Benders Decomposition in CPLEX 12.7.0

@CPLEX school 2017, Montreal

Benders decomposition: quick literature review

- Originally proposed for MILP by Benders [Benders, 1962]
- Generalized to convex MINLP by Geoffrion [Geoffrion, 1972]
- One of the methods of choice for large size stochastic LP and MILP
  - See, e.g., [Van Slyke and Wets, 1969, Carøe and Tind, 1998, Bodur et al., 2016, Hassanzadeh and Ralphs, 2016]
- But also effective on non-stochastic problems, as e.g.,
  - Network design
    - See, e.g., [Costa, 2006] for a survey paper
  - Linear and quadratic facility location
    - [Fischetti et al., 2015, Fischetti et al., 2016]
- **Benders decomposition in CPLEX 12.7.0**
  - Only MILP supported (no MIQP or MIQCP)

Benders decomposition: the basic idea

$$(\text{or-MILP})\begin{cases}\min \quad c^T x + d^T y \\ \quad A\,x \geq b \\ \quad T\,x + B\,y \geq e \\ \quad x \geq 0, x \text{ integer} \\ \quad y \geq 0\end{cases}$$

x: Primary variables (integer or continuous)

y: Secondary variables (continuous)

Linking constraints

- Benders reformulation:
  - Add variable $\eta = d^T y$ to represent the contribution of y to the objective function
  - Project onto the (x, η) space:

V, R: vertices and extreme rays of $Q = \{\pi \in \mathbb{R}^m : \pi^T B \leq d^T, \pi \geq 0\}$

$$(\text{re-MILP})\begin{cases}\min \quad c^T x + \eta \\ \quad A\,x \geq b \\ \quad v^T(e - T\,x) \leq \eta, \quad v \in V \\ \quad r^T(e - T\,x) \leq 0, \quad r \in R \\ \quad x \geq 0, \text{integer}\end{cases}$$

Benders decomposition: the basic idea

- The reformulated problem (re-MILP) can be solved by separating Benders cuts on the fly:

(master-MILP) $$\begin{cases} \min \ c^T x + \eta \\ A\,x \geq b \\ \{\,Benders\ cuts\} \\ x \geq 0, \text{integer} \end{cases}$$

(x^*, η^*)

$v^T(e - T\,x) \leq \eta$ (optimality cut)

$r^T(e - T\,x) \leq 0$ (feasibility cut)

(worker-LP) $$\begin{cases} \max \ \pi^T(e - Tx^*) \\ \pi^T B \leq d^T \\ \pi \geq 0 \end{cases}$$

Workers decomposition

- In some relevant cases (e.g., Stochastic MILP), the linking constraints decompose in $k \geq 2$ blocks

$$[T \quad B] = \begin{bmatrix} T_1 & B_1 & 0 & \dots & 0 \\ T_2 & 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_k & 0 & 0 & \dots & B_k \end{bmatrix}$$

- And (or-MILP) reads as

$$\text{(or-MILP)} \begin{cases} \min \ c^T x + d_1{}^T y_1 + \ \dots + \ d_k{}^T y_k \\ A\,x \geq b \\ T_i x + B_i\, y_i \geq e_i, \qquad i = 1, \dots, k \\ x \geq 0, x \text{ integer} \\ y_i \geq 0, \qquad i = 1, \dots, k \end{cases}$$

Workers decomposition

- In Benders reformulation, each block can be decomposed independently of the others

$$
\text{(re-MILP)} \begin{cases} \min \; c^T x + \eta_1 + \ldots + \eta_k \\ A\,x \geq b \\ {v_i}^T (e_i - T_i\,x) \leq \eta_i, & v_i \in V_i \; (i = 1, \ldots, k) \\ {r_i}^T (e_i - T_i\,x) \leq 0, & r_i \in R_i \; (i = 1, \ldots, k) \\ x \geq 0, \text{integer} \end{cases}
$$

