Appendix

Paper: Lagrangian Heuristics for Large-Scale Dynamic Facility Location with Generalized Modular Capacities

Authors: *Sanjay Dominik Jena, Jean-Franois Cordeau, Bernard Gendron*

This is the online appendix of the paper: Lagrangian Heuristics for Large-Scale Dynamic Facility Location with Generalized Modular Capacities. Section A illustrates how different multi-period facility location problems can be modeled using the Generalized Modular Capacity (GMC) formulation of the Dynamic Facility Location Problem with Generalized Modular Capacities (DFLPG). Section B provides the details for the generation of the test instances.

A Representation of Special Cases via the GMC Formulation

We now illustrate how three special cases can be modeled by using the GMC formulation. In the first problem, the size of the facility is chosen from a discrete set of capacity levels. Existing facilities may then be closed and reopened several times. In the second problem, capacities can be adjusted once a facility is constructed. At each facility, the capacity can be expanded or reduced from one capacity level to another. It is assumed that an expansion of ℓ capacity levels has always the same cost, regardless of the previous capacity level. These two problems are denoted as the *Dynamic Modular Capacitated Facility Location Problem with Closing and Reopening (DMCFLP_CR)* and the *Dynamic Modular Capacitated Facility Location Problem with Capacity Expansion and Reduction (DMCFLP_ER)*, respectively.

A subset of capacity change variables $y^{jt}_{\ell\ell'}$ is chosen to model these special cases. The cost coefficients $f^{jt}_{\ell\ell'}$ for these variables are based on the following fixed costs, defined to characterize the special cases:

- $c^c_{j\ell}$ and $c^o_{j\ell}$ are the costs to temporarily close and reopen a facility of size $\ell \geq 1$ at location j, respectively;
- $f^c_{j\ell}$ and $f^o_{j\ell}$ are the costs to reduce and to expand the capacity of a facility at location j by ℓ capacity levels, respectively;
- $F^o_{j\ell}$ is the cost to maintain an open facility of size ℓ at location j throughout one time period.

For the problem variant involving facility closing and reopening, we create an artificial capacity level $\overline{\ell}$ for each capacity level $\ell \in L\backslash\{0\}$. Capacity level $\overline{\ell}$ represents the state in which a facility of size ℓ is temporarily closed. At each time period $t \in T$ and location $j \in J$, we may find capacity transition decisions $y^{jt}_{\ell\ell'}$ that represent different types of operations (note that the costs for these decisions are usually composed by the cost to perform the capacity transition, as well as the maintenance cost for the new capacity level):

1. Facility construction and capacity expansion. The expansion of the capacity is represented by a capacity transition from capacity level $\ell \geq 1$ to any other capacity level $\ell' > \ell$. If the decision represents a facility construction, then ℓ is 0. The capacity is thus expanded by $\ell' - \ell$ capacity levels. The cost for this decision is set to $f^{jt}_{\ell\ell'} = f^o_{j(\ell'-\ell)} + F^o_{j\ell'}$.
2. Capacity reduction. The reduction of the capacity is represented by a transition from capacity level $\ell \geq 1$ to any other capacity level $\ell' < \ell$. The capacity is thus reduced by $\ell - \ell'$ capacity levels. The cost for this decision is set to $f^{jt}_{\ell\ell'} = f^c_{j(\ell-\ell')} + F^o_{j\ell'}$.
3. Maintaining the current capacity level. A facility may neither have its capacity expanded nor reduced. The cost is thus only composed of the maintenance cost, i.e., $f^{jt}_{\ell\ell} = F^o_{j\ell}$ if the capacity level represents an open facility, $f^{jt}_{\overline{\ell}\overline{\ell}} = 0$ if the capacity level represents a temporarily closed facility and $f^{jt}_{00} = 0$ if no facility exists.
4. Temporary closing. An open facility of size $\ell \geq 0$ can be temporarily closed, i.e., it changes to capacity level $\overline{\ell}$. The total cost is $f^{jt}_{\ell\overline{\ell}} = c^c_{j\ell}$.

