UNIVERSIDAD PONTIFICIA ICAI ICADE COMILLAS MADRID

upcomillas es

upcomillas es

Stochastic Dual Dynamic Programming

ESD.S30 Electric Power System Modeling for a Low Carbon Economy

Prof. Andres Ramos
http://www.iit.upcomillas.es/aramos/
Andres.Ramos@upcomillas.es
arght@mit.edu

MIT, Cambridge, MA, January 23-27, 2012

References

- Birge, J.R. and Louveaux, F. (1997) *Introduction to Stochastic Programming.* Springer-Verlag.
- Ermoliev, Y. and Wets, R.J-B. (eds.) *Numerical Techniques for Stochastic Optimization*. Springer-Verlag. Berlin, Germany. 1988.
- Kall, P. and Wallace, S.W. (1995) *Stochastic Programming.* John Wiley and Sons. (http://www.unizh.ch/ior/Pages/Deutsch/Mitglieder/Kall/bib/ka-wal-94.pdf)
- Infanger, G. *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs* Boyd & Fraser. 1994.
- Mayer, J. (1998) *Stochastic Linear Programming Algorithms. A Comparison Based on a Model Management System.* Gordon and Breach Science Publishers.
- Prekopa, A. (1995) *Stochastic Programming* Kluwer Academic Publishers.
- Ruszczynski, A. and Shapiro, A. (eds.) (2003) *Stochastic Programming* Elsevier.

- Alonso-Ayuso, A., E. Cerdá, L. F. Escudero, R. Sala (eds.) (2004) *Optimización bajo incertidumbre* Tirant lo Blanch. Valencia, España.
- Ramos, A., A. Alonso-Ayuso, G. Pérez (eds.) (2008) *Optimización bajo Incertidumbre* Universidad Pontificia Comillas. Madrid, España

Other resources

- Stochastic Programming Community Home Page. Stochastic Programming Resources (www, papers, tutorials, lecture notes, books) (http://stoprog.org/)
- Stochastic Programming Bibliography (http://www.eco.rug.nl/mally/spbib.html)
- Books on Stochastic Programming (http://stoprog.org/index.html?booksSP.html)
- STOCHASTIC PROGRAMMING E-PRINT SERIES (http://www.speps.org/)
- Optimization Online. Stochastic Programming submissions (http://www.optimization-online.org/ARCHIVE_CAT/STOCH/index.html)

- Red Temática de Optimización bajo Incertidumbre (ReTOBI) (http://www.optimizacionbajoincertidumbre.org/)

- International Conference in Stochastic Programming

2

Two-stage and multistage programming

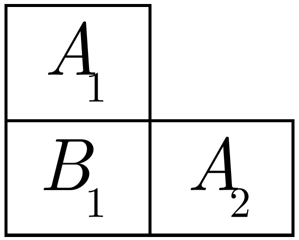
Two-stage (PL-2) and multistage (PL-P) linear programming

- Two-stage PL-2: decisions in two stages
- Multistage PL-P: decisions in multiple stages

- Stairway structure of the constraint matrix (block diagonal)
  - Each stage in only related with the previous one

- Problems of each stage are similar (they have the same structure)
- The matrix structure can be detected by visual inspection

Two-stage linear programming PL-2



$$
\begin{aligned}
\min_{x_1, x_2} (c_1^T x_1 &+ c_2^T x_2) \\
A_1 x_1 & &= b_1 \\
B_1 x_1 &+ A_2 x_2 &= b_2 \\
x_1, \quad & \quad x_2 &\geq 0
\end{aligned}
$$

Two-stage stochastic linear programming PLE-2

- O.F. minimizes first-stage costs and expected value of second-stage costs

$$
\begin{aligned}
\min_{x_1, x_2^\omega} \; & c_1^T x_1 + \sum_{\omega \in \Omega} p^\omega c_2^{\omega T} x_2^\omega \\
& A_1 x_1 && = b_1 \\
& B_1^\omega x_1 + A_2^\omega x_2^\omega && = b_2^\omega \\
& x_1, \quad x_2^\omega && \geq 0
\end{aligned}
$$

- If A_2^ω doesn't depend on ω it is called fixed resource
- Structure of the constraint matrix

A_1			
B_1^1	A_2^1		
B_1^2		A_2^2	
B_1^3			A_2^3

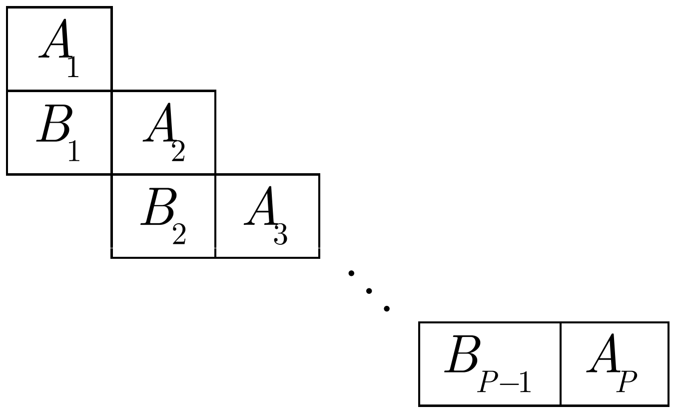
Deterministic equivalent problem

- State space is small

- Formulation of the deterministic equivalent problem

$$
\begin{aligned}
\min_{x_1, x_2^{\omega_1}, x_2^{\omega_2}, x_2^{\omega_3}} \; & c_1^T x_1 + p^{\omega_1} c_2^{\omega_1 T} x_2^{\omega_1} + p^{\omega_2} c_2^{\omega_2 T} x_2^{\omega_2} + p^{\omega_3} c_2^{\omega_3 T} x_2^{\omega_3} \\
& A_1 x_1 && = b_1 \\
& B_1^{\omega_1} x_1 + A_2^{\omega_1} x_2^{\omega_1} && = b_2^{\omega_1} \\
& B_1^{\omega_2} x_1 + A_2^{\omega_2} x_2^{\omega_2} && = b_2^{\omega_2} \\
& B_1^{\omega_3} x_1 + A_2^{\omega_3} x_2^{\omega_3} && = b_2^{\omega_3} \\
& x_1, \quad x_2^{\omega_1}, \quad x_2^{\omega_2}, \quad x_2^{\omega_3}, && \geq 0
\end{aligned}
$$

Multistage linear programming PL-P



$$\min_{x_p} \sum_{p=1}^{P} c_p^T x_p$$

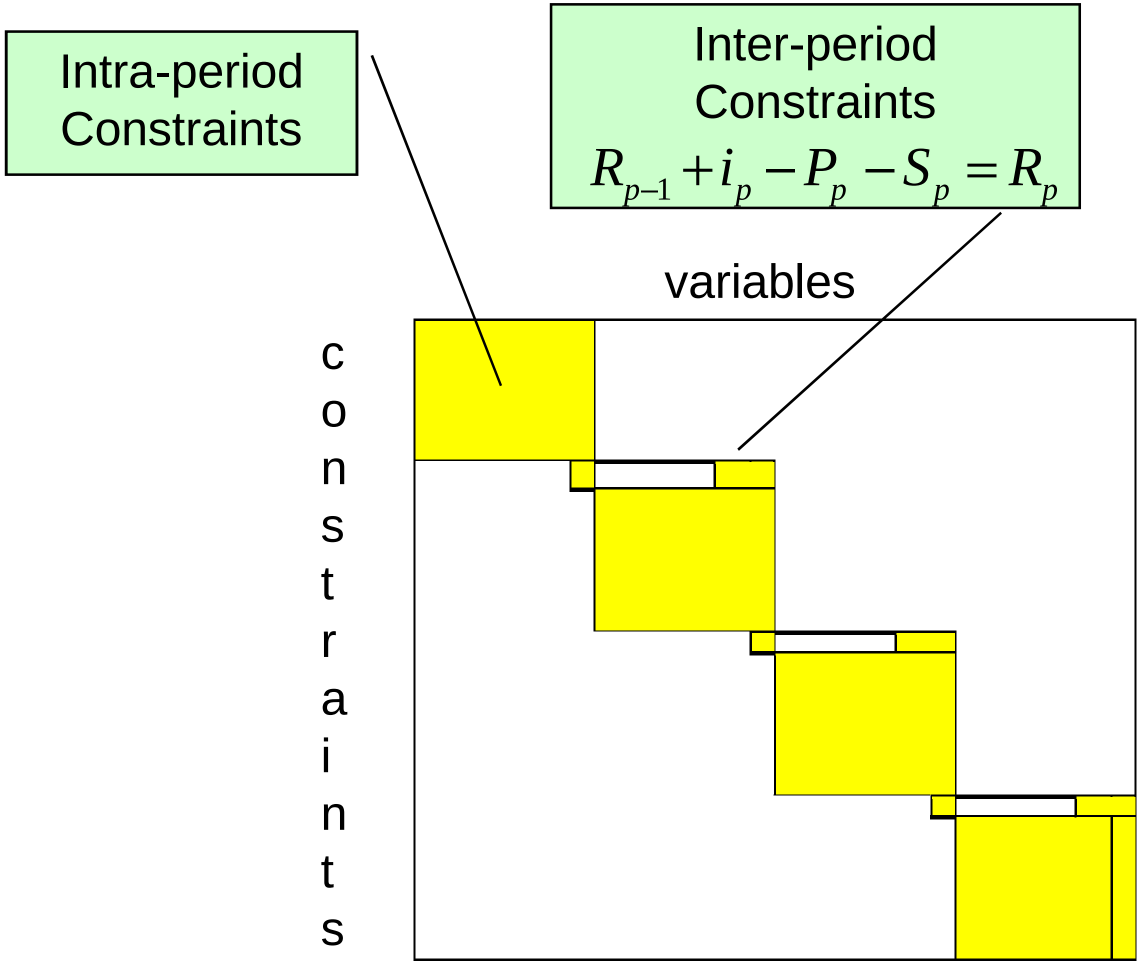
$$B_{p-1} x_{p-1} + A_p x_p = b_p \qquad p = 1, \ldots, P$$