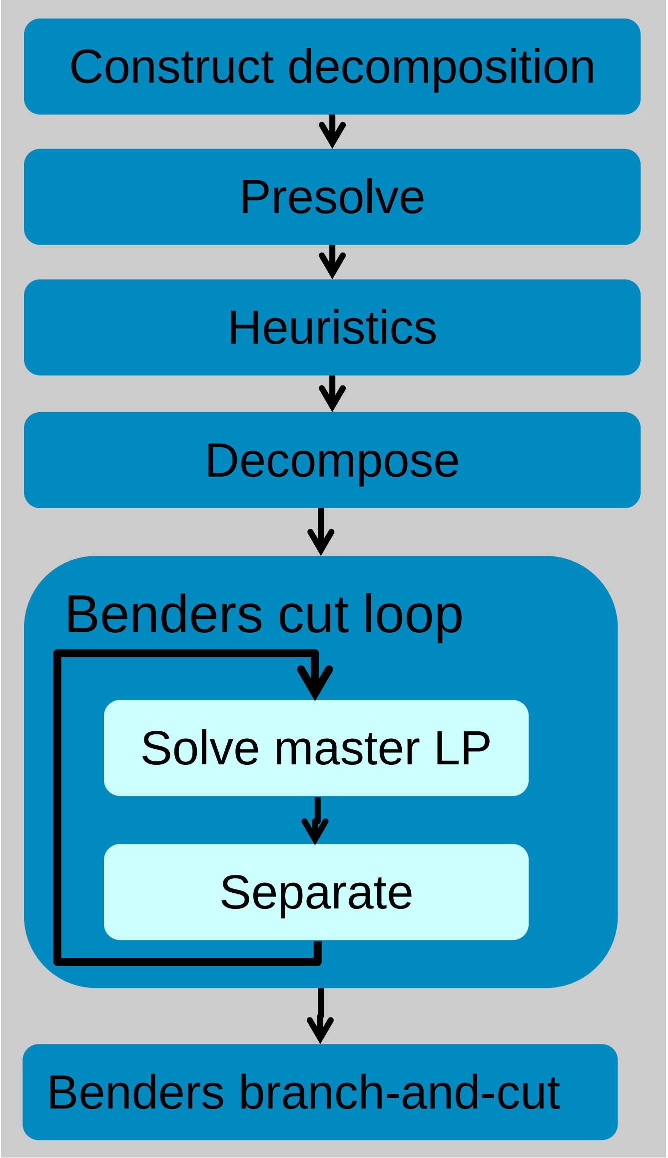
$(x^*, {\eta_i}^*)$ ↓ ↑

$${v_i}^T (e_i - T_i\,x) \leq \eta_i \quad \text{(optimality cuts)}$$

$${r_i}^T (e_i - T_i\,x) \leq 0 \quad \text{(feasibility cuts)}$$

$$
\text{(worker-LP)}_i \begin{cases} \max \; \pi^T (e_i - T_i\,x^*) \\ \pi^T B_i \leq {d_i}^T \\ \pi \geq 0 \end{cases}
$$

CPLEX Benders MIP Solver



- Decomposition
  - Provided through annotation APIs, or
  - Automatic
- Presolve
  - Preserves the block structure
- Heuristics
  - On the complete MIP
  - May run in parallel to the Benders cut loop
- Benders cut loop
  - Tries to solve the LP relaxation of projected MIP to optimality
- Benders branch-and-cut
  - Benders cuts must be separated as lazy constraints, to cut integer solutions
  - Also separated for fractional solutions

Using Benders: annotation APIs and “benders strategy” parameter

- Specify the block structure with (new) CPLEX annotation APIs
  - For each variable, specify the block which the variable belongs to
    - $i = 0$ for primary (i.e., master) variables x
    - $i \geq 1$ for secondary variables in block i (i.e., variables y_i)
  - Available with all CPLEX APIs
  - Can be used to write/read annotation to/from an xml file
- Or ask CPLEX to automatically construct a (simple) decomposition

	Parameter CPXPARAM_Benders_Strategy
1 (USER)	Use block decomposition as provided from the annotation.
2 (WORKERS)	Use given annotation only to identify primary (i.e., master) variables x and secondary variables y. Automatic workers decomposition: if possible, second stage matrix B is decomposed in disjoint blocks.
3 (FULL)	Ignore any given annotation and automatically construct a simple decomposition: integer variables in master block, continuous variables in workers, plus workers decomposition.
-1 (OFF)	Ignore given annotation (if any) and run conventional branch and cut.
0 (AUTO), default	If no annotation is provided, apply conventional branch and cut (no Benders). If annotation is provided, do the same as strategy 2 (WORKERS).

