CIRRELT

CIRRELT-2023-16

Facility Location: A Guide to Modeling and Solving Complex Problem Variants via Lagrangian Relaxation

Sanjay Dominik Jena

April 2023

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

Facility Location: A Guide to Modeling and Solving Complex Problem Variants via Lagrangian Relaxation

Sanjay Dominik Jena*

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Analytics, Operations and Information Technologies Department, School of Management, Université du Québec à Montréal

Abstract. Facility Location problems fit a large variety of practical planning contexts and are among the most studied combinatorial optimization problems. While these problems may become quite complex in certain applications, they are particularly well tackled by mathematical decomposition via Lagrangian Relaxation. This paper provides a guide to modeling and solving complex facility location problem variants via Lagrangian Relaxation. It first reviews the problem variants successfully tackled by Lagrangian Relaxation. It then guides the development of strong mixed-integer programming formulations for a variety of problem variants, covering multi-period models, a wide range of capacity constraints, modular facility structures, facility relocation, and parameter uncertainty. Finally, it discusses how such variants can be efficiently solved via Lagrangian Relaxation, capable of providing tight lower and upper bounds in short computing times.

Keywords: Lagrangian relaxation, facility location, modeling and solution guide

Acknowledgements. This paper is dedicated to Bernard Gendron, my former Ph.D. supervisor, colleague and friend. His passion and relentless search for stronger model formulations and more efficient solution methods are reflected within all developments here summarized. Bernard is not only a co-author of many of the here cited papers. Having strongly shaped my reasoning and passion for research makes him implicitly part of any other of my research outcomes in the domains of mixed-integer formulations, facility location and Lagrangian relaxation. I would also like to thank all colleagues and students who have inspired my research on these topics in one way or another, in particular, Ivan Contreras, Jean-François Cordeau, Antonio Frangioni, Enrico Gorgone, Peter Schütz and Sarka Stadlerova.

---

* Corresponding author: san.jena@gmail.com

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2023

1 Introduction

Facility location problems aim at finding the optimal locations to place facilities that are required to satisfy customer demand. Among the most simple and classical problem variants is the *Capacitated Facility Location Problem*, which considers as input a set of customer demands, along with a set of candidate facility locations, production capacities and transportation costs. The problem consists in selecting a subset of the candidate locations to construct production facilities, from which the customer demand is then served. As such, the problem balances the trade-off between investment and operational costs. While constructing too many facilities is likely to be costly, too few facilities may result in large (and therefore expensive) transportation distances to serve customer demands.

Facility location problems are among the most studied NP-hard combinatorial optimization problems, given that they realistically represent the planning contexts in many application domains, including production planning, telecommunication (see, e.g. Chardaire and Sutter, 1996), health care (see, e.g. Vahidnia et al., 2009), education (see, e.g. Antunes et al., 2009) and forestry (see, e.g. Jena et al., 2015b). While initial facility location problems were rather simpler, the degree of realism to which facility location problems have been modeled has drastically increased over the years. Planning problems nowadays may involve the selection of production equipment and capacity, multiple time-periods, the adjustment of production capacity over time, the relocation of facilities, multiple commodities, refined cost functions, the representation of parameter uncertainty and, in the context of more complex supply chains, distributions networks spanning over several echelons.

Even though multi-purpose optimization solvers have greatly evolved, the complexity and scale of the planning problems have outgrown their capacity to solve such problems in sufficiently short computing times. As a remedy, mathematical decomposition has been successfully applied to a variety of combinatorial optimization problems. In the context of facility location problems, Lagrangian relaxation has shown to be an efficient approach, if applied correctly. However, given the variety of facility location problems, developing an appropriate optimization model and the right ingredients required for a Lagrangian heuristic can be a challenging task. This paper aims at guiding the reader through the process of developing both an appropriate optimization model and the corresponding Lagrangian heuristic. To this end, we will first focus on the development of a strong optimization model. We will here focus on Mixed-integer Programming models, which have been by far the predominant modelling paradigm in this domain, particularly due to its flexibility. We then explain how Lagrangian relaxation can be applied and how its ingredients can be tailored to the specific problem variant at hand.

Lagrangian Relaxation. Lagrangian Relaxation, also referred to as Lagrangian decomposition, belongs to the class of mathematical decomposition methods. As such, it has been developed on the premise of exploiting the structure of the constraint coefficient matrix of the optimization model. While each method has its advantages and disadvantages, Lagrangian decomposition tends to be particularly promising when a large part of the constraint matrix is organized in a block structure, i.e., most of the constraints have non-zero coefficients organized in blocks, while only a few constraints link these blocks. The latter type of constraints, those that have non-zero coefficients simultaneously in several blocks, may be called the "complicating constraints". Lagrangian relaxation aims at relaxing the complicating constraints, and transfers them into the objective function by penalizing their violation multiplied by weights, called the *Lagrangian dual values* (or, multipliers) of the relaxed constraints. The resulting optimization problem is called the *Lagrangian subproblem*. Given that the complicating constraints have been eliminated from the constraint matrix, the problem can be decomposed into independent blocks. Ideally, solving each of these blocks is now easy and can be performed by a specialized algorithm. In most facility location problems, the only constraints that link the different blocks are the demand constraints, one for each candidate facility location. Relaxing the demand constraints therefore yields subproblems that are easy to solve, since the opening decision for each facility can be made independently from other locations.

The solution of the Lagrangian subproblem is, however, unlikely to be feasible to the original problem. Most likely, the relaxed constraints are not satisfied. Theoretically, a set of Lagrangian dual values exist such that the solution to the Lagrangian subproblem is also perfectly feasible to the original problem. This set of optimal Lagrangian dual values is the result of solving the so-called *Lagrangian dual problem*, which optimizes over the space of Lagrangian dual values. In practice, the optimal Lagrangian dual values are

hardly found. Instead, one aims at converging to a set of Lagrangian dual value such that the violation of the constraints is as small as possible.

Throughout the iterative process of improving the Lagrangian duals, the Lagrangian solution may be used to derive *solutions that are feasible* to the original problem. Naturally, the challenge here lies within the reconstitution of the relaxed constraints. As such, the employed process is an heuristic. Further, given that the Lagrangian subproblem is a relaxation of the original (minimization) problem, the optimal solution to any Lagrangian subproblem constitutes a *lower bound* to the original problem. A Lagrangian heuristics therefore typically produces both upper bounds (i.e., feasible solutions) and lower bounds. Unless this procedure is embedded in an exact algorithm (such as Branch-and-Bound), it remains an heuristic. In order to provide a bound on the quality of the found upper bounds, it is important to find the best possible lower bound. This lower bound is tyically impacted by the strength of the underlying formulation. It is therefore desirable to use a formulation that is as strong as possible.

Throughout this paper, we aim at illustrating how to solve the Lagrangian dual, and how to tailor the solution of the Lagrangian subproblem and the generation of feasible upper bounds to the various facility location problem variants. For an in-depth discussion of Lagrangian Relaxation itself, as well as valuable related insights, we refer the reader to the excellent introduction of Guignard (2003).

A review of Lagrangian Relaxation for Facility Location. Facility location problems have been solved by Lagrangian Relaxation since over 30 years. To the best of our knowledge, up to today, Beasley (1993) has been the only work proposing a framework to apply Lagrangian Relaxation to location problems. The authors considered the classical Capacitated Facility Location Problem and explored the two principal relaxations, which, until today, remain the most promising: relaxing the capacity constraints, or relaxing the demand constraints. Relaxing the capacity constraints mostly makes sense when tackling a simple location problem, in which no additional capacity decisions are involved. The resulting subproblem is an uncapacitated location problem, which itself is NP-hard. As a result, relaxing the demand constraints is a more popular approach, which allows, in most cases, to decompose the problem by candidate facility locations.

Table 1 reviews most of the papers that applied Lagrangian relaxation to some sort of facility location problem. In particular, it illustrates the characteristics of the tackled facility location problems. While not necessarily exhaustive, the table aims at illustrating how both the complexity of the planning problems and their corresponding Lagrangian relaxations have evolved over time.

We next explain the notation used in Table 1 to characterize the planning problems. Throughout the abbreviations used to define the characteristics of the planning problem, lower-case letters indicate that the characteristic results in an easier-to-solve problem, while upper-case letters indicate that the problem becomes more difficult to solve (typically, since it becomes more complex). The first column refers to the planning horizon: a problem has either a **(s)**ingle or **(M)**ultiple time-periods. The second column refers to the network structure over which commodities are routed. We here focus on **(s)**imple 2-echelon problems, i.e., those with two levels, typically constituted by a layer of facilities and a layer of customers. For the sake of completeness, we also include a few works with **(M)**ultiple layers (including, for example, a layer of warehouses) and those with a **(G)**eneral network structure. The following column indicates whether a **(s)**ingle or **(M)**ultiple commodities are considered. The next set of columns indicates what kind of capacity restrictions facilities are subject to. If none of the subcolumns is checked, then the problem is uncapacitated. Capacitated problems either have a **(f)**ixed pre-defined capacity or allow to select the capacity from a pre-defined set of capacity **(L)**evels. Some works also allow to install **(M)**ultiple facilities at the same location. Finally, in multi-period settings, the capacity of facilities may be **(E)**xpanded or **(R)**educed once, or even be **(A)**djusted several times throughout the planning horizon. Problems may also involve some particular **(I)**ndustrial capacity constraints in order to represent the application context more realistically. The next column indicates whether customers can be served by several facilities or only by a single one, also known as **(m)**ulti- and **(S)**ingle-sourcing, respectively. The latter involves that allocation variables are binary, which significantly complicates the solution of the problem. Single-sourcing also complicates the solution via Lagrangian Relaxation, since the demand allocation for each facility in the Lagrangian subproblem cannot be simply solved as a transportation problem. The function used to describe the production costs is indicated in the next column: either it is a **(s)**imple linear cost-function, or it is more **(C)**omplex, often involving piecewise linear functions. The next two columns indicate whether the problem allows for facility

Paper	Time per.	Echelons	Commodities	Capacity constr. f	L	M	E	R	I	A	Sourcing	Cost func.	Relocation	Uncertainty	Relaxed constr.
Guignard-Spielberg and Kim (1983)	s	s	s	✓							S	s			S
Barcelo et al. (1990)	s	s	s	✓							S	s			D
Sridharan (1991)	s	s	s	✓							m	s			D
Beasley (1993)	s	s	s	✓							m	s			D v C
Shulman (1991)	M	s	s	✓			✓				m	s			D
Chardaire and Sutter (1996)	M	s	s								m	s			D
Holmberg and Ling (1997)	s	s	M		✓						m	C			D
Agar and Salhi (1998)	s	s	s	✓							S	s			
Hinojosa et al. (2000)	M	M	M	✓							m	s			D
Correia and Captivo (2003)	s	s	s		✓						m	s			D
Wu et al. (2006)	s	s	M	✓		✓					m	s			D
Correia and Captivo (2006)	s	s	s		✓						S	s			D
Hinojosa et al. (2008)	s	M	M	✓							m	s			D & F
Li et al. (2009)	s	s	M	✓							m	s			C
Diabat et al. (2011)	M	s	s								S	s			D
Görtz and Klose (2012)	s	s	s	✓							m	s			D
Ghodsi (2012)	s	s	s	✓							m	s		✓	C & N
Gendron et al. (2016)	s	M	s								m	s			C
Jena et al. (2016)	M	s	M		✓		✓	✓	✓	✓	m	s	✓		D & R
Jena et al. (2017)	M	s	M		✓		✓			✓	m	s			D
Marín et al. (2018)	M	s	s	✓						✓	m	s		✓	D
Cabezas et al. (2021)	s	s	s								m	s		✓	D
Kadri et al. (2022)	s	G	M	✓							m	s			F
Štádlerová et al. (2023)	M	s	M		✓		✓				m	C		✓	D

Table 1: Characteristics of location problems tackled by Lagrangian relaxation

relocation and the representation of parameter uncertainty. Finally, the last set of columns indicates the constraints relaxed within the Lagrangian decomposition: **(D)**emand constraints, **(C)**apacity constraints, **(F)**low conservation constraints (only in multi-echelon problems), **(R)**elocation linking constraints (only when relocation is allowed) or **(N)**on-anticipativity constraints (only in the case of stochastic programming).

The indicated characteristics of the considered planning problems demonstrate that Lagrangian Relaxation has been consistently applied to new and more complex problem variants. In particular, complicating features, such as multiple time-periods, multiple commodities and more complex cost functions have been tackled more commonly in recent years. Most of the works have assumed predefined fixed production capacities for the facilities. In the last 20 years, more complex variants have been considered and solved, including the choice of the capacity level and the possibility of several facilities at the same location. Capacity expansion, reduction, and the adjustment of capacities along time are more recent integrations. Most problems also focus on the multi-sourcing problem variant, which is easier to solve than the single-sourcing variant (and also more common in the real planning practices). The proposed Lagrangian algorithms mostly relax the demand constraints. When problem variants are particularly complex, such as those including general network structures, relocation or parameter uncertainty, the corresponding complicating constraints often have also been relaxed. Literature on problem variants integrating such characteristics is rather recent.

