AIMMS Modeling Guide - Facility Location Problem

This file contains only one chapter of the book. For a free download of the complete book in pdf format, please visit www.aimms.com.

AIMMS B.V.
Diakenhuisweg 29-35
2033 AP Haarlem
The Netherlands
Tel.: +31 23 5511512

AIMMS Inc.
11711 SE 8th Street
Suite 303
Bellevue, WA 98005
USA
Tel.: +1 425 458 4024

AIMMS Pte. Ltd.
55 Market Street #10-00
Singapore 048941
Tel.: +65 6521 2827

AIMMS
SOHO Fuxing Plaza No.388
Building D-71, Level 3
Madang Road, Huangpu District
Shanghai 200025
China
Tel.: ++86 21 5309 8733

Email: info@aimms.com
WWW: www.aimms.com

This documentation was typeset by AIMMS B.V. using LATEX and the LUCIDA font family.

Chapter 22

A Facility Location Problem

This chapter

This chapter considers the problem of selecting *distribution centers* along with their associated customer zones. For small and medium-sized data sets, the mathematical model is a straightforward mixed-integer programming formulation and can easily be solved with standard solvers. However for large data sets, a decomposition approach is proposed. This chapter explains the Benders' decomposition technique and applies it to the facility location problem.

References

The example in this chapter is based on "Multicommodity Distribution System Design by Benders Decomposition" ([Ge74]) by Geoffrion and Graves.

Keywords

Integer Program, Mathematical Reformulation, Mathematical Derivation, Customized Algorithm, Auxiliary Model, Constraint Generation, Worked Example.

22.1 Problem description

Distribution system design

A commonly occurring problem in distribution system design is the optimal location of intermediate distribution centers between production plants and customer zones. These intermediate facilities (temporarily) store a large variety of commodities that are later shipped to designated customer zones.

Basic problem in words

Consider the situation where several commodities are produced at a number of plants with known production capacities. The demands for each commodity at a number of customer zones are also known. This demand is satisfied by shipping via *intermediate distribution centers*, and for reasons of administration and efficiency, each customer zone is assigned exclusively to a single distribution center. For each center there is a lower as well as an upper limit on the total throughput (of all commodities). There is also a fixed rental charge and a per unit throughput charge associated with each distribution center. In addition, there is a variable unit cost of shipping a commodity from a plant to a customer zone through a distribution center. This cost usually includes the unit production cost.

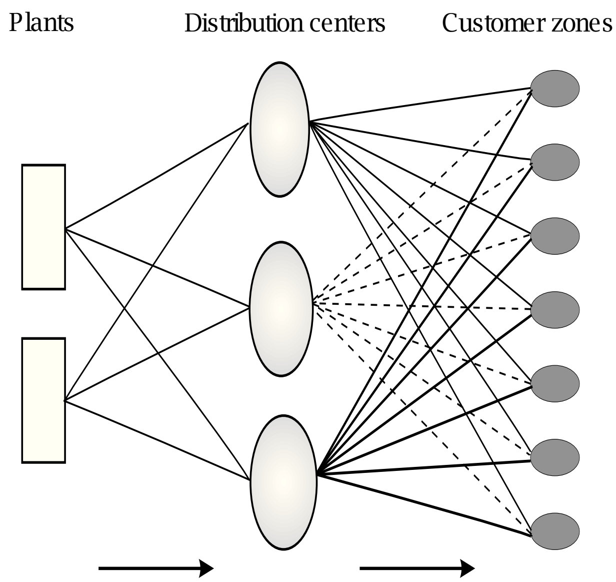

Figure 22.1: Commodity distribution scheme

Decisions to be made

The facility location problem is shown schematically in Figure 22.1. It has the property that the main decisions are of type yes/no. The problem is to determine which distribution centers should be selected, and what customer zones should be served by the selected distribution centers. The optimum solution is clearly dependent on the pattern of transportation flows for all commodities. It is assumed that the time frame under consideration is sufficiently long to motivate good decision making.

Cost minimization

The decisions described in the previous paragraphs are to be made with the objective to meet the given demands at minimum total distribution and production cost, subject to plant capacities and distribution center throughput requirements.

Problem extensions

This chapter formulates and solves the above problem description. However in real-world applications, there may be additional constraints which require some specialized formulation. Some possibilities are mentioned below.

- The throughput capacity in a particular distribution center can be treated as a decision variable with an associated cost.
- Top management could impose an a priori limit on the number of distribution centers, or express preferences for particular logical combinations of such centers (not A unless B, not C and D, etc.).
- Similarly, there could be an a priori preference for certain logical combinations of customer zones and distribution centers (if Center A is open, then Zone 2 must be assigned, etc.).