Using Benders with CPLEX interactive

Read decomposition from annotation

```
read myprob.lp
read myanno.ann
optimize
```

OR

Let CPLEX decompose the model

```
read myprob.lp
set benders strategy 3
optimize
```

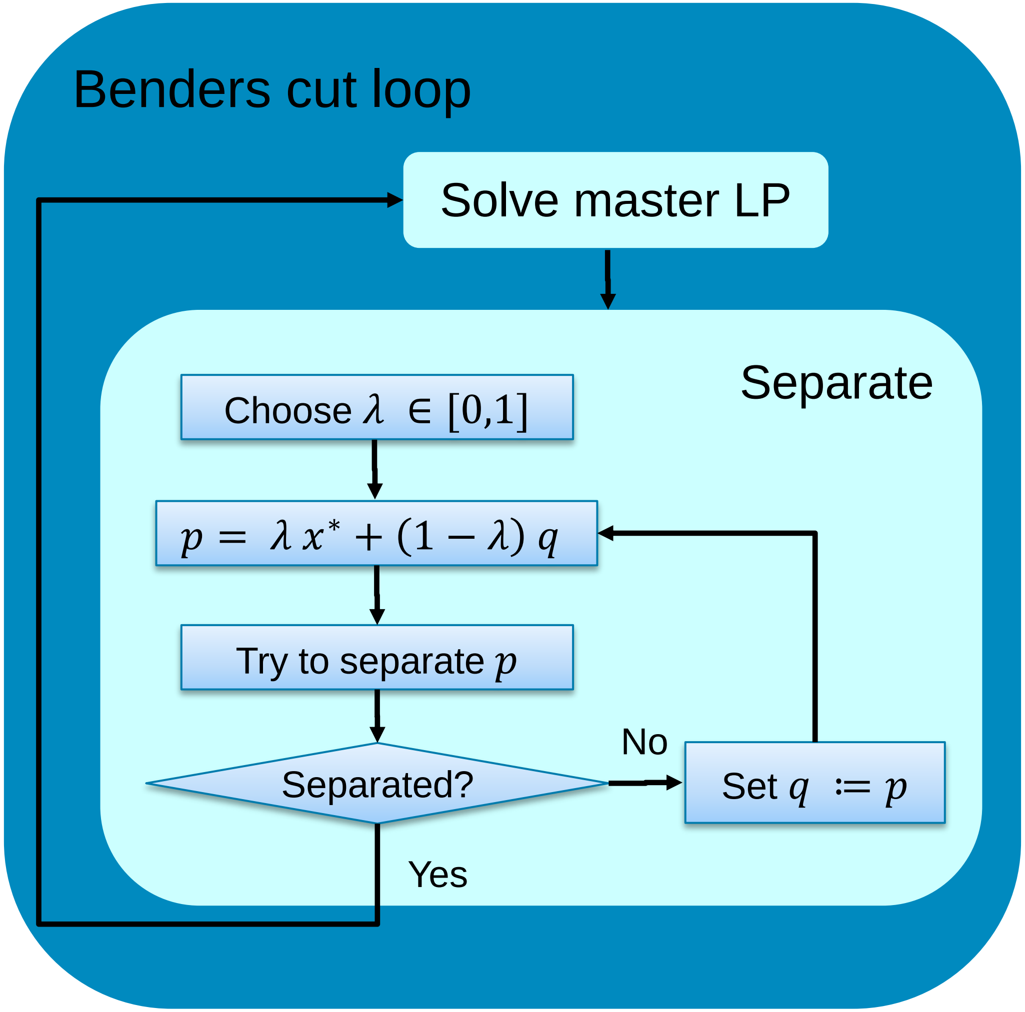
	Parameter CPXPARAM_Benders_Strategy
1 (USER)	Use block decomposition as provided from the annotation.
2 (WORKERS)	Use given annotation only to identify primary (i.e., master) variables x and secondary variables y. Automatic workers decomposition: if possible, second stage matrix B is decomposed in disjoint blocks.
3 (FULL)	Ignore any given annotation and automatically construct a simple decomposition: integer variables in master block, continuous variables in workers, plus workers decomposition.
-1 (OFF)	Ignore given annotation (if any) and run conventional branch and cut.
0 (AUTO), default	If no annotation is provided, apply conventional branch and cut (no Benders). If annotation is provided, do the same as strategy 2 (WORKERS).