Contributions, Scope and Outline. Given the diversity of problem variants and Lagrangian heuristics proposed in the last decades, this paper aims at providing a systematic guide on how to model complex facility location problems and solve them by means of Lagrangian heuristics. As such, it synthesizes content from three principal references (Jena et al., 2016, 2017; Štádlerová et al., 2023), but also provides a few new suggestions that have not yet been considered in the literature (e.g., a formulation to model piecewise linear cost functions that does not make assumptions on the shape of the function).

The paper focuses on modeling most of the characteristics above, but restricts to the case of multi-sourcing and 2-echelon problems. As such, we do not discuss the modeling or resolution of problems with multiple echelons or general transportation networks in supply chains (see, e.g., Allen et al., 2022). We therefore also exclude hub-location problems, which have been successfully tackled by Lagrangian relaxations (see, e.g., Contreras et al., 2009). Finally, it is worthwhile mentioning that Lagrangian Relaxation has also been applied to location problems based on discrete-choice models. A general framework is presented by Pacheco Paneque et al. (2022).

The structure of this paper is as follows. Section 2 first introduces a general problem formulation that accounts for multiple time-periods, multiple commodities, multiple capacity levels and the adjustment of capacity along time. The following subsections then discuss special cases and extend this model to account for complex production cost functions, modular facility structures, facility relocation and parameter uncertainty. Section 3 then reviews Lagrangian Relaxation and its components (i.e., the Lagrangian subproblem, the Lagrangian Dual, and the generation of feasible upper bounds) in the context of the base formulation. The following subsections then explain how these components can be adapted to the various problem variants. Finally, Section 4 concludes the paper.

2 Modeling Facility Location

We first review a general model in Section 2.1 that generalizes a variety of facility location problems and discuss the modeling elements that render such model stronger than alternative models. Section 2.2 then illustrates how the use of the model's parameters results in different special cases. Following, Section 2.3 discusses the use of more complex (specifically, piecewise linear) objective functions. Section 2.4 models complex facility structures, which may be partially and temporarily activated or closed. Section 2.5 focuses on a model extension that enables facility relocation. Finally, Section 2.6 elaborates on the representation of parameter uncertainty and the integration of different scenarios via stochastic programming.

2.1 A General Model

This section first reviews a general formulation (Jena et al., 2017) that encompasses a variety of different facility location problems, including those with multiple commodities, multiple time-periods, multiple capacity levels and allowing to adjust the capacity level throughout the planning horizon.

Facilities may be constructed at candidate facility locations indicated by a set J. A facility is constructed at one of the eligible capacities, given within a discrete set $L = \{0, 1, \ldots, \bar{\ell}\}$ of possible capacity levels, where level 0 indicates that no facility exists. Customer demand points are given by set I. The planning horizon contains a total of $\bar{t}$ time periods, included in set $T = \{1, 2, \ldots, \bar{t}\}$. Without loss of generality, we assume that the beginning of period $t + 1$ corresponds to the end of period $t \in T$. In addition, we make the assumption that all facility openings, closings and capacity changes are implemented at the beginning of a time period. Customer demands may be specified for different commodities, given within set P. The demand of customer $i \in I$ for commodity $p \in P$ in period $t \in T$ is denoted by d^t_{ip}, while the cost to serve one unit of commodity $p \in P$ from facility $j \in J$ operating at capacity level $\ell \in L$ to customer $i \in I$ during period $t \in T$ is denoted by $g^{jt}_{i\ell p}$. The capacity of a facility of size $\ell \in L$ at location $j \in J$ is given by u^j_ℓ (with $u^j_0 = 0$). The cost matrix $f^{jt}_{\ell_1\ell_2}$ describes the combined cost to change the capacity level of a facility at location $j \in J$ from $\ell_1 \in L$ to $\ell_2 \in L$ at the beginning of period $t \in T$ and to operate the facility at capacity level $\ell_2 \in L$ throughout time period $t \in T$. Furthermore, we let $\ell^j \in L$ be the initial capacity level of an existing facility at location $j \in J$.

The mathematical formulation uses binary variables $y^{jt}_{\ell_1\ell_2}$ equal to 1 if and only if the facility at location $j \in J$ changes its capacity level from $\ell_1 \in L$ to $\ell_2 \in L$ at the beginning of period $t \in T$. For a facility $j \in J$ open at capacity level $\ell \in L$, set $L^-(j, t, \ell) \subseteq L$ defines the capacity levels to which the capacity may be changed at the beginning of period $t \in T$. In a similar fashion, set $L^+(j, t, \ell) \subseteq L$ defines the capacity levels from which the capacity may be changed to level $\ell \in L$ at facility $j \in J$ at the beginning of period $t \in T$. Depending on the application context, only a limited set of capacity changes may be eligible in practice. As will be explained in the next section, a careful definition of sets L^- and L^+ therefore enables the model to represent different problem variants. The continuous allocation variables $x^{jt}_{i\ell p}$ denote the fraction of the demand of customer $i \in I$ for commodity $p \in P$ in period $t \in T$ that is served from a facility of size $\ell \in L$ located at $j \in J$. The resulting model is known as the *Generalized Modular Capacities (GMC)* formulation:

$$\text{(GMC)} \quad \min \sum_{j\in J}\sum_{t\in T}\sum_{\ell_1\in L}\sum_{\ell_2\in L^-(j,t,\ell_1)} f^{jt}_{\ell_1\ell_2} y^{jt}_{\ell_1\ell_2} + \sum_{i\in I}\sum_{j\in J}\sum_{\ell\in L}\sum_{p\in P}\sum_{t\in T} g^{jt}_{i\ell p} d^t_{ip} x^{jt}_{i\ell p} \tag{1}$$

$$s.t. \sum_{j\in J}\sum_{\ell\in L} x^{jt}_{i\ell p} = 1 \;\; \forall i \in I, \;\; \forall p \in P, \;\; \forall t \in T \tag{2}$$

$$\sum_{i\in I}\sum_{p\in P} d^t_{ip} x^{jt}_{i\ell p} \le \sum_{\ell_1\in L^+(j,t,\ell)} u^j_\ell y^{jt}_{\ell_1\ell} \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall t \in T \tag{3}$$

$$\sum_{\ell_1\in L^+(j,t-1,\ell)} y^{j(t-1)}_{\ell_1\ell} = \sum_{\ell_2\in L^-(j,t,\ell)} y^{jt}_{\ell\ell_2} \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall t \in T\setminus\{1\} \tag{4}$$

$$\sum_{\ell_2\in L^-(j,t,\ell^j)} y^{j1}_{\ell^j\ell_2} = 1 \;\; \forall j \in J \tag{5}$$

$$x^{jt}_{i\ell p} \le \sum_{\ell_1\in L} y^{jt}_{\ell_1\ell} \;\; \forall i \in I, \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall p \in P, \;\; \forall t \in T \tag{6}$$

$$x^{jt}_{i\ell p} \ge 0 \;\; \forall i \in I, \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall p \in P, \;\; \forall t \in T \tag{7}$$

$$y^{jt}_{\ell_1\ell_2} \in \{0, 1\} \;\; \forall j \in J, \;\; \forall \ell_1 \in L, \;\; \forall \ell_2 \in L^-(j, t, \ell), \;\; \forall t \in T. \tag{8}$$

The objective function (1) minimizes the total cost for changing the capacity levels, maintaining open facilities and serving the demand. Constraints (2) are the demand constraints for the customers. Constraints (3) are the capacity constraints at the facilities. For each location $j \in J$, the model tracks the current capacity level throughout the planning horizon. Constraints (4) link the capacity change variables in consecutive time periods, while Constraints (5) initialize the sizes of the facilities at the beginning of the planning horizon, ensuring that exactly one capacity level is selected. We may refer to such constraints as capacity flow conservation constraints, given that, for each facility location, a flow of 1 unit is conserved throughout the network to specify the capacity level defined over the various time periods.

Constraints (6), called the *Strong Inequalities (SI)*, ensure that no demand can be served from a facility of size $\ell \in L$ at period $t \in T$ if no such facility is available during period $t \in T$. Since the capacity constraints are also enforcing the same requirements, the SIs are redundant to the MIP model. Similar inequalities, which can be seen as a special case of flow cover inequalities (Padberg et al., 1983), are used in many facility

location and network design problems (e.g., Gendron, 2011; Chouman et al., 2016). In both facility location and network design formulations, such SIs tend to drastically improve the LP relaxation bound.

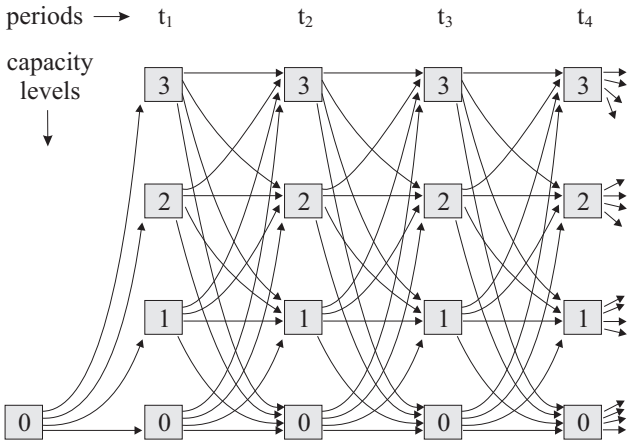


Figure 1: Network flow structure of capacity changes for a given facility.

Figure 4 illustrates the network that, for a given facility, manages the open capacity level at each of the time-periods. A flow of 1 unit is sent through the network by activating binary variables $y^{jt}_{\ell_1\ell_2}$. The use of such detailed variables may seem overly complicated on first sight, as there is a large number of binary variables if L is large. A more classical formulation would consider the simpler binary variables y^{jt}_{ℓ} that take value 1 if the facility $j \in J$ has capacity level $\ell \in L$ at throughout time period $t \in T$. In addition, such formulation uses additional binary variables indicating the expansion or the reduction of ℓ capacity levels. Both formulations are compared in Jena et al. (2015a). While the formulation using variables with a single ℓ-index consists of less variables, it tends to provide a larger integrality gap (i.e., it provides a weaker LP relaxation) than the GMC formulation. The main reason being that, in the GMC formulation, the OF coefficients $f^{jt}_{\ell_1\ell_2}$ incorporate the costs for both changing the capacity level and for maintaining the facility open at that level throughout the time-period. Constraints (4) are defined for each $\ell \in L$. This is not possible in the simpler formulation, where the LP relaxation solution may mix construction and maintenance variables with lowest costs. When developing a Lagrangian heuristic, one is therefore particularly interested in strong formulations which may allow the Lagrangian subproblem to provide stronger lower bounds.

Another advantage of using the more complex GMC formulation lies in its ability to represent more refined capacity change costs. While the simpler formulation represents only the costs to increase or decrease the capacity level by ℓ levels, the GMC formulations can encode more details within $f^{jt}_{\ell_1\ell_2}$, allowing, for example, to have different costs for a capacity expansion from level 1 to 3 then for an expansion from level 2 to 4, even though, in both cases, the capacity is increased by two levels.

2.2 Special Cases

The formulation presented above may be unnecessarily complex for several applications. By careful design of the sets used within this formulation, several special cases can be modelled. In particular, the single-commodity variant is modelled when set P holds only one commodity. The single-period problem variant is modelled when set T has only a single time-period. In this case, the set of eligible capacity levels for facility $j \in J$ is solely defined by $L^-(j, t = 0, \ell)$. If the problem is uncapacitated, i.e., facilities are assumed to have infinite capacity, Constraints (3) are not required. The strong inequalities (6) are then sufficient to ensure that only opened facilities can be used. Finally, if L holds only one capacity level in addition to capacity level 0, the model represents a problem variant with fixed capacity levels.

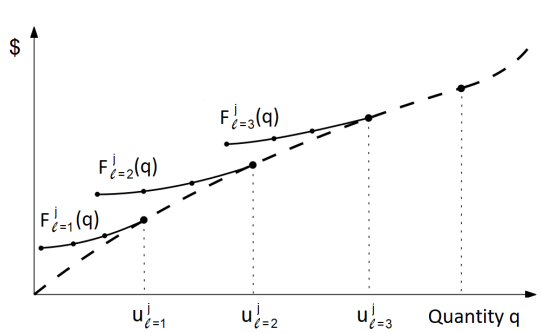
In the general case, if $L^- = L$ and $L^+ = L$, the formulation allows to expand and reduce from any capacity level to any other capacity level. Jena et al. (2015a) explore two problem variants as special cases. In one problem variant, facilities can be constructed at any capacity level. Once constructed, they can be temporarily closed when idle in order to avoid high maintenance costs. However, the available capacity at a given location cannot be changed. The other problem variant allows for expanding and reducing capacity

throughout the planning horizon. Here, a temporary closing is not possible, and one needs to completely shut down the facility (to level 0) in order to completely avoid maintenance costs. Both problem variants, as well as many more, can be modelled by appropriately defining L^- and L^+.

