Computers & Operations Research 163 (2024) 106395

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Multi-period facility location and capacity expansion with modular capacities and convex short-term costs ☆

Šárka Štádlerová *, Peter Schütz, Asgeir Tomasgard

Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, 7491 Trondheim, Norway

ARTICLE INFO

Keywords:
Multi-period facility location
Capacity expansion
Lagrangian relaxation

ABSTRACT

In this paper, we consider a multi-period facility location problem with capacity expansion motivated by the real-world problem of establishing hydrogen production infrastructure in Norway. The problem is formulated using modular capacities that capture economies of scale in production costs. The costs of opening a facility are represented by concave long-term costs, while the production costs of each capacity level are given by convex short-term costs. In our model, we allow only one expansion during the planning horizon, and have to observe limits on minimum production quantities. The objective is to minimize the sum of investment, expansion, production, and distribution costs while satisfying customer demand. To solve the problem we implement a solution method based on Lagrangian relaxation. The lower bound is calculated using a dynamic programming approach. To obtain an upper bound solution, we develop a greedy heuristic that converts the solution to the Lagrangian dual into a feasible solution. The approach is tested on different problem instances based on real-world data. The results show that our solution method based on Lagrangian relaxation outperforms Gurobi in terms of run time for all tested instances. Our Lagrangian based approach also always finds good or even near-optimal solutions, whereas Gurobi fails to find feasible solutions for some of the larger instances.

1. Introduction

The problem presented in this paper is motivated by the challenge of designing the hydrogen supply chain for maritime transportation in Norway. Designing the production network for satisfying hydrogen demand can be formulated as a facility location problem with a capacity extension, i.e. finding the optimal decisions regarding opening, expanding and operating production facilities such that the cost of satisfying demand is minimized. Hydrogen production is subject to economies of scale. These economies of scale depend on the size of the facility and the utilization of the installed capacity (Hirth et al., 2019). We model economies of scale by means of modular capacities subject to a general long-term cost function for investment and expansion and a convex short-term production cost function that depends on the installed capacity, see also Štádlerová and Schütz (2021). Due to increasing hydrogen demand during the planning horizon, we only allow expanding capacity. Closing facilities is not permitted. Due to the limited planning horizon and costly investments, we only allow a single capacity expansion. The production process is characterized by minimum production quantities, implying that an open facility always has to be operated. In this sense, our problem differs from the problem presented in Shulman (1991) and Jena et al. (2016, 2017) where the authors allow modular capacity adjustments in each time period and have no requirements on minimum production quantities. We further generalize the multi-period facility location and capacity expansion problem by distinguishing between long-term investment and expansion costs and short-term production costs. For each modular capacity, we have a convex piecewise linear short-term cost function, so that the production costs depend on both installed capacity and capacity utilization. However, the short-term production costs are independent of whether the capacity was reached by opening or by expansion.

The class of multi-period facility location problems with capacity expansion belongs to the group of NP-hard problems (Shulman, 1991). Thus, they are hard to solve for larger instances and we need efficient solution methods to obtain a good feasible solution. Methods based on Lagrangian relaxation and Benders decomposition perform well and their advantage is that they can provide information about the solution quality compared to heuristics.

In this paper, we solve a multi-period facility location problem with capacity expansion with general long-term costs and convex short-term costs in the objective function. The model solved here is the same as

☆ This work was performed within MoZEES, a Norwegian Center for Environment-friendly Energy Research (FME), co-sponsored by the Research Council of Norway (project number 257653) and 40 partners from research, industry and the public sector.

* Corresponding author.
E-mail address: sarka.stadlerova@ntnu.no (Š. Štádlerová).

https://doi.org/10.1016/j.cor.2023.106395
Received 27 May 2022; Received in revised form 24 July 2023; Accepted 16 August 2023
Available online 19 August 2023

the one presented in Štádlerová and Schütz (2021), but in contrast to that paper, we focus here on developing an efficient solution method based on Lagrangian relaxation. We show that the Lagrangian dual can be solved efficiently with a dynamic programming algorithm. Even with a simple greedy heuristic, we are able find good feasible solutions for more test instances than the commercial solver. The results also show that our Lagrangian heuristics outperforms commercial software in terms of run time when finding good, but not necessarily optimal solutions. For small instances, commercial software can find optimal solutions but may fail to find any feasible solution for large instances.

The remainder of this paper is structured as follows: we provide a literature review on the modelling of multi-period facility location problems and their solution in Section 2, and the mathematical formulation of the problem in Section 3. The solution method is presented in Section 4 and the case description is provided in Section 5. Computational results and conclusions are discussed in Sections 6 and 7, respectively.

2. Literature review

We structure the literature review into two main parts. In Section 2.1, we focus on literature related to the modelling approach of multi-period facility location and expansion problems and we also discuss the techniques how to incorporate economies of scale into the model. Then, we review different methods for solving this type of problem in Section 2.2.

2.1. Modelling approach

We organize the review on the modelling approach into two main groups. First, we discuss facility location and capacity expansion with modular capacities and then models with continuous capacities where the maximum capacity of a facility is given. Then, we review the modelling of economies of scale with modular capacities, and with a continuous objective.

Shulman (1991) and Dias et al. (2007b) study a multi-period plant location problem with modular capacities and with discrete expansion where a plant is modelled as a set of facilities. In these papers, capacity expansion is achieved by building an additional facility at the same location, while capacity reduction leads to the closing of facilities. The production costs are defined for each facility and depend only on facility type and quantity produced in the facility. Modular capacities are also used in the work by Jena et al. (2015, 2016, 2017), and Štádlerová and Schütz (2021). However, in these papers, the capacity expansion is modelled as a modification of existing facilities. In the work by Jena et al. (2015, 2016, 2017), capacity expansion and reduction, as well as closing and reopening, are allowed multiple times and Jena et al. (2016) present also the option of relocating capacities. On the other hand, the model by Štádlerová and Schütz (2021) allows for only one expansion during the planning horizon.

A model with continuous capacities and limited maximum capacity is presented by Hinojosa et al. (2008), Behmardi and Lee (2008), and Torres-Soto and Üster (2011). Hinojosa et al. (2008) present a slightly different modelling approach where a set of existing facilities and the target number of operating facilities at the end of the planning horizon is given. Initial facilities can be closed during the planning horizon but reopening is not allowed. Facilities that are opened during the planning horizon cannot be closed anymore. In contrast to the other papers we refer to, Behmardi and Lee (2008) formulate their problem as a profit maximization model, where demand does not have to be satisfied. Torres-Soto and Üster (2011) present a model where facilities can be opened and closed multiple times in response to varying demand and provide a comparison of problems with and without facility relocation. See also the review by Melo et al. (2006) for an overview of mathematical modelling frameworks for dynamic facility location and expansion models, covering different facility modification strategies. An overview over facility location and supply chain network design problems with focus on hydrogen can be found in Li et al. (2019).

Economies of scale in production processes often come from sharing the investment and expansion costs over more units of the produced product. As a result, higher production quantities lead to lower unit costs (Haldi and Whitcomb, 1967). A modular formulation of a capacitated facility location model allows for modelling of economies of scale as shown in Correia and Captivo (2003). They present non-linear costs dependent on capacity as they split investment and operating costs and provide specific unit operating costs for each facility size. Modular capacities capturing economies of scale are also used by Jena et al. (2015, 2016, 2017), Correia and Melo (2021), and Štádlerová and Schütz (2021). Another approach capturing economies of scale for a problem with discrete capacities is the use of a piecewise linear staircase cost function that also enables to model different production costs at different capacity levels (Holmberg, 1994).

The work by Van den Broek et al. (2006) and Schütz et al. (2008) differs from the previous papers as they have a continuous and differentiable objective. The authors study a facility location problem with a non-linear, non-convex, and non-concave objective function. Their cost function can be considered as a combination of non-linear costs depending on installed capacity from Correia and Captivo (2003) combined with the linear staircase cost approximation presented by Holmberg (1994).

2.2. Solution methods

Lagrangian relaxation is a well established technique for solving multi-period facility location and expansion problems. In general, the demand constraint is relaxed and then, the problem becomes separable in facility locations. Shulman (1991) shows that the resulting subproblems can be solved as shortest path problems. The Lagrangian dual is then solved by means of a subgradient method. Jena et al. (2016, 2017) present a similar solution method and also compare the use of subgradient and bundle methods for solving the Lagrangian dual.

Hinojosa et al. (2008) apply Lagrangian relaxation with a subgradient method to solve a model with inventory constraints. They relax the demand and flow conservation constraints to obtain subproblems separable in facility locations. Li et al. (2009) combine Lagrangian relaxation with a tabu search approach to improve the upper bound. They solve a problem that combines facility location and multicommodity flow distribution with transshipment points.

Castro et al. (2017) solve the multi-period facility location with Benders decomposition. The problem is decomposed into a master problem and a subproblem. The master problem contains the binary integer variables and provides an opening schedule of facilities while the subproblem consisting of continuous variables provides an optimal demand allocation. Torres-Soto and Üster (2011) compare both solution methods: Lagrangian relaxation and Benders decomposition, for a facility location and relocation problem. They show that the performance of the solution methods depends on the input data structure.

Arostegui et al. (2006) provide a comparison of tabu search, simulated annealing and genetic algorithms for facility location problems. Tabu search is also used by Melo et al. (2012) to solve a facility relocation problem. A primal–dual heuristic which aims to build feasible primal solutions based on admissible dual solutions is used in Dias et al. (2006, 2007a). The authors study a problem with minimum and maximum production requirements and show that these requirements make the problem more complex. Sauvey et al. (2020) develop a two-step heuristic to solve a multi-period facility location problem with modular capacities and delayed demand satisfaction. In the first step, their heuristic constructs feasible solutions that are further improved in the second step. A comparison of a tabu search and genetic algorithm for the facility location and capacity expansion problem studied in Jena et al. (2015) is provided by Silva et al. (2021). The authors show that for some instances, heuristics can find optimal or nearly optimal solutions.

For more examples of multi-period facility location and supply chain design as well as solution methods for these problems, see also the reviews by Melo et al. (2009), Arabani and Farahani (2012), and Nickel and Saldanha da Gama (2019).

3. Model formulation

We study a multi-period facility location problem with capacity expansion. We aim to find the optimal schedule for opening and expanding production facilities, including the optimal technology choice. Demand must be met exactly, shortfall or storage of excess production are not allowed. The objective is to minimize the discounted sum of investment and expansion costs, production costs, and distribution costs over the planning horizon.

3.1. Modelling approach

The set of available technologies and modular capacities is independent of the facility location. For each capacity and technology, we have a specific short-term production cost function that represents the costs at different utilization levels of the installed capacity. If a facility is opened, requirements regarding minimal utilization apply and these must be met in each period. Each customer has a specific yearly demand that has to be satisfied. The demand is non-decreasing during the planning horizon. Satisfying customer demand incurs distribution costs that depend on the distance between the customer and the facility. Note that a facility may not be able to serve all customers.