5. Reopening a closed facility. A temporarily closed facility of size $\ell \geq 1$ can be reopened, i.e., it changes its capacity level from $\bar{\ell}$ to ℓ. The total cost for this decision is $f^{jt}_{\bar{\ell}\ell} = c^o_{j\ell} + F^o_{j\ell}$.

The DMCFLP_CR is represented by transition decisions of type 1 (for construction only), 3, 4 and 5. We denote the resulting model as the *CR-GMC* formulation. The DMCFLP_ER is represented by transition decisions of type 1, 2 and 3. The resulting model is denoted as the *ER-GMC* formulation.

A third problem variant can be considered, which combines both features of the two special cases. It is denoted as the *Dynamic Modular Capacitated Facility Location Problem with Closing/Reopening and Capacity Expansion/Reduction (DMCFLP_CR_ER)*. The problem variant is modeled by using the transition decisions of type $1-5$ presented above. However, these decisions allow only one operation, for example either capacity reduction or facility closing, at each time period. In practice, it is very likely that one may want to reduce or expand the capacity before closing or after reopening a facility at the same time period. We may therefore consider four additional decision types that represent combinations of such operations:

(a) A facility is reopened at level ℓ and its capacity is expanded to level $\ell' > \ell$ at the same time period.

(b) A facility is reopened at level ℓ and its capacity is reduced to level $\ell' < \ell$ at the same time period.

(c) The capacity of a facility at level ℓ is expanded to level $\ell' > \ell$ and the facility is closed right after.

(d) The capacity of a facility at level ℓ is reduced to level $\ell' < \ell$ and the facility is closed right after.

By making the realistic assumption that the costs for closing and reopening a facility are non-decreasing as the size of the facility increases, we may discard two of the four possibilities.

Proposition A.1 *Let $c^o_{j\ell} \leq c^o_{j(\ell+1)}$ and $c^c_{j\ell} \leq c^c_{j(\ell+1)}$ for $\ell = 0, 1, 2, \ldots, (q-1)$, then there is at least one optimal solution that does neither use decisions of type (b) nor of type (c).*

PROOF. Note that case (c) may only occur in two situations: either the facility stays closed until the end of the planning horizon or the facility is reopened at a later moment. If the facility stays closed, then closing it at level ℓ is at most as expensive as combined capacity expansion and closing as suggested in case (c): $c^c_{j\ell} \leq f^o_{j(\ell'-\ell)} + c^c_{j\ell'}$. If the facility is closed at the beginning of time period t_1, but it will be reopened at the beginning of period $t_2 > t_1$, then the corresponding costs using case (c) are given by: $C^c = c^c_{j\ell} + f^o_{j(\ell'-\ell)} + c^o_{j\ell'} + F^o_{j\ell'}$. However, the same solution may be reproduced by closing the facility at level ℓ and expanding its capacity only after it has been reopened using case (a), which corresponds to the following costs: $C^a = c^c_{j\ell} + c^o_{j\ell} + f^o_{j(\ell'-\ell)} + F^o_{j\ell'}$. Now, because $c^o_{j\ell} \leq c^o_{j\ell'}$, we have: $C^a \leq C^c$. Therefore, a solution using case (a) is at most as expensive as a solution using case (c).

The same can be shown for the relation between cases (d) and (b), where reducing the capacity before temporary closing is at most as costly as reducing the capacity after temporary closing. ∎

We thus add only the transition decisions given by the cases (a) and (d) to the model:

6. Reopening and capacity expansion. A closed facility of capacity level ℓ is reopened and its capacity is expanded to level ℓ' (with $\ell < \ell'$).The cost for this decision, including the maintenance costs at capacity level ℓ' is thus set to $f^{jt}_{\bar{\ell}\ell'} = c^o_{j\ell} + f^o_{j(\ell'-\ell)} + F^o_{\ell'}$.

7. Capacity reduction and facility closing. An open facility reduces its capacity from level ℓ to level ℓ' (with $\ell > \ell'$) and is temporarily closed afterwards. The cost for this decision, including the maintenance costs at capacity level ℓ' is thus set to $f^{jt}_{\ell\bar{\ell}'} = f^c_{j(\ell-\ell')} + c^c_{j\ell'}$.