- If distribution centers happen to share common resources or facilities, there could be joint capacity constraints.

You are referred to Chapter 7 for ideas on how to model these special logical conditions.

22.2 Mathematical formulation

This section

This section presents the mathematical description of the facility location problem discussed in the previous section.

Qualitative model description

The objective and the constraints are described in the following qualitative model formulation.

Minimize: *total production and transport costs,*

Subject to:

- *for all commodities and production plants: transport must be less than or equal to available supply,*
- *for all commodities, distribution centers and customer zones: transport must be greater than or equal to required demand,*
- *for all distribution centers: throughput must be between specific bounds, and*
- *for all customer zones: supply must come from exactly one distribution center.*

Notation

The following notation will be used in this chapter:

Indices:

c	*commodities*
p	*production plants*
d	*distribution centers*
z	*customer zones*

Parameters:

S_{cp}	*supply (production capacity) of commodity c at plant p*
D_{cz}	*demand for commodity c in customer zone z*
$\overline{M}_d$	*maximum throughput at distribution center d*
$\underline{M}_d$	*minimum throughput at distribution center d*
R_d	*per unit throughput charge at distribution center d*
F_d	*fixed cost for distribution center d*
K_{cpdz}	*variable cost for production and shipping of commodity c, from plant p via distribution center d to customer zone z*

Variables:

x_{cpdz}	*nonnegative amount of commodity c shipped from plant p via distribution center d to customer zone z*

v_d	*binary to indicate selection of distribution center d*
y_{dz}	*binary to indicate that customer zone z is served by distribution center d*

Supply constraint

The supply constraint specifies that for each commodity c and each production plant p, the total amount shipped to customer zones via distribution centers cannot be more than the available production capacity,

$$\sum_{dz} x_{cpdz} \leq S_{cp} \qquad \forall c, p$$

Demand constraint

The demand constraint specifies that the demand for each commodity c in each zone z should be supplied by all plants, but only through the chosen distribution center y_{dz},

$$\sum_{p} x_{cpdz} \geq D_{cz} y_{dz} \qquad \forall c, d, z$$

Throughput constraints

The throughput constraints make sure that for each distribution center d the total volume of commodities to be delivered to its customer zones remains between the minimum and maximum allowed throughput,

$$\underline{M}_d v_d \leq \sum_{cpz} x_{cpdz} = \sum_{cz} D_{cz} y_{dz} \leq \overline{M}_d v_d \qquad \forall d$$

Allocation constraint

The allocation constraint ensures that each customer zone z is allocated to exactly one distribution center d.

$$\sum_{d} y_{dz} = 1 \qquad \forall z$$

Objective function

The objective function that is to be minimized is essentially the addition of production and transportation costs augmented with the fixed and variable charges for distribution centers and the throughput of commodities through these centers.

$$\textbf{Minimize:} \sum_{cpdz} K_{cpdz} x_{cpdz} + \sum_{d} [F_d v_d + R_d \sum_{cz} D_{cz} y_{dz}]$$

22.3 Solve large instances through decomposition

Black box approach

The facility location problem can be solved for small to medium sized data sets using any of the mixed integer programming solvers that are available through AIMMS. However, its solution process is based on a branch-and-bound approach and this can sometimes be improved if you add some constraints.

These constraints are redundant for the integer formulation but tighten the associated relaxed linear program solved at each node of the underlying branch-and-bound tree.

Redundant constraints

Two examples of such redundant constraints are:

$$\begin{aligned} y_{dz} &\leq v_d \qquad \forall d, z, \text{and} \\ \sum_d v_d &\leq L \end{aligned}$$

where L is a heuristically determined upper limit on the number of distribution centers to be opened (based on total demand). For your application, you may want to test if adding these constraints does indeed improve the solution process. In general, the benefit increases as the data set becomes larger.

Large instances

In some practical applications, it is not unusual for the number of commodities and customer zones to be in the order of 100's to 1000's. Under these conditions, it is possible that the internal memory required by the solver to hold the initial data set is insufficient. If there is enough memory for the solver to start the underlying branch-and-bound solution process, the number of nodes to be searched can be extremely large, and inefficient search strategies (such as depth-first search) may be required to keep the entire search tree in memory.

Decomposition

When your model uses an extremely large data set, you may consider re-examining your approach to the problem. One option is to decompose the problem into several smaller *subproblems* that are solved sequentially rather than simultaneously. The next section explains one such approach, namely Benders' decomposition. The technique is a powerful algorithmic-based approach and its application to solve large instances of the facility location problem will be detailed.

