optimality gaps (in %). The optimality gap is the percentage deviation between the best known feasible solution (upper bound) and the best known lower bound. Each instance is interrupted after one hour if an optimal solution has not yet been. The instances are solved for values $\rho = \{0.0, 0.5, 1.0\}$, due to the requirement to solve the nominal and strict robust problem before solving the light robust problem. The column names $\{P, A\} - \rho$ denote the presence or absence of the valid inequalities and the value for ρ. In case the deterministic problems ($\rho = 0$ or 1) have not been solved to optimality within an hour, we use the values $\overline{C}$ and $\widehat{C}$ obtained from the cases (either presence or absence of VI) that obtained the smallest optimality gap after an hour. This ensures that the problem at hand for $\rho = 0.5$ remains the same for both cases with and without the valid inequalities.

Table 8: Results obtained from benchmark instances of varying sizes with the addition and absence of valid inequalities.

Instance	$\lvert\mathcal{N}_C\rvert$	$\lvert\mathcal{N}_D\rvert$	P-0.0		A-0.0		P-0.5		A-0.5		P-1.0		A-1.0	
name			Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap
20-5-1a	16	5	97.98	0.00	94.44	0.00	32.57	0.00	99.50	0.00	25.78	0.00	26.13	0.00
20-5-1b	16	5	3.21	0.00	2.81	0.00	42.46	0.00	37.43	0.00	29.90	0.00	68.54	0.00
20-5-2a	16	5	47.52	0.00	44.17	0 .00	544.05	0.00	1194.26	0.00	14.86	0.00	12.72	0.00
20-5-2b	16	5	3.37	0.00	2.61	0.00	20.10	0.00	14.39	0.00	9.83	0.00	4.95	0.00
50-5-3	16	5	12.86	0.00	10.80	0.00	105.35	0.00	273.98	0.00	57.88	0.00	40.07	0.00
Avg			32.98	0.00	30.97	0.00	148.91	0.00	323.91	0.00	27.65	0.00	30.48	0.00
20-5-1a	20	5	497.80	0.00	418.98	0.00	1388.88	0.00	3600*	1.34	532.36	0.00	719.12	0.00
20-5-1b	20	5	27.61	0.00	25.89	0.00	706.45	0.00	1063.35	0.00	43.02	0.00	48.26	0.00
20-5-2a	20	5	751.57	0.00	887.43	0.00	41.35	0.00	2648.52	0.00	1521.04	0.00	2422.24	0.00
20-5-2b	20	5	23.07	0.00	11.50	0.00	88.84	0.00	1607.20	0.00	217.96	0.00	467.04	0.00
50-5-3	20	5	295.74	0.00	1570.28	0.00	198.95	0.00	2143.13	0.00	1458.45	0.00	2768.71	0.00
Avg			319.16	0.00	582.82	0.00	484.89	0.00	2212.44	0.27	754.57	0.00	1285.08	0.00
50-5-1	25	5	2464.07	0.00	3600*	1.45	3600*	45.36	3600*	64.45	1023.76	0.00	1724.13	0.00
50-5-1b	25	5	2225.78	0.00	3138.40	0.00	1092.57	0.00	3600*	100.00	3600*	4.61	3600*	4.29
50-5-2a	25	5	3600*	1.54	3600*	1.94	3600*	41.18	3600*	100.00	3600*	0.99	3600*	1.32
50-5-2b	25	5	242.13	0.00	230.17	0.00	3600*	25.18	3600*	100.00	803.11	0.00	1693.49	0.00
50-5-3	25	5	3600*	3.47	3600*	2.78	1013.95	0.00	3600*	38.78	3600*	2.36	3600*	2.32
Avg			2426.4	1.00	2833.71	1.24	2581.31	22.34	3600*	80.65	2525.37	1.59	2843.52	1.59

From Table 8 we can observe that the addition of the valid inequalities results on average in smaller computation times. Only for the instances containing 16 customers and $\rho = 0$ there is no improvement. For instances with 16, 20 and 25 customers, the number of benchmark instances where the presence of the valid inequalities finds an optimal solution faster is 5, 12 and 6, respectively, while the absence of valid inequalities is faster 10, 3 and 1 times, respectively. In the remaining 8 instances where both methods do not find the optimal solution within an hour, the case with valid inequalities provides, on average, smaller optimality gaps.