$$x_p \geq 0$$

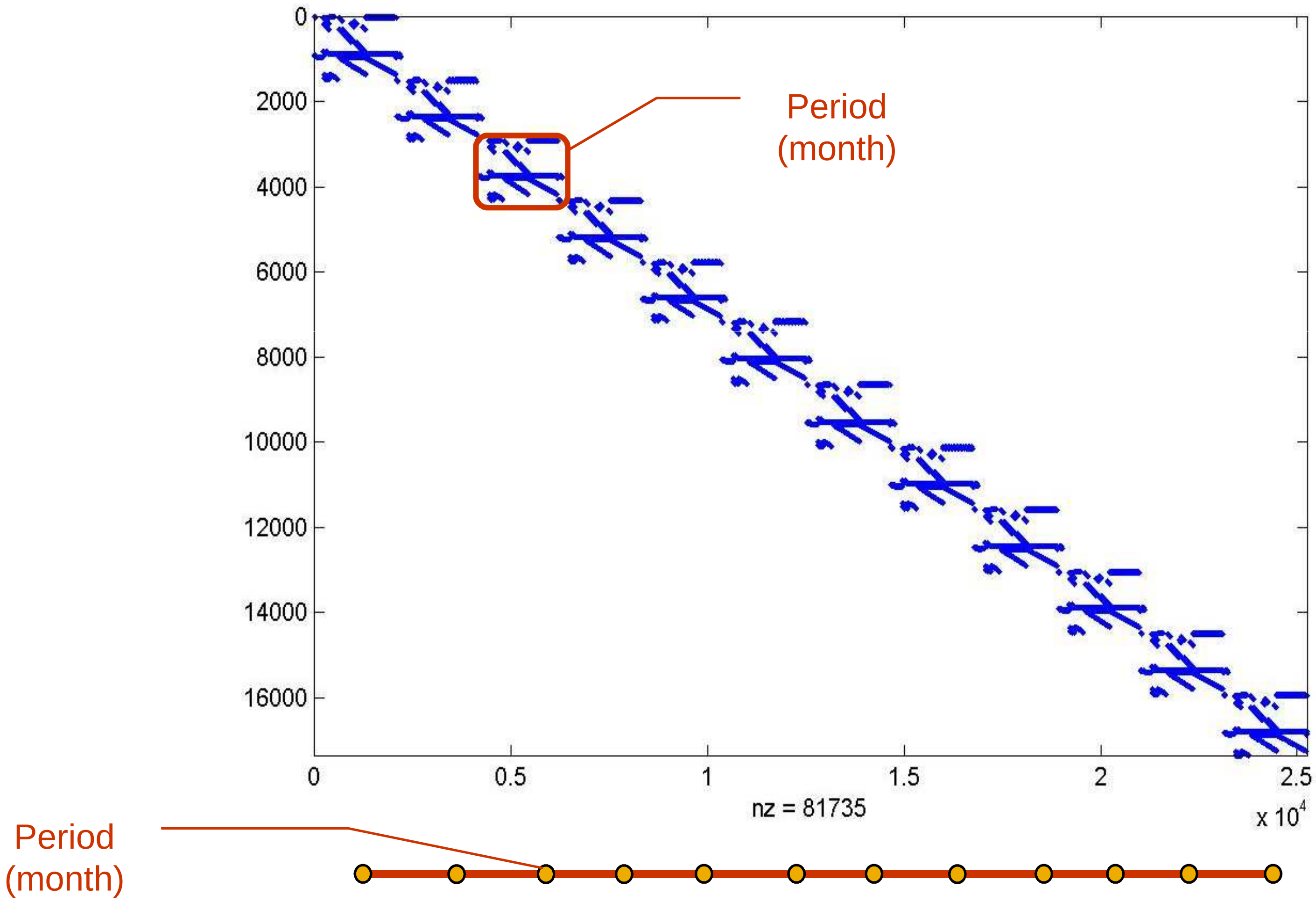
$$B_0 \equiv 0$$

Medium term hydrothermal scheduling problem: constraint matrix

R_{p-1}	reservoir level
i_p	inflow
P_p	hydro output
S_p	reservoir spillage



Medium term hydrothermal scheduling problem: constraint matrix



Multistage stochastic linear programming PLE-P

- O. F. minimizes expected costs of all the stages

$$\min_{x_p^{\omega_p}} \sum_{p=1}^{P} \sum_{\omega_p \in \Omega_p} p_p^{\omega_p} c_p^{\omega_p T} x_p^{\omega_p}$$

$$B_{p-1}^{\omega_p} x_{p-1}^{\omega_{p-1}} + A_p^{\omega_p} x_p^{\omega_p} = b_p^{\omega_p} \quad p = 1, \dots, P$$

$$x_p^{\omega_p} \geq 0$$

$$B_0^{\omega_1} \equiv 0$$

- Size grows exponentially with the number of scenarios
- Probabilities $p_p^{\omega_p}$ are conditioned
- Constraint matrix

A_1						
B_1^1	A_2^1					
B_1^2		A_2^2				
	B_2^1		A_3^1			
	B_2^2			A_3^2		
		B_2^3			A_3^3	
		B_2^4				A_3^4

3

Decomposition techniques

Decomposition techniques

- **Divide and conquer strategy**
  - Time division by periods
  - Spatial division by (thermal) units
  - Division by scenarios
- Allow the solution of huge problems (not directly solvable) with a certain structure by solving **iteratively small size** problems
- Objective function and feasible region have to be convex whenever to obtain dual variables is needed
- Dantzig-Wolfe 1960, Benders 1962, Geoffrion 1972 (generalized Benders)

Decomposition techniques: classification

- According to the **difficulties**
  - **Variables** (Benders)
  - **Constraints** (Dantzig-Wolfe or lagrangian relaxation)

- According to the **exchanged information** between master and subproblem
  - **Primal** (Benders)
  - **Dual** (Dantzig-Wolfe or lagrangian relaxation)

Coordinating mechanisms. Hydrothermal model

- **Primal (quantities)**
  - Master assigns an amount of water to release in each period or the reservoir levels at the end of the period
  - Each subproblem returns the marginal price (water value) associated to the use of the previous amount of water

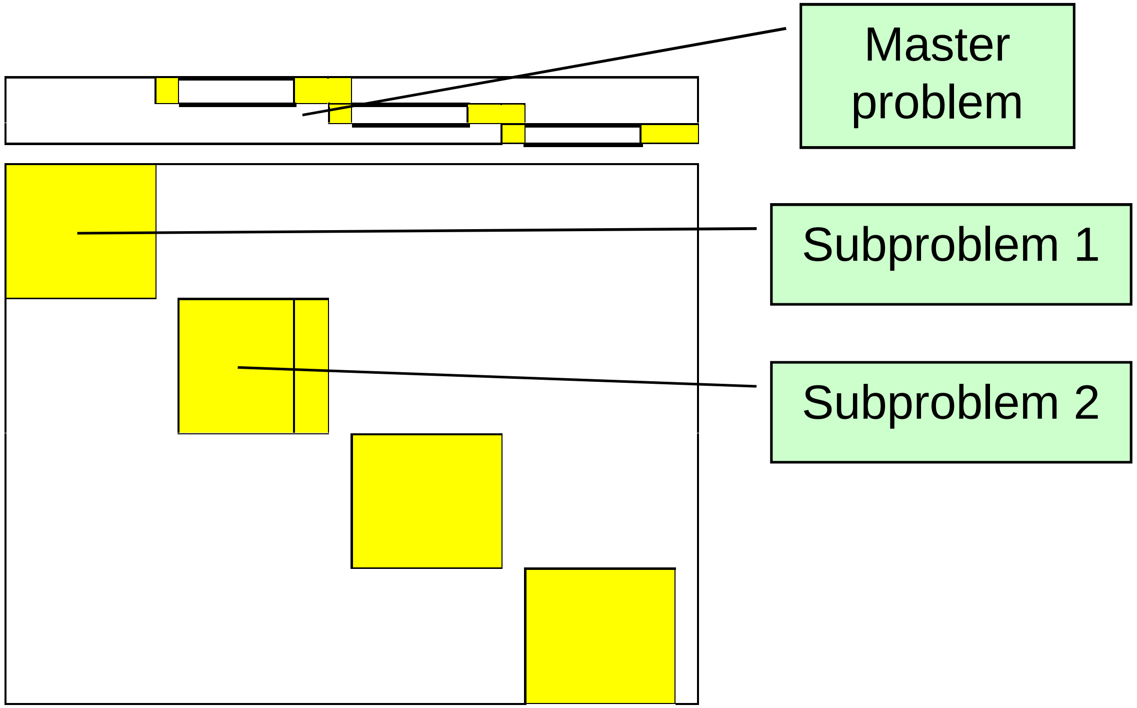
- **Dual (prices)**
  - Master gives a value to the water
  - Each subproblem returns the future cost function taking into account this value