22.4 Benders' decomposition with feasible subproblems

This section

This section presents the mathematical description of Benders' decomposition for the case with feasible subproblems. It is based on an abstract model that has been partitioned into an *easy* linear portion and a *difficult* nonlinear/integer portion. Once you understand the underlying decomposition theory plus the basic rules for writing dual linear programs described in the next section, you will be able to apply the Benders' decomposition approach to the facility location problem.

Initial problem $P(x, y)$

Consider the following minimization problem, which is referred to as $P(x, y)$:

Minimize:

$$c^T x + f(y)$$

Subject to:

$$\begin{aligned} Ax + F(y) &= b \\ x &\geq 0 \\ y &\in Y \end{aligned}$$

with $A \in \mathbb{R}^{m\times n}$, x and $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $y \in Y \subset \mathbb{R}^p$. Here, $f(y)$ and $F(y)$ may be nonlinear. Y can be a discrete or a continuous range.

Feasible subproblems $P(x|y)$

First, it is important to observe that for a fixed value of $y \in Y$ the problem becomes a linear program in terms of x. This is represented mathematically as $P(x|y)$. Next, it is assumed that $P(x|y)$ has a finite optimal solution x for every $y \in Y$. This may seem to be a rather restrictive assumption, but in most real-world applications this assumption is already met or else you can modify Y in such a way that the assumption becomes valid.

Equivalent reformulation $P_1(x, y)$

The expression for $P(x, y)$ can be written in terms of an equivalent nested minimization statement, $P_1(x, y)$:

$$\min_{y\in Y} \{f(y) + \min_x \{c^T x \mid Ax = b - F(y), \quad x \geq 0\}\}$$

Equivalent reformulation $P_2(u, y)$

This statement can be rewritten by substituting the dual formulation of the inner optimization problem (see Section 22.6), to get an equivalent formulation, $P_2(u, y)$:

$$\min_{y\in Y} \{f(y) + \max_u \{[b - F(y)]^T u \mid A^T u \leq c\}\}$$

Extreme point reformulation $P_3(u, y)$

The main advantage of the latter formulation is that the constraint set of the inner problem is independent of y. Furthermore, the optimal solution of the inner maximization problem is finite because of the explicit assumption that $P(x|y)$ has a finite optimal solution for every $y \in Y$. Such an optimal solution will always be at one of the *extreme points* $u \in U$. Therefore, the following equivalent formulation, $P_3(u, y)$, may be obtained:

$$\min_{y\in Y} \{f(y) + \max_{u\in U} [b - F(y)]^T u\}$$

This nested formulation of $P_3(x,y)$ can finally be re-written as a single minimization problem which is referred to as the *full master problem*, $P_4(y,m)$:

Full master $P_4(y,m)$

Minimize:

$$f(y) + m$$

Subject to:

$$\begin{aligned} [b - F(y)]^T u \le m \qquad & u \in U \\ y \in Y \qquad & \end{aligned}$$

In the full master problem it is important to observe that there is one constraint for each extreme point. It is true that there may be an enormous number in a problem of even moderate size. However, only a small fraction of the constraints will be binding in the optimal solution. This presents a natural setting for applying an iterative scheme in which a master problem begins with only a few (or no) constraints while new constraints are added as needed. This constraint generation technique is dual to the column generation scheme described in Chapter 20.

Solve $P(x,y)$ iteratively

From the full master problem, it is possible to define a *relaxed master problem* $M(y,m)$ which considers a subset B of the constraints U.

The relaxed master problem $M(y,m)$

Minimize:

$$f(y) + m$$

Subject to:

$$\begin{aligned} [b - F(y)]^T u \le m \qquad & u \in B \\ y \in Y \qquad & \end{aligned}$$

where B is initially empty and m is initially 0.

The Benders subproblem is $S(u|y)$, which solves for an extreme point u given a fixed value of $y \in Y$, can be written as the following maximization problem:

The subproblem $S(u|y)$

Maximize:

$$[b - F(y)]^T u$$

Subject to:

$$A^T u \le c$$

with $u \in \mathbb{R}^m$. $S(u|y)$ has a finite optimal solution, because of the assumption that $P(x|y)$ has a finite optimal solution for every $y \in Y$.

A flowchart of Benders' decomposition

Figure 22.2 presents a flowchart of Benders' decomposition algorithm for the case when all subproblems are feasible.

