Lecture 8 - Benders Decomposition

Lan Peng, Ph.D.

School of Management, Shanghai University, Shanghai, China

"Learning from one's mistakes."

1 Benders Decomposition

Benders Decomposition is usually applied for those problems that are easy to solve after fixing some of the variables. The general idea is to divide the problem into a master problem containing all complicated variables and subproblems containing simple variables. For example, given an MILP model as follows

$$\begin{aligned}
\text{(MILP)} \quad \min \quad & f^\top y + c^\top x \\
\text{s.t.} \quad & Ay \geq b \\
& By + Dx \geq d \\
& x \geq 0 \\
& \mathbf{y} \in \mathbb{Y} \subseteq \mathbb{Z}_+
\end{aligned}$$

In the formulation, decision variables x are separable, thus, for a given $\bar{y} \in \mathbb{Y}$, a subproblem is defined a follows

$$\begin{aligned}
\text{(Sub)} \quad \min \quad & f^\top \bar{y} + c^\top x \\
\text{s.t.} \quad & B\bar{y} + Dx \geq d \\
& x \geq 0
\end{aligned}$$

In which $f^\top \bar{y}$ is a constant for given $\bar{y}$, rewrite (Sub) as follows

$$\text{(Sub)} \quad \eta(\bar{y}) = \min_x \{c^\top x | Dx \geq d - B\bar{y}, x \geq 0\}$$

Notice that $\bar{y}$ remains in the constraints of (Sub), every time when a new $\bar{y}$ is given, the feasible region changes, which makes it difficult to keep track of the model. To facilitate our analysis, we take the dual of (Sub), which maintains the same feasible region every time when a new $\bar{y}$ is given. The dual of the subproblem is defined as follows

$$\text{(Dual-Sub)} \quad \eta(\bar{y}) = \max_\pi \{\pi^\top (d - B\bar{y}) | \pi^\top D \leq c, \pi \geq 0\}$$

Then, the equivalent formulation of the original problem is

$$
\begin{aligned}
\text{(Master)} \quad \min \quad & f^\top y + \eta \\
& Ay \geq b \\
& \eta \geq \max_{\pi}\{\pi^\top (d - By) | \pi^\top D \leq c, \pi \geq 0\} \\
& \mathbf{y} \in \mathbb{Y} \subseteq \mathbb{Z}_+
\end{aligned}
$$

Apply the Minkowski-Weyl Theorem to the (Dual-Sub), the feasible region $\{\pi|\pi^\top D \leq c, \pi \geq 0\}$ can be represented by a set of extreme points (e.p.) and a set of extreme directions (e.d.). Thus, let

- $\pi_e \in V$ be the set of all extreme points, and
- $\pi_d \in D$ be the set of all extreme directions

The original problem can be further rewritten as follows

$$
\begin{aligned}
\text{(Master)} \quad \min \quad & f^\top y + \eta \\
& Ay \geq b \\
& \eta \geq \pi_e^\top (d - By), \quad \forall \pi_e \in V \\
& 0 \geq \pi_d^\top (d - By), \quad \forall \pi_d \in D \\
& \eta \geq 0 \\
& \mathbf{y} \in \mathbb{Y} \subseteq \mathbb{Z}_+
\end{aligned}
$$

In which, $\eta \geq \pi_e^\top (d - By)$ are referred as optimality cuts, and $0 \geq \pi_d^\top (d - By)$ are referred as feasibility cuts. However, the set V and D are exponentially large, it is intractable if we have all feasibility cuts and optimality cuts added into the model. To reduce the computational burden, a Branch-and-Cut scheme algorithm framework is more widely implemented by researchers.

Define the initial restricted master problem which only contains the integer variables $\mathbf{y}$ as follows

$$
\begin{aligned}
\text{(RMP)} \quad \min \quad & f^\top y + \eta \\
& Ay \geq b \\
& \eta \geq 0 \\
& \mathbf{y} \in \mathbb{Y} \subseteq \mathbb{Z}_+
\end{aligned}
$$

Each iteration, an optimized $\bar{y}$ for the (RMP) will be given to the subproblem, which will create a (or multiple) optimality cut(s) or feasibility cut(s). Those cuts will be added into the model in a lazy manner. Each time when an optimality cut is found and added, the lower bound of the master problem will be lifted, otherwise, if a feasibility cut is found and added, the upper bound will be updated. The iteration terminates when the lower bound meet with the upper bound within an acceptable error.