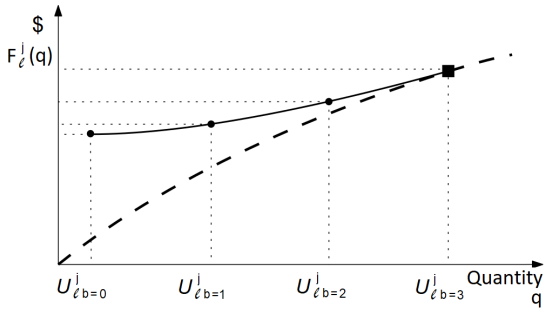
Finally, if minimum production quantities are required at the facilities, the total production quantity may exceed the amount transferred to the customers. Such excess capacity may incur penalties, which can be easily integrated into the model (see, e.g., Štádlerová et al., 2023).

2.3 Accounting for Complex Cost Functions

The facility location problem considered above makes the standard assumption that, for each capacity level $\ell \in L$, production costs are linear in function of the total production quantity, i.e., the per-unit production costs is constant for each level $\ell \in L$. Specifically, for a facility $j \in J$ and time period $t \in T$, parameter $g^{jt}_{i\ell p}$ defines both the per-unit production cost at capacity level $\ell \in L$ and the per-unit distribution costs for product $p \in P$ to customer $i \in I$. In practice, however, such costs may not be linear. In this case, a more accurate modeling of the cost function is particularly important for production costs (see, e.g., Christensen and Klose, 2021), which tend to be of higher order than transportation costs. In our case, such production costs may be defined differently for each capacity level. For example in the case of economies scale, the per-unit production costs would be a concave function.



(a) Separate functions for different capacity levels



(b) Piecewise-linear cost function for a specific capacity level $\ell \in L$

Figure 2: Exemple of a piecewise-linear production cost function (adapted from Štádlerová et al., 2023)

Štádlerová et al. (2022, 2023) consider a location problem in which equipment is placed to capture hydrogen. The type and quantity of capturing equipment installed defines the operating costs. As such, the proposed model uses different capacity levels, and for each level, a concave function is defined, representing the capturing costs for the total amount of hydrogen captured at that location. This cost function is approximated by a piecewise linear function. The total production costs are exemplified in Figure 2. Subfigure (a) illustrates the total production costs F that depend on both the selected capacity level $\ell \in L$ and the total production quantity q at that location. Subfigure (b) exemplifies a piecewise-linear function for a given capacity level $\ell \in L$: specifically, 3 line pieces defined by 4 breakpoints. While the authors propose a formulation tailored to the concave shape of the function, we will now show how any type of piecewise linear function can be used, assuming that the end of each linear piece is identical to the beginning of the next linear piece.

We separate the operational costs into two parts. Parameter $g^{jt}_{i\ell p}$ now exclusively refers to the per-unit transportation costs to distribute product $p \in P$ from facility $j \in J$ operating at level $\ell \in L$ to customer $i \in I$ during time period $t \in T$. In addition, we separately define the production costs by a cost function indicating the total costs to produce a quantity of q units. Specifically, for each $\ell \in L, \ell \geq 1$ and facility $j \in J$, we define a piecewise linear function $F^j_\ell(q)$ that determines the costs to produce a total quantity of q units at facility $j \in J$ operating at capacity level $\ell \in L$. This piecewise linear function consists of $\overline{b}^j_\ell$ line pieces, where $b = 1, \ldots, \overline{b}^j_\ell$ are the breakpoints of the production quantity defining the piecewise linear function. For each line piece $b \in [1, \overline{b}^j_\ell]$, the production quantity at the beginning of the piece is given by

$U^j_{\ell b}$, while the end of the piece is given by $U^j_{\ell b}$. As such, for a facility $j \in J$ operating at level $\ell \in L$, $U^j_{\ell(b=0)}$ is the minimum production level allowed, while $u^j_\ell = U^j_{\ell(b=\bar{b}^j_\ell)}$ is the maximum production level allowed.

Let $\overline{\mu}_{j\ell bt}$ be a binary variable that takes value 1 if facility $j \in J$ operating at level $\ell \in L$ produces a total quantity that falls into line piece b during time period $t \in T$. Further, let continuous variables $\mu^1_{j\ell bt}$ and $\mu^2_{j\ell bt}$ represent the linear combination between the start and the end of the selected line piece such that this linear combination corresponds to the total production quantity. The GMC model can be modified as follows in order to account for such complex production cost functions by adding the following constraints, where $[\bar{b}^j_\ell]$ denotes the sequential set $\left\{1, \ldots, \bar{b}^j_\ell\right\}$:

$$\sum_{b \in [\bar{b}^j_\ell]} \overline{\mu}_{j\ell bt} = \sum_{\ell_1 \in L} y^{jt}_{\ell_1 \ell} \quad \forall j \in J, \ \ \forall \ell \in L, \ \ \forall t \in T \tag{9}$$

$$\mu^1_{j\ell bt} + \mu^2_{j\ell bt} = \overline{\mu}_{j\ell bt} \quad \forall j \in J, \ \ \forall \ell \in L, \ \ b \in [\bar{b}^j_\ell], \ \ \forall t \in T \tag{10}$$

$$\sum_{i \in I} \sum_{p \in P} x^{jt}_{i\ell p} = \sum_{b \in [\bar{b}^j_\ell]} U^j_{\ell(b-1)} \mu^1_{j\ell bt} + U^j_{\ell b} \mu^2_{j\ell bt} \quad \forall j \in J, \ \ \forall \ell \in L, \ \ \forall t \in T \tag{11}$$

$$\overline{\mu}_{j\ell bt} \in \{0,1\} \quad \forall j \in J, \ \ \forall \ell \in L, \ \ b \in [\bar{b}^j_\ell], \ \ \forall t \in T \tag{12}$$

$$\mu^1_{j\ell bt}, \mu^2_{j\ell bt} \geq 0 \quad \forall j \in J, \ \ \forall \ell \in L, \ \ b \in [\bar{b}^j_\ell], \ \ \forall t \in T. \tag{13}$$

Constraints (9) ensure that a line piece is only selected if a corresponding facility operating at the correct capacity level exists. Constraints (10) set the linear combination between the two breakpoints defining the selected line piece. Constraints (11) require that the total production at a facility equals the production represented by the linear combination of the two breakpoints. Constraints (12) and (13) are the domain constraints.

The total production costs can now be computed based on the linear combination for the selected breakpoints. Hence, the following term has to be added to the Objective Function (1), computing the total production costs as the linear combination between the two breakpoints on the selected line-piece:

$$+ \sum_{j \in J} \sum_{\ell \in L} \sum_{b \in [\bar{b}^j_\ell]} \sum_{t \in T} F^j_\ell(U^j_{\ell(b-1)}) \mu^1_{j\ell bt} + F^j_\ell(U^j_{\ell b}) \mu^2_{j\ell bt}$$

A similar technique can be applied to model more complex transportation costs, if required.

2.4 Accounting for Modular Facility Structures

In specific applications, exclusively defining the capacity level may not be sufficient to accurately represent a more complex structure of the production facility. This is the case, for example, if a facility is composed by several modules (see, e.g., Wu et al., 2006; Alarcon-Gerbier and Buscher, 2022), each of which may have different capacities and production costs. Expanding (or reducing) the total production capacity at a location then requires a more refined modelling.

Jena et al. (2016) consider location decisions for forestry camps, each of which is composed of different hosting units. The total capacity given by different combinations of such units can be represented as capacity levels within the GMC model. However, in this planning context, idle units can further be temporarily closed independently from each other to avoid unnecessary maintenance costs. The variables used in the corresponding model therefore tracks two types of information at each location: the existing capacity level and the currently open capacity level. In particular, the authors use variables $y^{jt}_{\ell_1 \ell_2 n_1 n_2}$ that take value 1 if the facility at location $j \in J$ changes its existing capacity from level $n_1 \in L$ to level $n_2 \in L$ at the beginning of time period $t \in T$, while its available (open) capacity changes from level $\ell_1 \in L$ to level $\ell_2 \in L$. As a consequence, a total of $n_2 - \ell_2$ capacity levels are assumed to be temporarily unavailable (closed) during that time period. The total costs for the set of changes are encoded in the more complex parameter $f^{jt}_{\ell_1 \ell_2 n_1 n_2}$, which includes the cost to change the existing capacity level, the change of open capacity and the cost to maintain the facility open at level $\ell_2 \in L$ throughout the time period. In order to account for such complex decisions in the above defined GMC model, variables $y^{jt}_{\ell_1 \ell_2}$ have to be replaced by $\sum_{\ell_1 \in L} \sum_{\ell_2 \in L} y^{jt}_{\ell_1 \ell_2 n_1 n_2}$.

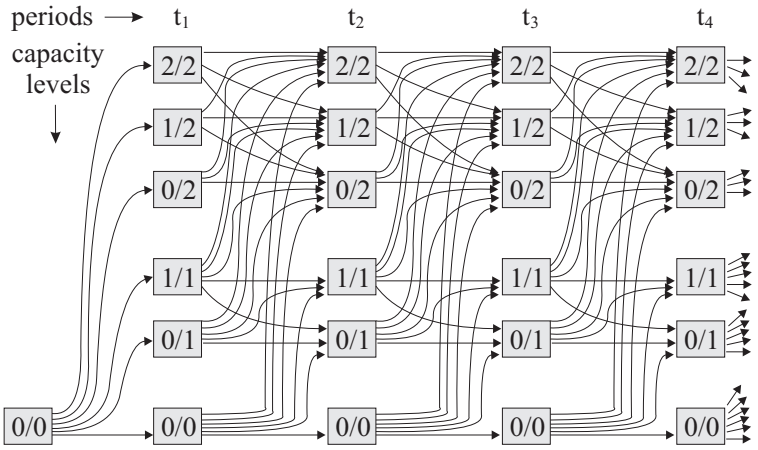


Figure 3: Network flow structure to manage partial facility closing and reopening. Each node indicates the level of open and existing capacity (Jena et al., 2016).

The resulting network structure to manage the open and existing capacity levels at a location is exemplified in Figure 3. For the sake of simplicity, this figure only illustrates the expansion of existing capacity levels (and not the reduction).

Within the model, capacity conservation equations (4) and (5) have to be replaced by the new capacity conservation constraints, which track the individual changes for each $n \in L$ and $\ell \in L$:

$$\sum_{\ell_1 \in L^+(j,t-1,\ell)} \sum_{n_1=\ell_1,\ldots,\overline{\ell}} y^{j(t-1)}_{\ell_1 \ell n_1 n} = \sum_{\ell_2 \in L^-(j,t,\ell)} \sum_{n_2=\ell_2,\ldots,\overline{\ell}} y^{jt}_{\ell \ell_2 n n_2}$$
$$\forall j \in J, \;\; \forall n \in L \backslash \{0\}, \;\; \forall \ell = 1,\ldots,n, \;\; \forall t \in T \backslash \{1\} \tag{14}$$

$$\sum_{\ell_2 \in L^-(j,t=0,\ell^j)} \sum_{n_2=\ell_2,\ldots,\overline{\ell}} y^{j1}_{(\ell_1=0)\ell_2(n_1=\ell^j)n_2} = 1 \qquad \forall j \in J. \tag{15}$$

Specifically, Equation (15) specifies that, at the beginning of the planning horizon, a facility at location $j \in J$ initially has no open capacity levels, even if a facility already exists at capacity level ℓ_j. Further note that the capacity flow conservation constraints (14) could be formulated using only one constraints for each location $j \in J$ and time period $t \in T$. However, using separated constraints for each existing level $n \in L$ and open level $\ell \in L$ tends to significantly strengthen the LP relaxation.

2.5 Accounting for Facility Relocation

In several application contexts, facilities may be relocated (see, e.g., Alarcon-Gerbier and Buscher, 2022) as opposed to shut down a facility at one location and construct a new facility at another location. A specific example is given by Jena et al. (2015b, 2016), where forestry camps, composed by mobile trailers, can be relocated closer to the new logging regions.