We consider capacity expansion being an expensive strategic decision. Since we assume a limited planning horizon, multiple expansions are not desired and we allow capacity expansion only once during the planning horizon. Once the facility is opened, it cannot be closed, it can be only expanded. Switching technologies is not permitted.

We model capacity expansion as a jump between the available capacities. Similar to Jena et al. (2015), the size of the expansion is only limited by the highest available capacity. The costs of expanding a facility are illustrated in Fig. 3.1(a). Let Q_k and Q_l be the production capacities installed at investment costs C_k and C_l, respectively. Capacity Q_l can also be achieved through investing in a facility of size Q_k and later expanding that facility to size Q_l. The corresponding expansion costs are E_{kl}, with $E_{kl} > C_l - C_k$. Expansion from capacity k towards capacity l implies $l > k$ and production capacity $Q_l > Q_k$. Thus, after expansion the facility operates with higher capacity than originally installed.

The short-term production cost function is modelled as a piecewise linear convex function. The short-term production cost function is convex by assumption based on Allen et al. (2012). Higher utilization of installed capacity leads to lower unit production costs. Fig. 3.1(b) shows the production costs F_k and F_l and corresponding short-term production functions $f_k(q)$ and $f_l(q)$ for capacities Q_k and Q_l, respectively. The short-term cost function depends only on the installed capacity, but it is independent of how this capacity was reached. We define quantities and costs at the breakpoints of the piecewise linear convex short-term cost function for each capacity level k. The lowest breakpoint represents the minimum production requirements for a given capacity and the highest breakpoint corresponds to the maximum capacity limit. By linear convex combination of these breakpoints, arbitrary quantities between the minimum and maximum production limit can be achieved and the corresponding costs can be computed. This modelling approach is also used in Štádlerová and Schütz (2021).

3.2. Optimization model

Let us first introduce the following notation:

Sets

$\mathcal{I}$	Set of possible facility locations
$\mathcal{J}$	Set of customer locations
$\mathcal{K}$	Sorted set of available discrete capacities, $\mathcal{K} = \{1, 2, \dots, \overline{K}\}$
$\mathcal{R}$	Set of available production technologies
$\mathcal{B}_{kr}$	Set of breakpoints of the short run cost function specific for installed capacity level $k \in \mathcal{K}$ and technology $r \in \mathcal{R}$
$\mathcal{T}$	Sorted set of time periods, $\mathcal{T} = \{1, 2, \dots, \overline{T}\}$

Parameters and coefficients

C_{kr}	Investment costs for capacity level $k \in \mathcal{K}$ and technology $r \in \mathcal{R}$;
D_{jt}	Demand at customer point $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$;
E_{klr}	Costs of expansion from capacity level $k \in \mathcal{K}$ to capacity $l \in \mathcal{K} : l > k$ for technology $r \in \mathcal{R}$;
F_{bkr}	Costs at breakpoint $b \in \mathcal{B}_{kr}$ of the short-run cost function given for capacity level $k \in \mathcal{K}$ and for technology $r \in \mathcal{R}$;
L_{ijt}	1 if demand at customer point $j \in \mathcal{J}$ can be served from facility at location $i \in \mathcal{I}$ in time period $t \in \mathcal{T}$, 0 otherwise;
Q_{bkr}	Production volume at breakpoint $b \in \mathcal{B}_{kr}$ of the short-run cost function, for point $k \in \mathcal{K}$ of the capacity function and technology $r \in \mathcal{R}$;
T_{ijt}	Transportation costs from facility at location $i \in \mathcal{I}$ to customer point $j \in \mathcal{J}$ in time period $t \in \mathcal{T}$;
y_{iklr0}	1, if there is initially an opened facility with installed capacity $k \in \mathcal{K}$, operating at capacity level $l \in \mathcal{K} : l \geq k$ with technology $r \in \mathcal{R}$ at facility location $i \in \mathcal{I}$, 0 otherwise;
δ_t	Discount factor in time period $t \in \mathcal{T}$;

Decision variables

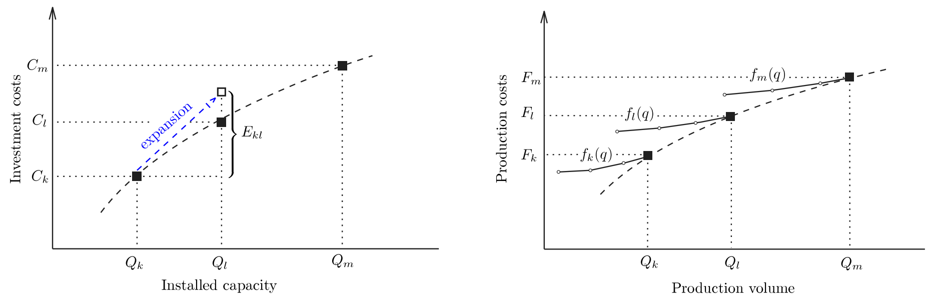
x_{ijlrt}	Amount of customer demand at point $j \in \mathcal{J}$ satisfied from facility at location $i \in \mathcal{I}$ operating at capacity level $l \in \mathcal{K}$ and using technology $r \in \mathcal{R}$ in time period $t \in \mathcal{T}$;
y_{iklrt}	1 if facility at location $i \in \mathcal{I}$ was opened with capacity level $k \in \mathcal{K}$ and is operated at capacity level $l \in \mathcal{K} : l \geq k$ using technology $r \in \mathcal{R}$ in time period $t \in \mathcal{T}$, 0 otherwise;
μ_{bilrt}	Weight of breakpoint $b \in \mathcal{B}_{lr}$ at facility location $i \in \mathcal{I}$ for capacity level $l \in \mathcal{K}$ and technology $r \in \mathcal{R}$ in time period $t \in \mathcal{T}$.

The problem is given as:

$$\begin{aligned}\min \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l\geq k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t C_{kr}\left(y_{iklrt} - y_{iklr(t-1)}\right) + \\ \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l>k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t E_{klr}(y_{iklrt} - y_{iklr(t-1)}) + \\ \sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{b\in\mathcal{B}_{lr}}\sum_{t\in\mathcal{T}} \delta_t F_{blr}\mu_{bilrt} + \\ \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t T_{ijt} x_{ijlrt},\end{aligned} \tag{1}$$

subject to:

$$\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l\geq k}\sum_{r\in\mathcal{R}} y_{iklrt} \leq 1, \qquad i \in \mathcal{I}, t \in \mathcal{T}, \tag{2}$$



(a) Investment and expansion costs (b) Production costs

Fig. 3.1. Short-term and long-term costs.

$$\sum_{l\in\mathcal{K}:l\geq k} y_{iklrt} \geq \sum_{l\in\mathcal{K}:l\geq k} y_{iklr(t-1)}, \qquad i\in\mathcal{I},k\in\mathcal{K},r\in\mathcal{R},t\in\mathcal{T}, \tag{3}$$

$$y_{iklrt} - y_{iklr(t-1)} \geq 0, \qquad i\in\mathcal{I},k\in\mathcal{K},l\in\mathcal{K}:l>k,r\in\mathcal{R},t\in\mathcal{T}, \tag{4}$$

$$\sum_{b\in\mathcal{B}_{lr}} \mu_{bilrt} = \sum_{k\in\mathcal{K}} y_{iklrt}, \qquad i\in\mathcal{I},l\in\mathcal{K},r\in\mathcal{R},t\in\mathcal{T}, \tag{5}$$

$$\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{K}} x_{ijlrt} = \sum_{l\in\mathcal{K}}\sum_{b\in\mathcal{B}_{lr}} Q_{blr}\mu_{bilrt}, \qquad i\in\mathcal{I},r\in\mathcal{R},t\in\mathcal{T}, \tag{6}$$

$$\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}} x_{ijlrt} = D_{jt}, \qquad j\in\mathcal{J},t\in\mathcal{T}, \tag{7}$$

$$x_{ijlrt} \leq L_{ijt}D_{jt}\sum_{k\in\mathcal{K}:k\leq l} y_{iklrt}, \qquad i\in\mathcal{I},j\in\mathcal{J},l\in\mathcal{K},r\in\mathcal{R},t\in\mathcal{T}, \tag{8}$$

$$y_{iklrt} \in \{0,1\}, \qquad i\in\mathcal{I},k\in\mathcal{K},l\in\mathcal{K}:l\geq k,r\in\mathcal{R},t\in\mathcal{T}, \tag{9}$$

$$x_{ijlrt} \geq 0, \qquad i\in\mathcal{I},j\in\mathcal{J},l\in\mathcal{K},r\in\mathcal{R},t\in\mathcal{T}, \tag{10}$$

$$\mu_{bilrt} \geq 0, \qquad i\in\mathcal{I},l\in\mathcal{K},r\in\mathcal{R},b\in\mathcal{B}_{lr},t\in\mathcal{T}. \tag{11}$$

The objective function (1) minimizes the discounted sum of investment costs, expansion costs, production costs, and distribution costs.

Restrictions (2) ensure that no more than one facility can be opened and operated at the given location. Once a facility is built, the only allowed modification is expansion, facility closing is not allowed. This is guaranteed by the constraints (3). Variable y_{iklrt} contains information about the initially installed capacity k as well as capacity l that is operated during period t. Note that $l \geq k$. With expansion in time period t', variable y_{ikkrt} becomes zero: $y_{ikkrt} = 0$ for $t \geq t'$ while variable y_{iklrt} becomes one: $y_{iklrt} = 1$ for $t \geq t'$. Inequalities (4) ensure that capacity index l can change only once, limiting the number of capacity expansions to one. Eqs. (5) link the correct short-term production cost function to the operated capacity l in period t and satisfy that production is allocated only to open facilities. In our model, demand must be met exactly and we do not allow for demand shortfall or storage. The requirement that the entire production has to be distributed to customers is expressed by (6). Constraints (6) also implicitly express the minimum production requirements as the production volumes at breakpoint $b \in \mathcal{B}_{kr}$ are strictly positive. However, this model formulation can be also used for problems without minimum production requirements if the production quantities at breakpoint $b \in \mathcal{B}_{kr}$ are defined starting from zero. The demand satisfaction is ensured by (7). Constraints (8) specify if facility i can serve customer j. Restrictions (9)–(11) are the binary and non-negativity requirements.

4. Solution method

In this section, we present our solution approach based on Lagrangian relaxation. In Section 4.1, we provide the formulation of the Lagrangian subproblem. The solution of the Lagrangian subproblem is further described in Section 4.2 before the procedure for updating the Lagrangian multipliers is discussed in Section 4.3. Finally, we present the heuristic for constructing a feasible solution in Section 4.4.

