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

\[
\begin{align*}
\min & \quad c^T x + q^T y \\
n & \quad Ax = b, \\
& \quad Tx + Wy = h, \\
& \quad x \geq 0, \\
& \quad y \geq 0.
\end{align*}
\]  

(1)

**ASS.** \( B_1 := \{ x : Ax = b, x \geq 0 \} \) is bounded and the problem has an optimal solution.

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

\[
f(x) = \min \{ q^T y : Wy = h - Tx, y \geq 0 \}
\]  

(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...
Proof (outline):

- **bounded below and piecewise linear (affine):** There are finitely many optimal basis $B$ chosen from $W$ such that
  \[ f(x) = q^T B^{-1}(h - Tx), \]
  where feasibility $B^{-1}(h - Tx) \geq 0$ is fulfilled for $x \in B_1$. Optimality condition $q^T B^{-1}W - q \leq 0$ does not depend on $x$.

We have an equivalent NLP problem

\[
\begin{align*}
\min & \quad c^T x + f(x) \\
\text{s.t.} & \quad Ax = b, \quad x \geq 0.
\end{align*}
\]

We solve the master problem (first-stage problem)

\[
\begin{align*}
\min & \quad c^T x + \theta \\
\text{s.t.} & \quad Ax = b, \\
& \quad f(x) \leq \theta, \\
& \quad x \geq 0.
\end{align*}
\]

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

Proof (outline):

- **convex:** let $x_1, x_2 \in B_1$ 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
  \[
  f(x) = \min \{ q^T y : Wy = h - Tx, \ y \geq 0 \} \leq q^T (\lambda y_1 + (1 - \lambda) y_2) = \lambda f(x_1) + (1 - \lambda) f(x_2).
  \]

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 $(h - T \tilde{x})^T \tilde{u} > 0$. For any feasible $x$ there exists some $y \geq 0$ such that $Wy = h - Tx$. If we multiply it by $\tilde{u}$

\[ \tilde{u}^T (h - T \tilde{x}) = \tilde{u}^T Wy \leq 0, \]

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

\[ \tilde{u}^T (h - T x) \leq 0 \]

the infeasible $\tilde{x}$ is cut off.
Algorithm – the optimality cut

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

\[
f(\hat{x}) = (h - Tx)^T \hat{u}.
\]

For arbitrary \( x \) we have

\[
f(x) = \sup_u \{(h - Tx)^T u : W^T u \leq q\}, \quad (9)
\]

\[
\geq (h - Tx)^T \hat{u}, \quad (10)
\]

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

\[
\hat{u}^T (h - Tx) \leq \theta.
\]

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

\[
\min c^T x + \theta \\
\text{s.t. } Ax = b, \\
\tilde{u}^T_l (h - Tx) \leq 0, \quad l = 1, \ldots, L, \\
\tilde{u}^T_k (h - Tx) \leq \theta, \quad k = 1, \ldots, K, \\
x \geq 0.
\]

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 \((x^*, \theta^*)\) 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

\[
\hat{\theta} \geq (h - T\hat{x})^T \hat{u} = f(\hat{x}).
\]

Thus, from 1. and 2. we obtain

\[
c^T x^* + \theta^* \geq c^T \hat{x} + \hat{\theta} \geq c^T x^* + \theta^*.
\]

Kall and Mayer (2005), Proposition 2.19
Benders optimality cuts

Recourse function

\[ f(x) = \min 2y_1 + 3y_2 \]
\[ \text{s.t. } y_1 + 2y_2 = 3 - x, \]
\[ 2y_1 - y_2 = 4 - 3x, \]
\[ y_1, y_2 \geq 0. \]  

(13)

Iteration 1

Set \( \theta = -\infty \) and solve master problem

\[ \min_x 2x \text{ s.t. } x \geq 0. \]  

(14)