From a modeling perspective, two different approaches can be taken to integrate facility relocation, visualized in Figure 4. A relocation can be modelled as a direct arc between an origin and destination location, as exemplified in Sub-figure (a). Such modelling allows for representing costs that are tailored to the specific origin-destination pair, e.g., based on the relocation distance. However, it also results in a large number of relocation decision variables. As an alternative, a facility may first be moved to an artificial hub-node, as illustrated in Sub-figure (b), and then be directed to its destination location. The latter approach assumes that the cost of facility relocation primary depends on the size of the facility, while the cost difference due to varying distances between the origin and destination locations are negligible.

We now explicitly model the latter option. To this end, binary variables $w^-_{jt\ell}$ and $w^+_{jt\ell}$ are added to the model, which take value 1 if a facility of capacity size $\ell \in L$ is removed from and relocated to, respectively, location $j \in J$ at the beginning of time period $t \in T$.

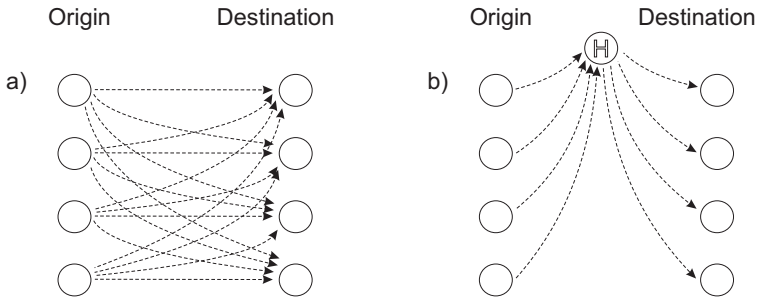


Figure 4: Two possible network flow architectures to model facility relocation.

These relocation variables have to be added to the capacity conservation constraints (4), which then write as follows $\forall j \in J\ \forall \ell \in L\ \forall t \in T \setminus \{1\}$:

$$\sum_{\ell_1 \in L^+(j,t-1,\ell)} y_{\ell_1 \ell}^{j(t-1)} + w_{jt\ell}^{+} = \sum_{\ell_2 \in L^-(j,t,\ell)} y_{\ell \ell_2}^{jt} + w_{jt\ell}^{-} \tag{16}$$

If the relocation of an existing facility at the first time-period is allowed, the corresponding $w_{jt\ell^j}^{-}$ variable needs to be added to the left-hand side of Equation (5). The following equalities ensure that for each facility of size $\ell \in L$ relocated to a location, there is exactly one facility of the same size removed from a location:

$$\sum_{j \in J} w_{jt\ell}^{+} = \sum_{j \in J} w_{jt\ell}^{-} \quad \forall \ell \in L \setminus \{0\}\,, \quad \forall t \in T \setminus \{1\}\,. \tag{17}$$

Finally, the relocation costs have to be added to the objective function. Assume that c_ℓ^R denotes the costs to relocate a facility of capacity size $\ell \in L$ from one location to another. The following term has to be added to the objective function (1):

$$+ \sum_{j \in J} \sum_{\ell \in L} \sum_{t \in T} \frac{c_\ell^R}{2} w_{jt\ell}^{+} + \sum_{j \in J} \sum_{\ell \in L} \sum_{t \in T} \frac{c_\ell^R}{2} w_{jt\ell}^{-}$$

Here, the relocation costs have been equally split onto the incoming and outgoing relocation variables. Given that Equality (17) ensures that any outgoing facility matches an incoming facility of the same size, the relocation cost can also be associated to only one of the two variables. However, in the case of a subsequent Lagrangian relaxation, splitting the costs over the two variables is preferred to ensure that both variables imply a usage cost.

If the relocation distance between specific location pairs has to be modelled, a single binary variable $w_{j_1 j_2 t \ell}$ with more specific relocation costs $c_{\ell j_1 j_2}^R$ can be used instead, indicating that the respective facility is relocated from location $j_1 \in J$ to location $j_2 \in J$. The capacity change network, illustrated in Figure 4(a), then uses variables that are directly used in Equations (4). As a consequence, Equalities (17) are no longer required. The use of such binary variables with more detailed information does, however, not improve the strength of the LP relaxation (Jena, 2014).

Note that the presented formulations make the assumption that relocation is instantaneous. If relocation is assumed to take one (or more) time-period(s), the $w_{jt\ell}^{+}$ variables have to be added to the capacity conservation constraints (16) of a later occurring time-period. The formulations discussed above can also be easily adapted to the case where only a part (e.g., certain units) of the facility is relocated.

2.6 Accounting for Uncertainty via Scenarios

Parameter uncertainty in facility location problems typically concerns the customer demands or the costs to provide available capacity and to serve the customers. While deterministic problem variants aim at minimizing the total costs required to satisfy the entire demand, the latter becomes ill defined when demand is uncertain. In this case, the main paradigm classes to deal with such uncertainty include robust optimization

(see, e.g., Ben-Tal et al., 2009) and stochastic optimization (see, e.g., Birge and Louveaux, 2011). Roughly speaking, the former aims at minimizing the total costs required to satisfy the worst possible demand outcome, while the latter aims at minimizing the average costs required to satisfy the possible demand outcomes. We here focus on the case of the latter, where we minimize the expected costs and demand uncertainty is assumed to be well represented by a set of demand scenarios. Such a case, where capacity decisions have to be made in the first stage, while demand allocation is done in the second stage, has more recently been handled by Lagrangian relaxation in the literature (see, e.g., Marín et al., 2018; Štádlerová et al., 2023). When optimizing expectation, it is rarely beneficial to aim at satisfying the entire demand of extreme cases that have a low probability of occurrence. It is therefore common to quantify the penalties induced by shortfall demand units (also known as the recourse).

We assume that set S contains a set of scenarios that sufficiently well represents the parameter uncertainty, with ω_s being the probability that scenario $s \in S$ will occur. For each scenario $s \in S$, d_{ip}^{st} represents the demand quantity for commodity $p \in P$ requested by customer $i \in I$ during time period $t \in T$. While other parameters may also be uncertain and depend on $s \in S$, for the sake of illustration, we here restrict to demand uncertainty. We assume a penalty cost in the case that the available capacity falls short of the requested demand. To this end, let f^S be the cost of one unit short of demand. The objective is then to minimize the expected total costs required to satisfy the various demand scenarios.

In a dynamic planning context, the set of eligible first-stage decisions directly depends on the application context, as well as the flow of information and decisions. Specifically, some of the capacity decisions (i.e., the opening of facilities at some time periods and, potentially, capacity adjustments) may have to be made in the first stage, while others can be postponed to the second stage, once demand quantities are more reliably known. The shape of the set of first-stage decisions defines the problem variant. At one extreme of the problem variants, all capacity decisions have to be made in the first stage. At another extreme, only a few facility opening decisions (e.g., those for the first time period) have to be made in the first stage, while the remainder can be tailored to the demand scenarios in the second stage. Such a flexible problem variant particularly makes sense if the planning is carried out dynamically in a rolling horizon fashion. For example, Štádlerová et al. (2023) requires that all facility openings, not matter at which time period, are decided in the first stage, while capacity expansions can be made in the second stage.

For the scope of this paper, it is impossible to cover all problem variants that may result from the various definitions of the first-stage decisions. We therefore here focus on the simplest case, where capacity decisions y are made in the first stage and remain the same, no matter the demand outcome in the second stage. The demand allocation decisions are then tailored to the specific demand scenario $s \in S$ in the second stage. Specifically, let $x_{i\ell p}^{sjt}$ be the proportion of demand d_{ip}^{st}, satisfied from the facility at location $j \in J$ open at capacity level $\ell \in L$. If minimum production quantities at the facilities are required in the problem, potential excess capacity will have to be defined separately for each demand scenario (see, e.g., Štádlerová et al., 2023).

We adapt the GMC formulation as follows. We define continuous variables $\tilde{x}_i^{st}$ as the unit amount of the demand of customer $i \in I$ at time period $t \in T$ and scenario s that is under-served. In the Objective Function (1), the corresponding penalties for shortfall quantities are added:

$$+ \sum_{i \in I} \sum_{s \in S} \sum_{t \in T} \omega_s f^S \tilde{x}_i^{st}$$

Furthermore, term $d_{ip}^{t} x_{i\ell p}^{sjt}$ in Objective Function (1) is replaced by $\sum_{s \in S} d_{ip}^{st} x_{i\ell p}^{sjt}$ to account for the scenario-dependent demand.

Demand constraints (2), capacity constraints (3) and SIs (6) are then replaced by their scenario-dependent counterparts (18), (19) and (20), respectively.

$$\sum_{j \in J} \sum_{\ell \in L} x_{i\ell p}^{sjt} + \tilde{x}_i^{st} = 1 \;\; \forall i \in I, \;\; \forall p \in P, \;\; \forall t \in T, \;\; \forall s \in S \tag{18}$$

$$\sum_{i \in I} \sum_{p \in P} d_{ip}^{st} x_{i\ell p}^{sjt} \leq \sum_{\ell_1 \in L^+(j,t,\ell)} u_\ell^j y_{\ell_1 \ell}^{jt} \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall t \in T, \;\; \forall s \in S \tag{19}$$

$$x_{i\ell p}^{sjt} \leq \sum_{\ell_1 \in L} y_{\ell_1 \ell}^{jt} \;\; \forall i \in I, \;\; \forall j \in J, \;\; \forall \ell \in L, \;\; \forall p \in P, \;\; \forall t \in T, \;\; \forall s \in S \tag{20}$$

$$\tilde{x}_i^{st} \geq 0 \;\; \forall i \in I, \;\; \forall t \in T, \;\; \forall s \in S. \tag{21}$$

As previously mentioned, such formulation assumes that facilities do not have minimum production requirements. In Section 2.3, we elaborate on how to incorporate capacity-level dependant minimum production requirements at facilities. In that case, under uncertain demand, additional variables have to be added, representing the produced excess capacity. Such production excess will then have to be penalized in the objective function. We refer the reader to the work of Štádlerová et al. (2023) for such a case.

3 Solution via Lagrangian Relaxation

Lagrangian decomposition relaxes complicating constraints and transfers their violation into the objective function. Two different relaxations are typically considered: relaxing the demand constraints (2) or relaxing the capacity constraints (3). The latter has been successfully attempted by several authors (Beasley, 1993; Li et al., 2009; Ghodsi, 2012; Gendron et al., 2016). However, the resulting subproblem is typically still NP-hard. For example, in the case of the classical CFLP, the resulting subproblem is an Uncapacitated facility location problem. While relaxing the capacity constraints may make sense in some particular situations, it is unlikely to provide competitive results when some of the key features of the problem involve the facility capacities, such as several capacity levels and capacity adjustment over time (as it is the case with the GMC).

We therefore here decide to relax the demand constraints (2) within the GMC formulation (see Section 2.1), which are the only constraints that provide a link among the different candidate facility locations. Such a relaxation also allows for a more unified approach that can be adjusted to the various extensions discussed throughout Sections 2.3 to 2.6.

Let α be the vector of Lagrange multipliers of demand constraints (2). After relaxing these constraints, transferring their violation into the objective function and rearranging its terms, we obtain the following Lagrangian subproblem:

$$\begin{aligned} L(\alpha) = \quad & \min \sum_{j\in J}\sum_{\ell_1\in L}\sum_{\ell_2\in L^-(j,t,\ell)}\sum_{t\in T} f^{jt}_{\ell_1\ell_2} y^{jt}_{\ell_1\ell_2} \\ & + \sum_{i\in I}\sum_{j\in J}\sum_{\ell\in L}\sum_{p\in P}\sum_{t\in T}(g^{jt}_{i\ell p} d^t_{ip} - \alpha^t_{ip}) x^{jt}_{i\ell p} + \sum_{i\in I}\sum_{p\in P}\sum_{t\in T}\alpha^t_{ip} \\ & s.t.\ (3)-(8). \end{aligned}$$

Solving this problem yields the Lagrangian solution, which is feasible for the original GMC formulation only if the demand violation $(1 - \sum_{j\in J}\sum_{\ell in L} x^{jt}_{i\ell p})$ equals 0 for all $i \in I$, $p \in P$ and $t \in T$. This the case only when the optimal Lagrange multipliers are found, which correspond to solving the so-called Lagrangian dual problem. Solving the Lagrangian dual to optimality is practically often impossible. Instead, it is common to iteratively improve the Lagrange multipliers such that the total violation decreases until a certain convergence criteria is met. Throughout this process, feasible solutions can be obtained by "repairing" the obtained Lagrangian solutions.

In the following sections, we elaborate on how to solve the Lagrangian subproblems efficiently, how to solve the Lagrangian dual problem and how to generate solutions that are feasible for the original problem. We will start with the case of the general GMC formulation, and then discuss the adjustments necessary to account for more complex problem variants (see Sections 2.3 - 2.6).