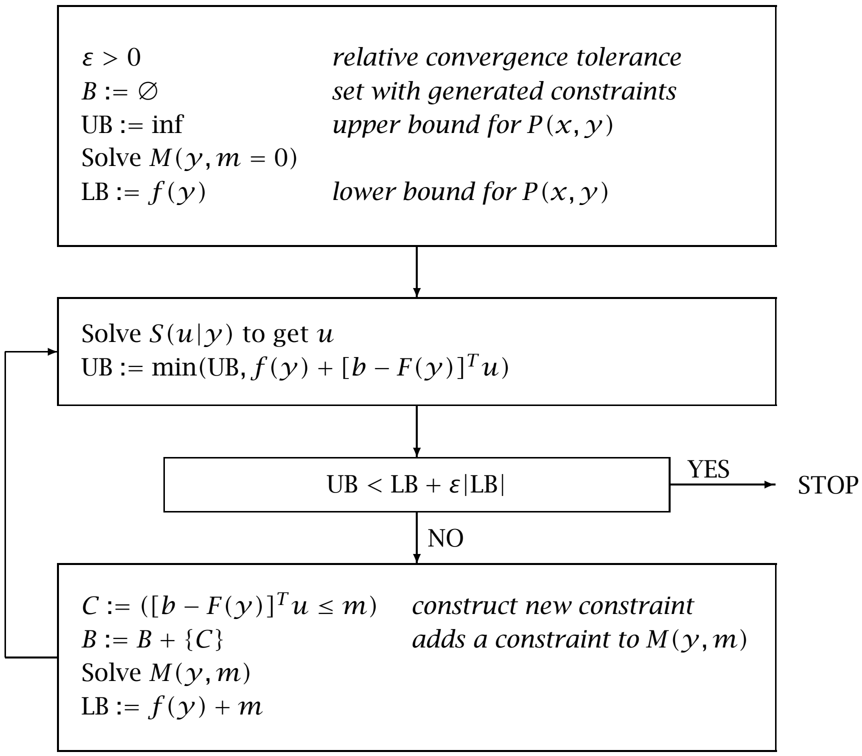

Figure 22.2: Benders' decomposition algorithm flowchart

The iterative process in words

Summarizing *Benders' decomposition* algorithm in words, the subproblem is solved for u given some initial $y \in Y$ determined by the master. Next, there is a simple test to determine whether a constraint involving u must be added to the master. If so, the master is solved to produce a new y as input to the subproblem which is solved again. This process continues until optimality (within a tolerance level) can be concluded.

Increasing lower bounds

Since B is a subset of U, the optimal value of the objective function of the relaxed master problem $M(y, m)$ is a lower bound on the optimal value of the objective function of the full master problem $P_4(y, m)$ and thus of $P(x, y)$. Each time a new constraint is added to the master, the optimal value of its objective function can only increase or stay the same.

Decreasing upper bounds

The optimal solution u plus the corresponding value of y of the subproblem $S(u|y)$, when substituted in the constraints of $P_3(u,y)$, produces an upper bound on the optimal value of $P_3(u,y)$ and thus of $P(x,y)$. The best upper bound found during the iterative process, can only decrease or stay the same.

Termination

As soon as the lower and upper bounds of $P(x,y)$ are sufficiently close, the iterative process can be terminated. In practice, you cannot expect the two bounds to be identical due to numerical differences when computing the lower and upper bounds. It is customary to set a relative sufficiently small tolerance typically of the order 1 to $\frac{1}{100}$ of a percent.

Premature termination

When the iterative process is terminated prematurely for whatever reason, the latest y-value is still feasible for $P(x|y)$. The current lower and upper bounds on $P(x,y)$ provide an indication on how far the latest solution (x,y) is removed from optimality.

22.5 Convergence of Benders' decomposition

Uniqueness assumption results in ...

For the moment, assume that for every iteration the extreme point u produced by solving the subproblem $S(u|y)$ is unique. Each such point will then result in the addition of a new constraint to the relaxed master problem.

... finite number of steps

As a result of the uniqueness assumption the iterative process will terminate in a finite number of steps. After all, there are only a finite number of extreme points. In the event that they have all been generated by solving $S(u|y)$ repeatedly, the resulting relaxed master problem $M(y,m)$ becomes equivalent to the full master problem $P_4(y,m)$ and thus the original problem $P(x,y)$.

... converging bounds

The sequence of relaxed master problems $M(y,m)$ produces a monotone sequence of lower bounds. In the event that after a finite number of steps the relaxed master problem $M(y,m)$ becomes the full master problem $P_4(y,m)$, the set B becomes equal to the set U. The corresponding lower bound is then equal to the original objective function value of $P_4(y,m)$ and thus of the original problem $P(x,y)$. At that moment both the optimal value of $f(y)$ and the optimal u-value of the subproblem make up the original solution of $P_3(u,y)$. As a result, the upper bound is then equal to the optimal objective function value of $P_3(u,y)$ and thus of the original problem $P(x,y)$.

... overall convergence

Based on the uniqueness assumption and the resulting convergence of lower and upper bounds in a finite number of steps, the overall convergence of Benders' decomposition is guaranteed. Termination takes place whenever the lower bound on the objective function value of $P(x,y)$ is equal to its upper bound.