Medium term hydrothermal scheduling problem

- Solvable by **Bd**, **DW-LR** or **nested Benders' decomposition**
  - Variables of hydro release complicate the solution $\Rightarrow$ Benders
  - Constraints of hydro release complicate the solution $\Rightarrow$ Dantzig-Wolfe, lagrangian relaxation
- Criterion:
  - Engineering: context dependent
  - Mathematical:
    - What complicates? (foreseeable number of iterations)
    - Respective size of master and subproblems

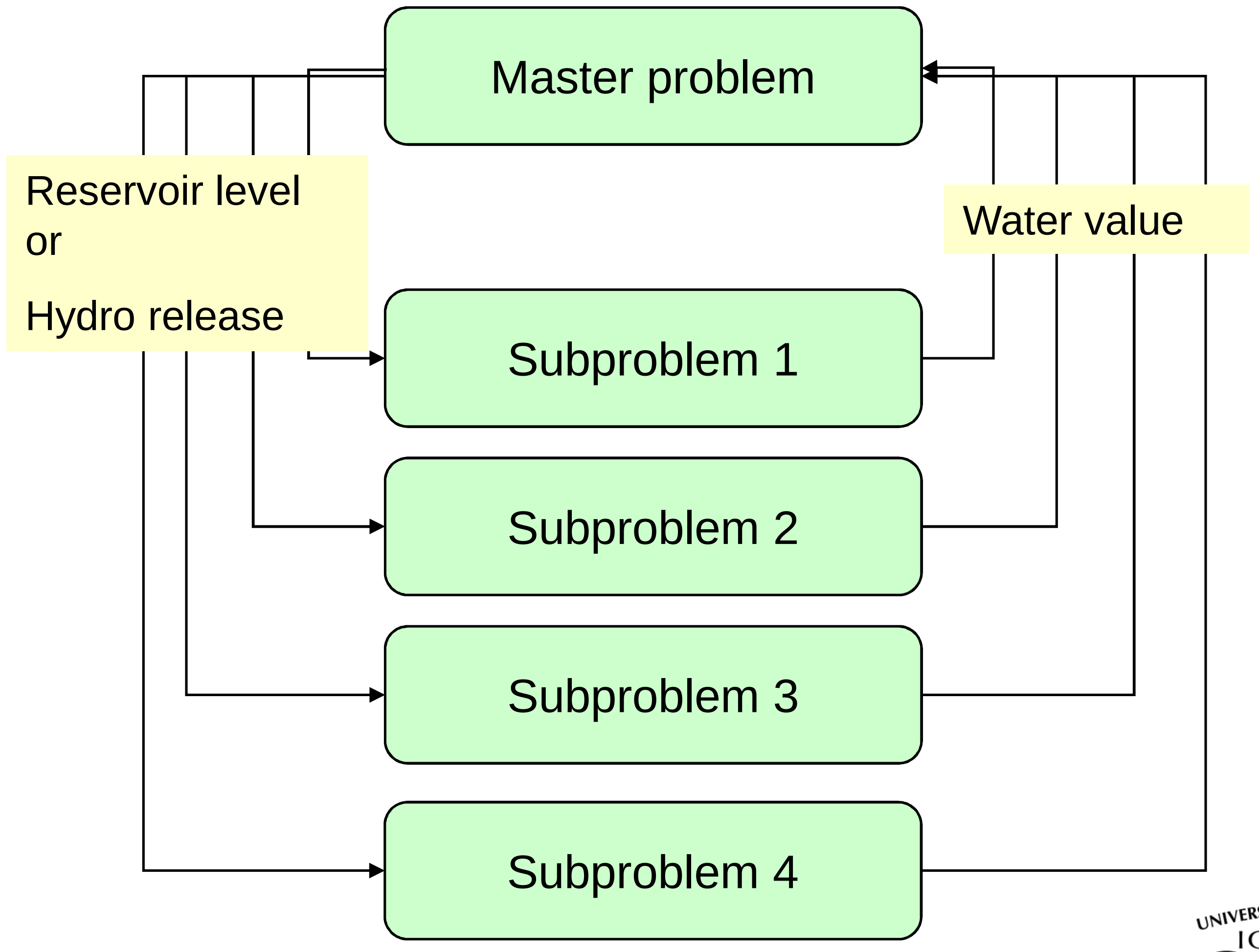
Algorithm: Benders

- Master problem: inter-period constraints
- Subproblem: intra-period constraints



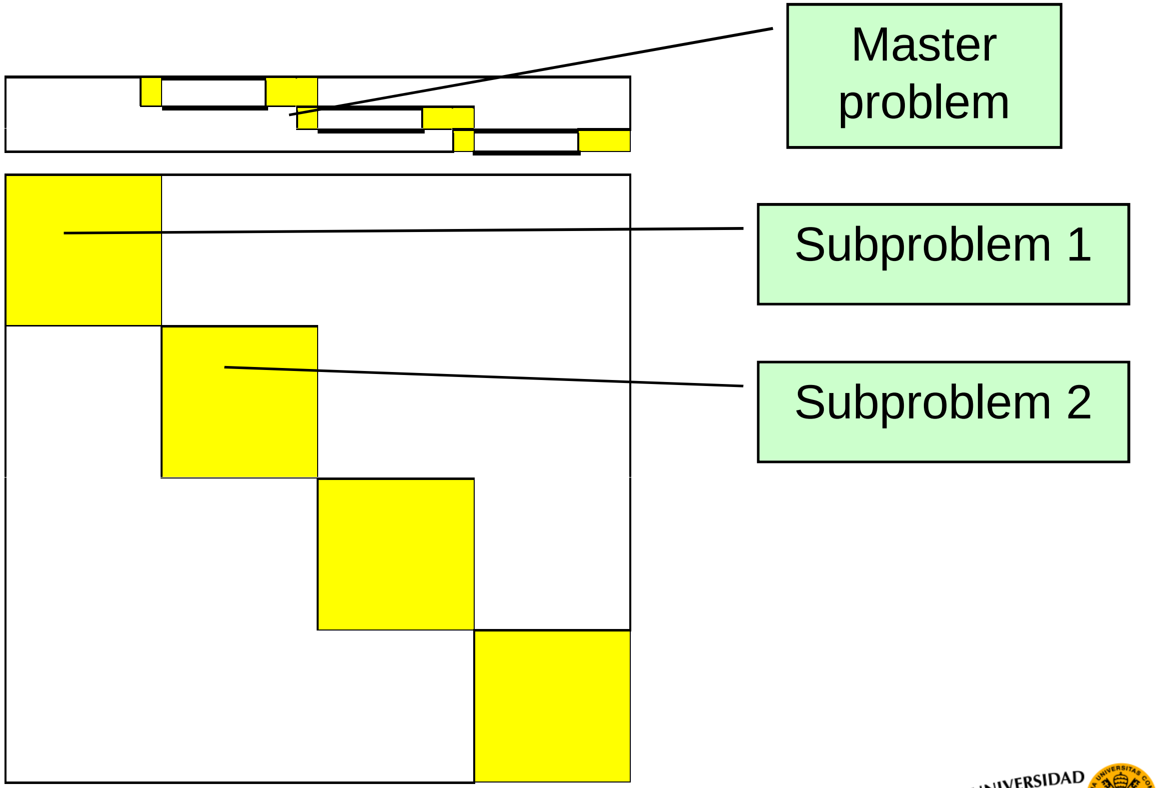
Benders' decomposition

- Reservoir level or hydro release is a given for the subproblem



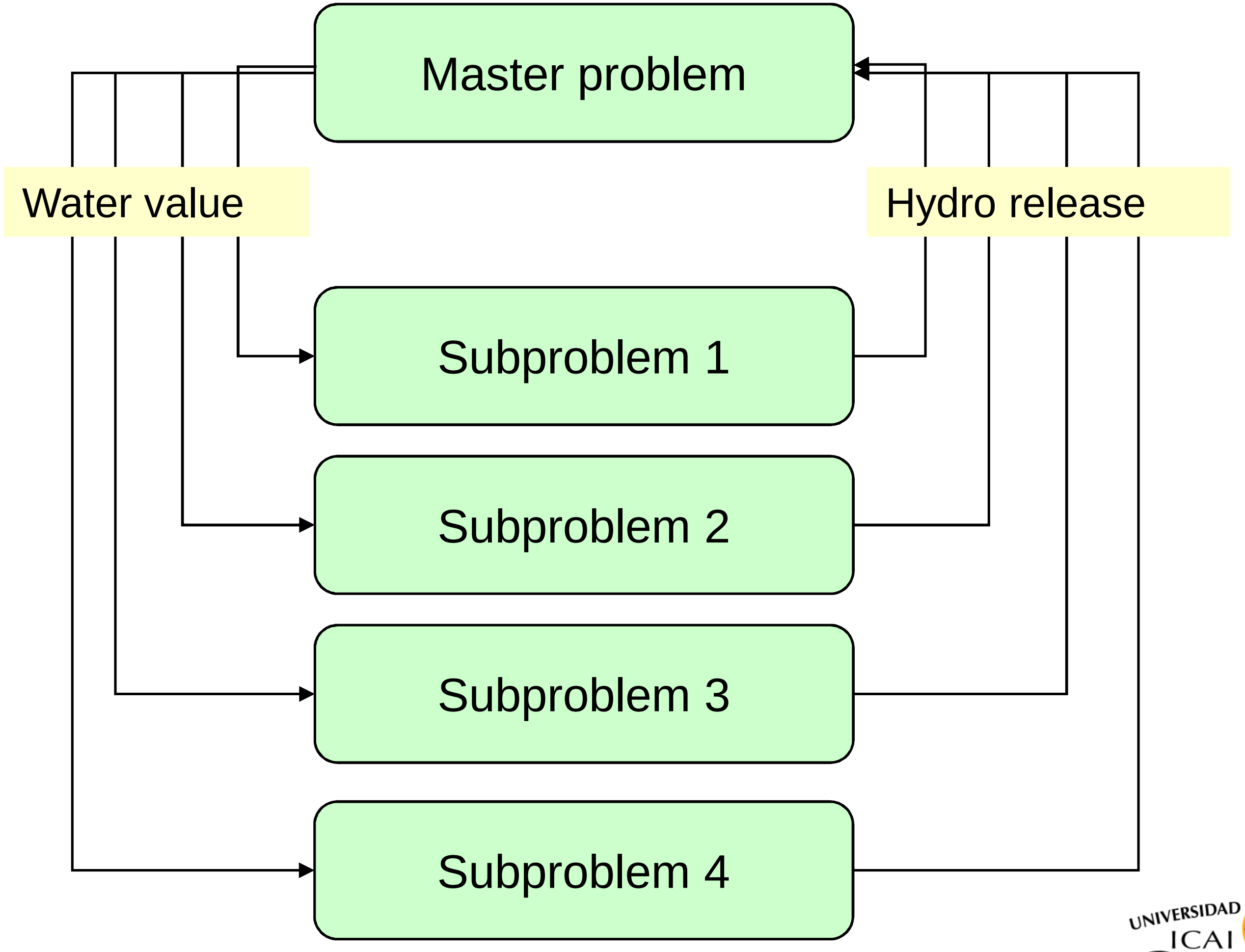
Algorithm: Dantzig-Wolfe or lagrangian relaxation