3.1 Solution of the Lagrangian Subproblem

In order to simplify the representation, let $\tilde{c}^{jt}_{i\ell p} = g^{jt}_{i\ell p} d^t_{ip} - \alpha^t_{ip}$ denote the modified costs for the $x^{jt}_{i\ell p}$ variables. Given that the demand constraints have been relaxed, the Lagrangian subproblem can be separated into one subproblem for each candidate facility location $j \in J$. The objective function of the Lagrangian subproblem can then be written as $L(\alpha) = \sum_{j\in J} L_j(\alpha) + \sum_{i\in I}\sum_{p\in P}\sum_{t\in T}\alpha^t_{ip}$. Here, $L_j(\alpha)$ corresponds to the part of

the objective function specific to candidate location $j \in J$, written as follows:

$$
\begin{aligned}
L_j(\alpha) = \quad & \min \sum_{\ell_1 \in L} \sum_{\ell_2 \in L^-(j,t,\ell_1)} \sum_{t \in T} f^{jt}_{\ell_1\ell_2} y^{jt}_{\ell_1\ell_2} + \sum_{i \in I} \sum_{\ell \in L} \sum_{p \in P} \sum_{t \in T} \tilde{c}^{jt}_{i\ell p} x^{jt}_{i\ell p} \\
& s.t. \sum_{i \in I} \sum_{p \in P} d^t_{ip} x^{jt}_{i\ell p} \leq \sum_{\ell_1 \in L^+(j,t,\ell)} u^j_\ell y^{jt}_{\ell_1\ell} \quad \forall \ell \in L, \ \ \forall t \in T \\
& \sum_{\ell_1 \in L^+(j,t-1,\ell)} y^{j(t-1)}_{\ell_1\ell} = \sum_{\ell_2 \in L^-(j,t,\ell)} y^{jt}_{\ell\ell_2} \quad \forall \ell \in L, \ \ \forall t \in T \backslash \{1\} \\
& \sum_{\ell \in L^-(j,t,\ell^j)} y^{j1}_{\ell^j\ell} = 1 \\
& x^{jt}_{i\ell p} \leq \sum_{\ell_1 \in L^+(j,t,\ell)} y^{jt}_{\ell_1\ell} \quad \forall i \in I, \ \ \forall \ell \in L, \ \ \forall p \in P, \ \ \forall t \in T \\
& x^{jt}_{i\ell p} \geq 0 \quad \forall i \in I, \ \ \forall \ell \in L, \ \ \forall p \in P, \ \ \forall t \in T \\
& y^{jt}_{\ell_1\ell_2} \in \{0,1\} \quad \forall \ell_1 \in L, \ \ \forall \ell_2 \in L^-(j,t,\ell_1), \ \ \forall t \in T.
\end{aligned}
$$

In the following sections, we will elaborate on how to solve the Lagrangian subproblem above, as well as those derived from the other proposed problem extensions.

3.1.1 Solving the subproblem for each candidate location

Subproblem $L_j(\alpha)$ can be solved independently for each facility location $j \in J$ (see, e.g. Jena et al., 2017). Each subproblem corresponds to the capacity planning over the planning horizon at that given location $j \in J$, which can be computed by solving a shortest-path problem defined on the acyclic time-capacity graph, as exemplified in Figure 4. The arc costs to change from one capacity level to another at the beginning of each time-period includes the costs to change capacity, maintain the facility open at the new capacity level, as well as the demand allocation to the customers (considering the modified costs $\tilde{c}^{jt}_{i\ell p}$). Specifically, the shortest-path network for facility location $j \in J$ is defined by nodes $(\ell, t), \ell \in L, t \in T$. The arc costs representing a capacity change from level $\ell_1 \in L$ to level $\ell_2 \in L$ at the beginning of period $t \in T$, maintenance costs at level $\ell_2 \in L$ and the optimal demand allocation is then given by $f^{jt}_{\ell_1\ell_2} + \hat{L}^\alpha_j(\ell, t)$. Here $\hat{L}^\alpha_j(\ell, t)$ represents the optimal demand allocation at a facility located at $j \in J$ and open at level $\ell \in L$ during time period $t \in T$ under multipliers α, and will be defined next.

Computing the optimal demand allocation. Computing the optimal demand allocation with costs $\hat{L}^\alpha_j(\ell, t)$ for each time period $t \in T$ and capacity level $\ell \in L$ minimizing the modified costs $\tilde{c}^{jt}_{i\ell p} = g^{jt}_{i\ell p} d^t_{ip} - \alpha^t_{ip}$ is equivalent to solving a continuous knapsack problem. This can be done in polynomial time. First, all demand nodes d^t_{ip} with non-positive allocation costs are sorted in increasing order of their adjusted per-unit transportation costs, given by $\tilde{c}^{jt}_{i\ell p}/d^t_{ip}$. Then, demand is allocated (respecting the sorted sequence) until either the demand of a customer $i \in I$ and commodity $p \in P$ is fully met or the capacity u^j_ℓ is filled.

The resulting shortest-path problem can be solved by standard algorithms, such as the Dijkstra algorithm, a minimum cost-flow network solver or Dynamic Programming. As mentioned earlier, the solutions obtained from solving the Lagrangian subproblem are likely infeasible to the original problem. In order to obtain feasible solutions that constitute an upper bound to the problem, one may aim at "repairing" the Lagrangian solutions. This will be discussed in Section 3.3.

Solution under special cases. The solution of the subproblem $L_j(\alpha)$ may simplify under some of the special cases discussed in Section 2.2. Naturally, having a single commodity instead of several commodities still requires to solve a shortest-path problem as described above. In contrast. in the case of a single-period problem, the optimal capacity level can easily be found by inspecting the costs for each level. If the problem is uncapacitated, the demand allocation described above requires to allocate to the facility all demand nodes with non-positive allocation costs. Finally, if set L only has a single capacity level in addition to level 0 and a facility cannot be closed once opened, the optimal time-period to open the facility can be identified by inspection: for each time-period, one considers the costs to open the facility and maintain it open until the

end of the planning horizon. If, however, facilities may be closed as well, or in the case of any other more complex capacity change rules induced by sets L^- and L^+, the subproblem is solved easiest as a shortest-path problem as previously described.

3.1.2 Solution in the case of Complex Cost Functions

When piecewise linear cost functions are used for the production costs that depend on the facility's capacity level $\ell \in L$, the Lagrangian subproblem is still be decomposed as previously shown. However, in order to find the optimal production schedule for each facility location, the demand allocation has to be adjusted. Specifically, the demand allocation, solved as a continuous knapsack problem, now has piecewise linear costs (see, e.g., Christensen and Klose, 2021). The objective to be minimized now also accounts for the piecewise-linear production costs given through the linear combination between μ^1 and μ^2, while g now exclusively refers to the distribution costs. As such, the continuous knapsack with piecewise-linear production costs minimizes the following term:

$$\sum_{j\in J}\sum_{\ell\in L}\sum_{t\in T}\sum_{b\in[\bar{b}^j_\ell]} F^j_\ell(U^j_{\ell(b-1)})\mu^1_{j\ell bt} + F^j_\ell(U^j_{\ell b})\mu^2_{j\ell bt} + \tilde{c}^{jt}_{i\ell p}$$

, where $\tilde{c}^{jt}_{i\ell p} = g^{jt}_{i\ell p} d^t_{ip} - \alpha^t_{ip}$.

Let $h^j_{\ell bt} = \frac{F^j_\ell(U^j_{\ell b}) - F^j_\ell(U^j_{\ell(b-1)})}{U^j_{\ell b} - U^j_{\ell(b-1)}}$ be the per-unit production costs at facility $i \in I$ that produces at capacity level $\ell \in L$ and breakpoint b of its piecewise linear cost function during time period $t \in T$. The marginal costs to produce and serve one additional unit while having selected line piece b on capacity level $\ell \in L$ is computed as $m^j_{\ell bt} = \tilde{c}^{jt}_{i\ell p} + h^j_{\ell bt}$.

We define one piecewise linear continuous knapsack for each capacity level $\ell \in L$ and time-period $t \in T$. Customer demands have to be added line piece by line piece. To this end, for a given $\ell \in L$ and $t \in T$, customer demands d^t_{ip} are sorted in non-decreasing order of their per-unit distribution costs $\tilde{c}^{jt}_{i\ell p}/d^t_{ip}$. Given that the distribution costs do not depend on b, the sorted sequence of customer-commodity pairs is the same no matter the line piece b. Starting with $b = 1$, customer demands are then allocated in that order either until the marginal cost is positive (i.e., $m^j_{\ell bt} > 0$) or until the capacity limit $U^j_{\ell b}$ of the current line piece b is reached. In the case of the latter, the next highest line piece $b = 2, \ldots, \bar{b}$ is considered.

If this iterative procedure reaches the maximum capacity $U^j_{\ell\bar{b}}$ of the last line piece $\bar{b}$, the facility cannot serve all profitable customer demands. On the other hand, if the minimum production capacity $U^j_{\ell(b=0)}$ has not been met by adding customers with marginal costs $m^j_{\ell bt} \leq 0$, customers with positive marginal costs have to be considered until the minimum production capacity is reached.

3.1.3 Solution in the case of Modular Facility Structures

In the context of modular facility structures that can be modelled by tracking both the existing capacity level and the capacity level that is currently open (i.e., available for production), the solution approach for the Lagrangian subproblem outlined in Section 3.1.1 can be extended. In this case, the shortest path problem is defined on a more complex capacity-time network, as illustrated in Figure 3.

Each node (ℓ, n) now represents the actual capacity that is available for use and the total capacity that exists at a given location. The arc costs to transition from one node to another additionally need to consider the costs to temporarily close or reopen a part of the existing capacity. Specifically, the shortest-path network for facility location $j \in J$ is defined by nodes (ℓ, n, t) with $\ell \in L, n = \ell, \ldots, q, t \in T$. As before, the arc costs represent a change from existing level $n_1 \in L$, open at level $\ell_1 \in L$, to existing level $n_2 \in L$ open at level $\ell_2 \in L$ at the beginning of period $t \in T$, as well as the maintenance costs at level $\ell_2 \in L$ and the optimal demand allocation. The arc transitioning costs are therefore given by $f^{jt}_{\ell_1\ell_2 n_1 n_2} + \hat{L}^\alpha_j(\ell, t)$, where $\hat{L}^\alpha_j(\ell, t)$ represents the costs of the optimal demand allocation as defined and computed in Section 3.1.1. As before, the resulting shortest path problem can be solved, for example, by dynamic programming. For further details concerning the solution procedure in the case of modular facility structures, we refer to Jena et al. (2016).

3.1.4 Solution in the case of Facility Relocation

If the context of the planning problem allows for facility relocation, both the demand constraints (2) and the relocation hub constraints (17) are considered complicating, since they sum over different facility locations. Two possibilities can be considered: either only relax the demand constraints or relax both sets of constraints.

In the former case, the problem cannot be decomposed by candidate facility location. One may add to the subproblem aggregated demand constraints, enforcing that the total production capacity available at each time-period is sufficient to satisfy the entire demand of that time-period. However, the problem cannot be solved combinatorially and has to be solved by a general-purpose MIP solver. The solution of the subproblem may therefore be rather slow. While the obtained solutions may give insights about beneficial facility relocations and yield interesting lower bounds, to the best of our knowledge, literature has not yet reported on competitive approaches for this type of relaxation.

In contrast, the latter case, relaxing both sets of contraints, has been successfully used in the literature. Here, the subproblem can be decomposed by facility locations (see Jena et al., 2016), each of which can be solved as a shortest-path problem in a similar fashion as previously discussed. In this case, additional Lagrangian multipliers β are added when relaxing the relocation hub constraints (17), and the following term is added in the objective function, indicating the violation of the relaxed constraints:

$$+ \sum_{\ell \in L \setminus \{0\}} \sum_{t \in T \setminus \{1\}} \left(\sum_{j \in J} w^{-}_{jt\ell} - \sum_{j \in J} w^{+}_{jt\ell} \right) \beta_{\ell t}$$

The resulting subproblem can be solved, as previously discussed, separately for each facility location. Note that the Lagrangian solutions obtained may not satisfy the relocation constraints, i.e., the number of removed facilities may not match the number of arriving facilities. The solutions must be "repaired" to be feasible to the original problem, which will be outlined in Section 3.3.3. Finally, we also note that a location problem that combines both facility relocation and modular facility structures requires a more subtle analysis in order to solve the Lagrangian subproblem. We refer the reader to the work of Jena et al. (2016) who have solved such a problem by means of Lagrangian relaxation.

3.1.5 Solution in the case of Uncertainty via Scenarios

When parameter-uncertainty is explicitly acknowledged in the model via demand scenarios, for the case where the entire capacity schedule has to be decided in the first stage, the solution of the Lagrangian subproblem changes only marginally. Specifically, scenario-dependent demand constraints (18) are relaxed, and Lagrangian multipliers α_{ipts} are used, now also depending on scenario $s \in S$.

