An introduction to Benders decomposition

Martin Branda

Charles University
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

COMPUTATIONAL ASPECTS OF OPTIMIZATION

Benders decomposition

Benders decomposition can be used to solve:

- linear programming
- mixed-integer (non)linear programming
- two-stage stochastic programming (L-shaped algorithm)
- multistage stochastic programming (Nested Benders decomposition)

Benders decomposition for two-stage linear programming problems

$$\begin{aligned} \min\ & c^T x + q^T y \\ \text{s.t. } & Ax = b, \\ & Tx + Wy = h, \\ & x \geq 0, \\ & y \geq 0. \end{aligned} \tag{1}$$

ASS. $\mathcal{B}_1 := \{x : \ Ax = b, x \geq 0\}$ is bounded and the problem has an optimal solution.

Benders decomposition

We define the **recourse function** (second-stage value function, slave problem)

$$f(x) \quad = \quad \min\{q^T y : \ Wy = h - Tx, \ y \geq 0\} \tag{2}$$

If for some x is $\{y : \ Wy = h - Tx, \ y \geq 0\} = \emptyset$, then we set $f(x) = \infty$.
The recourse function is piecewise linear, convex, and bounded below ...

Benders decomposition

Proof (outline):

- **bounded below and piecewise linear (affine)**: There are finitely many optimal basis B chosen from W such that

$$f(x) = q_B^T B^{-1}(h - Tx),$$

where feasibility $B^{-1}(h - Tx) \geq 0$ is fulfilled for $x \in \mathcal{B}_1$. Optimality condition $q_B^T B^{-1} W - q \leq 0$ does not depend on x.

Benders decomposition

Proof (outline):

- **convex**: let $x_1, x_2 \in \mathcal{B}_1$ and y_1, y_2 be such that $f(x_1) = q^T y_1$ and $f(x_2) = q^T y_2$. For arbitrary $\lambda \in (0,1)$ and $x = \lambda x_1 + (1-\lambda)x_2$ we have

$$\lambda y_1 + (1-\lambda)y_2 \in \{y : \ Wy = h - Tx, \ y \geq 0\},$$

i.e. the convex combination of y's is feasible. Thus we have

$$\begin{aligned} f(x) &= \min\{q^T y : \ Wy = h - Tx, \ y \geq 0\} && (3)\\ &\leq q^T(\lambda y_1 + (1-\lambda)y_2) = \lambda f(x_1) + (1-\lambda) f(x_2). && (4) \end{aligned}$$

Benders decomposition

We have an equivalent NLP problem

$$\begin{aligned} &\min\ c^T x + f(x)\\ &\text{s.t. } Ax = b, \qquad (5)\\ &\quad x \geq 0. \end{aligned}$$

We solve the master problem (first-stage problem)

$$\begin{aligned} &\min\ c^T x + \theta\\ &\text{s.t. } Ax = b,\\ &\quad f(x) \leq \theta, \qquad (6)\\ &\quad x \geq 0. \end{aligned}$$

We would like to approximate $f(x)$ (from below) ...

Algorithm – the feasibility cut

Solve

$$\begin{aligned} f(\hat{x}) &= \min\{q^T y : \ Wy = h - T\hat{x}, \ y \geq 0\} && (7)\\ &= \max\{(h - T\hat{x})^T u : \ W^T u \leq q\}. && (8) \end{aligned}$$

If the dual problem is unbounded (primal is infeasible), then there exists a growth direction $\tilde{u}$ such that $W^T \tilde{u} \leq 0$ and $(h - T\hat{x})^T \tilde{u} > 0$. For any feasible x there exists some $y \geq 0$ such that $Wy = h - Tx$. If we multiply it by $\tilde{u}$

$$\tilde{u}^T (h - T\hat{x}) = \tilde{u}^T W y \leq 0,$$

which has to hold for any feasible x, but is violated by $\hat{x}$. Thus by

$$\tilde{u}^T (h - Tx) \leq 0$$

the infeasible $\hat{x}$ is cut off.