Optimal solution \( \hat{x} = 0 \).
**Iteration 1**

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

$$\max \frac{3 - x}{u_1} + (4 - 3x)u_2$$

$$\text{s.t. } u_1 + 2u_2 \leq 2,$$
$$2u_1 - u_2 \leq 3.$$  

(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

$$(3 - x)\frac{8}{5} + (4 - 3x)\frac{1}{5} = 28/5 - 11/5x \leq \theta.$$  

**Iteration 2**

Add the optimality cut and solve

$$\min 2x + \theta$$

$$\text{s.t. } 28/5 - 11/5x \leq \theta,$$
$$x \geq 0.$$  

(16)

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

**Iteration 3**

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

$$\max \frac{3 - x}{u_1} + (4 - 3x)u_2$$

$$\text{s.t. } u_1 + 2u_2 \leq 2,$$
$$2u_1 - u_2 \leq 3.$$  

(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

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

**Iteration 3**

Add the optimality cut and solve

$$\min 2x + \theta$$

$$\text{s.t. } 28/5 - 11/5x \leq \theta,$$
$$4.5 - 1.5x \leq \theta,$$
$$x \geq 0.$$  

(18)

...
Two-stage stochastic programming problem

Probabilities $0 < p_s < 1$, $\sum_{s=1}^{S} p_s = 1$,

\[
\begin{align*}
\min & \ c^T x + \sum_{s=1}^{S} p_s q_s^T y_s \\
\text{s.t.} & \ Ax = b, \\
& W y_1 + T_1 x = h_1, \\
& W y_2 + T_2 x = h_2, \\
& \cdots \\
& W y_S + T_S x = h_S, \\
x \geq 0, & \quad y_s \geq 0, \quad s = 1, \ldots, S.
\end{align*}
\]

(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 $r_{ia}$ and probabilities $p_s = 1/S$, then we can use linear programming formulation

\[
\begin{align*}
\min & \ \xi + \frac{1}{(1-\alpha)S} \sum_{i=1}^{n} \left[ -\sum_{s=1}^{S} x_i r_{ia} - \xi \right], \\
\text{s.t.} & \ \sum_{i=1}^{n} x_i R_i \geq r_0, \\
& \sum_{i=1}^{n} x_i = 1, \ x_i \geq 0, \\
\end{align*}
\]

where $R_i = 1/S \sum_{s=1}^{S} r_{ia} \cdot \left[ \cdot \right]_+ = \max(\cdot, 0)$.

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 $K = \{1, \ldots, K_T\}$ with stages $K_t = \{K_{t-1} + 1, \ldots, K_t\}$ and probabilities $p_1, \ldots, p_T > 0$, $\sum_{n \in K_t} p_n = 1$.

- $a_n$ the ancestor of the node $n$.

- $D(n)$ the set of descendents of the node $n$.

- $T(n)$ the time stage of the node $n$. 

We set $F(M)(n) = \text{Master program}$.

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

Nested two-stage problem

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

$$F_n(x_{\text{a}_n}) = \min_{x_{\text{a}_n}, \vartheta_n} c_n^T x_{\text{a}_n} + \vartheta_n$$

s.t.

$$W_p x_{\text{a}_n} = h_p - T_p x_{\text{a}_n}.$$  

$$Q_n(x_{\text{a}_n}) = \sum_{m \in D(n)} \frac{p_m}{p_n} F_m(x_{\text{a}_n}).$$

For starting node ($n = 1$)

$$F_1 = \min_{x_1, \vartheta_1} c_1^T x_1 + \vartheta_1 \text{ s.t. } A x_1 = b, \vartheta_1 \geq Q_1(x_1),$$

$$Q_1(x_1) = \sum_{m \in D(1)} \frac{p_m}{p_n} F_m(x_1).$$

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

$$F_n(x_{\text{a}_n}) = \min_{x_{\text{a}_n}, \vartheta_n} c_n^T x_{\text{a}_n} + \vartheta_n \text{ s.t. } W_p x_{\text{a}_n} = h_n - T_p x_{\text{a}_n},$$

$$\vartheta_n \geq Q_n(x_{\text{a}_n}),$$

$$Q_n(x_{\text{a}_n}) = \sum_{m \in D(n)} \frac{p_m}{p_n} F_m(x_{\text{a}_n}).$$

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

$$F_n(x_{\text{a}_n}) = \min_{x_{\text{a}_n}, \vartheta_n} c_n^T x_{\text{a}_n} \text{ s.t. } W_p x_{\text{a}_n} = h_n - T_p x_{\text{a}_n}.$$
Extensions and applications
Nested Benders decomposition