Benders cuts separation: two-stage approach

- Given $(x^*, \eta_i{}^*)$, separate cuts for each block $i = 1, \dots, k$
  - Two-stage approach separation in the primal space
  - Normalization for feasibility cut selection, similar to [Fischetti et al., 2010]

Stage	Cut Generating LP (CGLP)	Dual of	Cut
1	$\begin{cases} \min\ d_i{}^T y_i \\ B_i\, y_i \geq e_i - T_i\, x^* \\ y_i \geq 0 \end{cases}$	$\begin{cases} \max\ \pi^T (e_i - T_i\, x^*) \\ \pi^T B_i \leq d_i{}^T \\ \pi \geq 0 \end{cases}$	optimality
2	$\begin{cases} \min\ s \\ B_i\, y_i + I\, s \geq e_i - T_i\, x^* \\ y_i \geq 0 \\ s \geq 0 \end{cases}$	$\begin{cases} \max\ \pi^T (e_i - T_i\, x^*) \\ \pi^T B_i \leq 0 \\ \mathbf{1}^T \pi \leq 1 \\ \pi \geq 0 \end{cases}$	feasibility

Benders cut loop: In-Out techniques



- **Basic idea:**
  - Find a point q in the interior of LP relaxation of (or-MILP)
  - Given x^* (optimal solution of current master LP), try to separate a point in the segment $[x^*, q]$
- **Goal:**
  - Break correlation between separation and optimization
  - Accelerate convergence of cut loop
- Proposed by [Ben-Ameur and Neto, 2007] for Network Design Problems
- Also used by [Fischetti et al., 2015] and [Fischetti et al., 2016] for Benders applied to Facility Location