4.1. Relaxed problem

We relax demand constraint (7) and define λ_{jt} as the matrix of Lagrangian multipliers. As a result, we obtain the following Lagrangian subproblem:

$$\begin{aligned}\min &\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l\geq k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t C_{kr}\left(y_{iklrt} - y_{iklr(t-1)}\right) + \\ &\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l>k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t E_{klr}(y_{iklrt} - y_{iklr(t-1)}) + \\ &\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{b\in\mathcal{B}_{lr}}\sum_{t\in\mathcal{T}} \delta_t F_{blr}\mu_{bilrt} + \\ &\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} (\delta_t T_{ijt} - \lambda_{jt})x_{ijlrt} + \\ &\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}} \lambda_{jt}D_{jt},\end{aligned} \tag{12}$$

subject to (2)–(6) and (8)–(11).

For given multiplies λ_{jt}, the expression $\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}} \lambda_{jt}D_{jt}$ is constant and the problem becomes separable in facility locations. We can therefore solve the problem for each facility location i independently. Let $g_i(\lambda)$ be the optimal value of the Lagrangian subproblem for the facility location i. The objective function (12) can then be reformulated as:

$$LR(\lambda) = \sum_{i\in\mathcal{I}} g_i(\lambda) + \sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}} \lambda_{jt}D_{jt}, \tag{13}$$

where

$$\begin{aligned}g_i(\lambda) = \min &\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l\geq k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t C_{kr}\left(y_{iklrt} - y_{iklr(t-1)}\right) + \\ &\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}:l>k}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} \delta_t E_{klr}(y_{iklrt} - y_{iklr(t-1)}) + \\ &\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{b\in\mathcal{B}_{lr}}\sum_{t\in\mathcal{T}} \delta_t F_{blr}\mu_{bilrt} + \\ &\sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{K}}\sum_{r\in\mathcal{R}}\sum_{t\in\mathcal{T}} (\delta_t T_{ijt} - \lambda_{jt})x_{ijlrt},\end{aligned} \tag{14}$$

subject to (2)–(6) and (8)–(11) for a given facility location i.

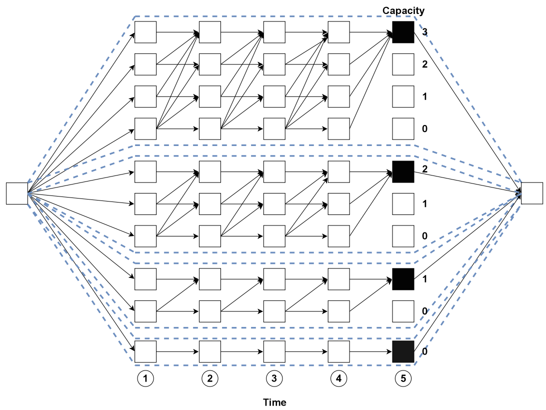


Fig. 4.1. Structure of our shortest path problems.

4.2. Solving the Lagrangian subproblem

The Lagrangian subproblems can be reformulated as a shortest path problem and solved as a dynamic programming problem. Note that for a given capacity, the problem in a single time period becomes continuous knapsack (Shulman, 1991). We first present the structure of our shortest path problem in Section 4.2.1 before discussing the costs of the continuous knapsack and the dynamic programming approach in Sections 4.2.2 and 4.2.3, respectively.

4.2.1. Shortest path problem

To solve the Lagrangian subproblem, we need to find both the optimal opening and expansion schedule and technology choice for each facility such that the objective (14) is minimized. A schedule is defined by the time of opening and expanding a facility as well as the associated capacities. In contrast to Shulman (1991) and Jena et al. (2016, 2017), we allow only one expansion during the whole planning horizon. Closing facilities, reducing capacity or changing technology, is not allowed, thus resulting in a different shortest path problem.

Our graph structure for a problem with three capacities and five time periods is illustrated in Fig. 4.1. Let target capacity $l^T \in \mathcal{K}$ be the capacity at the end of the planning horizon. Since only one capacity expansion is allowed, we can construct one subgraph for each possible target capacity that contains all allowed paths for reaching the target capacity at the end of the planning horizon. We also consider the alternative of not opening a facility at all. All subgraphs are depicted with a blue dashed line while the corresponding target capacity of each subgraph is indicated by the black squares in Fig. 4.1. There are two ways how to reach the target capacity: by opening a facility with target capacity right away, or by opening a smaller facility which has to be expanded towards the target capacity later during the planning horizon. Thus, for each path through the subgraph, at most two capacity changes are allowed: one for opening and the second one for expansion. To formulate the problem as a shortest path search in a single graph, we can connect all subgraphs with an artificial source and sink. The shortest path structure is the deterministic version of the expected shortest path presented in Štádlerová et al. (2023).

We obtain the optimal schedule for each technology separately and then select the schedule with the lowest costs. Thus, for the remainder of this Section, we omit the technology choice and refer to the investment and expansion costs without the technology subscript r.

4.2.2. Continuous knapsack problem

For the computation of the shortest path in the dynamic programming algorithm, we have to obtain the production and distribution costs of the optimal demand allocation in each period and for each capacity in the subgraph. For given installed capacity k, this problem corresponds to solving a continuous knapsack problem with piecewise linear costs (Amiri, 1997; Christensen and Klose, 2021). The costs of demand allocation for given facility i, capacity level k, technology r and time period t are denoted $K_{ikrt}(\lambda)$ and obtained from solving the continuous knapsack problem:

$$K_{ikrt}(\lambda) = \min \sum_{j\in\mathcal{J}} (T_{ij} - \lambda_{jt}) x_{ijkrt} + \sum_{b\in\mathcal{B}_{kr}} F_{bkr}\mu_{bikrt}, \tag{15}$$

subject to:

$$x_{ijkrt} \le L_{ij} D_{jt}, \qquad j \in \mathcal{J}, \tag{16}$$

$$\sum_{j\in\mathcal{J}} x_{ijkrt} = \sum_{b\in\mathcal{B}_{kr}} Q_{bk}\mu_{bikrt}, \tag{17}$$

$$\sum_{b\in\mathcal{B}_{kr}} \mu_{bikrt} = 1, \tag{18}$$

$$x_{ijkrt} \ge 0, \qquad j \in \mathcal{J}, \tag{19}$$

$$\mu_{bikrt} \ge 0, \qquad b \in \mathcal{B}_{kr}, \tag{20}$$

The approach for computing solution to the problem (15)–(20) is adopted from Schütz et al. (2008): For a given time period t, we first sort the customers according to increasing reduced costs $T_{ij} - \lambda_{jt}$. Note that the ordering of customers is independent of the capacity utilization level. For given facility i and technology r, the marginal costs of serving one additional demand unit is calculated as $q_{jkbt} = T_{ij} - \lambda_{jt} + u_{kb}$, where $u_{kb} = \frac{F_{kb+1} - F_{kb}}{Q_{kb+1} - Q_{kb}}$ represents the marginal production costs of the linepiece.

In general, we start allocating customers with the lowest negative reduced costs to facility i and continue adding customers until $q_{ikbt} > 0$ for the first time or until the limit of the installed capacity is reached. However, the marginal production costs depend on the linepiece b of the short-term cost function, i.e. capacity utilization. If adding a new customer and increasing capacity utilization causes a change of linepiece, we update the marginal cost q_{ikbt} and continue allocating customers as long as $q_{ikbt} \le 0$ and capacity is available. Due to the requirements on minimum production quantities, we have to fill the knapsack up to the minimum level even if that requires allocating customers with positive reduced costs.

The costs of the continuous knapsack consist only of production and reduced transportation cost for the allocated customer demand.

4.2.3. Dynamic programming approach

To find the optimal opening and expansion schedule for a given facility, we calculate the shortest path for each subgraph individually and then pick the one with the lowest costs. To limit the number of subgraphs and speed up solving the dynamic programming problem, we estimate a maximum target capacity l^{max}: We aim to satisfy only demand with negative marginal costs q_{jkbt}. But the exact value of marginal costs cannot be computed without the knowledge of installed capacity and utilization. We therefore neglect the production costs to find an upper bound on the target capacity and define the set of potential customers as customers with negative reduced costs: $\mathcal{J}_t^N := \{j \in \mathcal{J} | T_{ij} - \lambda_{jt} < 0\}$. We can then calculate the maximum production quantity q^{max} as:

$$q^{max} = \max_{t\in T} \sum_{j\in\mathcal{J}_t^N} D_{jt}.$$

Finally, we obtain the maximum target capacity l^{max} as the lowest capacity that is larger than the production quantity q^{max}.

The decision in period t is denoted $\hat{y}_t$ and determines the installed capacity in period t, k_t, where $\hat{y}_t \in \mathcal{K} \cup \{0\}$. The costs of allocating customers corresponding to the costs of the continuous knapsack for given facility, capacity, technology and time period depend on this decision and are denoted $K_{i\hat{y}_t r}(\lambda)$. The cost function when taking decision $\hat{y}_t \in \mathcal{K} \cup \{0\}$ in period t is denoted $C_t^d(\hat{y}_t, t)$. The function C_t^d is defined as:

$$C_t^d(\hat{y}_t, t) = \begin{cases} C_{\hat{y}_t r} + K_{i\hat{y}_t r}(\lambda) & \text{if} \quad k_{t-1} = 0 \wedge \hat{y}_t > 0, \quad \text{(a)} \\ E_{k_{t-1}\hat{y}_t r} + K_{i\hat{y}_t r}(\lambda) & \text{if} \quad k_{t-1} > 0 \wedge \hat{y}_t = l_T, \quad \text{(b)} \\ K_{i\hat{y}_t r}(\lambda) & \text{if} \quad \hat{y}_t = k_{t-1}, \quad \text{(c)} \\ 0 & \text{if} \quad \hat{y}_t = 0, \quad \text{(d)} \\ +\infty & \text{else} \quad \text{(e)}. \end{cases} \tag{21}$$

Eq. (21)(a) calculates the costs for the case of opening a new facility in period t. The total costs in period t then consist of opening costs $C_{i\hat{y}_t r}$ and the production and distribution costs given by the costs of the continuous knapsack $K_{i\hat{y}_t r}(\lambda)$. Eq. (21)(b) calculates the costs of expanding a facility as the sum of expansion costs $E_{k_{t-1}\hat{y}_t r}$ and the costs of continuous knapsack $K_{i\hat{y}_t r}(\lambda)$. Note that the facility has to be expanded to target capacity l_T, as only one expansion is allowed during the planning horizon. Constraint (21)(c) ensures that if the facility is open and there is no change in capacity, only the costs of the continuous knapsack $K_{i\hat{y}_t r}(\lambda)$ apply. As long the facility is not opened, the costs are equal to zero (21)(d). All other combinations of state and decision are not feasible and we, therefore, define the costs of these as $+\infty$ (21)(e).