Furthermore, three out of the five largest instances still have a lower bound equal to zero after an hour if no valid inequalities are used, resulting in an optimality gap of 100%. This implies that the valid inequalities help in finding stricter lower bounds. Recall that all valid inequalities can be applied to the light robust problem. In the nominal problem $\rho = 0$ and the strict robust problem $\rho = 1$, however, we can only include the symmetry breaking constraints and the two- and three-customer subtour elimination constraints. The remaining valid inequalities all relate to slack variables γ_{ij}. It can be observed that on average, the solving times show a larger improvement when $\rho = 0.5$. The same can be expected for every value $\rho \in (0, 1)$.

5.4 Results on LRP with uncertain customer set

We now consider the problem with deterministic demand values, but uncertainty whether customers have a positive demand or not. We first consider a problem with 10 potential customers and probability p to have a positive demand. If a customer has a positive demand, we still sample a known demand value $d_i \sim U[10, 30]\ \forall i \in \mathcal{N}_C$. We present the results for the problem with uncertainty in the customer set with $p = 0.5$ in Table 9.

Table 9: Results obtained from instances with uncertainty in the customer set.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	1452.505	0.000	98.156	72.886	0.319	24.85	8.979	2.076
0.1	1510.070	3.963	70.587	70.721	1.379	22.25	7.148	2.029
0.2	1576.903	8.564	56.464	63.211	2.037	15.30	4.172	1.877
0.3	1664.143	14.571	42.269	56.294	2.551	13.60	3.746	1.099
0.4	1743.340	20.023	28.299	56.218	2.857	10.75	2.889	1.205
0.5	1812.308	24.771	20.770	55.698	2.826	9.30	2.517	1.123
0.6	1870.316	28.765	15.629	55.285	2.370	8.20	2.139	1.146
0.7	1917.931	32.043	12.314	55.212	2.529	7.30	1.983	1.980
0.8	1977.356	36.134	9.744	53.763	2.408	5.95	1.530	0.998
0.9	2047.693	40.977	7.031	52.938	2.092	5.10	1.247	1.016
1.0	2246.570	54.669	0.000	43.554	1.503	0.00	0.000	1.078

Recall that the total sum of weighted slack variables in the case of uncertainty in the customer set denotes $\sum_i d_i \gamma_i$. This is equal to the total demand of customers that are not included in the problem. For instances of 10 potential customers, $p = 0.5$, and $\bar{d}_i \sim U[10, 30]$, as in Table 9, this can be seen with the total weighted sum of slack variables for $\rho = 0$ being close to 100. As expected, all potential customers are included in the strict robust problem. Furthermore, the running times are naturally smaller compared to the case of demand uncertainty for both stages. This is due to the smaller problem size. Since the number of customers in both the nominal problem and every sampled scenario are independently binomially distributed, i.e. $\sum_i a_i \sim \mathrm{B}(|\mathcal{N}_C|, p)$, we know that $E(\sum_i a_i) = p|\mathcal{N}_C|$.

According to the performance measures, the solutions are not as robust compared to previous results while the tightness ratios θ are smaller, see Table 5. This can also be devoted to the smaller instance sizes. A relatively large deviation in number of successes from the expected value is more common when the number of trials is low. We examine this by performing a similar experiment with a larger number of potential customers. Table 10 presents the results from the same test environment, with the exception that $|\mathcal{N}_C| = 12$. We can indeed observe a relatively better performance for larger values of ρ, while the values for θ are very similar. This indicates that for larger instances, we are able to find non-conservative solutions, that are less expensive than the optimal solution to the strict robust problem.

Table 10: Results obtained from instances with 12 potential customers with uncertainty in the customer set.

	(P1LR)					(P2)		
ρ	Tot. cost	% Wors.	Tot. slack	θ	Time	% Infeas.	% Uns.	Time
0.0	1583.900	0.000	117.168	74.534	0.512	30.25	8.899	8.656
0.1	1653.316	4.383	87.866	74.534	4.525	30.00	8.869	8.753
0.2	1773.729	11.985	64.312	64.464	8.565	19.35	4.590	3.097
0.3	1859.997	17.431	42.082	64.017	16.238	16.70	3.724	2.069
0.4	1950.700	23.158	31.105	64.017	22.021	13.75	2.734	2.111
0.5	2059.616	30.034	18.840	61.713	15.189	9.25	1.736	2.215
0.6	2152.380	35.891	9.539	61.760	13.192	8.90	1.707	2.203
0.7	2224.560	40.448	5.541	60.045	8.078	6.75	1.210	2.193
0.8	2337.945	47.607	5.284	55.338	21.477	4.10	0.647	2.551
0.9	2350.867	48.423	3.421	56.751	2.281	2.65	0.483	2.493
1.0	2592.477	63.677	0.000	42.857	2.977	0.00	0.000	2.470

6 Conclusion and further research

We have introduced a two-stage light robust problem formulation for the LRP with uncertainty in customer demand. To the best of our knowledge, this is the first time any type of robust optimization on an LRP has been studied. Using extensive random scenario sampling on a variety of problem settings, we have shown that the light robust problem can provide non-conservative solutions that achieve good performances in terms of satisfied demand. We have presented a set of valid inequalities designed specifically for light robust optimization applied to an LRP and have shown that it improves the performance of, in particular, larger-scaled widely used benchmark instances. However, a smaller computation time cannot be guaranteed. We also applied light robust optimization to a similar LRP with uncertainty in the customer set, which yielded comparable results in terms of performance as the case of uncertainty in customer demand.