B Test Instances

Instances for multi-period facility location problems essentially contain information about the customer demand for each time period, construction costs of the facilities and the costs to allocate demand between customers and facilities. The DFLPG and the three special cases additionally involve a detailed cost structure for the capacity changes. Due to the lack of openly available instance sets that include these properties, we generated a total of 540 instances, 180 for each capacity level, to test the presented models. These essentially extend the instances used by Jena et al. (2015a) by adding multiple commodities, the use of a cost matrix for capacity changes and a larger set of candidate facility locations. Some of these parameters have been adapted from the data for a real world application, which has been introduced in Jena et al. (2012). All problem instances have been made available as an online supplement. In the following we present how these instance properties are generated and which parameters are used.

B.1 Problem dimension

Instances were generated with different numbers of candidate facility locations $|J|$ and customers $|I|$, combining all pairs of $J \in \{50, 100, 150, 200, 250\}$ and $I \in \{|J|, 4 \cdot |J|\}$. To be precise, the instance dimensions are: *(50/50)*, *(50/200)*, *(100/100)*, *(100/400)*, *(150/150)*, *(150/600)*, *(200/200)*, *(200/800)*, *(250/250)* and *(250/1000)*.

B.2 Number of capacity levels

The number of capacity levels q also impacts the size of the models. Instances are generated with a maximum of 3, 5 and 10 capacity levels, which are assumed to be reasonable values for a broad variety of different application contexts.

The capacities u_ℓ^j are generated based on the total number of customers and are chosen such that the number of opened facilities in near optimal solutions strongly varies among the problem instances. Using CPLEX with a time limit of 24 hours and considering DMCFLP_CR_ER problem instances with $q = 10$ (only those for which optimality within 2% has been proven), the average, minimum and maximum percentages of locations with facility construction has been found to be 24%, 4% and 76%, respectively. The larger the set of customers, the higher is the capacity of each level. To be precise, the values of u_ℓ^j have been set in relation to some initial capacity $\overline{u}^j$ that depends on the number of customers covered in the instance, as reported in Table B.1.

# customers covered	$\overline{u}^j$
50	300
100	600
150	800
200	1,000
250	1,200
400	2,000
600	2,500
800	3,000
1,000	5,000

Table B.1: Total capacity of a facility at capacity level 1 in function of the problem size.

Then, for $q = 10$, the capacities are set as multiples of the initial capacity $\overline{u}^j$, i.e., $u_\ell^j = \ell \cdot \overline{u}^j$. To ensure that facilities have sufficiently large capacities in the cases of $q = 3$ and $q = 5$, we set the capacities to $u_\ell^j = 3 \cdot \ell \cdot \overline{u}^j,\ \ell = 1, \ldots, 3$ for $q = 3$, and to $u_\ell^j = 2 \cdot \ell \cdot \overline{u}^j,\ \ell = 1, \ldots, 5$ for $q = 5$. Note that we assume that the problem instances do not contain initially existing facilities, i.e., the initial capacity level of each facility is 0.

B.2.1 Number of time periods

All generated instances contain ten time periods, which is found to be sufficient to demonstrate capacity changes over time.

B.3 Customer/facility locations

For each of the different problem sizes, $|I|$ customer demand points have been randomly generated following a continuous uniform distribution, rounding the x and y coordinates to the next lowest integer value. The first $|J|$ points of $|I|$ customer locations have additionally been defined as candidate facility locations and therefore coincide with the customer demand points. The networks were generated on squares of sides with lengths 300, 380 and 450.

B.4 Demand allocation costs

Costs are divided into fixed and variable costs. Fixed costs are given by the construction of facilities and the change of their capacity levels. Variable costs are composed of the costs to produce and transport the commodities.

Transportation costs have been computed based on the Euclidean distance between the points, including a small modification that results in a slight clustering effect of the customers close to a facility. The transportation costs are composed of two components:

1. A cost that depends on the total distance, referred to as the vehicle cost. The vehicle cost is linear in function of the Euclidean distance between the two points on the network (5 monetary units per unit of distance).
2. A cost that depends on the travel time, referred to as the driver's payment. The driver's payment is 0 if the two points are within one-hour of transportation distance (assuming an average vehicle speed of 62 distance units per hour) and linear in function of the Euclidean distance if the two points are at more than one hour of driving distance (50 monetary units per hour).