4.3. Lagrangian multipliers

For given λ, we obtain a lower bound on the objective function by solving $LR(\lambda)$ (13). In order to obtain the best lower bound on the optimal objective value, we need to find the optimal value of λ. Initial tests showed that the standard subgradient method (Nemhauser and Wolsey, 1999) has a slow convergence (see also Jena et al., 2017; Schütz et al., 2009). Therefore, we solve the Lagrangian dual problem $LD = \max_\lambda LR(\lambda)$ by means of a cutting planes method with box constraints similar to Marsten et al. (1975) and Schütz et al. (2009). The tests have further shown that initializing the Lagrangian multipliers as $\lambda_{jt} = \min_{i\in\mathcal{I}} \delta_t T_{ijt}$ generally leads to faster convergence.

In each iteration m, we compute the coordinates of the subgradient ∇^m_{jt}, where $\nabla^m_{jt} = D_{jt} - \sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{K}} \sum_{r\in\mathcal{R}} x^m_{ijlrt}$ and x^m_{ijlrt} is the Lagrangian subproblem solution obtained for variables x in iteration m. We then define $L^m = LR(\lambda^m) - \sum_{j\in\mathcal{J}} \sum_{t\in\mathcal{T}} \lambda^m_{jt} \nabla^m_{jt}$ and obtain new Lagrangian multipliers by solving the following linear problem:

$$\max \phi \tag{22}$$

subject to

$$\phi \le L^g + \sum_{j\in\mathcal{J}} \sum_{t\in\mathcal{T}} \nabla^g_{jt} \lambda^{m+1}_{jt}, \qquad g = 1, \dots, m, \tag{23}$$

$$\lambda^{m+1}_{jt} \le \lambda^m_{jt} + \Delta^m_{jt}, \qquad j \in \mathcal{J}, t \in \mathcal{T}, \tag{24}$$

$$\lambda^{m+1}_{jt} \ge \lambda^m_{jt} - \Delta^m_{jt}, \qquad j \in \mathcal{J}, t \in \mathcal{T}, \tag{25}$$

$$\phi \in \mathbb{R}, \lambda^{m+1}_{jt} \in \mathbb{R} \tag{26}$$

Constraints (24) and (25) are the box constraints that limit how much the Lagrangian multipliers can change in each iteration. Changing the box size can considerably improve the convergence (Marsten et al., 1975). We adjust the size of the box Δ^m in each iteration. If the sign of the subgradient element changes compared to the previous iteration, we update the box size: $\Delta^m_{jt} = \alpha \Delta^{m-1}_{jt}$, where $0 < \alpha \le 1$. With this step, we reduce the size of the box in order to speed up the process of finding the optimal multipliers.

4.4. Lagrangian heuristic

In general, the solution of the relaxed problem is not feasible for the original problem (Nemhauser and Wolsey, 1999). Therefore, we develop a greedy heuristic for finding a feasible solution based on the solution of the relaxed problem to calculate an upper bound on the optimal solution. Fig. 4.2 illustrates the main structure of our heuristic.

We start by allocating customers to the facilities opened in the solution of the Lagrangian subproblem, see Algorithm 1 for details. Customers are first sorted according to increasing reduced costs. We then allocate the customers to the cheapest facility until we reach the utilization of the facility taken from the relaxed Lagrangian subproblem or until no more customers with unsatisfied demand can be allocated.

Algorithm 1: Assign_to_LB_capacities

Initialize from LB for all $t \in \mathcal{T}$: $y^{LB}_{iklrt}; x^{LB}_{ijlrt}; \mu^{LB}_{bilrt}$
$y_{iklrt} = y^{LB}_{iklrt}, i \in \mathcal{I}, k \in \mathcal{K}, l \in \mathcal{K} : l \ge k, r \in \mathcal{R}, t \in \mathcal{T}$
$x_{ijlrt} = 0, i \in \mathcal{I}, j \in \mathcal{J}, l \in \mathcal{K}, r \in \mathcal{R}, t \in \mathcal{T}$
Initialize available capacity as:
  $\phi_{it} = \sum_{b\in\mathcal{B}} \sum_{l\in\mathcal{K}} \sum_{r\in\mathcal{R}} Q_{blt} \mu^{LB}_{bilrt}, i \in \mathcal{I}, t \in \mathcal{T}$
Initialize used capacity: $\kappa_{it} = 0, i \in \mathcal{I}, t \in \mathcal{T}$
Initialize set of available facilities:
  $\mathcal{I}^A_t = \{i | \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{K}: l\ge k} \sum_{r\in\mathcal{R}} y_{iklrt} = 1\}, t \in \mathcal{T}$
Initialize set of partially satisfied customers in the LB:
  $\mathcal{J}^u_t = \{j | \sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{K}} \sum_{r\in\mathcal{R}} x^{LB}_{ijlrt} > 0\}, t \in \mathcal{T}$
Initialize unsatisfied demand: $d_{jt} = D_{jt}, j \in \mathcal{J}, t \in \mathcal{T}$
foreach $t \in \mathcal{T}$ **do**
  while $\mathcal{J}^u_t \ne \emptyset \wedge \mathcal{I}^A_t \ne \emptyset$ **do**
    foreach $i \in \mathcal{I}^A_t$ **do**
      foreach $j \in \mathcal{J}^u_t$ **do**
        if $L_{ijt} = 1 \wedge \kappa_{it} < \phi_{it}$ **then**
          $\Gamma_{ij} = T_{ij} - \lambda_{jt}$
    Sort all Γ_{ij} in non-decreasing order
    foreach Γ_{ij} **do**
      $l = \arg\max_{l\in\mathcal{K}} \sum_{k\in\mathcal{K}: k\le l} \sum_{r\in\mathcal{R}} y_{iklrt}$
      $r = \arg\max_{r\in\mathcal{R}} \sum_{k\in\mathcal{K}} y_{iklrt}$
      if $j \in \mathcal{J}^u_t \wedge \kappa_{it} < \phi_{it}$ **then**
        $x_{ijlrt} = x_{ijlrt} + \min\{d_{jt}, \phi_{it} - \kappa_{it}\}$
        update: $d_{jt} = d_{jt} - \min\{d_{jt}, \phi_{it} - \kappa_{it}\}$
        calculate used capacity as: $\kappa_{it} = \sum_{j\in\mathcal{J}} x_{ijlrt}$
        if $d_{jt} = 0$ **then**
          $\mathcal{J}^u_t = \mathcal{J}^u_t \setminus \{j\}$
        if $\kappa_{it} = \phi_{it}$ **then**
          $\mathcal{I}^A_t = \mathcal{I}^A_t \setminus \{i\}$
  Update set of all unsatisfied customers:
    $\mathcal{J}^u_t = \{j | \sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{K}} \sum_{r\in\mathcal{R}} x_{ijlrt} < D_{jt}\}$

After the initial allocation of customers to available facilities, we check for each time period if demand is satisfied, starting from the first period. If not all customers are satisfied for the given period, we iteratively expand the capacity of the production system in the general step `Expand capacity`. The process of expanding capacity consists of multiple steps: `Increase_utilization`, `Open_new`, and `Expand_existing`. After each step, we evaluate whether all customers are satisfied. If so, we move to the next time period. Otherwise, we continue with the next step of expanding available capacity.

For all three steps of the capacity extension process, we sort the facility locations according to how many customers with unsatisfied demand they can serve and start with the facility location that can serve most customers. In case of a tie, we choose the location with the lowest index. The first step, `Increase_utilization`, aims at increasing the utilization of the existing production facilities. In the first time period, based on the principles used for the initial assignment of customers to facilities, we increase capacity utilization beyond the level given by the solution to the Lagrangian subproblem until there are no longer unsatisfied customers or until the capacity limit of the facility is reached. In all other time periods, we first update the set of available facilities at the beginning of the time period and then allocate unsatisfied customers. Since opening of a new facility or expansion of

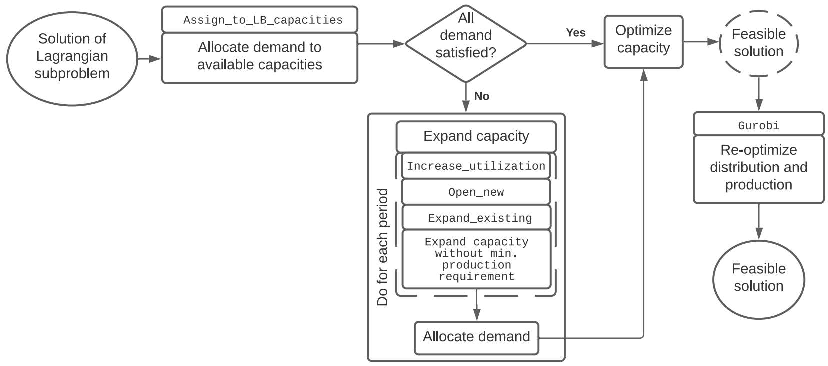


Fig. 4.2. General structure upper bound.

an existing facility in time period t affects the available capacity in all following time periods as well, this step is necessary to consider new capacities that were not available in the solution to the Lagrangian subproblem. During this step, the installed capacity is not changed. See Algorithm 2 for details.

Algorithm 2: Increase_utilization

if $t = 1$ **then**
  Initialize from `Assign_to_LB_capacities`: $\kappa_{it}; \phi_{it}; d_{jt}; y_{iklrt}; x_{ijlrt}; \mathcal{J}_t^u; \mathcal{I}_t^A; \Gamma_{it}$
else
  Initialize for each t from `Expand capacity` in $t-1$: $\kappa_{it}; \phi_{it}; d_{jt}; y_{iklrt}; x_{ijlrt}; \mathcal{J}_t^u; \Gamma_{it}$
  update $\mathcal{I}_t^A = \{i | \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{K}:l\geq k} \sum_{r\in\mathcal{R}} y_{iklrt} = 1\}$
while $\mathcal{J}_t^u \neq \emptyset \wedge \mathcal{I}_t^A \neq \emptyset$ **do**
  $i = \min\{\arg\max_{i\in\mathcal{I}_t^A} \sum_{j\in\mathcal{J}_t^u} L_{ijt}\}$
  $l = \arg\max_{l\in\mathcal{K}} \sum_{k\in\mathcal{K}:k\leq l} \sum_{r\in\mathcal{R}} y_{iklrt}$
  $r = \arg\max_{r\in\mathcal{R}} \sum_{k\in\mathcal{K}} y_{iklrt}$
  Set new available capacity ϕ_{it}: $\phi_{it} = \max_{b\in\mathcal{B}_{lr}} Q_{lbr}$
  Sort $j \in \mathcal{J}_t^u$ according to non-decreasing Γ_{ij}
  foreach $j \in \mathcal{J}_t^u$ **do**
    $x_{ijlrt} = x_{ijlrt} + \min\{d_{jt}, \phi_{it} - \kappa_{it}\}$
    Update: $d_{jt} = d_{jt} - \min\{d_{jt}, \phi_{it} - \kappa_{it}\}$
    Calculate used capacity as: $\kappa_{it} = \sum_{j\in\mathcal{J}} x_{ijlrt}$
    if $d_{jt} = 0$ **then**
      $\mathcal{J}_t^u = \mathcal{J}_t^u \setminus \{j\}$
    if $\kappa_{it} = \phi_i$ **then**
      $\mathcal{I}_t^A = \mathcal{I}_t^A \setminus \{i\}$

If there are still unsatisfied customers, we first open new facilities before expanding the capacity of existing facilities. The candidate locations for new facilities in time period t are all locations where no facility has been opened. We select the location that can serve most unsatisfied customers and check whether a facility is opened at this location in a future period. If there is such a facility, we compare the capacity needed for the current demand and the capacity suggested from the relaxed problem and pick up the higher one. If allocated demand is not sufficient to cover the minimum production requirements, the facility location is for this period removed from the set of available locations and we continue with the next candidate location. If there is no facility at this location, we allocate all possible demand to this location and install the lowest capacity sufficient to meet this demand and choose the potentially cheapest technology. The subroutine `Open_new` is described in Algorithm 3.

