Lecture 20 Stochastic Programming and Benders Decomposition

Can Li

ChE 597: Computational Optimization
Purdue University

History

- Proposed by Jacques F. Benders, a Dutch mathematician and Emeritus Professor of Operations Research at the Eindhoven University of Technology.
- Benders, Jacques F. "Partitioning procedures for solving mixed-variables programming problems." Numerische mathematik 4.1 (1962): 238–252.
- Reinvented by the stochastic programming community as the L-shaped method.
  Van Slyke, R. M., & Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming. SIAM journal on applied mathematics, 17(4), 638-663.

Motivation

- Solve optimization problems with "almost" block diagonal structure.

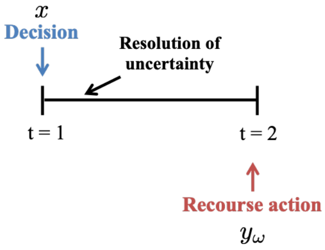
$$\begin{array}{cccc} x & y_{\omega 1} & y_{\omega 2} & y_{\omega 3} \\ \end{array}$$

$$\begin{bmatrix} T_{\omega 1} & W_{\omega 1} & & \\ T_{\omega 2} & & W_{\omega 2} & \\ T_{\omega 3} & & & W_{\omega 3} \end{bmatrix} \leq \begin{bmatrix} h_{\omega 1} \\ h_{\omega 2} \\ h_{\omega 3} \end{bmatrix}$$

- The rows correspond to constraints. The columns represent the variables in the constraint matrix. Once variables x are fixed, the rest of the problem is decomposable.
- This structure arise in two-stage stochastic programming.

Two-stage stochastic programming

Consider a decision maker who has to act in two consecutive stages. The first stage involves the choice of a decision vector x. Subsequently, some new information is obtained, and then, at the second stage, a new vector y of decisions is to be chosen. Regarding the nature of the obtained information, we assume that there are $|\Omega|$ possible scenarios, and that the true scenario is only revealed after x is chosen. We use ω to index the different scenarios. We index y_ω to represent the decisions taken in scenario ω.



Goal: choose x such that the expected cost in the second stage is minimized.

News Vendor problem

A news vendor goes to the publisher every morning and buys x newspapers at a price of c per paper. This number is usually bounded above by some limit u, representing either the news vendor's purchase power or a limit set by the publisher to each vendor. The vendor then walks along the streets to sell as many newspapers as possible at the selling price q. Any unsold newspaper can be returned to the publisher at a return price r, with $r < c$.
We are asked to help the news vendor decide how many newspapers to buy every morning. Demand for newspapers varies over days and is described by a discrete random variable d. d_ω describes the demand of newspaper in scenario ω. We assume there are $|\Omega|$ scenarios with equal probability.

Two stage stochastic program formulation

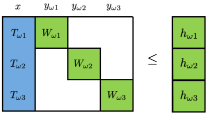
$$
\begin{aligned}
\min \quad & cx - \frac{1}{|\Omega|}\sum_{\omega\in\Omega}(qy_\omega + rz_\omega) \\
\text{s. t.} \quad & 0 \le x \le u \\
& \quad y_\omega \le d_\omega, \quad \forall \omega \in \Omega \\
& \quad y_\omega + z_\omega \le x, \quad \forall \omega \in \Omega \\
& \quad y_\omega, z_\omega \ge 0, \quad \forall \omega \in \Omega,
\end{aligned}
$$

where y_ω represents the number of newspapers sold to the customers, z_ω represents the number of newspapers return to the publisher in scenario ω.

Two-Stage Stochastic Linear Programs

$$
\begin{aligned}
\min z = & c^T x + \sum_{\omega \in \Omega} \tau_\omega \left[q_\omega^T y_\omega \right] \\
\text{s. t.} \quad & Ax = b, x \geq 0 \\
& T_\omega x + W_\omega y_\omega = h_\omega, \quad \forall \omega \in \Omega \\
& y_\omega \geq 0, \quad \forall \omega \in \Omega
\end{aligned}
$$

- $x \in \mathbb{R}^{n_1}$: first stage decisions, $y_\omega \in \mathbb{R}^{n_2}$ second stage decisions for scenario ω
- τ_ω probability of scenario ω
- $W_\omega \in \mathbb{R}^{m_2 \times n_2}$: *recourse matrix*
- $T_\omega \in \mathbb{R}^{m_2 \times n_1}$: *technology matrix*