Is uniqueness assumption true?

The entire convergence argument of the Benders' decomposition algorithm so far hinges on the uniqueness assumption, i.e. the assumption that all extreme points u produced by solving the subproblems $S(u|y)$ during the iterative process are unique as long as termination has not been reached. Assume that at some point during the iteration process *prior to termination*, the u-value produced by solving the subproblem is not unique. In this case, the lower bound is still strictly less than the upper bound, but the relaxed master problem will produce the same y value as in the previous iteration. The Benders' algorithm then cycles from here on and produces the same solution tuple $(\hat{u}, \hat{y})$ each solution. This tuple has the property that the current lower bound LB (obtained from the relaxed master problem) is

$$f(\hat{y}) + m$$

and that the current upper bound UB (with $\hat{u}$ obtained from the subproblem and substituted in the objective function of $P_3(\hat{u}, \hat{y})$) is at least as large as

$$f(\hat{y}) + [b - F(\hat{y})]^{\mathrm{T}}\hat{u}$$

Note that

$$m \geq [b - F(\hat{y})]^{\mathrm{T}}u, \qquad \text{for } u \in B$$

by construction. Note also that $\hat{u}$ is already in B due to cycling, which implies that

$$m \geq [b - F(\hat{y})]^{\mathrm{T}}\hat{u}$$

Combining the above leads to

$$\mathrm{LB} = f(\hat{y}) + m \geq f(\hat{y}) + [b - F(\hat{y})]^{\mathrm{T}}\hat{u} \geq \mathrm{UB}$$

which is a contradiction to the fact that prior to termination the lower bound is strictly less than the upper bound. This shows that the uniqueness assumption is true and that Benders' decomposition with feasible subproblems as described in this chapter will always converge.

22.6 Formulating dual models

This section

In order to apply the Benders' decomposition scheme, it is necessary to formulate the dual of $P(x|y)$. The rules for this step can be found in books on linear programming. For purposes of completeness and later reference these rules are summarized in this section in the form of typical examples.

Dual of a minimization problem

If a primal problem is stated as:

Minimize:

$$c_1x_1 + c_2x_2$$

Subject to:

$$
\begin{aligned}
& a_{11}x_1 + a_{12}x_2 \geq b_1 \\
& a_{21}x_1 + a_{22}x_2 = b_2 \\
& a_{31}x_1 + a_{32}x_2 \leq b_3 \\
& x_1 \geq 0, x_2 \geq 0
\end{aligned}
$$

then its dual problem is:

Maximize:

$$u_1 b_1 + u_2 b_2 + u_3 b_3$$

Subject to:

$$
\begin{aligned}
& a_{11}u_1 + a_{21}u_2 + a_{31}u_3 \leq c_1 \\
& a_{12}u_1 + a_{22}u_2 + a_{32}u_3 \leq c_2 \\
& u_1 \geq 0,\ u_2 \text{ free },\ u_3 \leq 0
\end{aligned}
$$

Dual of a maximization problem

If a primal problem is stated as:

Maximize:

$$c_1 x_1 + c_2 x_2$$

Subject to:

$$
\begin{aligned}
& a_{11}x_1 + a_{12}x_2 \geq b_1 \\
& a_{21}x_1 + a_{22}x_2 = b_2 \\
& a_{31}x_1 + a_{32}x_2 \leq b_3 \\
& x_1 \geq 0, x_2 \geq 0
\end{aligned}
$$

then its dual problem is:

Minimize:

$$u_1 b_1 + u_2 b_2 + u_3 b_3$$

Subject to:

$$
\begin{aligned}
& a_{11}u_1 + a_{21}u_2 + a_{31}u_3 \geq c_1 \\
& a_{12}u_1 + a_{22}u_2 + a_{32}u_3 \geq c_2 \\
& u_1 \leq 0, u_2 \text{ free }, u_3 \geq 0
\end{aligned}
$$

22.7 Application of Benders' decomposition

This section

Using the decomposition and duality theory of the previous sections, the facility location example can now be divided into a master problem and a dual subproblem.

Facility location problem $P(x, v, y)$

The original facility location problem can be summarized as follows.