Further research directions may include applying a similar light robust approach to the VRP. Since the LRP is a general version of the VRP, this might provide a similar performance on problems of a larger scale. In order to improve the computation time even further, it would be beneficial to introduce a heuristic to solve large instances of the LRP with uncertainty in demand. Within the field of robust LRPs, we might consider uncertainty in routing cost. This would provide a realistic model since travel times tend to deviate from their nominal values. Alternatively, we could use different types of robust optimization for the case with uncertainty in demand and compare the obtained performance and computation times. For example, we can consider a three-index flow formulation for which the constraints include summations over the uncertain demand values and introduce a budget uncertainty set.

A List of symbols

Notation	(*Domain*)	*Description*
Acronyms		
VRP		Vehicle routing problem
MDCVRP		Multi-depot capacitated vehicle routing problem
FLP		Facility location problem
LRP		Location-routing problem
MIP		Mixed-integer programming
Sets		
$\mathcal{N}_C$	$i, j \in \{1, \ldots, \|\mathcal{N}_C\|\}$	Set of customers.
$\mathcal{N}_D$	$k \in \{1, \ldots, \|\mathcal{N}_D\|\}$	Set of potential depots.
$\mathcal{N}$	$i, j \in \{1, \ldots, \|\mathcal{N}\|\}$	Set of all nodes, combines customers and potential depots.
Variables		
x_{ij}	$\{0, 1\}$	Equals 1 if node $i \in \mathcal{N}$ is followed by node $j \in \mathcal{N}$ on a route, 0 otherwise.
y_k	$\{0, 1\}$	Equals 1 if location $k \in \mathcal{N}$ is opened, 0 otherwise.
z_{ik}	$\{0, 1\}$	Equals 1 if customer $i \in \mathcal{N}_C$ is assigned to depot $k \in \mathcal{N}_D$, 0 otherwise.
u_{ij}	$\mathbb{R}_+$	The vehicle load when traveling from node $i \in \mathcal{N}$ to $j \in \mathcal{N}$.
v_k	$\mathbb{Z}_+$	The number of vehicles assigned to depot k.
γ_{ij}	$\mathbb{R}_+$	Slack variable.
Parameters		
c_{ij}	$\mathbb{R}_+$	Travel cost from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$.
f_k	$\mathbb{R}_+$	Cost for opening location $k \in \mathcal{N}_D$.
g_k	$\mathbb{R}_+$	Cost per vehicle assigned to location $k \in \mathcal{N}_D$.
p_i	$\mathbb{R}_+$	Cost per unit of unsatisfied demand for customer $i \in \mathcal{N}_C$.
d_i	$\mathbb{R}_+$	Demand of customer $i \in \mathcal{N}_C$.
$\bar{d}_i$	$\mathbb{R}_+$	Nominal demand of customer $i \in \mathcal{N}_C$.
$\hat{d}_i$	$\mathbb{R}_+$	Maximum absolute deviation in demand of customer $i \in \mathcal{N}_C$.
Q	$\mathbb{R}_{++}$	Vehicle capacity.
D_k	$\mathbb{R}_{++}$	Capacity of depot $k \in \mathcal{N}_D$.
ρ	$[0, 1]$	Relative deviation between nominal and strict robust total cost.
$\overline{C}$	$\mathbb{R}_+$	Nominal problem objective value
$\widehat{C}$	$\mathbb{R}_+$	Strict robust problem objective value
a_i	$\{0, 1\}$	Equals 1 if customer $i \in \mathcal{N}_C$ has positive demand, 0 otherwise.
Functions		
$F(y, v, d)$		Second stage cost function with uncertain demand.
$F(y, v, a)$		Second stage cost function with an uncertain customer set.

Bibliography

Albareda-Sambola, M., Fernández, E., and Laporte, G. (2007). Heuristic and lower bound for a stochastic location-routing problem. *European Journal of Operational Research*, 179(3):940–955.