Algorithm – the optimality cut

There is an optimal solution $\hat{u}$ of the dual problem such that

$$f(\hat{x}) = (h - T\hat{x})^T \hat{u}.$$

For arbitrary x we have

$$\begin{aligned} f(x) &= \sup_u \{(h - Tx)^T u : \ W^T u \le q\}, & (9)\\ &\ge (h - Tx)^T \hat{u}, & (10) \end{aligned}$$

because $\hat{u}$ is feasible for arbitrary x. From inequality $f(x) \le \theta$ we have the optimality cut

$$\hat{u}^T (h - Tx) \le \theta.$$

If this cut is fulfilled for actual $(\hat{x}, \hat{\theta})$, then STOP, $\hat{x}$ is an optimal solution.

Algorithm – master problem

We solve the **master problem with cuts**

$$\begin{aligned} &\min \ c^T x + \theta \\ &\text{s.t. } Ax = b, \\ &\quad \tilde{u}_l^T (h - Tx) \le 0, \ l = 1, \dots, L, \\ &\quad \tilde{u}_k^T (h - Tx) \le \theta, \ k = 1, \dots, K, \\ &\quad x \ge 0. \end{aligned} \qquad (11)$$

Algorithm

0. INIC: Set $\theta = -\infty$, $L = 0$, $K = 0$.
1. Solve the **master problem** to obtain $(\hat{x}, \hat{\theta})$.
2. For $\hat{x}$, solve the **dual of the second-stage** (recourse) problem to obtain
   - a direction of unbounded decrease (feasibility cut), $L = L + 1$,
   - or an optimal solution (optimality cut), $K = K + 1$.
3. STOP, if the current solution $(\hat{x}, \hat{\theta})$ fulfills the optimality cuts. Otherwise GO TO Step 1.

Convergence of the algorithm

There are finitely many extreme directions that can generate the feasibility cuts and finitely many (dual) feasible basis which can produce the optimality cuts.

Let (x^*, θ^*) be an optimal solution of the reformulated original problem.

1. The feasibility set of the master problem (6) is always contained in the feasibility set of the master problem with cuts (11) (no feasible solutions are cut).
2. The optimal solution $(\hat{x}, \hat{\theta})$ obtained by the algorithm is feasible for the master problem (6), because

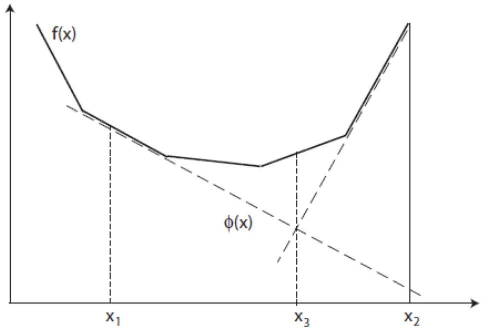
$$\hat{\theta} \ge (h - T\hat{x})^T \hat{u} = f(\hat{x}).$$

Thus, from 1. and 2. we obtain

$$c^T x^* + \theta^* \ge c^T \hat{x} + \hat{\theta} \ge c^T x^* + \theta^*.$$

Kall and Mayer (2005), Proposition 2.19

Benders optimality cuts



Kall and Mayer (2005)

Example

$$
\begin{aligned}
\min\ & 2x + 2y_1 + 3y_2 \\
\text{s.t. } & x + y_1 + 2y_2 = 3, \\
& 3x + 2y_1 - y_2 = 4, \\
& x,\ y_1,\ y_2 \geq 0.
\end{aligned} \tag{12}
$$

Example

Recourse function

$$
\begin{aligned}
f(x) = \min\ & 2y_1 + 3y_2 \\
\text{s.t. } & y_1 + 2y_2 = 3 - x, \\
& 2y_1 - y_2 = 4 - 3x, \\
& y_1,\ y_2 \geq 0.
\end{aligned} \tag{13}
$$

Iteration 1

Set $\theta = -\infty$ and solve master problem

$$
\min_x 2x \text{ s.t. } x \geq 0. \tag{14}
$$

Optimal solution $\hat{x} = 0$.

Iteration 1

Solve the dual problem for $\hat{x} = 0$:

$$\begin{aligned} \max_u \; & (3-x)u_1 + (4-3x)u_2 \\ \text{s.t. } & u_1 + 2u_2 \le 2, \\ & 2u_1 - u_2 \le 3. \end{aligned} \tag{15}$$

Optimal solution is $\hat{u} = (8/5, 1/5)$ with optimal value 28/5, thus no feasibility cut is necessary. We can construct an optimality cut

$$(3-x)8/5 + (4-3x)1/5 = 28/5 - 11/5x \le \theta.$$

Iteration 2

Add the optimality cut and solve

$$\begin{aligned} \min_{x,\theta} \; & 2x + \theta \\ \text{s.t. } & 28/5 - 11/5x \le \theta, \\ & x \ge 0. \end{aligned} \tag{16}$$

Optimal solution $(\hat{x}, \hat{\theta}) = (2.5455, 0)$ with optimal value 5.0909.