Minimize:

$$\sum_{cpdz} K_{cpdz} x_{cpdz} + \sum_{d} \{F_d v_d + R_d \sum_{cz} D_{cz} y_{dz}\}$$

Subject to:

$$\sum_{dz} x_{cpdz} \leq S_{cp} \qquad \forall (c, p) \qquad (1)$$

$$\sum_{p} x_{cpdz} = D_{cz} y_{dz} \qquad \forall (c, d, z) \qquad (2)$$

$$\sum_{d} y_{dz} = 1 \qquad \forall z \qquad (3)$$

$$\underline{M}_d v_d \leq \sum_{cz} D_{cz} y_{dz} \leq \overline{M}_d v_d \qquad \forall d \qquad (4)$$

$$v_d, y_{dz} \in \{0, 1\} \qquad (5)$$

$$x_{cpdz} \geq 0 \qquad (6)$$

How to decompose

To conduct a Benders' decomposition, it is first necessary to divide the variables and constraints into two groups. The binary variables v_d and y_{dz}, together with the constraints (3), (4) and (5) represent the set Y. The continuous variable x_{cpdz}, together with the constraints (1), (2) and (6) represent the linear part to be dualized. As detailed soon, σ and π are two dual variables introduced for constraints (1) and (2).

Model notation

In the description of the Benders' decomposition algorithm, the various models are indicated by $P(x, y)$, $M(y, m)$ and $S(u|y)$. The correspondence between the variables used here and those defined for the facility location problem is as follows.

- x used previously is equivalent to x used above,
- y used previously is equivalent to v and y used above, and
- u used previously is equivalent to σ and π used above.

As a result, the equivalent model indicators become $P(x, v, y)$, $M(v, y, m)$ and $S(\sigma, \pi | v, y)$, respectively.

Initial master model $M(v, y, m = 0)$

The initial master model $M(v, y, m = 0)$ can be stated as follows.

Minimize:

$$\sum_d \{F_d v_d + R_d \sum_{cz} D_{cz} y_{dz}\}$$

Subject to:

$$\sum_d y_{dz} = 1 \qquad \forall z \qquad (3)$$

$$\underline{M}_d v_d \le \sum_{cz} D_{cz} y_{dz} \le \overline{M}_d v_d \qquad \forall d \qquad (4)$$

$$v_d, y_{dz} \in \{0, 1\} \qquad (5)$$

Note that the initial master model does not yet contain any Benders' cuts (i.e. $m = 0$) and that it corresponds to solving:

$$\min_{y \in Y} f(y)$$

previously introduced in the original Benders' decomposition algorithm.

Problem to be dualized

The problem to be dualized, namely the equivalent of the inner optimization problem in $P_1(x, y)$ of Section 22.4, can now be stated as follows.

Minimize:

$$\sum_{cpdz} K_{cpdz} x_{cpdz}$$

Subject to:

$$\sum_{dz} x_{cpdz} \le S_{cp} \qquad \forall (c, p) \mid S_{cp} > 0 \qquad (1)$$

$$\sum_p x_{cpdz} = D_{cz} y_{dz} \qquad \forall (c, d, z) \mid y_{dz} = 1 \qquad (2)$$

$$x_{cpdz} \ge 0 \qquad (6)$$

Resulting subproblem $S(\sigma, \pi | v, y)$

By introducing two dual variables σ_{cp} and π_{cdz} corresponding to the two constraints (1) and (2) respectively, the dual formulation of the problem from the previous paragraph can be written in accordance with the rules mentioned in Section 22.6. Note that the dual variable σ_{cp} is only defined when $S_{cp} > 0$, and the dual variable π_{cdz} is only defined when $y_{dz} = 1$.

Maximize:

$$\sum_{cp} \sigma_{cp} S_{cp} + \sum_{cdz} \pi_{cdz} D_{cz} y_{dz}$$

Subject to:

$$\sigma_{cp} + \pi_{cdz} \le K_{cpdz} \qquad \forall (c,p,d,z)$$

$$\sigma_{cp} \le 0$$

$$\pi_{cdz} \text{ free}$$

Subproblem is always feasible

The question arises whether the above subproblem $S(\sigma, \pi | v, y)$ is always feasible for any solution of the initial master problem $M(v, y, m = 0)$. If this is the case, then the Benders' decomposition algorithm described in this chapter is applicable to the original facility location problem. Note that,

$$\sum_{p} S_{cp} \ge \sum_{z} D_{cz} \qquad \forall c$$

is a natural necessary requirement, and that

$$\sum_{z} D_{cz} \equiv \sum_{dz} D_{cz} y_{dz} \qquad \forall c$$

is an identity, because $\sum_d y_{dz} = 1$. These together imply that there is enough supply in the system to meet the demand no matter which distribution center d is used to serve a particular customer zone z.

Benders' cut to be added

The Benders' cut to be added each iteration is directly derived from the objective function of the above subproblem $S(\sigma, \pi | v, y)$ evaluated at the original solution (σ, π). This new constraint is of the form

$$\sum_{cp} \sigma_{cp} S_{cp} + \sum_{cdz} \pi_{cdz} D_{cz} y_{dz} \le m$$

where (σ_{cp}, π_{cdz}) are parameters and y_{dz} and m are unknowns.