2 A Uncapacitated Facilities Location Problem

2.1 Formulation

Consider the following facility location problem, where m is the number of potential facilities, n is the number of customers, and $\mathbb{Y}$ is the set of feasible plans of facility location plans, where $\mathbf{y} \in \mathbb{Y} \subseteq \{0,1\}^m$. c_{ij} is the cost for customer i to be assigned to facility j, d_j is the cost for opening facility j. The formulation is as following

Table 1: Sets and Parameters

Notations	**Description**
m	Number of potential facilities
n	Number of customers
$F = \{1, 2, \dots, m\}$	Set of potential facilities
$C = \{1, 2, \dots, n\}$	Set of customers
d_j	The cost of constructing facility j, where $j \in$.
c_{ij}	The cost for customer i to receive service from facility j, where $i \in C$ and $j \in F$.

Table 2: Decision variables

Notations	**Description**
y_j	Binary variables, takes 1 if facility j is decided to be constructed, where $j \in F$.
x_{ij}	Continuous variables, the percentage of demand for customer i to be fulfilled by facilities j, where $i \in C$ and $j \in F$.

$$
\begin{aligned}
\text{(FLP)} \quad \min \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{j=1}^{m} d_j y_j \\
\text{s.t.} \quad & \sum_{j=1}^{m} x_{ij} \geq 1, \quad \forall i \in C && (1) \\
& x_{ij} \leq y_j, \quad \forall i \in C, j \in F && (2) \\
& x_{ij} \geq 0, \quad \forall i \in C, j \in F && (3) \\
& y_j \in \{0, 1\}, \quad \forall j \in F && (4)
\end{aligned}
$$

In the formulation (FLP),

- Constraints (1) indicate that for each customer i, its request needs to be fulfilled.
- Constraints (2) indicate that a facility j can only be able to serve customer i if the facility is opened.
- Constraints (3) is the nonnegative constraints for $\mathbf{x}$.
- Constraints (4) defines $\mathbf{y}$ as binary variables.

2.2 Solution Approach

If y are fixed, i.e., $y = \bar{y} \in \mathbb{Y}$, the rest of the formulation will become an LP model with x_{ij} as the nonnegative decision variables.

$$
\begin{aligned}
\text{(Sub)} \quad \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{j=1}^{m} d_j\bar{y}_j \\
\text{s.t.} \quad & \sum_{j=1}^{m} x_{ij} \geq 1, \quad \forall i \in C \\
& x_{ij} \leq \bar{y}_j, \quad \forall i \in C, j \in F \\
& x_{ij} \geq 0, \quad \forall i \in C, j \in F
\end{aligned}
$$

If we take the dual of this new LP, we get

$$
\begin{aligned}
\text{(Dual-Sub)} \quad \max \quad & \sum_{i=1}^{n}(\lambda_i - \sum_{j=1}^{m} \bar{y}_j\pi_{ij}) + \sum_{j=1}^{m} d_j\bar{y}_j \\
\text{s.t.} \quad & \lambda_i - \pi_{ij} \leq c_{ij} \quad \forall i \in C, j \in F \\
& \lambda_i \geq 0 \quad \forall i \in C \\
& \pi_{ij} \geq 0 \quad \forall i \in C, j \in F
\end{aligned}
$$

Now get rid of the constant in the objective function of (Sub) and (Dual-Sub), which is $\sum_{j=1}^{m} d_j\bar{\mathbf{y}}$, we can then define

$$
\begin{aligned}
\eta(\bar{\mathbf{y}}) &= \min_{\mathbf{x}}\{\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} | \sum_{j=1}^{m} x_{ij} \geq 1, x_{ij} \leq \bar{y}_j, x_{ij} \geq 0, i \in C, j \in F\} \\
&= \max_{\pi}\{\sum_{i=1}^{n}(\lambda_i - \sum_{j=1}^{m} \bar{y}_j\pi_{ij}) | \lambda_i - \pi_{ij} \leq c_{ij}, \lambda_i \geq 0, \pi_{ij} \geq 0 i \in C, j \in F\}
\end{aligned}
$$