Iteration 2

Solve the dual problem for $\hat{x} = 2.5455$:

$$\begin{aligned} \max_u \; & (3-x)u_1 + (4-3x)u_2 \\ \text{s.t. } & u_1 + 2u_2 \le 2, \\ & 2u_1 - u_2 \le 3. \end{aligned} \tag{17}$$

Optimal solution is $\hat{u} = (1.5, 0)$ with optimal value 0.6818, thus no feasibility cut is necessary. We can construct an optimality cut

$$(3-x)1.5 + (4-3x)0 = 4.5 - 1.5x \le \theta.$$

Iteration 3

Add the optimality cut and solve

$$\begin{aligned} \min_{x,\theta} \; & 2x + \theta \\ \text{s.t. } & 28/5 - 11/5x \le \theta, \\ & 4.5 - 1.5x \le \theta, \\ & x \ge 0. \end{aligned} \tag{18}$$

...

Two-stage stochastic programming problem

Probabilities $0 < p_s < 1$, $\sum_s p_s = 1$,

$$\begin{aligned}
\min\ & c^T x + \sum_{s=1}^{S} p_s q_s^T y_s \\
\text{s.t. } & Ax = b, \\
& \begin{array}{cccccccc} W y_1 & & & & +T_1 x & = & h_1, \\ & W y_2 & & & +T_2 x & = & h_2, \\ & & \ddots & & \vdots & \vdots & \vdots \\ & & & W y_S & +T_S x & = & h_S, \end{array} \\
& x \geq 0,\ y_s \geq 0,\ s = 1, \dots, S.
\end{aligned} \tag{19}$$

One master and S "second-stage" problems – apply the dual approach to each of them.

Minimization of Conditional Value at Risk

If the distribution of R_i is discrete with realizations r_{is} and probabilities $p_s = 1/S$, then we can use **linear programming** formulation

$$\begin{aligned}
\min_{\xi, x_i}\ & \xi + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} [-\sum_{i=1}^{n} x_i r_{is} - \xi]_+, \\
\text{s.t. } & \sum_{i=1}^{n} x_i \overline{R}_i \geq r_0, \\
& \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0,
\end{aligned}$$

where $\overline{R}_i = 1/S \sum_{s=1}^{S} r_{is}$, $[\cdot]_+ = \max\{\cdot, 0\}$.

Conditional Value at Risk

Master problem

$$\begin{aligned}
\min_{\xi, x_i}\ & \xi + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} f_s(x, \xi), \\
\text{s.t. } & \sum_{i=1}^{n} x_i \overline{R}_i \geq r_0,\ \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0,
\end{aligned}$$

Second-stage problems

$$\begin{aligned}
f_s(x, \xi) = \min_{y}\ & y, \\
\text{s.t. } & y \geq -\sum_{i=1}^{n} x_i r_{is} - \xi, \\
& y \geq 0.
\end{aligned}$$

Solve the dual problems quickly ..

Multistage Stochastic Linear Programming

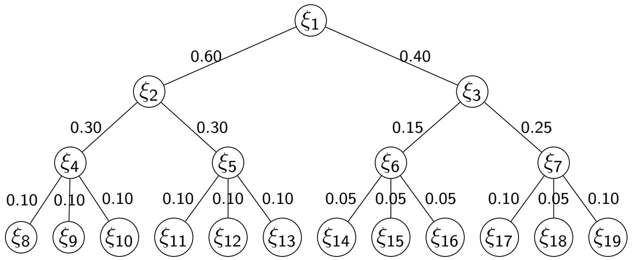
MSLiP=Multistage Stochastic Linear Programming - "nested Benders decomposition with added algorithmic features".

- Support of an arbitrary number of time periods and finite discrete distributions with Markovian structure.

Scenario TREE = a set of nodes $\mathcal{K} = \{1, \dots, K_T\}$ with stages $\mathcal{K}_t = \{K_{t-1} + 1, \dots, K_t\}$ and probabilities $p_1, \dots, p_T > 0$, $\sum_{n \in \mathcal{K}_t} p_n = 1$,

- a_n the ancestor of the node n,
- $\mathcal{D}(n)$ the set of descentants of the node n,
- $t(n)$ the time stage of the node n.