The last step in expanding capacity is to select facilities that should be expanded and determine their new capacity. The set of candidate locations is made up of facilities that have not been expanded before the current period. Again, we select the facility that can serve most customers with unsatisfied demand. Similar to the previous step, we check whether the facility is to be expanded at a later point in time.

Algorithm 3: Open_new

Initialize from `Increase_utilization`: t; d_{jt}; y_{iklrt}; $\mathcal{J}_t^u$
$Q^{min} = \min_{r\in\mathcal{R},b\in\mathcal{B}_{1r}} Q_{b1r}$
$Q^{max} = \max_{r\in\mathcal{R},l\in\mathcal{K},b\in\mathcal{B}_{lr}} Q_{blr}$
Define set of candidate facilities: $\mathcal{I}_t^{new} = \{i | \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{K}:l\geq k} \sum_{r\in\mathcal{R}} y_{iklrt} = 0\}$
while $\mathcal{J}_t^u \neq \emptyset \wedge \mathcal{I}_t^{new} \neq \emptyset$ **do**
  $i = \min\{\arg\max_{i\in\mathcal{I}_t^{new}} \sum_{j\in\mathcal{J}_t^u} L_{ijt}\}$
  Calculate necessary capacity as: $\omega_{it} = \sum_{j\in\mathcal{J}_t^u:L_{ijt}=1} d_{jt}$
  if $\omega_{it} < Q^{min}$ **then**
    $\mathcal{I}_t^{new} = \mathcal{I}_t^{new} \setminus \{i\}$
  else
    if $\sum_{k\in\mathcal{K}} \sum_{r\in\mathcal{R}} \sum_{\tau\in\mathcal{T}} y_{ikkr\tau} \geq 1$ **then**
      $k = \arg\max_{k\in\mathcal{K}} \sum_{r\in\mathcal{R}} \sum_{\tau\in\mathcal{T}} y_{ikkr\tau}$
      $r = \arg\max_{r\in\mathcal{R}} \sum_{\tau\in\mathcal{T}} y_{ikkr\tau}$
      if $\omega_{it} < Q_{1kr}$ **then**
        $\mathcal{I}_t^{new} = \mathcal{I}_t^{new} \setminus \{i\}$
      else
        if $\omega_{it} \leq Q^{max}$ **then**
          $k^m = \min\{\bar{k} \in \mathcal{K} | \max_{b\in\mathcal{B}} Q_{b\bar{k}r} \geq \omega_{it}\}$
        else
          $k^m = \max\{\bar{k} \in \mathcal{K}\}$
        $k' = \max\{k^m, k\}$
        Open new facility with capacity k' as: $y_{ik'k'rt'} = 1, t' \in \mathcal{T} : t' \geq t$ $\phi_{it'} = \max_{b\in\mathcal{B}} Q_{bk'r}, t' \in \mathcal{T} : t' \geq t$
    else
      foreach $r \in \mathcal{R}$ **do**
        $k' = \min\{k | \max_{b\in\mathcal{B}} Q_{bkr} \geq \omega_{it}\}$
        Calculate estimated costs: $E_r = C_{k'r} y_{ik'k'rt} + (\overline{T} - t + 1) \max_{b\in\mathcal{B}} F_{bk'r}$
      Choose technology r with lowest costs as: $r = \arg\min_{r\in\mathcal{R}} E_r$
      Open new facility with capacity k' as: $y_{iklrt'} = 0, k \in \mathcal{K}, l \in \mathcal{K} : l \leq k, r \in \mathcal{R}, t' \in \mathcal{T} : t' \geq t$ $y_{ik'k'rt'} = 1,\ t' \in \mathcal{T} : t' \geq t$ $\phi_{it'} = \max_{b\in\mathcal{B}} Q_{bk'r}, t' \in \mathcal{T} : t' \geq t$
    Run Algorithm 1 from line 10 with initialization: t; $\mathcal{J}_t^u$; $\mathcal{I}_t^A = \{i\}$; $\kappa_{it} = 0$; ϕ_{it}; $d_{jt}, j \in \mathcal{J}_t^u$
    $\mathcal{I}_t^{new} = \mathcal{I}_t^{new} \setminus \{i\}$

In that case we check if we can satisfy minimum production quantities of the new capacity with current demand. If these requirements are satisfied, we expand the facility early. Otherwise, we discard the facility for this period and move on to the next. If the facility is not expanded in the solution to the relaxed problem, we calculate the capacity that is required to serve all possible customers and expand the facility. Note that we can only expand once and therefore need to prevent early expansion with low capacity. Expansion of existing facilities is presented in Algorithm 4.

Algorithm 4: Expand_existing

Initialize from `Open_new`: $t; d_{jt}; y_{iklrt}; \mathcal{J}_t^u; Q^{max}$
Define set of candidate facilities:
  $\mathcal{I}_t^{exp} = \{i | \sum_{k\in\mathcal{K}} \sum_{r\in\mathcal{R}} y_{ikkrt} = 1 \wedge \sum_{k\in\mathcal{K}} \sum_{r\in\mathcal{R}} y_{ikkr(t-1)} = 1\}$
while $\mathcal{J}_t^u \neq \emptyset \wedge \mathcal{I}_t^{exp} \neq \emptyset$ **do**
  $i = \min\{\arg\max_{i\in\mathcal{I}_t^{exp}} \sum_{j\in\mathcal{J}_t^u} L_{ijt}\}$
  $k = \arg\max_{k\in\mathcal{K}} \sum_{r\in\mathcal{R}} y_{ikkrt}$
  $r = \arg\max_{r\in\mathcal{R}} y_{ikkrt}$
  if $\sum_{l\in\mathcal{K}:l>k} \sum_{\tau\in\mathcal{T}} y_{iklr\tau} \geq 1$ **then**
    $l = \arg\max_{l\in\mathcal{K}:l>k} \sum_{\tau\in\mathcal{T}} y_{iklr\tau}$
  else
    $l = k$
  Calculate necessary capacity as: $\omega_{it} = \sum_{j\in\mathcal{J}_t^u : L_{ijt}=1} d_{jt}$
  if $\omega_{it} < Q_{1lr}$ **then**
    $\mathcal{I}_t^{exp} = \mathcal{I}_t^{exp} \setminus \{i\}$
  else
    if $\omega_{it} \leq Q^{max}$ **then**
      $l^m = \min\{\bar{l} \in \mathcal{K} | \max_{b\in\mathcal{B}} Q_{b\bar{l}r} \geq \omega_{it}\}$
    else
      $l^m = \max\{\bar{l} \in \mathcal{K}\}$
    $l' = \max\{l^m, l\}$
    Expand facility to capacity l' as:
      $y_{iklrt'} = 0, k \in \mathcal{K}, l \in \mathcal{K} : l \geq k, r \in \mathcal{R}, t' \in \mathcal{T} : t' \geq t$
      $y_{ikl'rt'} = 1, t' \in \mathcal{T} : t' \geq t$
      $\phi_{it'} = \max_{b\in\mathcal{B}} Q_{bl'r}, t' \in \mathcal{T} : t' \geq t$
    Run Algorithm 1 from line 10 with initialization: t; $\mathcal{J}_t^u$;
      $\mathcal{I}_t^A = \{i\}$; $\kappa_{it} = 0$; ϕ_{it}; $d_{jt}, j \in \mathcal{J}_t^u$
    $\mathcal{I}_t^{exp} = \mathcal{I}_t^{exp} \setminus \{i\}$

If the previous steps are not sufficient to satisfy all customers, we further expand the capacity, neglecting the requirement on minimum production quantities. We start with the facility location that can serve most unsatisfied customers, open a new facility, and install the lowest capacity sufficient to satisfy demand. We then optimize a transportation and capacity utilization problem for currently open facilities using Gurobi to reallocate customers to satisfy minimum production requirements in the entire production system and update the set of unsatisfied customers. If even more capacity is needed, we further increase the expansion capacity of already expanded facilities. We start again with the location that can serve most unsatisfied customers. This step should not violate the feasibility of our solution because we proceed with this step at the moment when all facilities that could potentially satisfy the remaining unsatisfied customers are on their capacity limits. Therefore, there exist customers that can be reallocated to the expanded facility so that the minimum production requirements are satisfied in the whole production system.

After this step, we obtain a feasible solution. However, once all customers are satisfied, we check if we can reduce capacity at some of the facilities, as we may have installed more capacity than needed during the capacity expansion process. We reduce the surplus capacity of opened facilities by computing the costs for all combinations of opening capacities and time as well as expansion capacities and time that satisfy minimum production quantity requirements in each period. We then install the capacity that leads to the lowest costs. This solution is further improved in a final step by using Gurobi. We fix all binary variables based on the results of our heuristic. Then, we solve the original problem with only continuous variables related to demand allocation and capacity utilization. However, to further improve the run times, this final step is performed only when the solution improves or is no worse than 1.1 times the best solution so far. The parameter 1.1 has been chosen based on initial tests.

4.5. Restricted MIP

To improve the quality of the upper bound, Jena et al. (2017) construct a restricted mixed-integer model for their problem by exploiting information from their bundle method used to update the Lagrangian multipliers. We propose a similar approach based on the information obtained from solving the Lagrangian dual: Let α_g, $g = 1, \ldots, m$ be the dual variable belonging to constraint (23). In the optimal solution to problem (22)–(26), we have $\sum_{g=1}^{m} \alpha_g = 1$ and hence α_g can be understood as the probability that the solution to the relaxed problem from iteration g provides good decisions. We then calculate:

$$\gamma_{iklrt}^{p} = \sum_{g=1}^{m} \alpha_g y_{iklrt}^{g}, \qquad i \in \mathcal{I}, k \in \mathcal{K}, l \in \mathcal{K} : l \geq k, r \in \mathcal{R}, t \in \mathcal{T}, \tag{27}$$

where $y_{iklrt}^{g} = 1$, if the solution to the relaxed problem chooses to operate a facility at location i, with opening capacity k, operated capacity l and installed technology r in time period t in iteration g.