The (FLP) model can be equivalently rewritten as follows

$$
\begin{aligned}
\text{(FLP)} \quad &= \min_{\bar{y}\in\mathbb{Y}}\{\sum_{j=1}^{m} d_j\bar{y}_j + \eta(\bar{\mathbf{y}})\} \\
&= \min_{\bar{y}\in\mathbb{Y}}\{\sum_{j=1}^{m} d_j\bar{y}_j + \max_{\pi}\{\sum_{i=1}^{n}(\lambda_i - \sum_{j=1}^{m} \bar{y}_j\pi_{ij}) | \lambda_i - \pi_{ij} \leq c_{ij}, \lambda_i \geq 0, \pi_{ij} \geq 0 i \in C, j \in F\}\}
\end{aligned}
$$

Notice that for index i, the customer, there is no constraint among them, which is logically to be true since in reality customers are not related to each other.

$$
\begin{array}{llllll}
(\text{Dual-Sub}) & v(\hat{\mathbf{y}}) = \max & (\lambda_1 - \sum_{j=1}^{m} \hat{y}_j \pi_{1j}) + & (\lambda_2 - \sum_{j=1}^{m} \hat{y}_j \pi_{2j}) + \ldots + & (\lambda_n - \sum_{j=1}^{m} \hat{y}_j \pi_{nj}) + \sum_{j=1}^{m} d_j \hat{y}_j & \\
& \text{s.t.} & \lambda_1 - \pi_{1j} \leq c_{1j} & & & \forall j \in \{1, 2, \ldots, m\} \\
& & & \lambda_2 - \pi_{2j} \leq c_{2j} & & \forall j \in \{1, 2, \ldots, m\} \\
& & \ldots & & & \\
& & & & \lambda_n - \pi_{nj} \leq c_{nj} & \forall j \in \{1, 2, \ldots, m\} \\
& & \lambda_1 \geq 0 & & & \\
& & & \lambda_i \geq 0 & & \\
& & \ldots & & & \\
& & & & \lambda_n \geq 0 & \\
& & \pi_{1j} \geq 0 & & & j \in \{1, 2, \ldots, m\} \\
& & & \pi_{2j} \geq 0 & & j \in \{1, 2, \ldots, m\} \\
& & \ldots & & & \\
& & & & \pi_{nj} \geq 0 & j \in \{1, 2, \ldots, m\}
\end{array}
$$

Therefore, the due of subproblem can be further decomposed by i, for each i,

$$
\begin{aligned}
(\text{Dual-Sub-}i) \quad \eta_i(\bar{\mathbf{y}}) = \max \quad & \lambda_i - \sum_{j=1}^{m} \bar{y}_j \pi_{ij} \\
\text{s.t.} \quad & \lambda_i - \pi_{ij} \leq c_{ij}, \quad \forall j \in F \\
& \lambda_i \geq 0, \quad \forall i \in C \\
& \pi_{ij} \geq 0, \quad \forall i \in C, j \in F
\end{aligned}
$$

The (Dual-Sub-i) can easily be solved as following

$$
\begin{aligned}
& \bar{\lambda_i} = \min\{c_{ij}, j \in \mathbf{O}\} \\
& \begin{cases} \bar{\pi_{ij}} = 0, \quad j \in \mathbf{O} \\ \bar{\pi_{ij}} = \max\{0, \bar{\lambda_i} - c_{ij}\}, \quad j \in \mathbf{C} \end{cases}
\end{aligned}
$$

In which $\mathbf{O}$ is the set of indices j where $\bar{y}_j = 1$ and $\mathbf{C}$ is the set of indices j where $\bar{y}_j = 0$. The optimal value of dual variables can be interpreted in terms of facilities location problem. $\bar{\lambda_i}$ is the cost of serving customer i when $y = \bar{y}$ and $\bar{\pi_{ij}}$ is the reduction in the cost of serving customer i when facility j is opened and $y_i = \bar{y_i}$.

After solving the (Dual-Sub-i), we can find the objective function value as well as the dual variable values for (Dual-Sub).