Let $dist_{ij}$ denote the distance between facility location j and customer i. The costs to transport one unit of demand from facility j to customer i is therefore set to:

$$g_{ij}^{T} = C_p^t \cdot dist_{ij} + 50 \cdot \max\left(0, \frac{dist_{ij}}{62} - 1\right),$$

where C_p^T is set to 15, 10, 15, 10 and 15 for the first to the fifth commodity p, respectively, for the experiments that use up to five commodities. For the experiments that use up to 50 commodities (reported in Table 9), C_p^T for $p = 1, \ldots, 10$ has been set to 6, 3, 1, 4, 2, 7, 5, 2, 4, and 6. This sequence of values has then been repeated for the remaining values for $p = 11, \ldots, 50$.

The variable and fixed costs include economies of scale in function of the size of the facility. These costs are therefore described by concave cost functions, as explained in the following. The production costs for each unit served from a facility to a customer is defined as the cost to operate a facility and depends on the size of the facility. The cost to produce one commodity unit at capacity level 1 is set to 20.90 monetary units. At each higher capacity level, the production cost is 3% cheaper than at the previous level:

$$g_{j0}^{P} = 20.90$$

$$g_{j\ell}^{P} = 0.97 \cdot g_{j(\ell-1)}^{P}.$$

Note that the production costs are added to the transportation costs to determine the total demand allocation costs $g_{i\ell}^{jt}$ to serve the customer demands:

$$g_{i\ell}^{jt} = g_{ij}^{T} + g_{j\ell}^{P}.$$

B.5 Fixed costs

The construction cost, also referred to as capacity expansion cost, is set to 100,000 monetary units for a facility of level 1. Each additional capacity level is 10% cheaper than the previous one. The construction costs for facilities of different capacity levels are therefore computed according to the following formula:

$$f^o_{j1} = 100,000$$

$$f^o_{j2} = 190,000$$

$$f^o_{j\ell} = f^o_{j(\ell-1)} + 0.9 \cdot (f^o_{j(\ell-1)} - f^o_{j(\ell-2)})$$

The maintenance costs for a facility of a certain size are computed in a similar fashion. They are set relatively high to motivate capacity changes. The maintenance costs for a facility of capacity level 1 are set to 51,000 monetary units. The maintenance costs for each additional capacity level are 15% cheaper than the previous ones:

$$F^o_{j1} = 51,000$$

$$F^o_{j2} = 94,350$$

$$F^o_{j\ell} = F^o_{j(\ell-1)} + 0.85 \cdot (F^o_{j(\ell-1)} - F^o_{j(\ell-2)})$$

Fixed Costs for the Special Cases

For the three special cases, i.e., the DMCFLP_CR, DMCFLP_ER and the DMCFLP_CR_ER, the cost to reduce the capacity of a facility by ℓ capacity levels is set to 10% of the costs to expand the capacity of a facility by ℓ capacity levels.

Finally, the costs for reopening and closing existing facilities were taken from the input data of the previously mentioned industrial application introduced by Jena et al. (2015b). Although being strictly increasing, these costs do not necessarily represent economies of scale. The costs to reopen a closed facility of capacity level $1, \dots, 10$ are 3,138.34, 4,084.69, 4,924.58, 5,693.26, 7,085.07, 7,727.50, 8,342.34, 8,933.68, 10,057.70 and 10,594.80, respectively. The costs to close an open facility of capacity level $1, \dots, 10$ are 8,624.93, 11,595.80, 14,305.60, 16,836.50, 21,524.10, 23,727.90, 25,858.30, 27,925.70, 31,901.10 and 33,820.70, respectively.

Fixed Costs for the DFLPG

For the DFLPG, the construction costs are as indicated above, i.e., the costs to construct a facility of size ℓ and its maintenance costs at time period t are set to: $f^{jt}_{0\ell} = f^o_{j\ell} + F^o_{j\ell}$.