After the relaxation of demand constraints (18), the objective function of the Lagrangian subproblem $L_j(\alpha)$ additionally contains the following term, given that the shortfall variables $\tilde{x}_i^{st}$ are part of the relaxed constraints:

$$+ \sum_{i \in I} \sum_{s \in S} \sum_{t \in T} (\omega_s f^S - \alpha_{ipts}) \tilde{x}_i^{st}.$$

To solve the Lagrangian subproblem, the shortest path problem (see Section 3.1.1) throughout the time-capacity network can still be applied, given that the entire capacity schedule is assumed to remain the same for all demand scenarios. Two changes have to be made to compute the costs at each node. First, the objective function of the Lagrangian solution now has to take into consideration the term above for shortfall variables $\tilde{x}_i^{st}$. Second, the costs for the demand allocation within each of the time-capacity nodes within this network have to consider the different demand scenarios. To this end, for each time period $t \in T$ and capacity level $\ell \in L$, the continuous knapsack has to be solved for each demand scenario s.

Given the demand shortfall variables $\tilde{x}_i^{st}$ only appeared in the relaxed demand constraints (18), but not in the remaining constraints, they are now constrained only by their domain constraints (21). As such their solution values do not impact the solution values of the demand allocation variables x. The continuous knapsack can therefore be solved as previous explained in Section (3.1.1). In order to obtain valid lower bounds from the Lagrangian solution, one still requires to solve for the values of $\tilde{x}_i^{st}$, which can be computed by inspection (i.e., they are set either to 0 or to the corresponding customer demand).

Different subsets of second-stage decisions. As noted in Section 2.6, we here assume that all capacity decisions have to be made within the first stage of the two-stage stochastic optimization problem. In many applications, this is not the case. For example, in a dynamic context in which the optimization model is reoptimized at each time-period, only capacity decisions for the first (i.e., the next upcoming) time-period may have to be made in the first stage, while the remainder can be approximated in the second stage and will only be taken in the upcoming executions of the optimization model. In such cases, the Lagrangian subproblem for each facility location cannot be solved simply by solving a shortest path-problem. Instead, one needs to distinguish decisions that are made within the first stage and those that are made in the second stage, independently for each scenario. A shortest path problem then has to be solved separately for each scenario and for a given set of first-stage decisions. For a facility location that features both uncertainty-based scenarios and complex production cost functions, and that defines certain capacity decisions within the second stage of the stochastic optimization problem, we refer to the work of Štádlerová et al. (2023), who explain in detail the use of Lagrangian relaxation to solve such a problem.

3.2 Solution of the Lagrangian Dual Problem

For any given Lagrange multiplier α, the solution provided by the Lagrangian subproblem $L(\alpha)$ provides a lower bound to the original problem. The best possible lower bound z^* is obtained by solving the so-called Lagrangian dual problem:

$$z^* = \max_{\alpha} L(\alpha).$$

In theory, there exists a set of multipliers α such that the solution of the Lagrangian subproblem is also feasible to the original problem, i.e., the relaxed constraints are not violated. In practice, however, these multipliers are rather difficult to find and solving the Lagrangian dual exactly is typically challenging. Iterative approaches, aiming at converging to multipliers that maximize the Lagrangian dual, and therefore minimize the implicit violation of the relaxed constraints, has been shown to be an effective approach. These approaches can be divided into two classes: subgradient methods, which consider one subgradient at each iteration and bundle methods, which make use of a subset of subgradients. Both classes will be discussed below.

3.2.1 Subgradient Method

Subgradient methods iteratively adjust the Lagrangian multipliers into the direction that allows for a smaller violation of the relaxed constraints. They consist of two main ingredients: the subgradient direction and the stepsize. In the case where the demand constraints (2) have been relaxed, a subgradient direction γ_{ip}^{kt} for multiplier α_{ip}^{t} at iteration k can be easily computed as the violation of the relaxed constraints when considering the solution x of the current Lagrangian subproblem. It is computed as:

$$\gamma_{ip}^{kt} = 1 - \sum_{j \in J} \sum_{\ell \in L} x_{\ell p}^{ijt} \quad \forall i \in I, \quad \forall p \in P, \quad \forall t \in T.$$

For all subgradient directions, a common stepsize is used, representing the amount each Lagrangian multipliers will be adjusted into the subgradient direction. A common approach, suggested by Held et al. (1974) and successfully used in succeeding works (see, e.g., Shulman, 1991; Sridharan, 1991; Correia and Captivo, 2003; Jena et al., 2017), is to choose the stepsize λ^k at iteration k depending on how close the current solution value (which constitutes a lower bound) is to the best known upper bound $\hat{Z}$, scaled by scalar δ^k and the magnitude of all current subgradient directions:

$$\lambda^k = \delta^k \frac{\widehat{Z} - L^k(\alpha)}{\sum_{i \in I} \sum_{p \in P} \sum_{t \in T} (\gamma_{ip}^{kt})^2}.$$

The Lagrangian multipliers used within the next iteration are then updated as follows:

$$\alpha_{ip}^{(k+1)t} = \alpha_{ip}^{kt} + \lambda^k \gamma_{ip}^{kt} \quad \forall i \in I, \quad \forall p \in P, \quad \forall t \in T.$$

The performance and convergence may highly vary among problem instances and is known to be sensitive to the values of its parameters. For the scalar, it has been found beneficial (see, e.g., Jena et al., 2017) to use an initial value of δ^0 and a subsequent halving of the value every 25 consecutive iterations without improvement of the lower bound. Lagrangian multipliers α are typically initialized with 0.

Subgradient for relaxed relocation constraints. If the problem accounts for relocation, and the corresponding relocation linking constraints (17) have been relaxed, the subgradient method has to take into consideration the subgradients for the multipliers β. At the k-th iteration, a subgradient can be computed as follows, with variables w^- and w^+ fixed to the values found in the solution to the Lagrangian subproblem:

$$\mu^k_{\ell t} = \sum_{j \in J} w^-_{jt\ell} - \sum_{j \in J} w^+_{jt\ell} \;\; \forall \ell \in L \backslash \{0\}, \;\; \forall t \in T.$$

The computation of the step size has to be extended as follows:

$$\lambda^k = \delta^k \frac{\widehat{Z} - L^k(\alpha)}{\sum_{i \in I} \sum_{p \in P} \sum_{t \in T} (\gamma^{kt}_{ip})^2 + \sum_{\ell \in L \backslash \{0\}} \sum_{t \in T} (\mu^k_{\ell t})^2}.$$

Finally, the Lagrangian multipliers for the relaxed relocation linking constraints, used within the next iteration are then updated as follows:

$$\beta^{(k+1)}_{\ell t} = \beta^k_{\ell t} + \lambda^k \mu^{\ell t} \;\; \forall \ell \in L \backslash \{0\}, \;\; \forall t \in T.$$

Subgradient in the case of demand uncertainty. As the subgradient is computed as the violation of the relaxed constraint, it now has to account for shortfall variables $\tilde{x}$ in each of the scenarios $s \in S$, computed as follows:

$$\gamma^{kts}_{ip} = 1 - \sum_{j \in J} \sum_{\ell \in L} x^{ijt}_{\ell p} - \tilde{x}^{st}_i \;\; \forall i \in I, \;\; \forall p \in P, \;\; \forall t \in T, \;\; \forall s \in S.$$

The stepsize λ^k for the next iteration k now has to sum over all squared subgradients as before, but additionally including those for the different scenarios:

$$\lambda^k = \delta^k \frac{\widehat{Z} - L^k(\alpha)}{\sum_{i \in I} \sum_{p \in P} \sum_{t \in T} \sum_{s \in S} (\gamma^{kts}_{ip})^2}.$$

Finally, the computation of the next Lagrangian multipliers α is now also carried out separately for each scenario $s \in S$:

$$\alpha^{(k+1)ts}_{ip} = \alpha^{kts}_{ip} + \lambda^k \gamma^{kts}_{ip} \;\; \forall i \in I, \;\; \forall p \in P, \;\; \forall t \in T, \;\; \forall s \in S.$$

3.2.2 Bundle Methods

An alternative to classical subgradient methods are so-called Bundle methods, which use a subset $B \subseteq \{1, \ldots, k\}$ of the previous multiplier-subgradient tuples $< L(\alpha^m), \gamma^m >$ from any previous iteration $m \in [1, k]$, where k is the current iteration. In contrast to the subgradient method explained above, those methods do not rely on an upper bound.

We will here focus on two bundle methods that have been successfully applied to solve complex facility location problems.

Aggregated Bundle method. We illustrate the aggregated bundle method proposed by Frangioni and Gallo (1999) and Frangioni (2005), which has been successfully applied in several location (see, e.g., Jena et al., 2016, 2017) and network design problems (see, e.g., Frangioni and Gorgone, 2014). The method aims at finding a linear combination θ among the available subgradients by solving the following quadratic optimization problem:

$$\min_{\theta} \left\{ \tfrac{1}{2} \parallel \sum_{m \in B} \gamma^s \theta^m \parallel^2 + \tfrac{1}{R} E_B \theta; \; s.t. \; \sum_{m \in B} \theta^m = 1, \; \theta \geq 0 \right\},$$

where R is the so-called trust region, and $E_m = L(\alpha) + \gamma(\hat{\alpha} - \alpha) - L(\hat{\alpha})$ is the linearization error from the current point $\hat{\alpha}$.

The tentative ascent direction is then computed by the convex combination of the subgradients, using the convex multipliers θ. Alternatively, the dual problem can be solved to compute the ascent direction, or directly the new point. Frangioni and Gallo (1999) elaborate on this relationship in detail.

The solution values for θ^m do not only derive the next ascent direction, but may also be useful to construct feasible integer solutions. Given that high values of θ^m indicate that certain subgradients are judged more promising then others, one may want to pay close attention to the corresponding multipliers and the Lagrangian solution resulting from them. This relationship is briefly discussed in Section 3.3.5.

Boxstep method. Another method that has been successfully used to solve the Lagrandian dual for complex location problems (see, e.g., Schütz et al., 2009; Štádlerová et al., 2023) is the boxstep method (Marsten et al., 1975). The method directly computes the set of the next Lagrangian multipliers $\alpha^{(k+1)}$ by solving the following linear program, where $L^k = L(\alpha^k) - \sum_{i\in I}\sum_{p\in P}\sum_{t\in T}\alpha_{ip}^{kt}\gamma_{ip}^{kt}$:

$$
\begin{aligned}
&\max\ \phi \\
&s.t.\ \phi \le L^m + \sum_{i\in I}\sum_{p\in P}\sum_{t\in T}\alpha_{ip}^{(k+1)t}\gamma_{ip}^{mt} \quad m = 1,\dots,k \\
&\qquad \alpha_{ip}^{(k+1)t} \le \alpha_{ip}^{kt} + \Delta_{ip}^{kt} \ \ \forall i \in I,\ \ \forall p \in P,\ \ \forall t \in T \\
&\qquad \alpha_{ip}^{(k+1)t} \ge \alpha_{ip}^{kt} - \Delta_{ip}^{kt} \ \ \forall i \in I,\ \ \forall p \in P,\ \ \forall t \in T \\
&\qquad \phi \in \mathbb{R}, \alpha_{ip}^{(k+1)t} \in \mathbb{R}.
\end{aligned}
$$

Here, Δ_{ip}^{kt} defines the box, specific to each multiplier α_{ip}^{kt}. Typically, the box decreases in size as the algorithm converges. For example, Štádlerová et al. (2023) propose to decrease the boxsize by a predefined percentage whenever Δ_{ip}^{kt} changes its sign from one iteration to the next. Further, the authors also reset the box size if the multipliers do not change for three consecutive iterations in order to allow for escaping local optimal. Note that the ideal initial box size depends on the magnitude of the multiplier values and should be selected such that the convergence of the LB is fastest.

If several demand scenarios are used (as suggested in Section 2.6), the right-hand side of the first constraint of the LP of the boxstep method has to be adjusted such that the Lagrange multipliers are also summed over all demand scenarios $s \in S$. Similarly, the two remaining constraints have to be defined for each demand scenario $s \in S$.

3.3 Generation of Feasible Upper Bounds

At each iteration, the Lagrangian heuristic may make the attempt to construct an upper bound to the problem, which corresponds to a feasible solution to the original problem. When the subgradient method is used to solve the Lagrangian dual problem, this upper bound also directly impacts the computation of the step size and therefore the convergence of the algorithm.

The generation of good upper bounds may be time consuming. If the Lagrangian relaxation is used within another optimization framework, e.g., within a branch-and-bound algorithm to provide stronger lower bounds, the quality of the upper bounds may be less important. In contrast, if the main objective is to produce high quality feasible solutions that can be used in practical planning, then the generation of such solutions should be given much care.