The restricted master problem initialized with no additional constraints, as follows

$$
(\text{RMP}) \quad \min \quad \sum_{j=1}^{m} d_j y_j
$$

$$\text{s.t.} \quad \mathbf{y} \in \mathbb{Y}$$

There is only one case where the (Dual-Sub) problem will be unbounded, which would be when $\bar{\mathbf{y}} = \mathbf{0}$. In that case, a feasibility cut will be added into the master problem, as follows

$$0 \geq \sum_{i=1}^{n} (\bar{\lambda_i} - \sum_{j=1}^{m} y_j \bar{\pi_{ij}}) \tag{5}$$

For other cases, as long as there is at least one facility opened, the Subproblem is always feasible, therefore, a feasibility cut will be added into the master problem, as follows

$$\eta \geq \sum_{i=1}^{n} (\bar{\lambda_i} - \sum_{j=1}^{m} y_j \bar{\pi_{ij}})$$

The iteration continues until no more cut can be added into the master problem and we derive the optimal solutions.

3 Pareto-optimality cut

3.1 Choosing procedure

Definition 3.1 (dominance, pareto-optimal)**.** We say cut

$$z \geq hy + u^1(b - Dy) \succ z \geq hy + u^2(b - Dy) \tag{6}$$

if

$$hy + u^1(b - Dy) \geq hy + u^2(b - Dy), \quad \forall y \tag{7}$$

with at least one strict inequality. Further, we call a cut **pareto optimal** if it is not dominated by any other cut.

Notice that a Benders cut is determined by the vector u, so we want to find which of the produced multiple solutions to the subproblem are not dominated by others (i.e. which vectors produce pareto optimal cuts).

Define a **core point** of $\mathbb{Y}$, $y^0 \in ri(\mathbb{Y}^c)$, which can be any point of the relative interior of convex hull of $\mathbb{Y}$.

Assume that we have $\hat{y}$, the current solution to the master problem, and $v(\hat{y})$ is the optimal value of the subproblem. To find a pareto optimal cut among the multiple solutions u of the dual of the subproblem $f(\hat{y})$, we solve the following linear program:

$$\begin{aligned} \max \quad & h^T y^0 + u(b - Dy^0) && (8) \\ \text{s.t.} \quad & u(b - D\hat{y}) = v(\hat{y}) - h^T \hat{y} && (9) \\ & uA \leq c^T && (10) \end{aligned}$$

Constraints (9) and (10) ensure that we are searching only through alternative optimal solutions for the current subproblem.

Note that y^0 can be any arbitrary core point of $\mathbb{Y}$. In other words, choosing different core point and solving that problem we obtain different pareto optimal cut, among which we cannot tell which one is better.

Let us see why is that true. Suppose that y^0 is a core point of $\mathbb{Y}$ and $U(\hat{y})$ is the set of optimal solutions to the subproblem:

$$\max\{h^T\hat{y} + u(b - D\hat{y})\} \tag{11}$$

Let u^0 solve the problem:

$$\max\{h^T y^0 + u(b - Dy^0)\} \tag{12}$$

Now we can prove by contradiction that u^0 is not dominated by any other cut from $U(\hat{y})$. Suppose there exists $\bar{u}$, that dominates u^0:

$$h^T y + \bar{u}(b - Dy) \geq h^T y + u^0(b - Dy) \quad \forall y \in \mathbb{Y} \tag{13}$$

taking any point w of the convex hull of $\mathbb{Y}^c$ (which can be represented as a convex combination of a finite number of points from $\mathbb{Y}$) we have:

$$h^T w + \bar{u}(b - Dw) \geq h^T w + u^0(b - Dw) \quad \forall w \in \mathbb{Y}^c \tag{14}$$

note that with $y = \hat{y}$ in 13, $\bar{u}$ must be an optimal solution to 11. But then from 13 and (12) we have:

$$h^T y + \bar{u}(b - Dy) = h^T y + u^0(b - Dy) \quad \forall y \in \mathbb{Y} \tag{15}$$

and since $\bar{u}$ dominated u^0, there exists at least one $\bar{y}$ such that:

$$h^T\bar{y} + \bar{u}(b - D\bar{y}) > h^T\bar{y} + u^0(b - D\bar{y}) \tag{16}$$

3.2 Pareto Optimal Cuts for Facility Location Problem

Consider the previous facility location problem. If y are fixed, i.e., $y = \hat{y} \in \mathbb{Y}$, the rest of the formulation will become a LP model with x_{ij} as the nonnegative decision variables which can further be separated as a set of subproblems for each i,