Resulting master problem $M(v, y, m)$

By adding the Bender's cuts to the initial master problem $M(v, y, m = 0)$, the following regular master problem $M(v, y, m)$ can be obtained after introducing the set B of Benders' cuts generated so far. Note that the optimal dual variables σ and π have been given an extra index $b \in B$ for each Benders' cut.

Minimize:

$$\sum_{d} \{F_d v_d + R_d \sum_{cz} D_{cz} y_{dz}\} + m$$

Subject to:

$$\begin{aligned}
&\sum_{d} y_{dz} = 1 && \forall z \\
&\underline{M}_d v_d \leq \sum_{cz} D_{cz} y_{dz} \leq \overline{M} v_d && \forall d \\
&\sum_{cp} \sigma_{bcp} S_{cp} + \sum_{cdz} \pi_{bcdz} D_{cz} y_{dz} \leq m && \forall b \\
&v_d, y_{dz} \in \{0, 1\}
\end{aligned}$$

All ingredients available

At this point all relevant components of the original facility location problem to be used inside the Benders' decomposition algorithm have been presented. These components, together with the flowchart in Section 22.4, form all necessary ingredients to implement the decomposition algorithm in AIMMS.

22.8 Computational considerations

This section

The presentation thus far has mainly focussed on the theory of Benders' decomposition and its application to the particular facility location problem. Implementation issues have barely been considered. This section touches on two of these issues, namely *subproblem splitting* aimed at preserving primary memory, and *use of first-found integer solution* aimed at diminishing computational time. Whether or not these aims are reached, depends strongly on the data associated with each particular model instance.

Splitting subproblem $S(\sigma, \pi | v, y)$

The dual subproblem can be broken up and solved independently for each commodity c. This gives the advantage that the LP model is divided into $|c|$ smaller models which can all be solved independently. This can reduce memory usage, which is especially true when K_{cpdz} is stored on disk or tape. In this case, it is sufficient to read data into primary memory for only one c at the time. For each fixed commodity $\overline{c}$ the problem then becomes

Maximize:

$$\sum_{p} \sigma_{\overline{c}p} S_{\overline{c}p} + \sum_{dz} \pi_{\overline{c}dz} D_{\overline{c}z} y_{dz}$$

Subject to:

$$\sigma_{\overline{c}p} + \pi_{\overline{c}dz} \leq K_{\overline{c}pdz} \qquad \forall (p, d, z)$$

$$\sigma_{\overline{c}p} \leq 0$$

$$\pi_{\overline{c}dz} \quad \textit{free}$$

Joining solutions

After solving the above problem for each fixed commodity $\overline{c}$, the objective function value of the overall problem $S(\sigma, \pi | v, y)$, is then the sum over all commodities of the individual objective function values.

Use first integer solution

The Benders' cut is derived after finding the optimal integer solution to the master problem. In practice, finding optimal integer solutions can be extremely time consuming as most solution algorithms have difficulty proving that a perceived optimal solution is indeed the optimal solution. Finding a first integer solution is in general easier than finding an optimal integer solution. That is why an alternative implementation of Benders' decomposition can be proposed to take advantage of such a first integer solution.

Extra cut required

Consider the objective function value of the relaxed master problem for the first integer solution found. This value is not necessarily optimal and therefore cannot be considered as a valid lower bound for the original problem. Nevertheless, it will be treated as such. Now, the Benders' algorithm can terminate prematurely whenever this fake lower bound exceeds the current upper bound. In order to avoid premature termination in the presence of these fake lower bounds, the following constraint should be added:

$$f(y) + m \leq \mathrm{UB} - \epsilon, \qquad \epsilon \geq 0, \text{small}$$

This constraint makes sure that any fake lower bound resulting from the use of a first integer solution of the master problem cannot be greater than or equal to the current upper bound, and thus will never cause unwanted premature termination.

Convergence with first integer solutions

New Benders' cuts are added every iteration. Their presence together with the above constraint on the fake lower bound will eventually result in an empty integer solution space. This happens when the generated Benders' cuts are such that the true lower bound is greater than or equal to the upper bound minus ϵ. From original Benders' algorithm it follows that convergence has occurred and that the current upper bound provides equals the optimal objective function value. Thus, the alternative approach based on first integer solutions terminates when the modified master problem becomes infeasible and no longer produces a first integer solution.

22.9 A worked example

This section

In this section you will find a small and somewhat artificial example to illustrate the computational results of applying the Benders' decomposition approach to the facility location problem described in this chapter.

Network layout

In this example there are two production plants, three customer zones, and seven potential sites for distribution centers. Their coordinates are presented in Table 22.1, and are used to determine the transport cost figures as well as the map in Figure 22.3.