If $\gamma_{iklrt}^{p} \geq 0.85$, we fix the binary decision and set y_{iklrt} equal to one. We then solve the original problem with the fixed variables using Gurobi. Selecting the correct threshold for fixing the binary variables is important for the performance of this approach. If the threshold value is chosen too low, too many variables are fixed and no improvement can be achieved. In the other case, only a few variables are fixed and the problem is close to its original size, resulting in long solution times or even a computationally intractable problem. A threshold value of 0.85 worked best during the initial testing and also corresponds to the value suggested in Jena et al. (2017).

5. Case description

In this section, we present the input data used in our computational experiments. The input data is based on real-world data from Norway. We present the potential facility locations in Section 5.1 and the customer locations with associated demand in Section 5.2. The available technologies and their cost structure as well as transportation costs are discussed in Section 5.3.

5.1. Facility location

In the base case, we consider 17 candidate locations for hydrogen facilities on the Norwegian west coast. The candidate locations for hydrogen production are obtained from the interactive map set up by Ocean Hyway Cluster (2020b). These candidate locations are always present in all larger instances. We further present test instances with 34, 51, 102 facility locations. All candidate locations are in Norwegian ports thus if the number of potential facility locations increases, the distances between candidate locations decrease.

5.2. Demand

We use two demand scenarios: one scenario including demand from the maritime sector and one scenario with demand from the whole transportation sector. We have a planning horizon of 14 years and during this time the demand is non-decreasing. Demand development over time can be seen in Fig. 5.1. Fig. 5.1 shows that in the first three periods, the demand is equal in both scenarios as the demand comes only from the maritime sector. The main growth in the whole transportation sector scenario comes in two phases in periods 4 and 9 when hydrogen transition in heavy transport and long-distance bus transport is planned (DNV GL, 2019).

The maritime scenario considers 51 demand points in Norwegian ports and consists of the hydrogen demand for high-speed passenger ferries, car ferries and the coastal route Bergen–Kirkenes (Aarskog and Danebergs, 2020; Ocean Hyway Cluster, 2020a). The demand for the whole transportation sector consists of the maritime sector with its 51 demand points plus road traffic and the railway sector (DNV

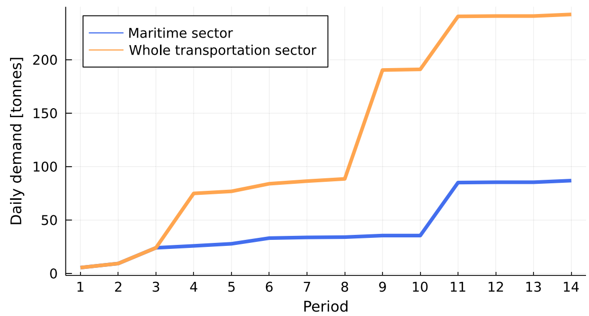


Fig. 5.1. Demand development during the planning horizon.

Table 5.1
Hydrogen distribution costs in [€/km/kg H_2].

Distance [km]	1–50	51–100	101–200	201–400	401–800	801–1000
Costs	0.00498	0.00426	0.00390	0.00372	0.00363	0.00360

GL, 2019). We divide the road traffic demand between 70 and 392 customers, respectively, proportionally according to the statistic about road traffic volumes in Norway (Statistics Norway, 2021). The scenario with 70 customers consists of the 51 ports plus 19 municipalities with the highest road traffic volumes and then the road traffic demand is divided between these 70 points. Considering 392 customers, we divide the road traffic demand between all Norwegian municipalities. We neglect municipalities with calculated demand lower than 10 kg daily and recalculate the demand allocation for the remaining customers.

5.3. Costs

The distribution costs per kilometer and kilogram of hydrogen are taken from Danebergs and Aarskog (2020). The distribution costs are valid for the appropriate interval as shown in Table 5.1. The distance limit between production facility and customer is 1000 km.

We consider two hydrogen production technologies: electrolysis (EL) and steam methane reforming with carbon capture and storage (SMR+). We approximate the facility capacity using 8, 16, or 24 discrete points for EL and 7, 15, or 23 points for SMR+. The difference between capacity levels is not constant in order to achieve overlapping production quantities and more choices for low capacities. For all capacities, we consider 4 short-term utilization levels. The short-term breakpoints are at 20%, 50%, 80% and 100% of installed production quantity, whereas the 20% utilization corresponds to the minimum production quantity for each capacity level. We use the model from Jakobsen and Åtland (2016) to calculate the cost. Investment and production costs at the highest utilization level for both technologies are shown in Fig. 5.2(a). We also use a S-shaped piecewise linear cost function for the production costs, modelling economies as well as diseconomies of scale, see Fig. 5.2(b). The main data for the case with 8 capacities is taken from Štádlerová and Schütz (2021). For simplicity, the discount factor δ_t is set to one for all periods.

We model the expansion as a more expensive alternative than opening a bigger facility right away. We obtain the expansion costs as the difference between the investment costs of the new capacity and the originally installed capacity plus an additional 10% markup.

6. Results

All calculations have been carried out on a computer with two 3.6 GHz Intel Xeon Gold 6244 CPU (8 core) processors and 384 GB RAM. Gurobi Optimizer version 9.1.2 is used as a commercial solver for comparison and for computing the optimal distribution and production quantities for a given set of facilities in the last step of our heuristic. Our algorithm is implemented in Julia 1.6.3. We enable parallelization at 32 threads in Julia as Gurobi optimizer also utilizes up to 32 threads on this computer.

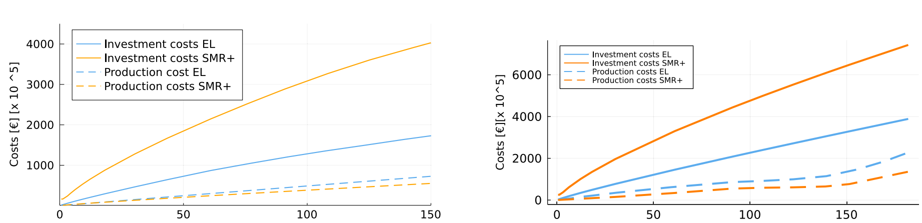
We denote our instances with the number of candidate locations for facilities (F), demand points (D) and capacity levels (C). For example, F17D51C8 represents an instance with 17 candidate locations, 51 demand points and 8 capacity levels. Capacity levels marked with asterisk denote instances with S-shaped costs.

6.1. Solution quality

To evaluate the performance and quality of our approaches, we compare the results of the Lagrangian relaxation (LR) and the restricted MIP (R-MIP) to the ones obtained from Gurobi (GUR). As quality criteria, we study optimality gap, run time, and the best upper bound. We stop the Lagrangian relaxation after 1000 iterations whereas R-MIP and Gurobi have run time limit of 24 h. In each LR iteration, we solve the Lagrangian subproblem, the Lagrangian dual and calculate the upper bound. In order to further improve the run time, we evaluated the upper bound in fixed intervals and only when the lower bound increases. However, we did not find equally good solutions with these approaches. When using R-MIP, we first perform 1000 iterations where we solve the Lagrangian subproblem and update the Lagrangian multipliers (but do not calculate an upper bound). Afterwards, we start solving the restricted MIP.

The main results are summarized in Table 6.1 for the instances with convex production costs and in Table 6.2 for the instances with S-shaped production costs. We provide the run time (in seconds) needed to achieve an optimality gap lower than 3% and 1.5% for LR and GUR. For the R-MIP, we only consider the best solution. Please note that the optimality gap reported for R-MIP in Tables 6.1 and 6.2 is the optimality gap for the original problem, using the lower bound from the Lagrangian relaxation to calculate the gap. In case the table field is empty, a solution with the specified gap is not found within the stopping criterion. We also show the total run time, best upper bound, and final gap for our algorithm, R-MIP and Gurobi. If R-MIP is solved to optimality within 24 h or if GUR finds the optimal solution within 24 h, we show the time needed to find this solution.

Lagrangian relaxation outperforms Gurobi in terms of run time for all problem instances where we can prove optimality gaps less than 3%. In particular, our algorithm finds good (but not necessarily



(a) Investment and production costs from Štádlerová & Schütz (2021)

(b) Modified investment and production costs with S-shaped form

Fig. 5.2. Investment and yearly production cost at the highest utilization level.

Table 6.1
Run times and optimality gap for our algorithm (LR), Gurobi (GUR) and restricted MIP (R-MIP) for the instances with convex production costs.

Instance	Time in seconds				Best upper bound			Total run time			End gap		
	Gap < 3%		Gap < 1.5%		× 10^6			[s]			[%]		
	LR	GUR	LR	GUR	LR	GUR	R-MIP	LR	GUR	R-MIP	LR	GUR	R-MIP
F17D51C8	31	121	43	121	578.77	577.94	577.95	273	**256**	277	0.90	**0.01**	0.65
F17D51C16	36	279	64	279	564.05	564.05	566.36	300	**9 795**	331	1.33	**0.01**	1.80
F17D51C24	48	1 587	485	2 428	562.73	558.42	582.20	710	**10 361**	362	1.49	**0.01**	4.77
F17D70C8	31	866	44	916	1590.75	1587.66	1587.66	344	**1 150**	346	0.60	**0.01**	0.35
F17D70C16	21	67	98	413	1540.69	1537.70	1537.90	405	**8 030**	435	0.93	**0.01**	0.90
F17D70C24	54	28 438	180	28 438	1539.08	1529.20	1559.72	789	**51 836**	500	1.15	**0.01**	2.40
F17D392C8	45	1 871	258	3 601	1356.69	1352.55	1396.68	1631	**6 140**	1 651	1.32	**0.01**	4.00
F17D392C16	60	16 545	985	16 545	1340.20	1334.80	1345.90	1910	86 400	1 690	1.46	0.28	1.95
F17D392C24	64	–	–	–	1343.70	–	1331.30	2192	86 400	86 400	1.87	–	0.91
F34D51C8	32	564	39	564	554.42	553.37	555.17	372	**767**	290	0.94	**0.01**	1.09
F34D51C16	32	292	162	292	538.09	533.68	537.62	834	**4 478**	2 218	1.46	**0.01**	1.65
F34D51C24	124	1 511	503	1 513	533.40	530.69	532.16	1439	86 400	945	1.24	0.06	1.05
F34D70C8	51	1 246	–	1 534	1501.09	1478.00	1523.71	430	**15 973**	701	2.45	**0.01**	3.80
F34D70C16	54	916	426	916	1466.35	1448.10	1456.54	777	86 400	1 895	0.98	0.10	1.23
F34D70C24	145	83 487	262	83 487	1455.48	1559.90	1445.90	1214	86 400	86 400	1.26	7.69	0.60
F34D392C8	106	34 441	1703	36 894	1315.37	1303.30	1306.07	2009	86 400	4 263	1.28	0.1	0.80
F34D392C16	265	–	–	–	1357.61	–	1398.17	2470	86 400	2 711	2.77	–	9.20
F34D392C24	349	–	–	–	1299.73	–	–	3003	86 400	86 400	2.44	–	–
F51D51C8	61	997	61	1 070	552.67	550.00	550.00	421	**2 797**	347	1.29	**0.01**	0.70
F51D51C16	127	725	970	1 757	532.66	528.30	534.62	1099	**45 444**	4 237	1.47	**0.01**	2.10
F51D51C24	112	2 918	–	3 557	530.35	526.20	529.74	2179	86 400	86 400	1.83	0.31	1.50
F51D70C8	206	260	–	1 096	1505.91	1467.10	1530.12	509	**17 917**	478	2.89	**0.01**	5.00
F51D70C16	207	2 768	425	3 749	1451.70	1437.60	1460.25	707	86 400	931	1.41	0.41	2.15
F51D70C24	190	31 337	–	33 300	1454.45	1444.98	1438.05	1970	86 400	58 822	1.90	1.03	0.77
F51D392C8	243	–	2130	–	1307.88	–	1304.20	2706	86 400	2 711	1.18	–	0.50
F51D392C16	939	–	–	–	1307.16	–	–	2913	86 400	86 400	2.61	–	–
F51D392C24	1277	–	–	–	1297.34	–	–	3521	86 400	86 400	2.89	–	–
F102D392C8	1152	–	–	–	1090.02	–	1079.40	2726	86 400	61 050	2.66	–	1.18
F102D392C16	1566	–	–	–	1290.82	–	–	4807	86 400	86 400	2.82	–	–
F102D392C24	4649	–	–	–	1282.15	–	1292.48	6720	86 400	56 080	2.96	–	3.72