In both cases, Lagrangian heuristics typically aim at exploiting the solutions obtained from the Lagrangian subproblem in each of the iterations. These solutions are unlikely to be feasible, given that certain constraints have been relaxed (in our case, the demand constraints, and possibly the relocation linking constraints). However, when a set of "good" multipliers are used, the violation of the constraints may be small, and the provided solution may be close to a feasible one, or at least contain certain information that is helpful to approximate the optimal solution to the original problem. It therefore makes sense to attempt to "repair" such solutions such that they become feasible to the original problem, as will be outlined throughout Sections 3.3.1 to 3.3.4. While such heuristics may be sufficient and provide high quality solutions for more simple

problem variants, they are unlikely to find close-to optimal solutions in more complex problem variants. In this case, techniques aiming at improving the feasible solutions found may be employed, which will be discussed in Section 3.3.5.

3.3.1 Repairing Infeasibility of the Lagrangian Solutions

As previously explained, given that the demand constraints have been relaxed, the demand allocation from facilities to customers as proposed by the solution of the Lagrangian subproblem is most likely infeasible to the original problem. We now aim at repairing such infeasibility in order to obtain a valid upper bound.

A feasible solution to the here considered problem consists of two components:

- *A facility opening schedule* given by $\ell(j,t) \in L$, indicating a capacity level for each facility location $j \in J$ and each time period $t \in T$, as suggested by the Lagrangian solution. This opening schedule may be appropriate to efficiently serve demands in the original problem. However, given that demand constraints have been relaxed, capacity may also either be short or in excess at certain time periods. To ensure feasibility, additional capacity may have to be opened at certain time periods (either by opening new facilities or by increasing capacity at existing facilities). In addition, excess capacity may be reduced to further reduce maintenance costs and therefore improve the planning solution.
- *A demand allocation* from facilities to customers. The Lagrangian solution provides a tentative demand allocation, which may either exactly meet customer demands d^t_{ip}, under-serve (i.e., the flow sent from facilities does not cover the entire demand) or over-serve them (i.e., too much flow is being sent). Two solutions are common to correct unsatisfied demand: either the tentative demand allocation is repaired such that all demand is met exactly, or the tentative demand is ignored and demand is allocated optimally from scratch.

An attempt to "repair" one component of the Lagrangian solution will inevitably affect the other component, as the available capacity impacts the demand allocation and the demand allocation limits the possible capacity schedules. There are multiple ways how to approach this interplay. Given that computational efficiency is a key concern within the generation of upper bounds, below we consider a greedy approach to ensure sufficient capacity and repair the demand allocation. The demand allocation may be reoptimized afterwards, if required.

Greedy capacity increase and demand correction. We first focus on the greedy approach proposed by Jena et al. (2017) (and original proposed by Shulman (1991) for a simpler problem variant) that aims at correcting missing production capacity, along with a greedy correction of under- or over-served demand. Let Σ_1, Σ_2 and Σ_3 denote three subsets, containing the demand triplets $< i,p,t >$ that are exactly met, over-served and under-served, respectively. They are defined as follows:

$$\Sigma_1 = \left\{ < i,p,t >: \sum_{j \in J} x^{jt}_{i\overline{\ell(j,t)}p} = 1 \right\}, \Sigma_2 = \left\{ < i,p,t >: \sum_{j \in J} x^{jt}_{i\overline{\ell(j,t)}p} > 1 \right\}$$

$$\text{and } \Sigma_3 = \left\{ < i,p,t >: \sum_{j \in J} x^{jt}_{i\overline{\ell(j,t)}p} < 1 \right\}.$$

Based on this tentative demand allocation, the greedy approach proceeds as follows:

1. *Reduce redundant demand allocation*: more flow than required may currently be sent to certain customers. In an attempt to reduce redundant flows that is the most expensive, one may first sort, for each demand triplet $< i,p,t > \in \Sigma_2$, the allocation from facilities $j \in J$ in decreasing order of their allocation costs $g^{jt}_{i\ell p}$. Excess demand allocation is then iteratively reduced from one facility at a time until the requested demand quantity is met exactly.
2. *Increase capacity* in the case of capacity shortfall: if the total available capacity is insufficient to cover the total demand at certain time periods, the available capacity has to be increased. A naive approach

may identify facilities already opened at some other time-periods and increase the capacity for the critical time-period to the same capacity level. If this is not possible, any random facility may be opened. More elaborate approaches may want to consider the transportation costs to under-served customers.

3. *Complete missing demand allocation*: for each demand triplet $< i, p, t > \in \Sigma_3$, feasible allocations from facilities $j \in J$ are first sorted in increasing order of their allocation costs $g_{i\ell p}^{jt}$ (considering, of course, only facilities $j \in J$ with left-over capacity at time period $t \in T$). Missing demand is then iteratively increased from one facility at a time until the requested demand is met exactly.

4. *Reduce capacity* in the case of capacity excess: once the demand allocation is feasible, a corresponding optimal capacity schedule may be computed, just as done when solving the Lagrangian subproblem, but considering only capacity levels that are sufficiently high to serve the allocated demand.

More elaborate approaches. While the approach above may not provide optimal demand allocation, it is quite fast and can therefore be carried out at each iteration. Alternatively, once sufficient production capacity is ensured, an optimal demand allocation may be derived relatively quickly. This corresponds to a transportation problem in a bipartite graph, with production capacity as supply and customers as demand nodes. While this may be solved with several algorithms, from an implementation standpoint, an efficient way is to use specialized network solvers (such as the network algorithm from CPLEX), which can be reoptimized quickly with updated supply, demand and transportation costs.

In simple problem variants, the above greedy approach may be sufficient to generate upper bounds of sufficiently high quality (see, e.g. Jena et al., 2017). In more complex problem variants, the trade-off between different time-periods, costs for capacity adjustment, maintenance and demand allocation becomes too complex. Even more elaborate efforts to increase capacity may therefore not be effective when considered isolated from the other decisions, and therefore unnecessarily increase computing time. In such circumstances, a wiser choice may be to aim at improving existing feasible solutions within an improvement phase, where the interplay between these decisions may be jointly considered (see, e.g. Jena et al., 2016). One possibility is the construction of a restricted MIP model, which will be explained in Section 3.3.5.

Upper bound generation for special cases. The outlined upper bound procedure is naturally also valid for special cases of the GMC. In particular, with a single time-period or predefined capacities (as opposed to capacity levels), the procedures described above simplify trivially and significantly. Finally, note that the approaches outlined above also exactly apply to one of the proposed extensions. Specifically, in the case of modular facility structures (see Section 2.3), the capacity opening schedule is then given by a pair $(\ell(j,t), n(j,t))$ with $\ell \in L, n \in L$, indicating the available and existing capacity levels for each facility location $j \in J$ and each time period $t \in T$.

3.3.2 Repairing in the case of Complex Cost Functions

As a result of relaxing the demand constraints, the optimal demand allocation has been computed separately for each of the candidate facility locations by solving a piecewise-linear continuous knapsack. When generating feasible upper bounds, the demand allocation has to be solved by taking into consideration all available facilities at once. Given that the total production costs $F_\ell^j(U_{\ell q}^j)$ now depend on the total production quantity q at that facility (i.e., it has a piecewise-linear cost function), the optimal demand allocation cannot be computed by solving a transportation problem (or a minimum cost flow problem). As such, Lagrangian heuristics for planning problems that contain such cost functions resort to general purpose MIP solvers to solve for the optimal demand allocation (see, e.g., Štádlerová et al., 2023).

3.3.3 Repairing in the case of Facility Relocation

When relocation linking constraints have been relaxed, the greedy approach presented above can be extended by an adding a step carried out at the beginning of the procedure. This step aims at matching the proposed incoming and outgoing facility relocations of the different capacity sizes at each time period. Specifically, for each facility $j' \in J$ for which the Lagrangian solution suggests an outgoing relocation of size $\ell \in L$ at time period $t \in T$ (i.e., $w_{j't\ell}^- = 1$), one needs to find another location $j'' \in J, j' \neq j''$ with an incoming relocation

decisions of the same size and at the same time period (i.e., $w^+_{j''t\ell} = 1$). While some of the proposed incoming and outgoing relocation decisions can be adopted into a feasible solution, some of those decisions cannot be matched. Any procedure to carry out such matching should therefore ensure that the final capacity schedule remains coherent and feasible, and consider the resulting costs of the matching. Specifically:

- Not matching a suggested outgoing relocation typically implies that the facility will have to be shut down in order to converge to a feasible capacity schedule. Similarly, not matching a suggested incoming relocation implies that a facility will have to be constructed, given that the proposed capacity schedule assumes the availability of such a facility. This has some implications, as illustrated next.
- If two outgoing relocations from the same location j' have been selected at different time periods t' and t'', but there is no incoming relocation to this location in between these two time periods (for example, because this incoming relocation decision has not been matched), then it is implicit that a new facility had to be constructed before t'' (but after t') such that it can actually be relocated somewhere else. Note that this likely incurs unnecessarily high costs, since constructing a facility at location j' only to relocate it to location j'' afterwards is likely to be more expensive then directly constructing at location j''.
- Similarly, if two incoming relocations to j'' have been selected for different time periods, but there is no outgoing relocation in between these time periods, then this implies that the facility that has been relocated here first has been shut down in order to make place for the second incoming facility. Most likely, this also is sub-optimal.

A greedy approach (similar to the one used by Jena et al. (2016)) may iterate through the time periods, count the number of matches of each size $\ell \in L, \ell \geq 1$ at each time period, and then arbitrarily select some of those facilities to select these matches. If the costs for facility shutdown and construction are not the same at all locations, outgoing relocations may be prioritized for locations where shutdown is expensive, and incoming relocation may be prioritized for locations where construction is expensive.

Optimal matching. A more complex approach, which has not yet been considered in the literature, would be to optimize these matching decisions. The corresponding matching may be formulated as a restricted variant of the original MIP formulation, where only a fraction of decisions is used. Specifically, demand allocation is ignored. The resulting problem is a min-cost network flow problem, which aims at deciding whether the suggested outgoing $w^-_{jt\ell}$ and incoming $w^+_{jt\ell}$ relocations should be adopted, and which capacity expansions and reductions should be used in case one of the suggested relocations is not adopted. Capacity expansion decisions are only possible when an incoming capacity relocation has been suggested, and capacity reduction decisions are only possible when an outgoing capacity relocation has been suggested.

While such an optimal matching may be carried out computationally efficiently, it has to be noted that the relocation decisions, along with certain capacity changes may also be reoptimized in an improvement phase, as exemplified in Section 3.3.5.

3.3.4 Repairing in the case of Uncertainty via Scenarios

The upper bound procedure outlined in Section 3.3.1 can easily be adapted to the case where a set of demand scenarios is used, but all capacity decisions are taken in the first stage of the stochastic planning problem. Specifically, the costs for the demand allocation has to be computed over the given demand scenarios according to their probabilities. While, ideally, all demand scenarios are considered in order to obtain a realistic estimation of the expected demand allocation costs when modifying the capacity schedule, a subset can be used in order to speed up the procedure. For instance, Štádlerová et al. (2023) use four reference scenarios to approximate the expected demand allocation costs, corresponding to those that have maximum, minimum, mean and medium total demand among all scenarios. In particular, the scenario with maximum total demand ensures that the designed capacity schedule has sufficient capacity to satisfy all demand.

3.3.5 Improvement Strategies

The feasible solutions derived from the Lagrangian solutions may be of sufficiently high quality when the problem is not too complex (see, e.g., Shulman, 1991; Correia and Captivo, 2003; Wu et al., 2006). In more complex problem variants, the synergies and interaction between the various decisions may be too high, and the derived solutions may not be of sufficient quality. In this case, improvement heuristics can be employed to improve upon existing solutions. A variety of improvement heuristics has been proposed in the literature, including:

- Local and tabu searches: starting from a specific feasible solution, search procedures, exploring manually defined neighborhoods tailored to the problem structure, can be employed to iteratively improve the solution (see, e.g., Correia and Captivo, 2006; Hinojosa et al., 2008; Li et al., 2009).
- Iterative Matheuristics such as Local Branching (Fischetti and Lodi, 2003): starting from a feasible solution, a maximum number of variables may change their solution variables in order to allow for iterative improvements.
- Restricted MIP models: considering only a subset of the original decisions may sufficiently reduce the problem size such that it can be solved quickly (see, e.g., Jena et al., 2017).

In contrast to the restricted MIP, the first two types of heuristics are iterative procedures. Matheuristics and the restricted MIP have the advantage that search neighbourhoods do not have to be tailored to the problem structure. All of such methods are typically time-intensive. The feasible solution on which such methods are applied should therefore be chosen carefully. They may, for example, be employed at each iteration at which a new best upper bound has been found, in order to further polish this bound. In the simplest case, the improvement method is added as a second phase once the Lagrangian relaxation has sufficiently converged.