City	Type	X-coord.	Y-coord.
Arnhem	Production plant	191	444
Rotterdam	Production plant	92	436
Amsterdam	Distribution center	121	488
The Hague	Distribution center	79	454
Utrecht	Distribution center	136	455
Gouda	Distribution center	108	447
Amersfoort	Distribution center	155	464
Zwolle	Distribution center	203	503
Nijmegen	Distribution center	187	427
Maastricht	Customer zone	175	318
Haarlem	Customer zone	103	489
Groningen	Customer zone	233	582

Table 22.1: Considered cities and their coordinates

Commodities

A total of two commodities are considered. The corresponding supply and demand data for the production plants and the customer zones are provided in Table 22.2, and are specified without units.

	Product A		Product B	
City	S_{cp}	D_{cz}	S_{cp}	D_{cz}
Arnhem	18		18	
Rotterdam	15		40	
Maastricht		8		9
Haarlem		9		10
Groningen		7		11

Table 22.2: Supply and demand data

Throughput data

For each distribution center the minimal and maximal throughput data, together with the associated throughput cost figures, are displayed in Table 22.3

d	$\underline{M}_d$	$\overline{M}_d$	R_d	F_d
Amsterdam	2	20	5.0	180
The Hague		20	7.0	130
Utrecht		14	3.0	60
Gouda		20	5.5	150
Amersfoort		21	6.0	140
Zwolle		17	7.0	150
Nijmegen		16	3.5	100

Table 22.3: Distribution throughput data

Cost determination

The transport cost values K_{cpdz} are based on distance according to the following formula:

$$\begin{aligned} K_{cpdz} \quad = \quad & \Big(\sqrt{(X_p - X_d)^2 + (Y_p - Y_d)^2} + \\ & \sqrt{(X_z - X_d)^2 + (Y_z - Y_d)^2}\Big)/100 \end{aligned}$$

Note that these cost values are the same for both products, and could have been written as K_{pdz}.

Solution

In the optimal solution 'The Hague', 'Gouda' and 'Amersfoort' are selected as distribution centers. 'Haarlem' is served from 'The Hague', 'Maastricht' is served from 'Gouda', and 'Groningen' is served from 'Amersfoort'. The optimal flows through the network are presented in Table 22.4. The graphical representation of the optimal flows is displayed in Figure 22.3. The corresponding total production and transport costs amount to 828.9408. This optimal solution was obtained with an optimality tolerance of $\varepsilon = 0.0001$ and a total of 15 Benders' cuts.

c	p	d	z	x_{cpdz}
product A	Arnhem	Gouda	Maastricht	2
product A	Arnhem	Amersfoort	Groningen	7
product A	Rotterdam	The Hague	Haarlem	9
product A	Rotterdam	Gouda	Maastricht	6
product B	Arnhem	Amersfoort	Groningen	11
product B	Rotterdam	The Hague	Haarlem	10
product B	Rotterdam	Gouda	Maastricht	9

Table 22.4: Optimal flows

Figure 22.3: Optimal commodity flows

Final comments

The computational performance of the Benders' decomposition method in terms of solution times is inferior when compared to solving the model as a single mathematical program. Nevertheless, the decomposition method provides a solution approach for extremely large model instances, with the added advantage that a feasible solution is available at any iteration. A premature termination of the algorithm (for instance, when the upper bound remains nearly constant) may very well lead to a good near-optimal solution. This observation applies to the data instance provided in this section.

22.10 Summary

In this chapter a facility location problem was translated into a mixed-integer mathematical program. A Benders' decomposition approach was introduced to support the solution of large model instances. The theory underlying the decomposition method with feasible subproblems was first introduced, and subsequently applied to the facility location model. A flowchart illustrating the general Benders' decomposition algorithm was presented as the basis for an implementation in Aimms. A small data set was provided for computational purposes.

Exercises

22.1 Implement the facility location model described in Section 22.2 using the example data presented in Tables 22.1, 22.2 and 22.3.

22.2 Implement the same model by using the Benders' decomposition approach described in Section 22.4 and further applied in Section 22.7. Verify whether the solution found with Aimms is the same as the one found without applying Benders' decomposition.

22.3 Implement the Benders' decomposition approach based on using the first integer solution found during the solution of the relaxed master model as described in Section 22.8. In Aimms you need to set the option `Maximal_number_of_integer_solutions` to 1 in order for the MIP solver to stop after it has found the first feasible integer solution.

Bibliography

[Ge74] A.M. Geoffrion and G.W. Graves, *Multicommodity distribution system design by benders decomposition*, Management Science **20** (1974), no. 5, 822–844.