Benders performance evaluation: the test bed

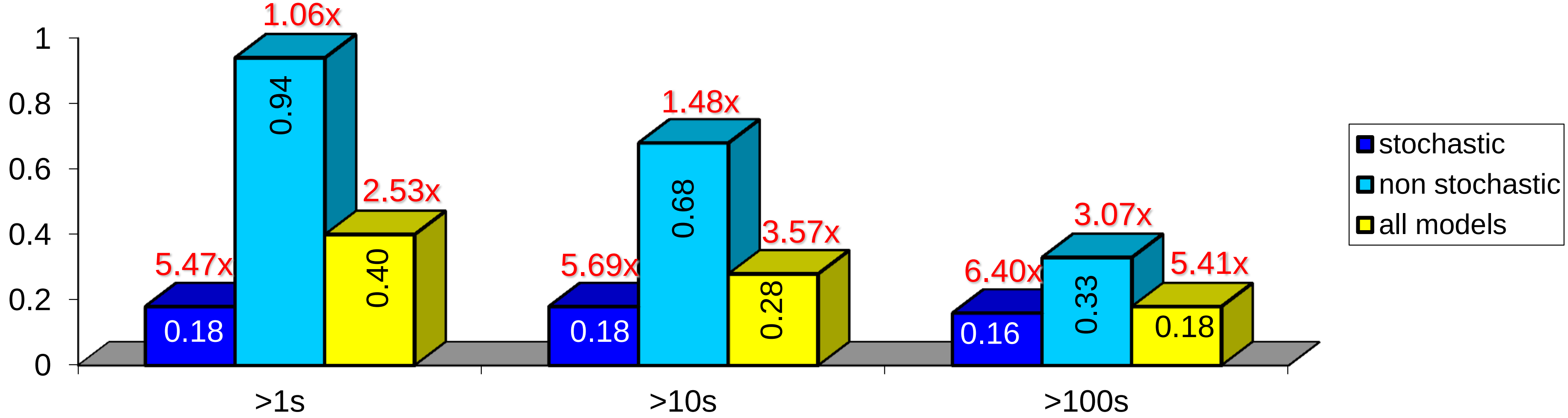
- **Stochastic MILP** (189 models in total)
  - Stochastic Capacitated Facility Location (SCFL)
    - 32 models from [Loveaux, 1986], also used by [Bodur et al., 2016]
    - 50 customers, 25 – 50 facilities, 250 – 500 scenarios
    - Number of non-zeroes after presolve: 560,595 – 2,248,396
  - Stochastic Network Interdiction Problem (SNIP)
    - 70 models, from [Pan and Morton, 2008], also used by [Bodur et al., 2016]
    - 320 binary variables, 456 scenarios, each scenario is a network-LP with 5290 constraints and 830 variables
    - Number of non-zeroes after presolve: 1,827,210 – 1,883,640
  - Stochastic fixed charge multi-commodity network design (SCMND)
    - 70 models from [Crainic et al., 2014]
    - Networks with 60 to 120 vertices and 10 to 40 commodities, 96 or 320 scenarios of demand and varying correlations between the demands of different commodities.
    - Number of non-zeroes after presolve: 197,760 – 2,994,600
  - Stochastic models from our internal library (17 models in total)
    - Unit Commitment, Optimal Power Flow, Network Design

Benders performance evaluation: the test bed

- **Non-stochastic MILP** (201 models in total)
  - Capacitated and Uncapacitated Facility Location (CFL and UFL)
    - 128 models generated as in [Fischetti et al., 2015] and [Fischetti et al., 2016]
    - From 30 facilities and 120 customers (toy models) up to 580 facilities and 2120 customers
    - Number of non-zeroes after presolve: 10,800 – 3,170,986
  - Network Expansion (NEXP)
    - 60 models from [Atamturk et al., 2001]
    - Network expansion problems with step functions capacity installation cost. Networks with between 50 and 150 vertices, 20% edge density and between 2 and 8 steps in the cost function.
    - Number of non-zeroes after presolve: 801 – 40,359
  - From our internal library (13 models in total)
    - Unit Commitment, Long term nuclear power plant maintenance planning, Network Design

CPLEX 12.7.0: Benders B&C performance evaluation

Benders B&C compared to “regular” B&C on all models



	Stochastic			Non stochastic		
	# models	# wins/losses	# timeout (regular B&C / Benders B&C)	# models	# wins/losses	# timeout (regular B&C / Benders B&C)
> 1sec	168	130/37	76/17	152	61/80	6/4
> 10 secs	163	126/37	76/17	86	52/32	6/4
> 100 secs	149	117/32	76/17	44	30/13	6/4

Date: 6 November 2016
Testset: 390 models
Machine: Intel X5650 @ 2.67GHz, 24 GB RAM, 12 threads
Timelimit: 10,000 sec

CPLEX 12.7.0: Benders B&C performance evaluation

	Benders B&C on "> 1 sec"					
	# models	# wins / losses	# timeout (regular B&C / Benders B&C)	Benders speedup / slowdown	Regular B&C Secs (geomean)	Benders B&C Secs (geomean)
SNIP	69	69/0	69/0	+28.10x	10,000	355.9
SCFL	32	32/0	1/0	+19.33x	452.3	23.4
SMCND	52	15/37	1/17	-3.60x	235.1	846.8
Other stochastic	15	14/0	5/0	+6.53x	262.2	40.2
NEXP	44	2/40	0/2	-5.50x	6.0	33.0
CFL and UFL	101	57/36	5/0	+2.42x	18.6	7.7
Other non stochastic	7	2/4	1/2	-1.61x	113.5	183.3