Benders reformulation

$$\min z = c^T x + \sum_{\omega \in \Omega} \tau_\omega \eta_\omega(x)$$

where $\eta_\omega(x)$ is called the *recourse function*.

$$\begin{aligned} \eta_\omega(x) = \min_{y_\omega} \quad & q_\omega^T y_\omega \\ \text{s.t.} \quad & W_\omega y_\omega = h_\omega - T_\omega x \\ & y_\omega \geq 0 \end{aligned}$$

$$\begin{aligned} \eta_\omega(x) = \max_{p_\omega} \quad & p_\omega^T (h_\omega - T_\omega x) \\ \text{s.t.} \quad & p_\omega^T W_\omega \leq q_\omega^T \end{aligned}$$

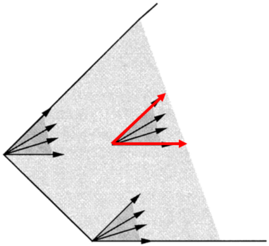
We assume the dual is feasible. This is always true when the primal is bounded.

Two cases can occur

- When the dual is unbounded, the primal is infeasible. There exists an extreme ray r_ω^j such that $(r_\omega^j)^T (h_\omega - T_\omega x) > 0$
- When the dual is feasible and bounded, the optimal solution is obtained at an extreme point of the polyhedron $\{p_\omega | p_\omega^T W_\omega \leq q_\omega^T\}$

Recall the definitions of extreme point and extreme ray

- Extreme point $\Leftrightarrow$ basic feasible solution: at a feasible solution $x^* \in \mathbb{R}^n$, n linearly independent constraints are active
- Extreme ray: Let $C \subseteq \mathbb{R}^n$ be given by $\{x \in \mathbb{R}^n : Ax \geq 0\}$ for some $A \in \mathbb{R}^{m \times n}$. Let $d \in C$ be nonzero. Let $A^{=}x = 0$ denote the subsystem of $Ax \geq 0$ consisting of all the inequalities active at d. Then d is an extreme ray of C if and only if $\operatorname{rank}(A^{=}) = n - 1$.



Benders reformulation

$$\eta_\omega(x) = \max_{p_\omega} \quad p_\omega^T (h_\omega - T_\omega x)$$
$$\text{s.t.} \;\; p_\omega^T W_\omega \leq q_\omega^T$$

Note that the feasible region of the dual does not depend on x.

- For the dual to be bounded (primal feasible), we must have

$$\left(r_\omega^j\right)^T (h_\omega - T_\omega x) \leq 0 \quad \forall j \in \mathcal{J}$$

  where r_ω^j, $j \in \mathcal{J}$ is the set of extreme rays of the polyhedral cone $\{r_\omega | r_\omega^T W_\omega \leq 0^T\}$

- If the dual is bounded (primal feasible), then we have

$$\eta_\omega(x) = \max_{i \in \mathcal{I}} \left(p_\omega^i\right)^T (h_\omega - T_\omega x)$$

  where p_ω^i, $i \in \mathcal{I}$ is the set of extreme points of the polyhedron $\{p_\omega | p_\omega^T W_\omega \leq q_\omega^T\}$.

Benders reformulation

$$
\begin{aligned}
\min \quad & c^T x + \sum_{\omega \in \Omega} \tau_\omega \eta_\omega \\
\text{s.t.} \quad & Ax = b, x \geq 0 \\
& \eta_\omega \geq \left(p_\omega^i\right)^T (h_\omega - T_\omega x) \quad \forall i \in \mathcal{I}, \omega \in \Omega \quad \text{optimality cuts} \\
& \left(r_\omega^j\right)^T (h_\omega - T_\omega x) \leq 0 \quad \forall j \in \mathcal{J}, \omega \in \Omega \quad \text{feasibility cuts}
\end{aligned}
$$

- A reformulation of the original problem in the space of x variables.
- The number of extreme rays and extreme points can be very large, making the Benders reformulation expensive to solve. However, only a small fraction of them are active in the Benders reformulation.
- The L-shaped algorithm (Benders decomposition) is an iterative algorithm to find the active cuts.