Scenario tree



For example $a(12) = 5$, $\mathcal{D}(6) = \{14, 15, 16\}$, $t(4) = 3$.

Nested formulation of the discrete MSLP

For starting node ($n = 1$)

$$\begin{aligned} F_1 &= \min_{x_1, \vartheta_1} \left\{ c_1^T x_1 + \vartheta_1 \;\; s.t.\; Ax_1 = b,\; \vartheta_1 \geq Q_1(x_1) \right\}, \\ Q_1(x_1) &= \sum_{m \in \mathcal{D}(1)} \frac{p_m}{p_n} F_m(x_1). \end{aligned}$$

For nested stages $n = 2, \ldots, K_{T-1}$

$$\begin{aligned} F_n(x_{a_n}) &= \min_{x_n, \vartheta_n} \left\{ c_n^T x_n + \vartheta_n \;\; s.t.\; W_n x_n = h_n - T_n x_{a_n}, \right. \\ &\qquad \left. \vartheta_n \geq Q_n(x_n) \right\}, \\ Q_n(x_n) &= \sum_{m \in \mathcal{D}(n)} \frac{p_m}{p_n} F_m(x_n). \end{aligned}$$

For final stage $n = K_{T-1} + 1, \ldots, K_T$

$$F_n(x_{a_n}) = \min_{x_n} \left\{ c_n^T x_n \; s.t.\; W x_n = h_n - T_n x_{a_n} \right\}.$$

Nested two-stage problem

(M)(n) Master program = n-th nested two-stage problem:

$$\begin{aligned} F_n(x_{a_n}) = \min_{x_n, \vartheta_n} &\; c_n^T x_n + \vartheta_n \\ & s.t. \\ W_n x_n &= h_n - T_n x_{a_n}, \\ \vartheta_n &\geq Q_n(x_n), \text{ convex constraint}, \\ Q_n(x_n) &= \textstyle\sum_{m \in \mathcal{D}(n)} \frac{p_m}{p_n} F_m(x_n). \end{aligned}$$

$F_1 = F_1(x_{a_1})$, where we set $x_{a_1} = 0$, $W_1 = A$ and $h_1 = b$.
We set $\vartheta_n = 0$ for $n = K_{T-1} + 1, \ldots, K_T$.

Relaxed Master problem

(RM)(n) Relaxed Master program, $n = 1, \ldots, K_T$:

$$\begin{aligned} \widetilde{F}_n(x_{a_n}) = \min_{x_n, \vartheta_n} &\; c_n^T x_n + \vartheta_n \\ & s.t. \\ W_n x_n &= h_n - T_n x_{a_n}, \\ F_n x_n &\geq f_n, && \text{feasibility cuts} \\ D_n x_n + 1\vartheta_n &\geq d_n, && \text{optimality cuts.} \end{aligned}$$

$\widetilde{F}_1 = \widetilde{F}_1(x_{a_1})$, where we set $x_{a_1} = 0$, $W_1 = A$ and $h_1 = b$.
(RM)(n), $n = K_{T-1} + 1, \ldots, K_T$, compensatory bounds ϑ_n and cuts are not involved.

Dual problem

(RD)(n) Dual problem to the relaxed master problem (RM)(n), $n = 2, \dots, K_T$:

$$\max_{\pi_n, \alpha_n, \beta_n, \lambda_n, \mu_n} \pi_n^T (h_n - T_n x_{a_n}) + \alpha_n^T f_n + \beta_n^T d_n$$

$$\begin{aligned} &s.t. \\ \pi_n^T W_n + \alpha_n^T F_n + \beta_n^T D_n &= c_n, \\ 1^T \beta_n &= 1, \\ \alpha_n, \beta_n &\geq 0, \\ \pi_n & \quad \text{unrestricted}. \end{aligned}$$

We set $\alpha_n, \beta_n = 0$ for $n = K_{T-1} + 1, \dots, K_T$

Algorithm MSLiP

(0)

- Set $\vartheta_n^{(0)} = 0$ for all $n = 1, \dots, K_{T-1}$,
- Solve

$$x_1^{(0)} = \arg\min_{x_1} \{ c_1^T x_1 \;\; s.t. \; Ax_1 = b \}.$$

Algorithm MSLiP

(1)