- Master problem: inter-period constraints
- Subproblem: intra-period constraints

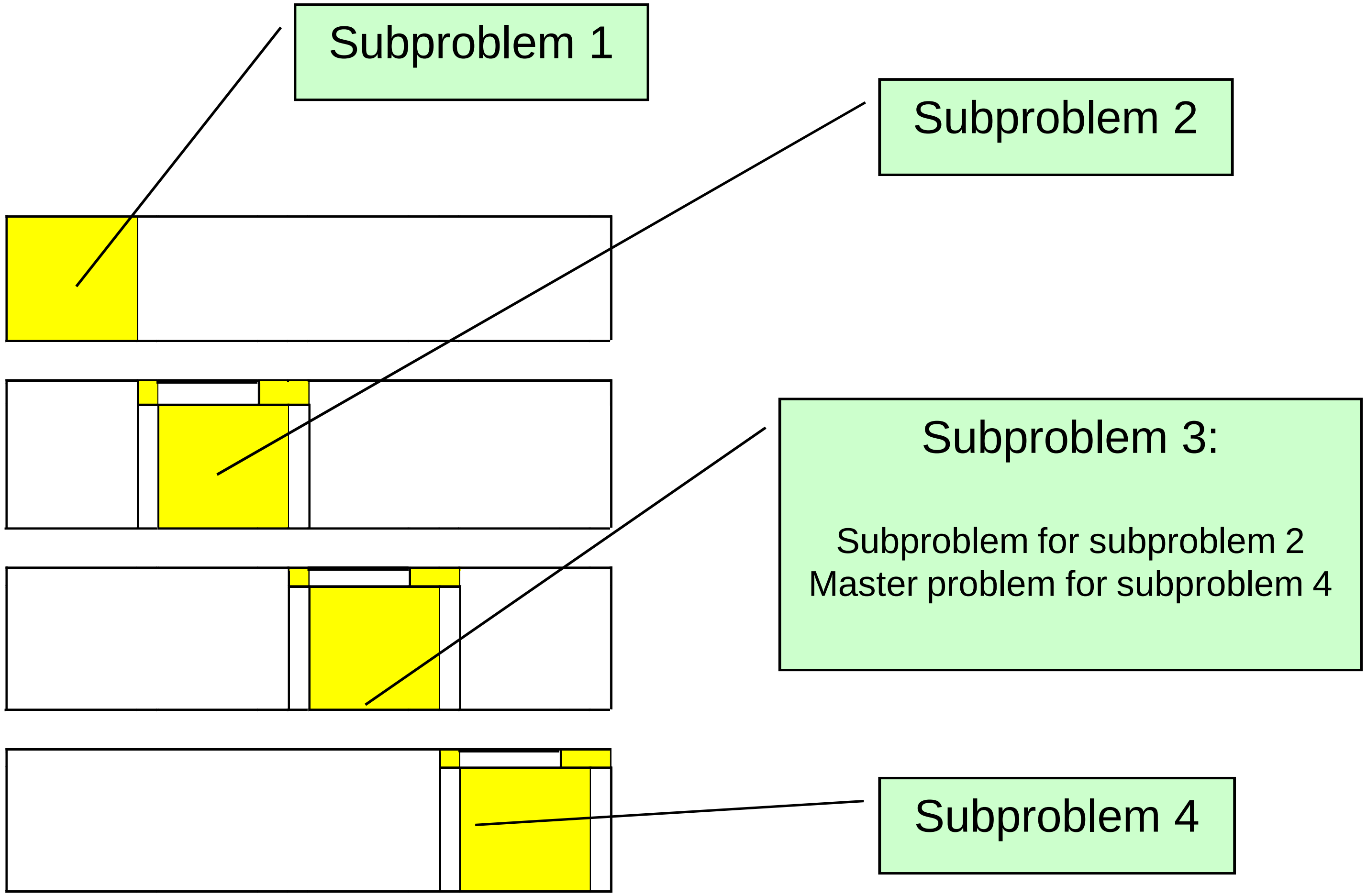


DW or LR decomposition

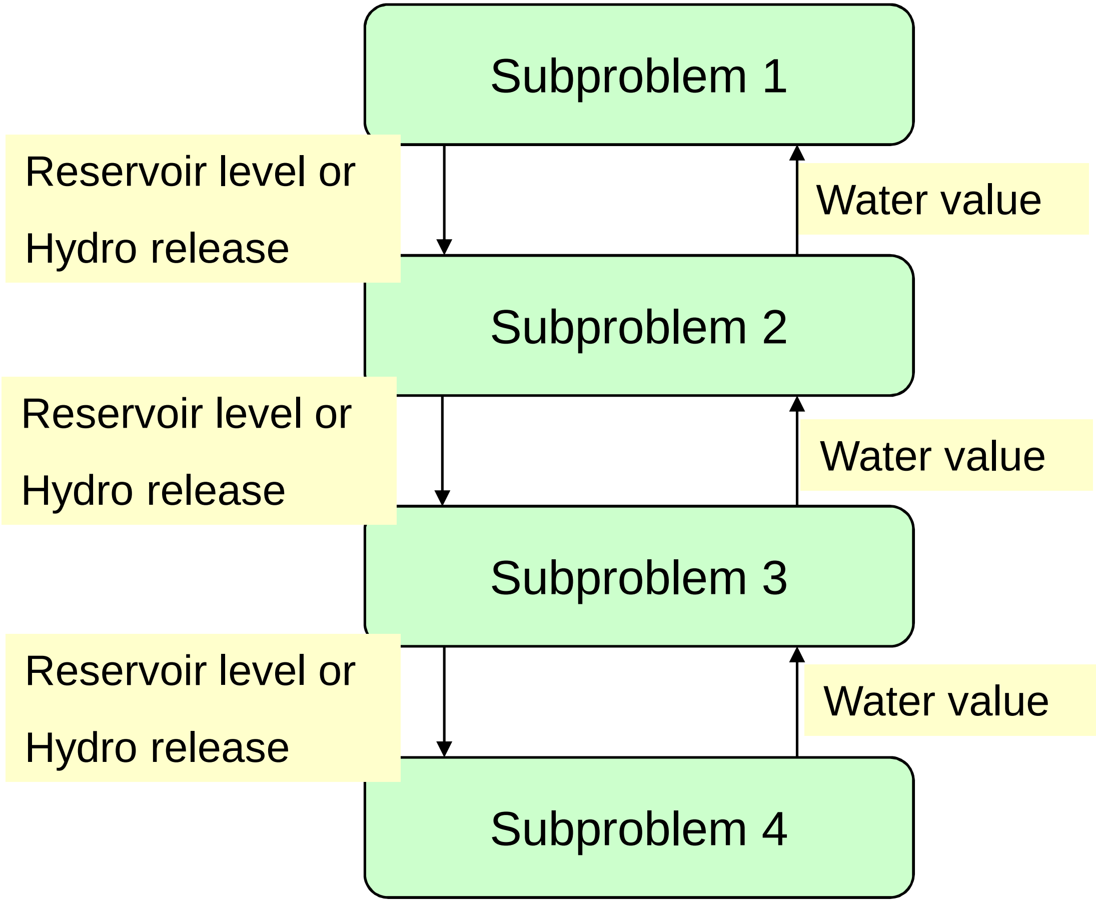
- Reservoir level is a variable for the subproblem



Algorithm: nested Benders' decomposition



Nested Benders' decomposition



4

Benders' decomposition

When to use Benders' decomposition?