Benders master problem

The Benders master problem at iteration k,

$$
\begin{aligned}
\min \quad & c^T x + \sum_{\omega \in \Omega} \tau_\omega \eta_\omega \\
\text{s.t. } & Ax = b, x \geq 0 \\
& \eta_\omega \geq \left(p_\omega^i\right)^T (h_\omega - T_\omega x) \quad \forall i \in \mathcal{I}^k, \omega \in \Omega \quad \text{optimality cuts} \\
& \left(r_\omega^j\right)^T (h_\omega - T_\omega x) \leq 0 \quad \forall j \in \mathcal{J}^k, \omega \in \Omega \quad \text{feasibility cuts}
\end{aligned}
$$

- $\mathcal{I}^k$, $\mathcal{J}^k$ are the subsets of extreme points and extreme rays at iteration k. At $k = 0$, we can assume $\mathcal{I}^0 = \emptyset$, $\mathcal{J}^0 = \emptyset$
- The solution to the master problem at iteration k is denote as x^k.

Benders subproblem

After the Benders master problem is solved, we fix $x = x^k$ to check whether there are violated feasibility/optimality cuts by solving the following Benders subproblem,

$$\begin{aligned} \eta_\omega(x^k) = \min_{y_\omega} \quad & q_\omega^T y_\omega \\ \text{s.t.} \quad & W_\omega y_\omega = h_\omega - T_\omega x^k \\ & y_\omega \geq 0 \end{aligned}$$

$$\begin{aligned} \eta_\omega(x^k) = \max_{p_\omega} \quad & p_\omega^T \left(h_\omega - T_\omega x^k \right) \\ \text{s.t.} \quad & p_\omega^T W_\omega \leq q_\omega^T \end{aligned}$$

- These subproblem is decomposable by scenario.
- In practical implementation, usually we just formulate it as the primal form.

Benders subproblem to generate a feasibility cut

If the primal subproblem is infeasible, one can solve the following feasibility subproblem to generate a feasibility cut.

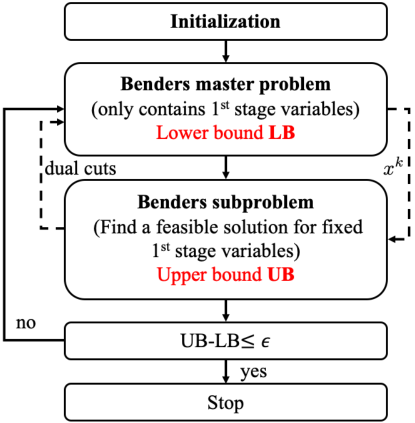
$$
\begin{aligned}
\min_{y_\omega, s_\omega^+, s_\omega^-} \quad & e^T s_\omega^+ + e^T s_\omega^+ \\
\text{s.t.} \quad & W_\omega y_\omega + s_\omega^+ - s_\omega^- = h_\omega - T_\omega x^k \\
& y_\omega \geq 0, s_\omega^+ \geq 0, s_\omega^- \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\max_{p_\omega} \quad & p_\omega^T \left(h_\omega - T_\omega x^k \right) \\
\text{s.t.} \quad & p_\omega^T W_\omega \leq 0^T \\
& -e \leq p_\omega \leq e
\end{aligned}
$$

where $s_\omega^+, s_\omega^- \in \mathbb{R}^{m_2}$ are the slack variables added to the primal equality constraints. The objective is to minimize the sum of all the slack variables.

The dual problem will return an extreme ray of the polyhedral cone $\{r_\omega | r_\omega^T W_\omega \leq 0^T\}$.

Benders decomposition



Geometric interpretation of Benders reformulation

Projection of a high dimensional polyhedron to a lower dimension.
The original polyhedron set

$$
\begin{aligned}
P = \Big\{(x, \eta_\omega, y_\omega, \omega \in \Omega) | \quad & \eta_\omega = q_\omega^T y_\omega \\
& Ax = b, x \geq 0 \\
& \quad T_\omega x + W_\omega y_\omega = h_\omega, \quad \forall \omega \in \Omega \\
& \quad y_\omega \geq 0, \quad \forall \omega \in \Omega \Big\}
\end{aligned}
$$

onto the x space.

$$
\begin{aligned}
\mathrm{Proj}_x(P) = \Big\{ x | Ax = b, x \geq 0 & \\
& \eta_\omega \geq \left(p_\omega^i\right)^T (h_\omega - T_\omega x) \quad \forall i \in \mathcal{I}, \omega \in \Omega \\
& \left(r_\omega^j\right)^T (h_\omega - T_\omega x) \leq 0 \quad \forall j \in \mathcal{J}, \omega \in \Omega \Big\}
\end{aligned}
$$

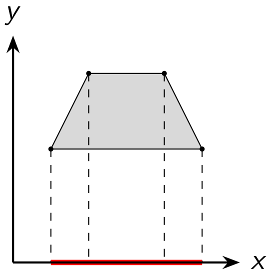
Projection of a Set

Definition

The projection of a set $S \subseteq \mathbb{R}^{n+m}$ onto the $x \in \mathbb{R}^n$ space is defined as the set of all x coordinates of points in S.