- SNIP
  - Unsolvable with regular B&C
  - Benders B&C: All 70 models in > 100 secs, 6 models in > 1k secs, 1 timeout
- SCFL, CFL, UFL
  - Strong LP relaxation, but much harder to solve in the original space
  - Benders cut loop converges very fast and with very few cuts to LP optimum
- SMCND, NEXP
  - Weak LP relaxation, integrality cuts much more effective in the original space, heuristics more effective in the original space
  - Benders cut loop: slow convergence to the LP optimum

CPLEX 12.7.0: Benders B&C performance evaluation

SMCND instances

- A corner case in favor of regular B&C
  - Regular B&C
    - Presolved problem size: # rows / # cols / # nonzeroes = 32,160 / 204,060 / 617,760
    - Gap left at the end of root node: 0.01% w.r.t. optimal solution, 0.33% w.r.t. incumbent solution
    - Problem solved in 3 nodes and 54.02 secs
  - Benders B&C
    - Problem size after Benders cut loop: # rows / # cols / # nonzeroes = 2,583 / 156 / 84,008
    - Gap left at the end of root node: 8.40% w.r.t. optimal solution, 14.69% w.r.t. incumbent solution
    - Branching not effective (numerical issues): time limit (10k secs) hit after 105,934 nodes with 4.29% gap left
- A corner case in favor of Benders B&C
  - Regular B&C
    - Gap left at the end of root node: 9.39% w.r.t. optimal solution, 10.96% w.r.t. incumbent solution
    - Problem solved in 1,083 nodes and 4,958.66 secs
  - Benders B&C
    - Gap left at the end of root node: 23.34% w.r.t. optimal solution, 24.30% w.r.t. incumbent solution
    - Branching is effective: problem solved in 16,854 nodes and 370.94 secs

CPLEX 12.7.0: Benders B&C performance evaluation

- **SNIP models, an example**
  - Regular B&C
    - Presolved problem size: # rows / # cols / # nonzeroes = 761,179 / 128,800 / 1,827,210
    - Time to solve the first LP relaxation: 959.10 secs
    - Elapsed time at the end of root node: 9,327.62 secs (gap left w.r.t. optimal solution is 15.41%)
    - Time limit hit after 4 nodes with 20.77% gap left (w.r.t. best available solution)
  - Benders B&C
    - Problem size after Benders cut loop: # rows / # cols / # nonzeroes = 2,738/ 664 / 6,196
    - Elapsed time at the end of root node: 131.7 secs (gap left w.r.t. optimal solution is 31.23%)
    - Problem solved in 104,446 nodes and 254.77 secs

CPLEX 12.7.0: Benders B&C performance evaluation

CFL and UFL instances

- Regular B&C is in troubles with more than 240 facilities and 2120 customers:
  - 240 facilities and 2120 customers: 1 instance solved in 728.7 secs, 3 instances solved in more than 1k secs, 1 time limit
  - 580 facilities and 2120 customers: 1 instance solved in 960.4 secs, 4 out of memory
- Benders B&C solves all the above 10 instances with time varying from 34.4 to 72.6 secs
- UFL, an instance with 240 facilities and 2120 customers
  - Regular B&C
    - Presolved problem size: # rows / # cols / # nonzeroes = 489,600 / 489,839 / 1,464,720
    - Time to solve the first LP relaxation: 164.12 secs
    - Elapsed time at the end of root node: 289.33 secs (gap left w.r.t. optimal solution is 5.42%)
    - Time limit hit after 2,511 nodes with a gap left of 1.42% (w.r.t. best available solution)
  - Benders B&C
    - Problem size after Benders cut loop: # rows / # cols / # nonzeroes = 1,637 / 2,279/ 17,578
    - Elapsed time at the end of root node: 36.6 secs (gap left w.r.t. optimal solution is 5.42%)
    - Problem solved in 17,928 nodes and 72.61 secs

IBM®

Legal Disclaimer