- Variables x_1 complicate the solution of the problem
- Implicitly $n_1 << n_2$
- Number of iterations related with n_1
- Matrix structure induces separability of subproblems
- Master and subproblem have different nature
  - Master in discrete variables (MIP)
  - Subproblem with nonlinear (convex) objective function (NLP)
- Benders' decomposition needs convex o.f. and convex feasible region of the subproblem

Relaxed master problem and subproblem

- Master: first stage + cuts

$$
\begin{aligned}
&\min_{x_1,\theta_2} c_1^T x_1 + \theta_2 \\
&A_1 x_1 = b_1 \\
&\pi_2^{lT} B_1 x_1 + \theta_2 \ge f_2^l + \pi_2^{lT} B_1 x_1^l \qquad l = 1,\dots,j \\
&x_1 \ge 0
\end{aligned}
$$

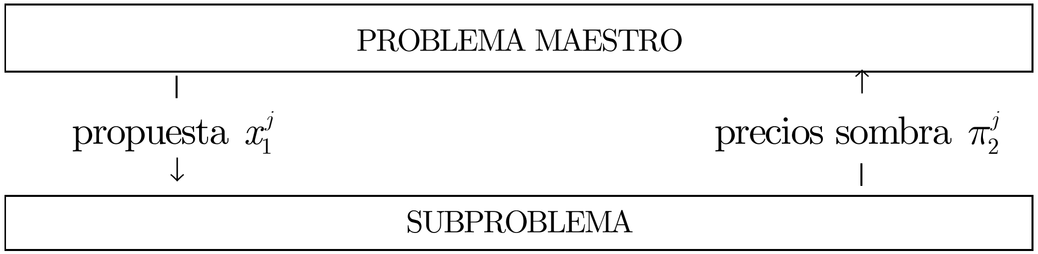
- Subproblem: second stage with known decisions of the first stage

$$
\begin{aligned}
f_2^j = \quad &\min_{x_2} c_2^T x_2 \\
&A_2 x_2 = b_2 - B_1 x_1^j \qquad : \pi_2^j \\
&x_2 \ge 0
\end{aligned}
$$

Benders' master and subproblem

- Master
  - One cut is added in each iteration
  - Each cut defines a new feasible region
  - Optimal solution of previous iteration becomes infeasible
  - It is worthy to use dual simplex method
  - Size $(m_1 + j) \times (n_1 + 1)$
  - It can be nonconvex (MIP, NLP)
- Subproblem
  - Each iteration modifies the RHS of the constraints
  - It is worthy to use primal simplex method (if size is adequate)
  - Size $m_2 \times n_2$
  - It must be convex (LP, NLP)

Benders' algorithm (i)



- Upper bound of the optimal value of the o.f. of problem PL-2
$$\overline{z} = c_1^T x_1^j + c_2^T x_2^j$$
- Lower bound of the problem, value obtained by the o.f. of the relaxed master problem
$$\underline{z} = c_1^T x_1^j + \theta_2^j$$
- Convergence condition
$$\frac{\left|\overline{z} - \underline{z}\right|}{\left|\overline{z}\right|} = \frac{\left|c_2^T x_2^j - \theta_2^j\right|}{\left|c_1^T x_1^j + c_2^T x_2^j\right|} \le \varepsilon$$

or repetition of the last master proposal

Benders' algorithm (ii)

- Successive approximation of the second-stage objective function by cuts.
- Benders' cuts (cutting planes, support hyperplanes) are an outer linearization of the recourse function.
- Lower bound is monotonous increasing.

  Upper bound is not necessarily monotonous decreasing.
  - Upper bound is the minimum of previous upper bounds
- In the first iteration the value of x_1^0 can be fixed, if the problem nature is known, or by solving the master problem without cuts $\theta_2 = 0$
- In each iteration we have a quasi-optimal feasible solution

Benders' algorithm (iii)

1. Initialization: $j=0$ $\underline{z}=-\infty$ $\overline{z}=\infty$ $\varepsilon=10^{-4}$
2. Solving the master problem

$$\begin{aligned}&\min_{x_1,\theta_2} c_1^T x_1 + \theta_2\\ &A_1 x_1 = b_1\\ &\pi_2^{lT} B_1 x_1 + \delta_1^l \theta_2 \geq \pi_2^{lT} b_2 \qquad l=1,\dots,j\\ &x_1 \geq 0\end{aligned}$$

$$\begin{aligned}&\min_{x_1,\theta_2} c_1^T x_1 + \theta_2\\ &A_1 x_1 = b_1\\ &\pi_2^{lT} B_1 x_1 + \delta_1^l \theta_2 \geq f_2^l + \pi_2^{lT} B_1 x_1^l \qquad l=1,\dots,j\\ &x_1 \geq 0\end{aligned}$$

Determine the solution (x_1^j, θ_2^j) and the lower bound

If no optimality cuts $\theta_2 = 0$

3. Solving the subproblem of sum of infeasibilities

$$\begin{aligned}f_2^j = \quad &\min_{x_2,v^+,v^-} e^T v^+ + e^T v^-\\ &A_2 x_2 + I v^+ - I v^- = b_2 - B_1 x_1^j \qquad : \pi_2^j\\ &x_2, v^+, v^- \geq 0\end{aligned}$$

If $f_2^j \geq 0$ infeasibility cut

If $f_2^j = 0$ go to step 4.

4. Solving the Benders' subproblem

$$\begin{aligned}f_2^j = \quad &\min_{x_2} c_2^T x_2\\ &A_2 x_2 = b_2 - B_1 x_1^j \qquad : \pi_2^j\\ &x_2 \geq 0\end{aligned}$$

Obtain x_2^j and update the upper bound.

3. If stopping rule is met

If not go to step 2.

$$\frac{|\overline{z}-\underline{z}|}{|\overline{z}|} = \frac{\left|c_2^T x_2^j - \theta_2^j\right|}{\left|c_1^T x_1^j + c_2^T x_2^j\right|} \leq \varepsilon$$

Fixed cost transportation problem

- Complete problem

$$\min_{x_{ij},y_{ij}} \sum_{ij} \left(c_{ij} x_{ij} + f_{ij} y_{ij} \right)$$

$$\sum_{j} x_{ij} \le a_i \quad \forall i$$

$$\sum_{i} x_{ij} \ge b_j \quad \forall j$$

$$x_{ij} \le M_{ij} y_{ij} \quad \forall ij$$

$$x_{ij} \ge 0, y_{ij} \in \{0,1\}$$

Flows

Investment decisions

Capacity of each origin

Demand of each destination

Flow can pass only for installed connections

Fixed cost transportation problem

- Master

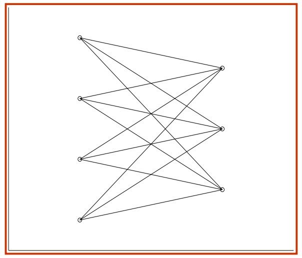
$$\min_{y_{ij},\theta} \theta + \sum_{ij} f_{ij} y_{ij}$$
$$\delta^l \theta + \sum_{ij} \pi_{ij}^l M_{ij} y_{ij} \geq f^l + \sum_{ij} \pi_{ij}^l M_{ij} y_{ij}^l \quad l = 1,\ldots,k$$
$$y_{ij} \in \{0,1\}$$

- Subproblem

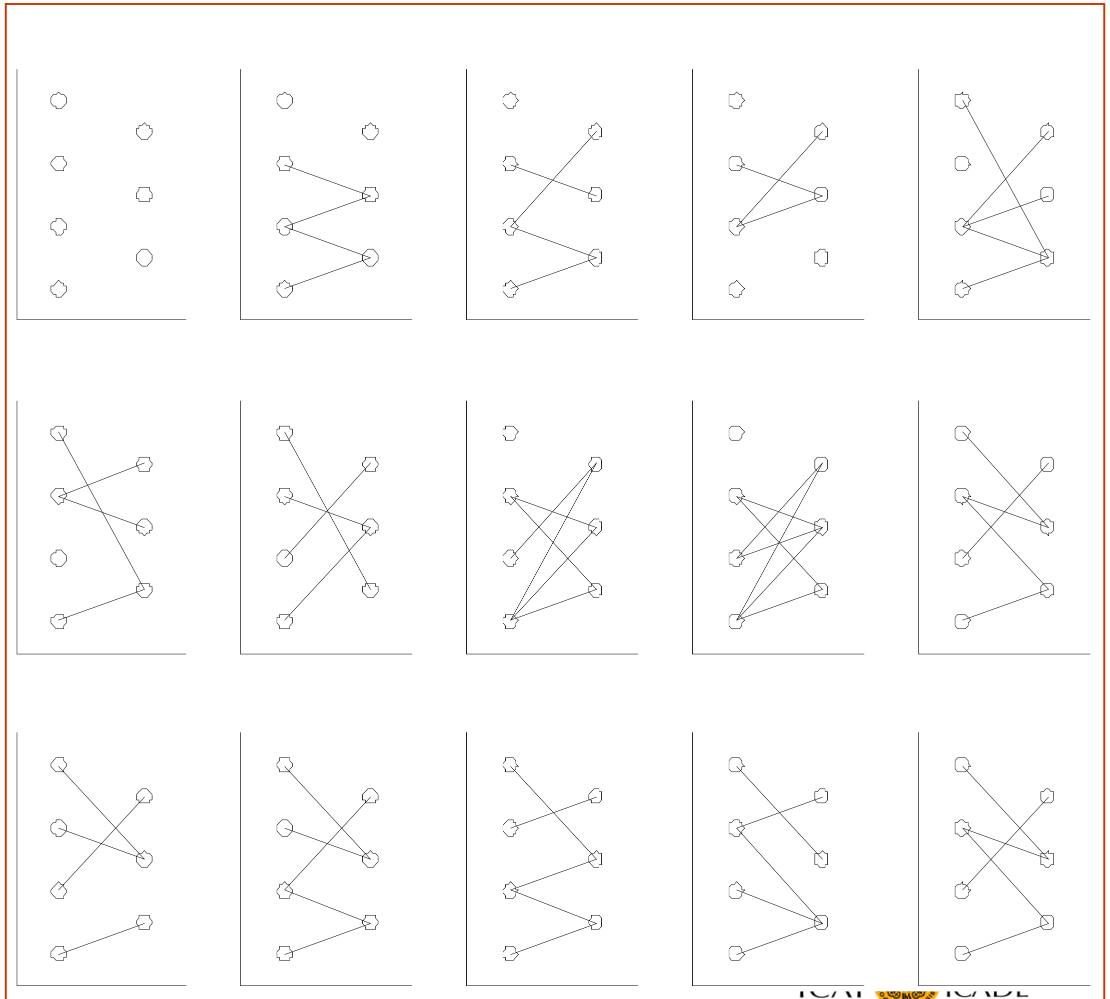
$$\min_{x_{ij}} \sum_{ij} c_{ij} x_{ij}$$
$$\sum_j x_{ij} \leq a_i \quad \forall i$$
$$\sum_i x_{ij} \geq b_j \quad \forall j$$
$$x_{ij} \leq M_{ij} y_{ij}^k \quad \forall ij \quad : \pi_{ij}^k$$
$$x_{ij} \geq 0$$