- Solve the dual problem (RD)(m) to the (RM)(m), $\forall m \in \mathcal{D}(n)$. We get
  - dual optimal solution $(\pi_m^*, \alpha_m^*, \beta_m^*), \forall m \in \mathcal{D}(n)$,
  - or feasible extreme direction $(\pi_{m(j)}^j, \alpha_{m(j)}^j, \beta_{m(j)}^j)$ in which the dual problem to the subproblem $m(j) \in \mathcal{D}(n)$ is unbounded, i.e.

$$\pi_{m(j)}^j (b_{m(j)} - W_m x_n) + \alpha_{m(j)}^j f_m > 0.$$

$\rightarrow$ **feasibility cut** of the feasible set of (MR)(n):

$$\underbrace{\pi_{m(j)}^j W_m}_{(F_n)_{j\cdot}} x_n \geq \underbrace{\pi_{m(j)}^j b_{m(j)} + \alpha_{m(j)}^j f_m}_{(f_n)_j}.$$

Algorithm MSLiP

(2)

- If $\vartheta_n < Q_n(x_n) \rightarrow$ **optimality cut** of the feasible set of (MR)(n)

$$\overbrace{\sum_{m \in \mathcal{D}(n)} p_m \pi_m^i T_m}^{(D_n)_{i\cdot}} x_n + \vartheta_n \geq$$

$$\geq \underbrace{\sum_{m \in \mathcal{D}(n)} p_m \left[\pi_m^i h_m + \alpha_m^i f_m + \beta_m^i d_m \right]}_{(d_n)_i}.$$

- Else if $\vartheta_n \geq Q_n(x_n)$ then we have optimal solution x_n of (MR)(n).

Fast-forward-fast-back (FFFB)

- FORWARD pass ($t = 1, \dots, T$, $n = K_t - 1, \dots, K_t$) terminates by:
  - infeasibility of the relaxed master program (RM)(n) $\rightarrow$ add feasibility cut to (RM)(a_n) & BACKTRACKING,
  - obtaining optimal solutions $\hat{x}_n$ for all $n = 1, \dots, K_T$ $\rightarrow$ BACKWARD pass.
- BACKTRACKING ($n \rightarrow a_n$) terminates by:
  - feasibility of the relaxed master program (RM)(a_n) $\rightarrow$ FORWARD pass,
  - reaching the root node with an infeasible (RM)(1) $\rightarrow$ MSLP is infeasible.
- BACKWARD pass always goes through all nodes (adding optimality cuts if necessary).
  - No optimality cuts have been added $\rightarrow$ optimal solution,
  - else $\rightarrow$ FORWARD pass.

MSLiP

- The algorithm (FFFB) terminates in a **finite number of iterations**.
- If termination occurs after BACKWARD pass then the current solution is optimal.
- **Validity of**
  - feasibility cuts $\sim$ feasible solutions of (M)(n) are not cut off.
  - optimality cuts $\sim$ objective function of (RM)(n) yields a lower bound to the objective function (M)(n).
- Cuts generated by the algorithm are valid.
-

$$\text{"}\,\widetilde{F}_1^{(BACKWARD)} \le F_1 \le \widetilde{F}_1^{(FORWARD)}\text{"}$$

QDECOM

= Quadratic DECOMposizion, regularizing quadratic term in the objective (two-stage).
(RMQ) Relaxed Master program

$$\begin{aligned}
\widetilde{F} \;=\; \min_{x,\vartheta^m} c^T x_n + \sum_{m\in\mathcal{D}} p_m \vartheta^m &\;+\; \frac{1}{2}\left\|x - x^{(i-1)}\right\|^2 \\
s.t.& \\
Ax &= b, \\
Fx &\ge f, \\
D^m x + 1\vartheta^m &\ge d^m, \forall m \in \mathcal{D}.
\end{aligned}$$

Literature

- L. Adam: Nelinearity v úlohách stochastického programování: aplikace na řízení portfolia. Diplomová práce MFF UK, 2011. (IN CZECH)
- J.F. Benders (1962): Partitioning procedures for solving mixed-variables programming problems, Numerische Mathematik 4(3), 238–252.
- P. Kall, J. Mayer: Stochastic Linear Programming: Models, Theory, and Computation. Springer, 2005.