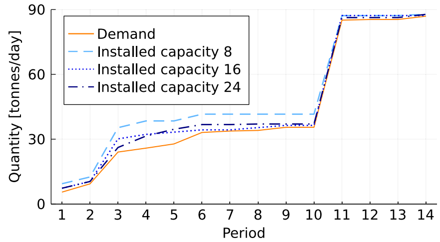
Table 6.2
Run times and optimality gap for our algorithm (LR), Gurobi (GUR) and restricted MIP (R-MIP) for the instances with S-shaped production costs.

Instance	Time in seconds				Best upper bound			Total run time			End gap		
	Gap < 3%		Gap < 1.5%		× 10^6			[s]			[%]		
	LR	GUR	LR	GUR	LR	GUR	R-MIP	LR	GUR	R-MIP	LR	GUR	R-MIP
F17D392C16*	138	14 871	1205	23 313	1356.70	1349.70	1403.94	1694	86 400	4 170	1.47	0.45	4.83
F34D51C16*	162	3 557	673	4 069	554.85	550.82	552.63	810	86 400	55 761	1.49	0.27	1.53
F34D70C16*	55	2 145	574	2 145	1488.22	1479.20	1479.75	690	86 400	86 400	1.41	0.16	0.79
F34D392C16*	137	–	–	–	1316.13	1371.45	1320.18	2515	86 400	2 286	1.87	9.63	9.00
F51D51C16*	133	934	–	1 141	550.82	546.74	551.18	1016	86 400	968	2.00	0.60	2.10
F51D70C16*	232	4 593	–	4 593	1486.20	1460.90	1466.10	913	86 400	86 400	2.31	0.22	0.95
F51D392C16*	956	–	–	–	1313.30	–	1299.20	3120	86 400	86 400	2.10	–	1.04
F102D392C16*	958	–	–	–	1302.14	–	–	4652	86 400	86 400	2.11	–	–

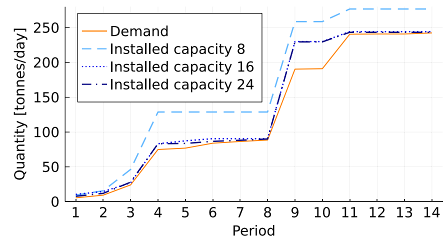
optimal) solutions much faster than Gurobi. For the instances where we find solutions within 1.5% of optimality with both our algorithm and Gurobi, our algorithm reaches this gap on average 30 times faster than Gurobi. Note however, that the run time advantage is less pronounced for smaller instances than for large instances.

We see that Gurobi can find optimal solutions for 13 of 38 instances. These instances are highlighted with a bold face. For instances, where Gurobi found the optimal solution within 24 h, we see from the best upper bound in Table 6.1 that the objective from the Lagrangian heuristic is on average less than 0.7% higher than the optimal solution. However, Lagrangian relaxation is 20 times faster than Gurobi. Gurobi fails to find feasible solutions for 11 of the instances (9 instances in Table 6.1 and 2 instances in Table 6.2) while our algorithm finds a solution with a gap < 3% for all tested instances. Note that the solutions found by the Lagrangian relaxation are no worse than 1.5% compared to the solutions found by Gurobi with the exception of the instance F51D70C8 where the difference is 2.6% and for two instances (F34D70C24 and F34D392C16*), we can find better solutions.

Tables 6.1 and 6.2 show that using the R-MIP, we can find feasible solutions for 33 out of 38 instances. For 6 of the instances out, R-MIP finds better solutions than LR and Gurobi. In general, the run times of the Lagrangian relaxation are lower for larger instances, whereas the R-MIP is faster for smaller instances. The reason for this is that restricted MIP is solved quickly and only once, while the Lagrangian relaxation calculates the upper bound in each iteration. The Lagrangian relaxation also finds better solutions than the R-MIP for 23 instances. Compared to Gurobi, the R-MIP run times are, in general, faster with exception of the smallest instance F17D51C8. For eleven of the instances both Gurobi and R-MIP run for 24 h. Even if the R-MIP can find an optimal or near-optimal solution, the proven optimality gap is still positive due to the Lagrangian lower bound. Note that even if the restricted MIP is solved to optimality, the best bound on restricted MIP is not the bound of the original problem. In general, the R-MIP finds feasible solutions with a cost similar to the ones found by the Lagrangian heuristic, but often has longer run time and worse proven optimality gap.

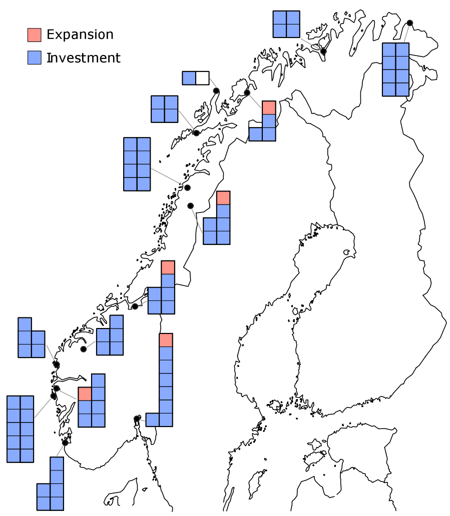


(a) Maritime demand and 51 demand points

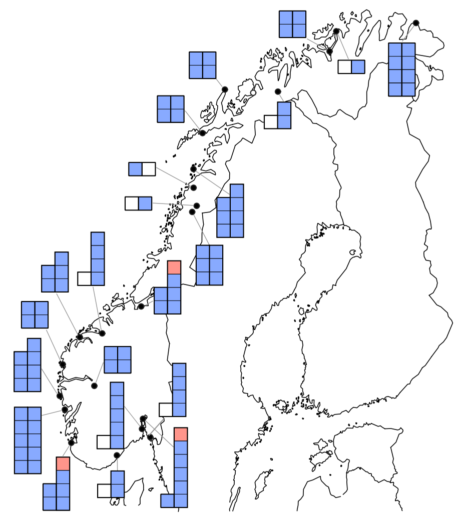


(b) All transportation demand and 70 demand points

Fig. 6.1. Comparison: Installed capacity and demand for instances with 17 candidate locations.



(a) Solution structure for instances F17D51C8 and F17D70C8



(b) Solution structure for instances F34D51C8 and F34D70C8

Fig. 6.2. Solution structure: opened facilities and theirs capacity. Left column illustrates the problem with 51 customers and right column the problem with 70 customers.

When analyzing the LR run times in Table 6.1 in more detail, we see that increasing the number of candidate locations has a relatively low impact on the run times. One of the reasons for that might be the parallel implementation for calculating the lower bound. We enable parallelization with 32 threads and use one thread to solve the subproblem for a single candidate location. We use all available threads only when solving instances with 32 or more candidate locations, which might explain the small differences in run time.

When comparing the run times for instances with 16 capacity levels, we observe that instances with S-shaped costs (see Table 6.2) are harder to solve using R-MIP and GUR than the corresponding instances with convex costs (see Table 6.1). Gurobi cannot find an optimal solution for any of these instances and for two of them, it cannot find any feasible solution, while the Lagrangian relaxation is still capable of finding good feasible solutions for all instances. The R-MIP finds better solution than Lagrangian relaxation for three instances. The run time in these instances however, is 24 h.

The instances with 70 or more demand points have a higher demand level compared to the instances with 51 demand points. However, despite the larger number of demand points, there is no clear effect on run time. For some instances, the higher number of customers even seems to have a positive impact on the run time. An explanation for this might be that due to higher demand, it is easier to satisfy the minimum production requirements and thus we find good solutions faster. If we increase the number of customers further from 70 to 392 (5.6 times) without changes in the total demand level, we observe an increase in run time. However, in most test instances, the run time increases less than 5.6 times.

We also analyze the impact of the number of capacity levels on the run time. The time needed to solve instances with 8 and 16 capacities is quite stable for most instances despite doubling the number of capacities. However, when increasing the number of capacity levels further to 24, we see a considerable increase in run time. With 16 available capacity levels, the installed capacity of the optimal solution approximates the demand much better than the optimal solution with 8 available capacities, see Fig. 6.1. Due to this, our algorithm finds good solutions faster. When increasing the number of capacities to 24, we cannot achieve a much better fit of the installed capacity and the additional capacity levels therefore only increase the computational complexity of the problem, causing the increase in run time.

6.2. Solution structure

To analyze the impact of different instances on the locations for hydrogen production in Norway, we compare the LR results for instances with 8 discrete capacities and convex production costs, namely F17D51C8, F17D70C8, F34D51C8, and F34D70C8. Fig. 6.2 shows the location and size of the opened hydrogen for the different instances. In Fig. 6.2(a), we compare the solutions for 17 candidate locations, whereas the solutions for 34 candidate locations are shown in Fig. 6.2(b). The evolution of installed capacity over time for the instances with 17 facilities can be seen in Fig. 6.1 for 51 and 70 customers.