Case study. Solution

- Possible arcs



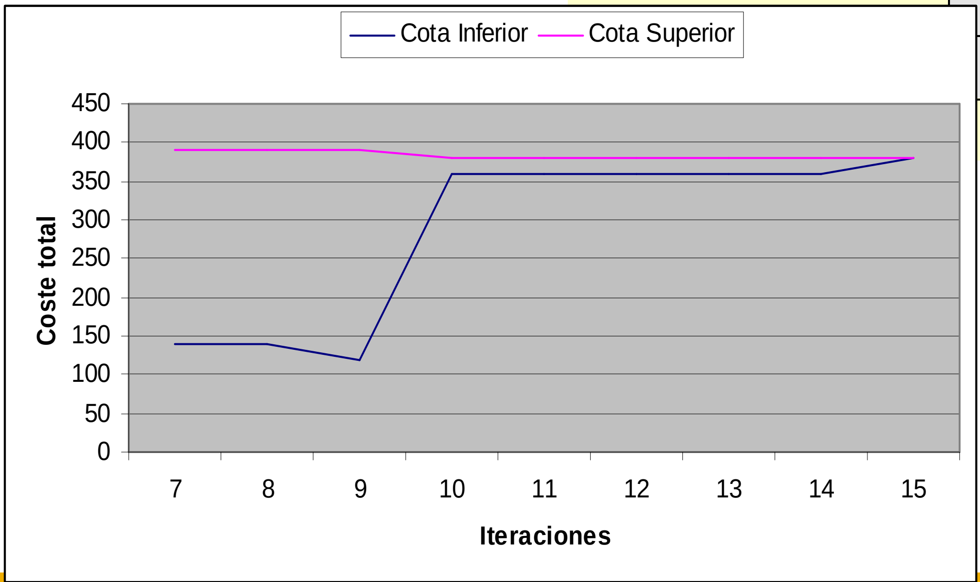
- Solutions along decomposition



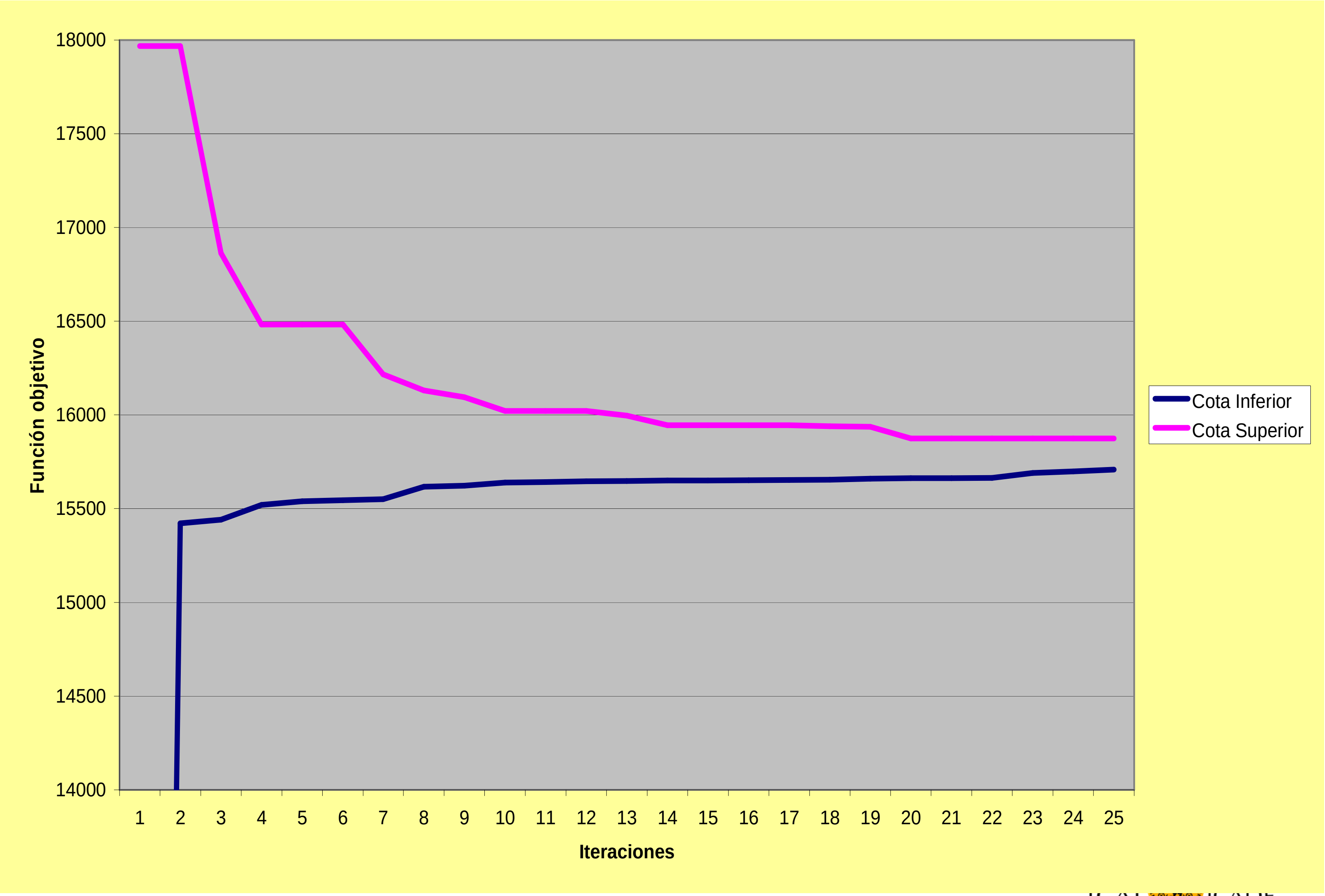
Case study. Convergence

- Lower bound is not increasingly monotonous because there are infeasible iterations

Iteración	Cota Inferior	Cota Superior
1 a 6	$-\infty$	∞
7	140	390
8	140	390
9	120	390
10	360	380
11	360	380
12	360	380
13	360	380
14	360	380
15	380	380



Convergence of hydrothermal scheduling model



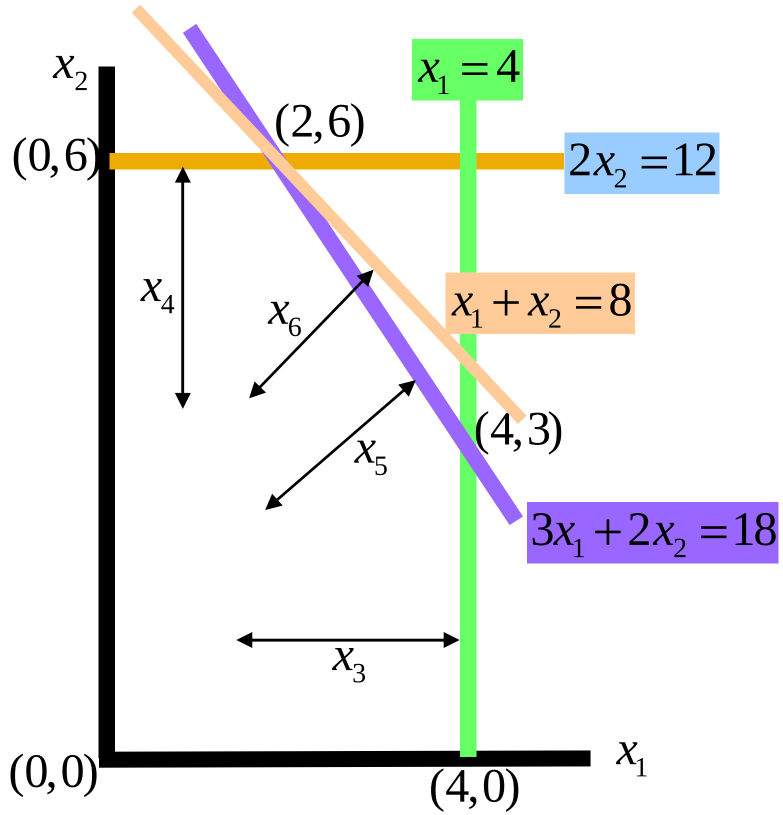
Achtung! Achtung!