$$\begin{array}{lll} \text{(Dual-Sub-}i\text{)} & \max & \lambda_i - \sum_{j=i}^{m} \hat{y}_j \pi_{ij} \\ & \text{s.t.} & \lambda_i - \pi_{ij} \leq c_{ij} \quad \forall j \in \{1, 2, \ldots, m\} \\ & & \lambda_i \geq 0 \\ & & \pi_{ij} \geq 0 \qquad\qquad j \in \{1, 2, \ldots, m\} \end{array}$$

and each subproblem can be solved as following

$$\begin{array}{ll} \hat{\lambda}_i = \min\{c_{ij}, j \in \mathbf{O}\} & \\ \hat{\pi}_{ij} = 0 & \text{if } j \in \mathbf{O} \\ \hat{\pi}_{ij} = \max\{0, \hat{\lambda}_i - c_{ij}\} & \text{if } j \in \mathbf{C} \end{array}$$

then, the optimal objective function for the ith separated subproblem is $v_i(\hat{\mathbf{y}}) = \hat{\lambda}_i$. From our previous discussion, for each i, we can solve the following problem (Pareto-Sub-i) to get (part of) the Pareto-optimal cut $\mathbf{u}_i = (\lambda_i, \pi_{i1}, \pi_{i2}, \cdots, \pi_{im})$. (The Pareto-optimal cut is $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)$)

$$
\begin{aligned}
\text{(Pareto-Sub-}i\text{)} \quad \max \quad & \lambda_i - \sum_{j=i}^{m} y_j^0 \pi_{ij} \\
\text{s.t.} \quad & \lambda_i - \sum_{j=i}^{m} \hat{y}_j \pi_{ij} = \hat{\lambda}_i \\
& \lambda_i - \pi_{ij} \leq c_{ij} \qquad \forall j \in \{1, 2, \ldots, m\} \\
& \lambda_i \geq 0 \\
& \pi_{ij} \geq 0 \qquad j \in \{1, 2, \ldots, m\}
\end{aligned}
$$

where, as mentioned, $\mathbf{y}^0 \in ri(\mathbb{Y}^c)$. By observation, this problem can be solved as following. First, replace the objective function with the first equality constraint, then the formulation can be rewritten as

$$
\begin{aligned}
\text{(Pareto-Sub-}i\text{)} \quad \max \quad & \hat{\lambda}_i + \sum_{j=i}^{m} (\hat{y}_j - y_j^0) \pi_{ij} \\
\text{s.t.} \quad & \lambda_i - \pi_{ij} \leq c_{ij} \qquad \forall j \in \{1, 2, \ldots, m\} \\
& \lambda_i \geq 0 \\
& \pi_{ij} \geq 0 \qquad j \in \{1, 2, \ldots, m\}
\end{aligned}
$$

For those $j \in \mathbf{O}$, $\hat{y}_j = 1$, $\hat{y}_j - y_j^0 > 0$, for those π_{ij} we want to maximize them, however, we have $\lambda_i - \hat{\lambda}_i = \sum_{j \in \mathbf{O}} \pi_{ij}$, i.e., summation of all such π_{ij} is constrained, to maximize the objective function, we can choose the one where $\hat{y}_j - y_j^0$ is the most.

For those $j \in \mathbf{C}$, $\hat{y}_j = 0$, $\hat{y}_j - y_j^0 < 0$, for those π_{ij} we want to minimize them, therefore $\pi_{ij} = \max\{0, \lambda_i - c_{ij}\}$.

Therefore, the objective function value will be a function of λ_i, or $f(\lambda_i)$ as

$$
f(\lambda_i) = \hat{\lambda}_i + \max_j \{\hat{y}_j - y_j^0\}(\lambda_i - \hat{\lambda}_i) + \sum_{j \in \mathbf{C}} (\hat{y}_j - y_j^0) \max\{0, \lambda_i - c_{ij}\} \tag{17}
$$

This becomes a piece-wise linear function for λ_i, we can find the optimal value of λ_i easily, and once it is found, π_{ij} can be derived by $\max\{0, \lambda_i - c_{ij}\}$, therefore for a given i we can get $\mathbf{u}_i = (\lambda_i, \pi_{i1}, \pi_{i2}, \cdots, \pi_{im})$. Repeat this for n times, we can get the Pareto-optimal cut $\mathbf{u}$.