When studying the consequences of different demand levels, we see for both the instances with 17 and 34 candidate locations that the solution for 70 demand points is characterized by considerable production capacity in south-eastern Norway. This is not surprising as the largest share of the demand from land-based transportation is in this region. We also observe that the total capacity of the facilities located in south-western Norway does not change that much. The stable production capacity indicates that hydrogen intended for maritime transportation continues to be produced at these facilities.

Increasing the number of candidate locations to 34 has little impact on the number of open facilities for the instances with 51 demand points, even though the locations change slightly and total costs are reduced by approximately 4%. For the instances with 70 demand points, the difference is much more apparent: The number of facilities in south-eastern Norway increases from 1 to 4. Another consequence of this increase in the number of facilities is that these facilities are smaller. One of the reasons for the increased number of facilities in south-eastern Norway is that the locations of these facilities are not available in the original set of 17 candidate locations. It is worth noting that increasing the number of candidate locations further to 51 and 102 locations does not increase the number of opened facilities. Our analysis shows that for 51 and 102 candidate locations, there are many alternative solutions with only slightly different cost level.

Both observations mentioned above for comparing demand level and the number of candidate locations indicate that the economies of scale present in the production cost of hydrogen are not sufficient to justify higher distribution costs. The model, therefore, chooses to open smaller facilities, closer to the demand points rather than centralizing production and distributing hydrogen over larger distances.

7. Conclusions

We have presented a solution method for the multi-period facility location and capacity expansion problem with a limited number of expansions. This work is motivated by the real-world problem of locating hydrogen production facilities in Norway which requires considering the limits on minimum production quantities and two production technologies. Our solution method is based on Lagrangian relaxation and combined with a heuristic to build a feasible solution from the solution of the relaxed problem.

Our test instances have different sizes with respect to the number of candidate locations, demand points and capacity levels. Our algorithm outperforms Gurobi in terms of run time for all instances. For instances, where we find solutions within 1.5% of optimality with both, our algorithm and Gurobi, we can prove this optimality gap on average 13 times faster than Gurobi. However, for some of the largest instances, Gurobi even fails to find a feasible solution.

In terms of objective function value, Gurobi can find optimal solutions for 13 of 38 instances. However, this applies mainly to smaller instances and comes at a cost of considerably longer run times. The solutions provided by our algorithm will be good enough for most practical purposes where run time requirements supersede the need for optimal solutions. If better solutions are indeed needed, improving the greedy heuristic for finding an upper bound should be considered.

Among exact methods, a promising direction for future research is to combine the Lagrangian relaxation approach with a branch-and-bound procedure or to implement Benders decomposition. Further, different heuristic approaches can be studied to improve the quality of the solution. In future research, a more general model formulation allowing also for capacity reduction and closing of facilities can be considered.

CRediT authorship contribution statement

Šárka Štádlerová: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Peter Schütz:** Writing – review & editing, Supervision, Project administration. **Asgeir Tomasgard:** Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

Acknowledgements

We would like to thank the anonymous reviewers for their thorough review of the manuscript. We sincerely appreciate their valuable comments and suggestions, which helped us to improve the quality of the manuscript.

References

Aarskog, F.G., Danebergs, J., 2020. Estimation of energy demand in the norwegian high-speed passenger ferry sector towards 2030. IFE/E-2020/003, Halden, Norway.

Allen, B.T., Doherty, N., Weigelt, K., Mansfield, E., 2012. Managerial Economics, 8 W.W. Norton & Company, Inc.

Amiri, A., 1997. Solution procedures for the service system design problem. Comput. Oper. Res. 24 (1), 49–60.

Arabani, A.B., Farahani, R.Z., 2012. Facility location dynamics: An overview of classifications and applications. Comput. Ind. Eng. 62 (1), 408–420.

Arostegui, Jr., M.A., Kadipasaoglu, S.N., Khumawala, B.M., 2006. An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. Int. J. Prod. Econ. 103 (2), 742–754.

Behmardi, B., Lee, S., 2008. Dynamic multi-commodity capacitated facility location problem in supply chain. In: Fowler, J., Mason, S. (Eds.), Proceedings of the 2008 Industrial Engineering Research Conference. Institute of Industrial and Systems Engineers (IISE), pp. 1914–1919.

Castro, J., Nasini, S., Saldanha da Gama, F., 2017. A cutting-plane approach for large-scale capacitated multi-period facility location using a specialized interior-point method. Math. Program. 163 (1–2), 411–444.

Christensen, T.R.L., Klose, A., 2021. A fast exact method for the capacitated facility location problem with differentiable convex production costs. European J. Oper. Res. 292 (3), 855–868.

Correia, I., Captivo, M.E., 2003. A Lagrangean heuristic for a modular capacitated location problem. Ann. Oper. Res. 122 (1), 141–161.

Correia, I., Melo, T., 2021. Integrated facility location and capacity planning under uncertainty. Comput. Appl. Math. 40 (5), 1–36.

Danebergs, J., Aarskog, F.G., 2020. Future compressed hydrogen infrastructure for the domestic maritime sector. IFE/E-2020/006, Halden, Norway.

Dias, J., Captivo, M.E., Clíma, J., 2007a. Dynamic multi-level capacitated and uncapacitated location problems: an approach using primal-dual heuristics. Oper. Res. 7 (3), 345–379.

Dias, J., Captivo, M.E., Clímaco, J., 2006. Capacitated dynamic location problems with opening, closure and reopening of facilities. IMA J. Manag. Math. 17 (4), 317–348.

Dias, J., Captivo, M.E., Clímaco, J., 2007b. Dynamic location problems with discrete expansion and reduction sizes of available capacities. Investig. Operacional 27 (2), 107–130.

DNV GL, 2019. Produksjon og bruk av hydrogen i norge. Rapport 2019-0039, Oslo, Norway, (in Norwegian).

Haldi, J., Whitcomb, D., 1967. Economies of scale in industrial plants. J. Polit. Econ. 75 (4, Part 1), 373–385.

Hinojosa, Y., Kalcsics, J., Nickel, S., Puerto, J., Velten, S., 2008. Dynamic supply chain design with inventory. Comput. Oper. Res. 35 (2), 373–391.

Hirth, M., Hove, K., Janzen, D., Eide, P., Helland, P., Østvik, I., Ryberg, T., Ødegard, J., 2019. Norwegian future value chains for liquid hydrogen. NCE Maritime CleanTech, Report liquid hydrogen 2019, Stord, Norway.

Holmberg, K., 1994. Solving the staircase cost facility location problem with decomposition and piecewise linearization. European J. Oper. Res. 75 (1), 41–61.

Jakobsen, D., Åtland, V., 2016. Concepts for large scale hydrogen production. Master's thesis. Department of Energy and Process Engineering, NTNU, Trondheim, Norway.

Jena, S.D., Cordeau, J.-F., Gendron, B., 2015. Dynamic facility location with generalized modular capacities. Transp. Sci. 49 (3), 484–499.

Jena, S.D., Cordeau, J.-F., Gendron, B., 2016. Solving a dynamic facility location problem with partial closing and reopening. Comput. Oper. Res. 67, 143–154.

Jena, S.D., Cordeau, J.-F., Gendron, B., 2017. Lagrangian heuristics for large-scale dynamic facility location with generalized modular capacities. INFORMS J. Comput. 29 (3), 388–404.

Li, J., Chu, F., Prins, C., 2009. Lower and upper bounds for a capacitated plant location problem with multicommodity flow. Comput. Oper. Res. 36 (11), 3019–3030.

Li, L., Manier, H., Manier, M.-A., 2019. Hydrogen supply chain network design: An optimization-oriented review. Renew. Sustain. Energy Rev. 103, 342–360.

Marsten, R.E., Hogan, W.W., Blankenship, J.W., 1975. The boxstep method for large-scale optimization. Oper. Res. 23 (3), 389–405.

Melo, M.T., Nickel, S., Saldanha da Gama, F., 2006. Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Comput. Oper. Res. 33 (1), 181–208.

Melo, M.T., Nickel, S., Saldanha da Gama, F., 2009. Facility location and supply chain management–a review. European J. Oper. Res. 196 (2), 401–412.

Melo, M., Nickel, S., Saldanha da Gama, F., 2012. A tabu search heuristic for redesigning a multi-echelon supply chain network over a planning horizon. Int. J. Prod. Econ. 136 (1), 218–230.

Nemhauser, G.L., Wolsey, L.A., 1999. Integer and Combinatorial Optimization. vol. 55, John Wiley & Sons, Hoboken.

Nickel, S., Saldanha da Gama, F., 2019. Multi-period facility location. In: Laporte, G., Nickel, S., Saldanha da Gama, F. (Eds.), Location Science. Springer, Cham, pp. 303–326.

Ocean Hyway Cluster, 2020a. 2030 Hydrogen demand in the norwegian domestic maritime sector. OHC HyInfra project, Workpackage C: Mapping future hydrogen demand.

Ocean Hyway Cluster, 2020b. Interactive map - potential maritime hydrogen in Norway. OHC HyInfra project, Workpackage C: Mapping future hydrogen demand.

Sauvey, C., Melo, T., Correia, I., 2020. Heuristics for a multi-period facility location problem with delayed demand satisfaction. Comput. Ind. Eng. 139, 106171.

Schütz, P., Stougie, L., Tomasgard, A., 2008. Stochastic facility location with general long-run costs and convex short-run costs. Comput. Oper. Res. 35 (9), 2988–3000.

Schütz, P., Tomasgard, A., Ahmed, S., 2009. Supply chain design under uncertainty using sample average approximation and dual decomposition. European J. Oper. Res. 199 (2), 409–419.

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

Silva, A., Aloise, D., Coelho, L.C., Rocha, C., 2021. Heuristics for the dynamic facility location problem with modular capacities. European J. Oper. Res. 290 (2), 435–452.

Štádlerová, Š., Jena, S.D., Schütz, P., 2023. Using Lagrangian relaxation to locate hydrogen production facilities under uncertain demand: a case study from Norway. Comput. Manag. Sci. 20 (10).

Štádlerová, Š., Schütz, P., 2021. Designing the hydrogen supply chain for maritime transportation in Norway. In: Mes, M., Lalla-Ruiz, E., Voß, S. (Eds.), Computational Logistics. Springer, Cham, pp. 36–50.

Statistics Norway, 2021. Statistics Norway: 12579: Road traffic volumes. Online; accessed 02.11.2021, https://www.ssb.no/en/statbank/table/12579/.

Torres-Soto, J.E., Üster, H., 2011. Dynamic-demand capacitated facility location problems with and without relocation. Int. J. Prod. Res. 49 (13), 3979–4005.

Van den Broek, J., Schütz, P., Stougie, L., Tomasgard, A., 2006. Location of slaughterhouses under economies of scale. European J. Oper. Res. 175 (2), 740–750.