- Degeneration in LP problems
  - In real cases it is frequent to find multiple optima (degeneration in dual problem) with the same or different basis. Given that decomposition techniques are based on dual variables you must be very careful in its computation
  - For example, in hydrothermal scheduling model formulated as LP it can exist spatial degeneration (system can produce with one plant or another) and temporal (system can produce now or in the future)

Primal degeneration

- Variable x_6 is degenerated (basic variable with value 0)

$$\begin{array}{lllllll}
\min z = -3x_1 - 5x_2 & & & & & & \\
x_1 & & +x_3 & & & & = 4 \\
 & 2x_2 & & +x_4 & & & = 12 \\
3x_1 & +2x_2 & & & +x_5 & & = 18 \\
x_1 & +x_2 & & & & +x_6 & = 8 \\
x_1, & x_2, & x_3, & x_4, & x_5, & x_6 & \geq 0
\end{array}$$



5

Nested Benders' decomposition

Nested Benders' Decomposition (i)

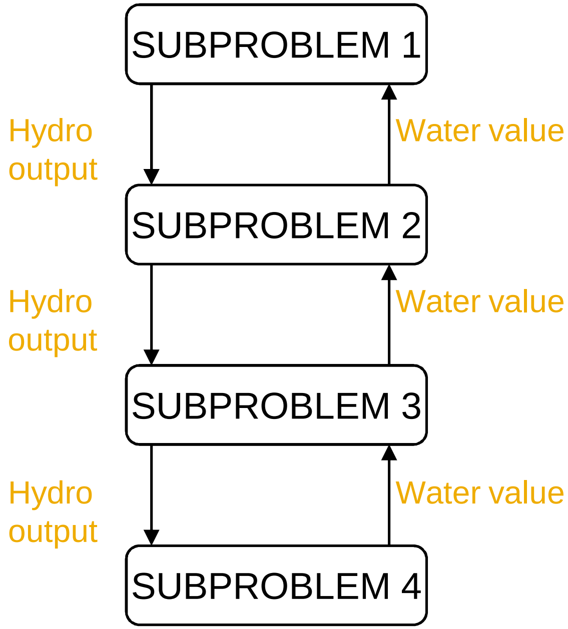
- Recursive application of decomposition technique.
- Let us see the PL-P problem:

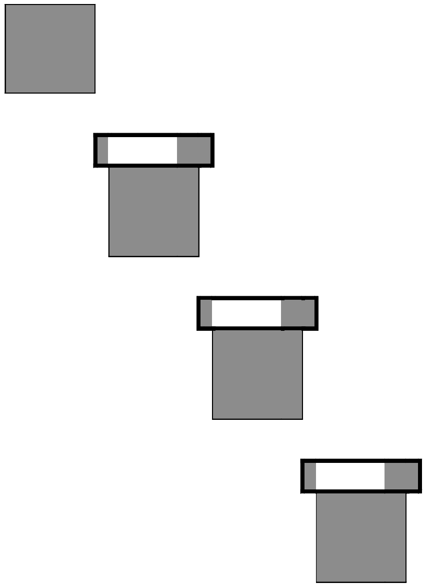
$$\min_{x_p} \sum_{p=1}^{P} c_p^T x_p$$
$$B_{p-1} x_{p-1} + A_p x_p = b_p \quad p = 1, \ldots, P$$
$$x_p \geq 0$$
$$B_0 \equiv 0$$

- We apply Benders' decomposition:
  - Stage 1 master, stage 2 to *P* subproblem
  - We decompose the subproblem that begins in stage 2
    - Stage 2 master, stage 3 to *P* subproblem
    - We decompose the subproblem that begins in stage 3
      - Stage 3 master, stage 4 to *P* subproblem
      - We decompose the subproblem that begins in stage 4

Nested Benders' Decomposition (ii)

- In stage *p*
  - The problem of this stage is solved
  - As a master it receives cuts from *p*+1 and passes the solution to *p*+1,
  - As a subproblem it builds cuts from *p*-1 and receives the solution from *p*-1.





Nested Benders' Decomposition

- Generic problem to solve

$$
\begin{aligned}
&\min_{x_p,\theta_{p+1}} c_p^T x_p + \theta_{p+1} \\
&A_p x_p = b_p - B_{p-1} x_{p-1}^l && : \pi_p \\
&\pi_{p+1}^{lT} B_p x_p + \theta_{p+1} \geq q_p = \pi_{p+1}^{lT} b_{p+1} + \eta_{p+1}^{lT} q_{p+1} && : \eta_p \quad l = 1,\dots,j \\
&x_p \geq 0 \\
&\theta_{P+1} \equiv 0 \\
&B_0 \equiv 0 \\
&\pi_{P+1}^l \equiv 0 \\
&\eta_{P+1}^l \equiv 0
\end{aligned}
$$

$$
\begin{aligned}
&\min_{x_p,\theta_{p+1}} c_p^T x_p + \theta_{p+1} \\
&A_p x_p = b_p - B_{p-1} x_{p-1}^l && : \pi_p \\
&\pi_{p+1}^{lT} B_p x_p + \theta_{p+1} \geq f_{p+1}^l + \pi_{p+1}^{lT} B_p x_p^l && : \eta_p \quad l = 1,\dots,j \\
&x_p \geq 0 \\
&\theta_{P+1} \equiv 0 \\
&B_0 \equiv 0 \\
&\pi_{P+1}^l \equiv 0 \\
&\eta_{P+1}^l \equiv 0
\end{aligned}
$$

- Problem converges when first stage does

9

Decomposition in two-stage and multistage stochastic programming

Two-stage stochastic linear programming PLE-2

- O.F. minimizes first-stage costs and expected value of second-stage costs

$$
\begin{array}{llr}
\min\limits_{x_1, x_2^\omega} c_1^T x_1 + \sum\limits_{\omega \in \Omega} p^\omega c_2^{\omega T} x_2^\omega & & \\
A_1 x_1 & & = b_1 \\
B_1^\omega x_1 & + A_2^\omega x_2^\omega & = b_2^\omega \\
x_1, & x_2^\omega & \geq 0
\end{array}
$$

- If A_2^ω doesn't depend on ω it is called fixed resource
- Structure of the constraint matrix

A_1			
B_1^1	A_2^1		
B_1^2		A_2^2	
B_1^3			A_2^3

Deterministic equivalent problem

- State space is small
- Formulation of the deterministic equivalent problem

$$\begin{array}{lllll}
\min\limits_{x_1, x_2^{\omega_1}, x_2^{\omega_2}, x_2^{\omega_3}} & c_1^T x_1 + p^{\omega_1} c_2^{\omega_1 T} x_2^{\omega_1} + p^{\omega_2} c_2^{\omega_2 T} x_2^{\omega_2} + p^{\omega_3} c_2^{\omega_3 T} x_2^{\omega_3} & & & \\
A_1 x_1 & & & & = b_1 \\
B_1^{\omega_1} x_1 & + A_2^{\omega_1} x_2^{\omega_1} & & & = b_2^{\omega_1} \\
B_1^{\omega_2} x_1 & & + A_2^{\omega_2} x_2^{\omega_2} & & = b_2^{\omega_2} \\
B_1^{\omega_3} x_1 & & & + A_2^{\omega_3} x_2^{\omega_3} & = b_2^{\omega_3} \\
x_1, & x_2^{\omega_1}, & x_2^{\omega_2}, & x_2^{\omega_3}, & \geq 0
\end{array}$$

- In Benders' decomposition subproblem results separable and has the same structure in the constraints

$$c_2 = \begin{pmatrix} p^{\omega_1} c_2^{\omega_1} \\ p^{\omega_2} c_2^{\omega_2} \\ p^{\omega_3} c_2^{\omega_3} \end{pmatrix} \quad B_1 = \begin{pmatrix} B_1^{\omega_1} \\ B_1^{\omega_2} \\ B_1^{\omega_3} \end{pmatrix} \quad A_2 = \begin{pmatrix} A_2^{\omega_1} & & \\ & A_2^{\omega_2} & \\ & & A_2^{\omega_3} \end{pmatrix} \quad b_2 = \begin{pmatrix} b_2^{\omega_1} \\ b_2^{\omega_2} \\ b_2^{\omega_3} \end{pmatrix} \quad x_2 = \begin{pmatrix} x_2^{\omega_1} \\ x_2^{\omega_2} \\ x_2^{\omega_3} \end{pmatrix}$$

Decomposition in PLE-2

- Master monocut

$$\min_{x_1,\theta_2} c_1^T x_1 + \theta_2$$
$$\begin{aligned} A_1 x_1 &= b_1 \\ \sum_{\omega\in\Omega} p^\omega \pi_2^{\omega l T} B_1^\omega x_1 + \theta_2 &\geq \sum_{\omega\in\Omega} p^\omega \pi_2^{\omega l T} b_2^\omega \qquad l = 1,\dots,j \\ x_1 &\geq 0 \end{aligned}$$

and multicut

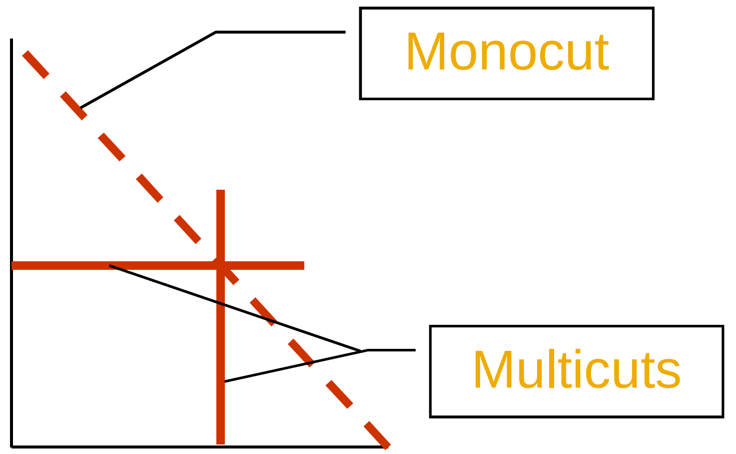
$$\min_{x_1,\theta_2^\omega} c_1^T x_1 + \sum_{\omega\in\Omega} p^\omega \theta_2^\omega$$
$$\begin{aligned} A_1 x_1 &= b_1 \\ \pi_2^{\omega l T} B_1^\omega x_1 + \theta_2^\omega &\geq \pi_2^{\omega l T} b_2^\omega \qquad \omega\in\Omega \qquad l = 1,\dots,j \\ x_1 &\geq 0 \end{aligned}$$