Dual problem

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

\[
\max_{\pi_n, \alpha_n, \beta_n, \lambda_n, \mu_n} \pi_n^T (b_n - T_n x_n) + \alpha_n^T f_n + \beta_n^T d_n
\]

s.t.

\[
\pi_n^T W_n + \alpha_n^T F_n + \beta_n^T D_n = c_n,
\]

\[
1^T \beta_n = 1,
\]

\[
\alpha_n, \beta_n \geq 0,
\]

\[
\pi_n \text{ unrestricted}.
\]

We set \(\alpha_n, \beta_n = 0\) for \(n = K_T - 1 + 1, \ldots, K_T\).

Algorithm MSLiP

(0) Set \(\vartheta_n^{(0)} = 0\) for all \(n = 1, \ldots, K_T - 1\).

(1) Solve the dual problem (RD)(n) to the (RM)(n), \(\forall n \in D(n)\).
- dual optimal solution \((\pi_n^*, \alpha_n^*, \beta_n^*)\), \(\forall n \in D(n)\).
- or feasible extreme direction \((\pi_n^{(j)}, \alpha_n^{(j)}, \beta_n^{(j)})\) in which the dual problem to the subproblem \(m(j) \in D(n)\) is unbounded, i.e.

\[
\pi_m^{(j)} (b_m x_n) - W_m x_n + \alpha_m^{(j)} f_m + \beta_m^{(j)} d_m > 0.
\]

→ feasibility cut of the feasible set of (MR)(n):

\[
\pi_m^{(j)} W_m x_n \geq \pi_m^{(j)} b_m x_n + \alpha_m^{(j)} f_m + \beta_m^{(j)} d_m.
\]

(2) If \(\vartheta_n < Q_n(x_n)\) → optimality cut of the feasible set of (MR)(n)

\[
\sum_{m \in D(n)} \rho_m \pi_m^{(n)} T m x_n + \vartheta_n \geq \sum_{m \in D(n)} \rho_m \left[ \pi_m^{(n)} h_m + \alpha_m^{(n)} f_m + \beta_m^{(n)} d_m \right].
\]

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

Martin Branda (KPMS MFF UK) 2021-03-15 32 / 39

Martin Branda (KPMS MFF UK) 2021-03-15 33 / 39

Martin Branda (KPMS MFF UK) 2021-03-15 34 / 39

Martin Branda (KPMS MFF UK) 2021-03-15 35 / 39
Extensions and applications

Nested Benders decomposition

Fast-forward-fast-back (FFFB)

- **FORWARD pass** ($t = 1, \ldots, T$, $n = K_t - 1, \ldots, K_t$) 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 $n = 1, \ldots, K_T \rightarrow$ 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.
  $$\tilde{F}(\text{BACKWARD})_1 \leq F_1 \leq \tilde{F}(\text{FORWARD})_1$$

QDECOM

- Quadratic DECOMposition, regularizing quadratic term in the objective (two-stage).
- (RMQ) Relaxed Master program
  $$\tilde{F} = \min_{x,\nu} c^T x_a + \sum_{m \in D} p m \nu^m + \frac{1}{2} \| x - x(l-1) \|^2$$
  s.t. $$Ax = b,$$
  $$Fx \geq f,$$
  $$D^m x + 1^m \nu^m \geq d^m, \forall m \in D.$$ 

Literature