Mathematically, if $S \subseteq \mathbb{R}^{n+m}$, then the projection of S onto the x space, denoted as $\text{Proj}_x(S)$, is given by:

$$\text{Proj}_x(S) = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^m, (x, y) \in S\}$$



Projection of the trapezoid onto the x-axis

Projection of a polyhedron

$$Q := \{(u,x) \in \mathbb{R}^p \times \mathbb{R}^q : Au + Bx \leq b\}$$

where A, B and b have m rows, the projection of Q into $\mathbb{R}^q$, or into the x-space, is defined as

$$\mathrm{Proj}_x(Q) := \{x \in \mathbb{R}^q : \exists u \in \mathbb{R}^p : (u,x) \in Q\}.$$

Projection Theorem

$$\mathrm{Proj}_x(Q) = \{x \in \mathbb{R}^q : (vB)x \leq vb, v \in \mathrm{extr}\, W\}$$

where $\mathrm{extr}\, W$ denotes the extreme rays of the *projection cone*

$$W := \{v : vA = 0, v \geq 0\}$$

Remark It can be shown applying the projection theorem to P can reproduce the result of Benders reformulation.

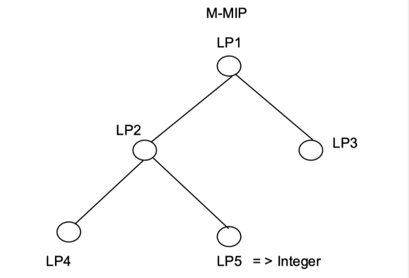
Benders decomposition with first stage (mixed)-integer variables

$$
\begin{aligned}
\min z = & c^T x + \sum_{\omega \in \Omega} \tau_\omega \left[q_\omega^T y_\omega \right] \\
\text{s. t.} \quad & Ax = b, x \in \{0,1\}^{n_1} \times \mathbb{R}_+^{n_1'} \\
& T_\omega x + W_\omega y_\omega = h_\omega, \quad \forall \omega \in \Omega \\
& y_\omega \geq 0, \quad \forall \omega \in \Omega
\end{aligned}
$$

- The master problem is an MILP.
- The subproblem is still an LP.
- The form of the feasibility and optimality cuts do not change.
- In each iteration, a branch and bound solver needs to be used to solve the MILP.

Single-tree Benders implementation

- Similar to the single-tree OA algorithm (QG), a single-tree algorithm can be used to accelerate Benders decomposition with mixed-integer first stage variables.
- The cuts are typically added using callback functions. We will have a full lecture on how to use callback functions.
- Check this blog for details
  Benders Decomposition Then and Now by Paul A. Rubin
  `https://orinanobworld.blogspot.com/2011/10/benders-decomposition-then-and-now.html`



Solve Benders subproblems
To add feasibility/optimality cuts

Integer L-shaped method

Consider two-stage stochastic program with pure binary first stage variables and (mixed)-integer second stage variables. In the literature, it is also called stochastic programs with *(mixed)-integer recourse*.

$$
\begin{aligned}
\min z = & c^T x + \sum_{\omega \in \Omega} \tau_\omega \left[q_\omega^T y_\omega \right] \\
\text{s. t.} \quad & Ax = b, x \in \{0,1\}^{n_1} \\
& T_\omega x + W_\omega y_\omega = h_\omega, \quad \forall \omega \in \Omega \\
& y_\omega \in \{0,1\}^{n_2} \times \mathbb{R}_+^{n_2'}, \quad \forall \omega \in \Omega
\end{aligned}
$$

The master problem is an IP. The subproblem is MILP. We cannot generate cuts from directly MILP (strong duality no longer holds). How to generate valid Benders cuts?