Baumgartner, A., Bauschert, T., Blzarour, A. A., and Reddy, V. S. (2017). Network slice embedding under traffic uncertainties—a light robust approach. In *Network and Service Management (CNSM), 2017 13th International Conference on*. IEEE.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

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

Ben-Tal, A. and Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13.

Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3):411–424.

Bertsimas, D. and Caramanis, C. (2010). Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations research*, 52(1):35–53.

Bertsimas, D. and Thiele, A. (2006). Robust and data-driven optimization: Modern decision-making under uncertainty. *INFORMS Tutorials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*, 3.

Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer Science & Business Media.

Cao, E., Lai, M., and Yang, H. (2014). Open vehicle routing problem with demand uncertainty and its robust strategies. *Expert Systems with Applications*, 41(7):3569–3575.

Carrizosa, E., Goerigk, M., and Schöbel, A. (2017). A biobjective approach to recoverable robustness based on location planning. *European Journal of Operational Research*, 261(2):421–435.

Contardo, C., Cordeau, J.-F., and Gendron, B. (2013). A computational comparison of flow formulations for the capacitated location-routing problem. *Discrete Optimization*, 10(4):263–295.

Drexl, M. and Schneider, M. (2015). A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, 241(2):283–308.

Fischetti, M. and Monaci, M. (2009). Light robustness. In *Robust and Online Large-Scale Optimization*, pages 61–84. Springer.

Fischetti, M., Salvagnin, D., and Zanette, A. (2009). Fast approaches to improve the robustness of a railway timetable. *Transportation Science*, 43(3):321–335.

Goerigk, M., Schmidt, M., Schöbel, A., Knoth, M., and Müller-Hannemann, M. (2013). The price of strict and light robustness in timetable information. *Transportation Science*, 48(2):225–242.

Goerigk, M. and Schöbel, A. (2016). Algorithm engineering in robust optimization. In *Algorithm engineering*, pages 245–279. Springer.

Gounaris, C. E., Wiesemann, W., and Floudas, C. A. (2013). The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693.

Karaoglan, I., Altiparmak, F., Kara, I., and Dengiz, B. (2011). A branch and cut algorithm for the location-routing problem with simultaneous pickup and delivery. *European Journal of Operational Research*, 211(2):318–332.

Koç, Ç., Bektaş, T., Jabali, O., and Laporte, G. (2016). The fleet size and mix location-routing problem with time windows: Formulations and a heuristic algorithm. *European Journal of Operational Research*, 248(1):33–51.

Kuhn, K., Raith, A., Schmidt, M., and Schöbel, A. (2016). Bi-objective robust optimisation. *European Journal of Operational Research*, 252(2):418–431.

Laporte, G., Louveaux, F., and Mercure, H. (1989). Models and exact solutions for a class of stochastic location-routing problems. *European Journal of Operational Research*, 39(1):71–78.

Laporte, G., Nobert, Y., and Arpin, D. (1986). An exact algorithm for solving a capacitated location-routing problem. *Annals of Operations Research*, 6(9):291–310.

Liebchen, C., Lübbecke, M., Möhring, R., and Stiller, S. (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In *Robust and Online Large-Scale Optimization*, pages 1–27. Springer.

Miller, C. E., Tucker, A. W., and Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, 7(4):326–329.

Min, H., Jayaraman, V., and Srivastava, R. (1998). Combined location-routing problems: A synthesis and future research directions. *European Journal of Operational Research*, 108(1):1–15.

Nagy, G. and Salhi, S. (2007). Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2):649–672.

Perl, J. and Daskin, M. S. (1985). A warehouse location-routing problem. *Transportation Research Part B: Methodological*, 19(5):381–396.

Prins, C., Prodhon, C., and Calvo, R. W. (2006). Solving the capacitated location-routing problem by a grasp complemented by a learning process and a path relinking. *4OR*, 4(3):221–238.

Prodhon (2006). Classical instances for LRP. `http://prodhonc.free.fr/Instances/instances_us.htm`. Online; accessed 20-07-2018.

Prodhon, C. and Prins, C. (2014). A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1):1–17.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences*, 42(1):43–47.

Salhi, S. and Rand, G. K. (1989). The effect of ignoring routes when locating depots. *European Journal of Operational Research*, 39(2):150–156.

Schöbel, A. (2014). Generalized light robustness and the trade-off between robustness and nominal quality. *Mathematical Methods of Operations Research*, 80(2):161–191.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157.

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

Yanıkoğlu, İ., Gorissen, B., and den Hertog, D. (2017). Adjustable robust optimization–a survey and tutorial. *Available online at ResearchGate*.

Zeng, B. and Zhao, L. (2013). Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461.