The costs to change capacity levels for this problem are based on a cost matrix, and, therefore, differ from the costs for capacity expansion and reduction shown above for the special cases. The cost to completely remove a facility is set to 25% of the construction cost of a facility of the same size: $f^{jt}_{\ell 0} = f^o_{j\ell}/4$.

Finally, the cost to change the capacity level from $\ell \geq 1$ to $\ell' \geq 1$ is set to the difference of their construction costs, scaled by 50%:

$$f^{jt}_{\ell\ell'} = \begin{cases} 1.5 \cdot (f^o_{j\ell'} - f^o_{j\ell}) \text{ , if } \ell < \ell' \\ 1.5 \cdot (f^o_{j\ell} - f^o_{j\ell'}) \text{ , if } \ell > \ell'. \end{cases}$$

B.6 Demand distribution

We consider two different demand scenarios. In both scenarios, the demand for each of the customers is randomly generated and randomly distributed over time. The two scenarios differ in their total demand summed over all customers in each time period. In the first scenario (*regular*), the total demand is similar in each time period. We set the average demand for a customer to 12 units per time period. The total demand for all customers is therefore approximately $10 \cdot |I|$ units at each time period. The second scenario (*irregular*) assumes that the total demand follows strong variations along time and is therefore different at each time period. In this scenario, the total demand for all customers is multiplied by a random distortion factor at each time period. This random distortion factor is set to the absolute value of a normal random variable with mean value 1.0 and standard deviation 0.6 (note that this procedure produced distortion factors from 0.14 to 2.24). Let $totDem_t$ be the total customer demand for time period t, computed as explained above for one of the two scenarios.

We now explain how the individual demands for each of the customers are generated and distributed on the different time periods such that its total sum equals approximately the value of $totDem_t$ at each of the time periods. For all customers and all time periods, the total demand covers approximately $12 \cdot |I| \cdot |T|$ units. In a first step, this total demand is randomly distributed on each of the customers. In a second step, each customer demand is distributed on different time periods:

1. Let $totRemDem$ denote the total demand for all customers and time periods that has not yet been allocated to any customer. Furthermore, let $numRemCust$ indicate the number of customers that have not yet been allocated any demand. For each customer, its total demand for all time periods, denoted to $totJDem_j$, is computed as a random normal variable with a mean $\mu = totRemDemand/numRemCust$ and standard deviation $\sigma = \mu/2$. Note that, throughout our instance generation, this method did not produce any negative value.

2. The total demand for each customer, $totJDem_j$ is then divided into four equal parts. One part of the demand is allocated to a time period that is randomly selected following a uniform distribution. Each of the other three parts is allocated to the time period t that has the highest gap between the total demand yet allocated to period t and its value $totDem_t$.

The demands for the second to fifth commodity are computed based on the demand of the first commodity. To be precise, the demand d_{ip}^t for $p \geq 2$ is computed as $d_{ip}^t = d_{i1}^t \cdot rand(1.0, 0.2) \cdot avgDem_p/avgDem_1$, where $avgDem_1 = 10$, $avgDem_2 = 6$, $avgDem_3 = 9$, $avgDem_4 = 5$, $avgDem_5 = 8$ and $rand(1.0, 0.2)$ is a random variable with normal distribution, a mean of 1.0 and a standard deviation of 0.2.

The problem instances with up to 50 commodities have been generated in a similar fashion. Here, the average demands for the first ten commodities have been set to 10, 6, 3, 1, 8, 4, 7, 2, 5 and 9, respectively. Each additional ten commodities possess the same average demand as he first ten commodities. Finally, each demand has been scaled by multiplying by $\frac{2}{|P|}$.

Note that the choice of allocating demand to only a few of the time periods is motivated by the aforementioned industrial application in the forest industry, where each logging region is harvested, on average, about four seasons over the ten-period planning horizon. Furthermore, it results in a geographically more dispersed distribution of the demand which creates the need to adjust capacities at the facilities.

References

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

Jena, S. D., J.-F. Cordeau, B. Gendron. 2015b. Modeling and solving a logging camp location problem. *Annals of Operations Research* **232**(1) 151–177.