LP relaxation-based cuts

Generate cuts from the LP relaxation of the second-stage problem.

$$\begin{aligned} \eta_\omega(x^k) = \min_{y_\omega} \quad & q_\omega^T y_\omega \\ \text{s.t.} \quad & W_\omega y_\omega = h_\omega - T_\omega x^k \\ & y_\omega \in [0,1]^{n_2} \times \mathbb{R}_+^{n_2'} \end{aligned}$$

- These cuts are valid since the constraints in the LP relaxation are valid for the original integer recourse problem.
- However, these cuts along does not guarantee convergence due to the integrality gap.

Integer cuts

Solve the subproblem with the integrality constraints. Typically, one would add all the possible LP-relaxation-based cuts before starting to add integer cuts. We denote the iteration number of adding the integer cuts by r.

$$\begin{aligned} \eta_\omega(x^r) = \min_{y_\omega} \quad & q_\omega^T y_\omega \\ \text{s.t.} \quad & W_\omega y_\omega = h_\omega - T_\omega x^r \\ & y_\omega \in \{0,1\}^{n_2} \times \mathbb{R}_+^{n_2'} \end{aligned}$$

Let the optimal objective be η_ω^r. Define

$$B_r = \{i | i \in 1, \ldots, n_1, x_i^r = 1\} \quad N_r = \{i | i \in 1, \ldots, n_1, x_i^r = 0\}$$

Then the following integer cut is valid

$$\eta_\omega \geq (\eta_\omega^r - L) \left(\sum_{i \in B_r} x_i - \sum_{i \in N_r} x_i \right) - (\eta_\omega^r - L)(|B_r| - 1) + L$$

where L is a trivial lower bound of η_ω.

Proof of validity

$$\eta_\omega \geq (\eta_\omega^r - L)\left(\sum_{i\in B_r} x_i - \sum_{i\in N_r} x_i\right) - (\eta_\omega^r - L)\left(|B_r| - 1\right) + L$$

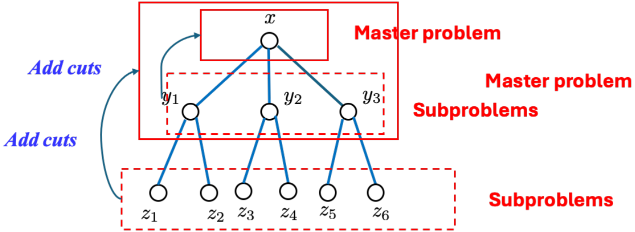
where L is a trivial lower bound of η_ω.

Proof. The quantity $\sum_{i\in B_r} x_i - \sum_{i\in N_r} x_i$ is always less than or equal to $|B_r|$. It takes the value $|B_r|$ only when x is the r^{th} feasible solution. Now, when $\sum_{i\in B_r} x_i - \sum_{i\in N_r} x_i$ is equal to $|B_r|$, the right-hand side takes the value η_ω^r, and otherwise the right-hand side takes a value less than or equal to L. A set of valid cuts is therefore obtained by imposing one such constraint for each first-stage feasible solution.

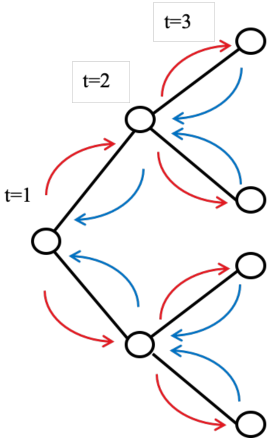
Remark the integer cut usually cuts off one solution at a time and can be ineffective.

Nested Benders decomposition

Idea: recursively apply Benders decomposition to problems with a multi-level tree structure.



Nested Benders decomposition



Forward Simulation works as nested Benders **master problems**:

- Finds feasible solutions.
- Provides **upper bound**.

Backward recursion works as nested **subproblems**:

- Updates cost-to-go function **(adds Benders cuts)**
- Provides **lower** bound.

References

- Bertsimas, D., & Tsitsiklis, J. N. (1997). Introduction to linear optimization. Belmont, MA: Athena scientific.
- Birge, J. R., & Louveaux, F. (2011). Introduction to stochastic programming. Springer Science & Business Media.
- Rahmaniani, R., Crainic, T. G., Gendreau, M., & Rei, W. (2017). The Benders decomposition algorithm: A literature review. European Journal of Operational Research, 259(3), 801-817.
- Laporte, G., & Louveaux, F. V. (1993). The integer L-shaped method for stochastic integer programs with complete recourse. Operations research letters, 13(3), 133-142.
- Birge, J. R. (1985). Decomposition and partitioning methods for multistage stochastic linear programs. Operations research, 33(5), 989-1007.