Restricted MIP Model. While tailored improvement heuristics may be very efficient, restricted MIP models have the advantage that they are flexible, easy to implement and consider the complex synergies among the different types of decisions. In other words, they may be able to consider neighbourhoods that are manually hard to design. Selecting the decisions available in the restricted model is a delicate trade-off: making available too few decisions may not allow for a significant improvement, while making available too many decisions may make the resulting model hard to solve. In order to define a restricted MIP to the problem at hand, one may first identify the part of the capacity schedule decisions that should be fixed (i.e., the facilities, along with their capacity levels for certain time periods). For locations and time periods where the schedule is not fixed, one then defines the available capacity levels, as well as the eligible capacity transitions (i.e., the eligible changes between different capacity levels). Clearly, it is crucial to identify decisions that may be part of the optimal solution to the original problem. Luckily, in this regard, the iterative process of the Lagrangian Relaxation may be helpful. Two options can be considered (Jena et al., 2017):

1. Consider decisions selected by the **Lagrangian solutions**: Assume that a total of n^{iter} number of iterations has been performed. Let $n^C_{j\ell t}$ be the number of Lagrangian solutions where capacity level $\ell \in L$ has been selected at location $j \in J$ and time period $t \in T$. Further, let L^R_{jt} be the set of capacity levels that will be made available in the restricted MIP at location $j \in J$ and time period $t \in T$. **Capacity opening decisions may be fixed**, if they appear in at least $100 \times pFix$ percent of all Lagrangian solutions, where $pFix$ is a predefined parameter from $]0, 1]$. Specifically, a capacity at location $j \in J$ and time period $t \in T$ is fixed to $\ell \in L$, i.e., $L^R_{jt} = \{\ell\}$, if $n^C_{j\ell t}/n^{iter} \geq pFix$. As **available capacity levels** at each location $j \in J$ and period $t \in T$, one may use those that have been the most popular. That is, if L^R_{jt} has not been fixed in the previous step, it can be composed by the n^S capacity levels that have appeared most often in the Lagrangian solutions, where n^S is a predefined parameter. Finally, the set of **eligible capacity transitions** (corresponding to variables $y^{jt}_{\ell_1\ell_2}$) may, in the most generous case, be defined as all transitions among the defined sets of available capacity levels, i.e., $\ell_1 \in L^R_{jt}$, $\ell_2 \in L^R_{j(t+1)}$ $\forall j \in J, \forall t \in T$.

2. Decisions endorsed by the aggregated **Bundle method**: the aggregated bundle method (see Section 3.2.2) provides a weight θ^m for each of the previous Lagrangian multipliers and solutions $m \in B$. This weight can be seen as a probability that the corresponding multipliers, and hence the associated solution, lead to a good opening schedule. This information can be exploited in various ways. A straightforward approach computes the probability $\tilde{y}_\ell^{jt} = \sum_{m \in B} \theta^m \overline{y}_\ell^{mjt}$, where $\overline{y}_\ell^{mjt}$ takes value 1 if the Lagrangian solution of Bundle member m selects capacity level $\ell \in L$ for location $j \in J$ at time period $t \in T$. A restricted MIP model can then be constructed: for each location $j \in J$ and time period $t \in T$, a capacity decision is fixed if $\tilde{y}_\ell^{jt} \geq pFix$. If a decision is not fixed, the set L_{jt}^R of available capacity levels is composed of the n^S capacity levels with highest probabilities $\tilde{y}_\ell^{jt}$.

4 Conclusions

Even though facility location problems have been investigated since over 50 years, they still remain an active component of contemporary research, given their strategic importance in most supply chains. In this paper, we have reviewed the chronological development of solution methods based on Lagrangian relaxation to tackle facility location problems, which have become increasingly complex over time.

We have reviewed a strong formulation, that allows for multiple time periods, multiple commodities, multiple capacity levels and the adjustment of capacity along time. We have also explicitly addressed extensions that account for complex production cost functions, modular structures of facilities, and the incorporation of parameter uncertainty. As such, we have reviewed most of the relevant features in modern facility location planning problems. We have then shifted focus to the efficient solution of such problem variants via Lagrangian relaxation: specifically, the solution of the Lagrangian subproblem, the solution of the Lagrangian dual problem and the generation of feasible upper bounds. As such, we have provided a systematic guide to modeling and solving complex facility location problems, that will, hopefully, be useful to practitioners and researchers that aim at solving problem variants composed of any combination of the here discussed problem characteristics.

While Lagrangian relaxation has certainly received less than attention than other decomposition methods in recent years, more recent literature suggests that this approach remains powerful even when faced with complex problem variants. While such problems are likely to become even more complex in the future in order to account for more details of the practical planning problems, certain contemporary applications demonstrate the need to integrate location problems in more complex planning contexts. For example, the design of carbon-dioxide transportation and storage networks translates into an integrated location and network design problem, where both the capturing and storage facilities, as well as the pipelines, have to be selected from a set of available (capacity) sizes. Such problems remain challenging to solve. In particular, demands may not be specified separately for different customer locations. Instead, one may have to attain a single global target level. In such case, the popular strong inequalities, which have been key to developing strong formulations, become ineffective, which makes it difficult to obtain strong bounds. Research on such problems is eagerly needed, and is therefore attractive from an academic and a practical point of view.

References

Agar MC, Salhi S (1998) Lagrangean heuristics applied to a variety of large capacitated plant location problems. Journal of the Operational Research Society 49(10):1072–1084

Alarcon-Gerbier E, Buscher U (2022) Modular and mobile facility location problems: A systematic review. Computers & Industrial Engineering p 108734

Allen RC, Avraamidou S, Butenko S, Pistikopoulos EN (2022) Solution strategies for integrated distribution, production, and routing problems arising in modular manufacturing. Technical report

Antunes AP, Berman O, Bigotte Ja, Krass D (2009) A location model for urban hierarchy planning with population dynamics. Environment and Planning A 41(4):996–1016

Barcelo J, Hallefjord A, Fernandez E, Jörnsten K (1990) Lagrangean relaxation and constraint generation procedures for capacitated plant location problems with single sourcing. OR Spektrum 12(2):79–88

Beasley JE (1993) Lagrangean heuristics for location problems. European Journal of Operational Research 65(3):383–399

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization, vol 28. Princeton university press

Birge JR, Louveaux F (2011) Introduction to stochastic programming. Springer Science & Business Media

Cabezas X, García S, Martin-Barreiro C, Delgado E, Leiva V (2021) A two-stage location problem with order solved using a lagrangian algorithm and stochastic programming for a potential use in covid-19 vaccination based on sensor-related data. Sensors 21(16):5352

Chardaire P, Sutter MC A Costa (1996) Solving the dynamic facility location problem. Networks 28(2):117–124

Chouman M, Crainic TG, Gendron B (2016) Commodity representations and cutset-based inequalities for multicommodity capacitated fixed-charge network design. Transportation Science p forthcoming

Christensen TRL, Klose A (2021) A fast exact method for the capacitated facility location problem with differentiable convex production costs. European Journal of Operational Research 292(3):855–868

Contreras I, Díaz JA, Fernández E (2009) Lagrangean relaxation for the capacitated hub location problem with single assignment. OR spectrum 31:483–505

Correia I, Captivo EM (2006) Bounds for the single source modular capacitated plant location problem. Computers & Operations Research 33(10):2991–3003

Correia I, Captivo ME (2003) A Lagrangean heuristic for a modular capacitated location problem. Annals of Operations Research 122(1):141–161

Diabat A, Richard JP, Codrington CW (2011) A Lagrangian relaxation approach to simultaneous strategic and tactical planning in supply chain design. Annals of Operations Research 203(1):55–80

Fischetti M, Lodi A (2003) Local branching. Mathematical programming 98:23–47

Frangioni A (2005) About Lagrangian Methods in Integer Optimization. Annals of Operations Research 139(1):163–193

Frangioni A, Gallo G (1999) A bundle type dual-ascent approach to linear multicommodity min-cost flow problems. INFORMS Journal on Computing 11(4):370–393

Frangioni A, Gorgone E (2014) Bundle methods for sum-functions with "easy" components: Applications to multicommodity network design. Mathematical Programming 145(1):133–161

Gendron B (2011) Decomposition methods for network design. Procedia - Social and Behavioral Sciences 20:31–37

Gendron B, Khuong PV, Semet F (2016) A lagrangian-based branch-and-bound algorithm for the two-level uncapacitated facility location problem withguin single-assignment constraints. Transportation Science 50(4):1286–1299

Ghodsi G (2012) A lagrangian relaxation approach to a two-stage stochastic facility location problem with second-stage activation cost. Master's thesis, University of Waterloo

Görtz S, Klose A (2012) A simple but usually fast branch-and-bound algorithm for the capacitated facility location problem. INFORMS Journal on Computing 24(4):597–610

Guignard M (2003) Lagrangean Relaxation. Top 11(2):151–228

Guignard-Spielberg M, Kim S (1983) A strong Lagrangian relaxation for capacitated plant location problems. Tech. rep., Department of Statistics Technical Report

Held M, Wolfe P, Crowder HP (1974) Validation of subgradient optimization. Mathematical Programming 6(1):62–88

Hinojosa Y, Puerto J, Fernández F (2000) A multiperiod two-echelon multicommodity capacitated plant location problem. European Journal of Operational Research 123(2):271–291

Hinojosa Y, Kalcsics J, Nickel S, Puerto J, Velten S (2008) Dynamic supply chain design with inventory. Computers & Operations Research 35(2):373–391

Holmberg K, Ling J (1997) A Lagrangean heuristic for the facility location problem with staircase costs. European Journal of Operational Research 97(1):63–74

Jena SD (2014) Dynamic facility location with modular capacities: Models, algorithms and applications in forestry. PhD thesis, Université de Montréal

Jena SD, Cordeau JF, Gendron B (2015a) Dynamic facility location with generalized modular capacities. Transportation Science 49(3):484–499

Jena SD, Cordeau JF, Gendron B (2015b) Modeling and solving a logging camp location problem. Annals of Operations Research 232(1):151–177

Jena SD, Cordeau JF, Gendron B (2016) Solving a dynamic facility location problem with partial closing and reopening. Computers & Operations Research 67:143–154

Jena SD, Cordeau JF, Gendron B (2017) Lagrangian heuristics for large-scale dynamic facility location with generalized modular capacities. INFORMS Journal on Computing 29(3):388–404

Kadri A, Koné O, Gendron B (2022) A lagrangian heuristic for the multicommodity capacitated location problem with balancing requirements. Computers & Operations Research 142:105720

Li J, Chu F, Prins C (2009) Lower and upper bounds for a capacitated plant location problem with multicommodity flow. Computers & Operations Research 36(11):3019–3030

Marín A, Martínez-Merino LI, Rodríguez-Chía AM, Saldanha-da Gama F (2018) Multi-period stochastic covering location problems: Modeling framework and solution approach. European journal of operational research 268(2):432–449

Marsten RE, Hogan WW, Blankenship JW (1975) The boxstep method for large-scale optimization. Operations Research 23(3):389–405

Pacheco Paneque M, Gendron B, Sharif Azadeh S, Bierlaire M (2022) A lagrangian decomposition scheme for choice-based optimization. Computers & Operations Research 148:105985

Padberg MW, Van Roy TJ, Wolsey LA (1983) Valid linear inequalities for fixed charge problems. Operations Research 33(4):842–861

Schütz P, Tomasgard A, Ahmed S (2009) Supply chain design under uncertainty using sample average approximation and dual decomposition. European journal of operational research 199(2):409–419

Shulman A (1991) An algorithm for solving dynamic capacitated plant location problems with discrete expansion sizes. Operations Research 39(3):423–436

Sridharan R (1991) A lagrangian heuristic for the capacitated plant location problem with side constraints. Journal of the Operational Research Society 42(7):579–585

Štádlerová Š, Aglen TM, Hofstad A, Schütz P (2022) Locating hydrogen production in norway under uncertainty. In: Computational Logistics: 13th International Conference, ICCL 2022, Barcelona, Spain, September 21–23, 2022, Proceedings, Springer, pp 306–321

Štádlerová Š, Jena SD, Schütz P (2023) Using lagrangian relaxation to locate hydrogen production facilities under uncertain demand: A case study from norway. Computational Management Science

Vahidnia MH, Alesheikh AA, Alimohammadi A (2009) Hospital site selection using fuzzy AHP and its derivatives. Journal of environmental management 90(10):3048–56

Wu L, Zhang X, Zhang J (2006) Capacitated facility location problem with general setup cost. Computers & Operations Research 33(5):1226–1241