- Subproblem

$$\min_{x_2^\omega} c_2^{\omega T} x_2^\omega$$
$$A_2^\omega x_2^\omega = b_2^\omega - B_1^\omega x_1^l \qquad : \pi_2^\omega$$
$$x_2^\omega \geq 0$$

Monocut vs. multicut

- Monocut $(n_1+1)\text{x}(m_1+\text{j})$.
  Multicut $(n_1+\Omega)\text{x}(m_1+j\Omega)$
- Multicut convenient when m_2 is large and Ω no much larger than n_1. Requires less Benders iterations but more cumbersome
- Multicut approximates independently each scenario. Monocut approximates the weighted sum of scenarios



Multistage stochastic linear programming PLE-P

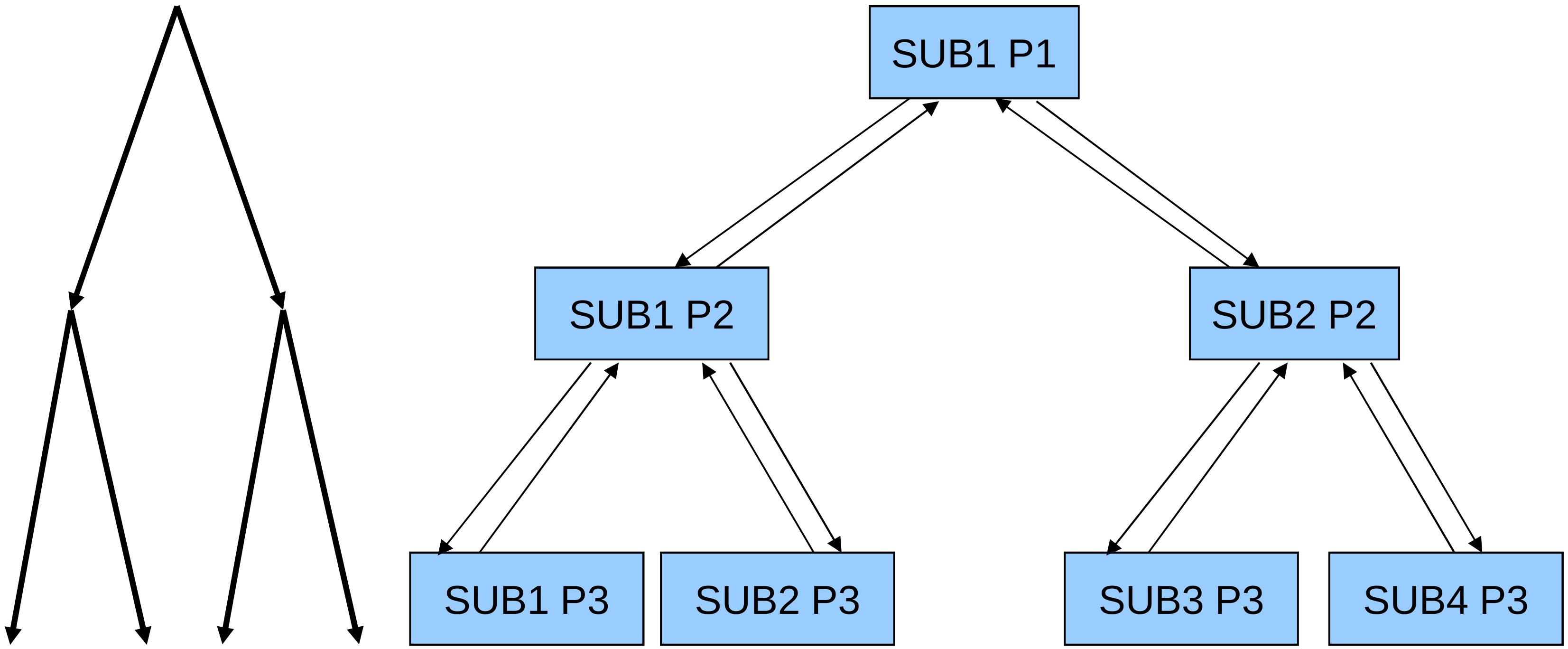
- O.F. minimizes expected costs of all the stages

$$
\begin{aligned}
&\min_{x_p^{\omega_p}} \sum_{p=1}^{P} \sum_{\omega_p \in \Omega_p} p_p^{\omega_p} c_p^{\omega_p T} x_p^{\omega_p} \\
&B_{p-1}^{\omega_p} x_{p-1}^{\omega_{p-1}} + A_p^{\omega_p} x_p^{\omega_p} = b_p^{\omega_p} \quad p = 1, \dots, P \\
&x_p^{\omega_p} \geq 0 \\
&B_0^{\omega_1} \equiv 0
\end{aligned}
$$

- Probabilities $p_p^{\omega_p}$ are conditional
- Constraint matrix

A_1						
B_1^1	A_2^1					
B_1^2		A_2^2				
	B_2^1		A_3^1			
	B_2^2			A_3^2		
		B_2^3			A_3^3	
		B_2^4				A_3^4

Multistage stochastic problem. Nested Benders' decomposition



Stochastic multistage decomposition

Step 0 Set $I_t^{\xi_t} = J_t^{\xi_t} = 0$. Set $\theta_t^{\xi_t} \equiv 0$ at the initial iteration

Step 1 Forward pass:

Repeat for $t = 1,\ldots,T$

- Repeat for each node ξ_t of stage t
  - Solve $(RP_t^{\xi_t})$
  - If feasible: obtain solution $x_t^{\xi_t}$
    - If $t = 1$ obtain lower bound $\underline{z} = v(RP_1^{\xi_1})$
  - If infeasible: stop forward pass, set $T' = t$ and go to Step 4

Step 2 Upper bound computation:

Evaluate objective function of the complete problem with the primal solutions so far obtained. $\overline{z} = v(P)$

Step 3 (stopping rule)

If $\overline{z} - \underline{z} < tol$ stop, $x_t^{\xi_t}$ is optimal solution, else go to Step 4

Step 4 Backward pass

Repeat for $t = T',\ldots,1$

- Repeat for each node ξ_t of stage t
  - Solve $(RP_t^{\xi_t})$
  - If feasible: obtain objective $\theta_t^{\xi_t,i} = v(RP_t^{\xi_t})$ and dual values $\pi_t^{\xi_t,i}$
    - Augment $I_t^{\xi_t} = I_t^{\xi_t} + 1$
  - If infeasible: obtain sum of infeasibilities $\tilde{\theta}_t^{\xi_t,j}$ and dual values $\tilde{\pi}_t^{\xi_t,j}$
    - Augment $J_t^{\xi_t} = J_t^{\xi_t} + 1$

Go to step 1

Improvements in decomposition techniques

Improvements in decomposition techniques (i)

- Optimization method used for the problems

  Subproblems solved many times with modifications. In Benders' decomposition the master adds constraints and the subproblem change the constraint RHS.

  – Simplex dual method is the initial candidate. Try simplex method or interior point method.
  – Warm start: Use of previous bases (option BRATIO in GAMS).
  – Use of initial point taken from the deterministic equivalent problem for a scenario.

- Advanced start procedures generate preliminary cuts prior to initiating a formal Benders algorithm

Improvements in decomposition techniques (ii)

- Tree traversing strategies: Ways to traverse through the tree from the root to the leaves. The “best” tree traversing strategy will properly balance the quality of the cuts (and hence the lower bound) with the computational effort required to generate them
  - Fast-pass: from 1 to P and from P-1 to 1
  - Shuffle: solves the stage with largest error between lower and upper bounds. It is centered in final stages, never goes backward until the error of a stage is bounded (*fast-forward*)
  - Cautious: goes forward when the error in a stage is small enough. It is centered in initial stages, never goes forward until the error of a stage is bounded (*fast-backward*)

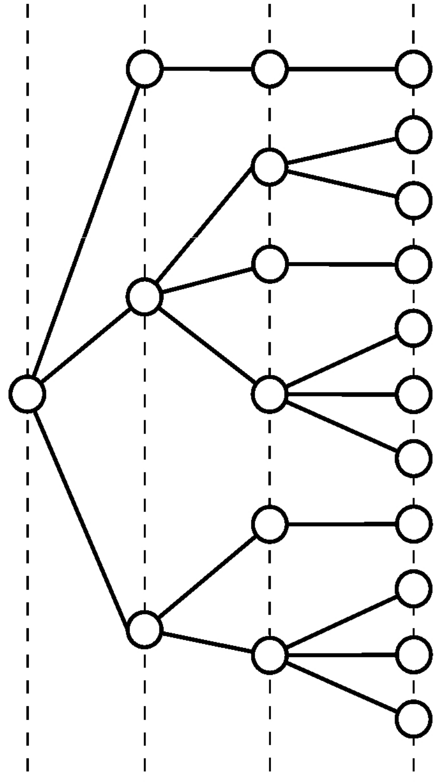
D.P. Morton *An enhanced decomposition algorithm for multistage stochastic hydroelectric scheduling* Annals of Operations Research 1996 Vol. 64 p. 211 - 235

Improvements in decomposition techniques (iii)

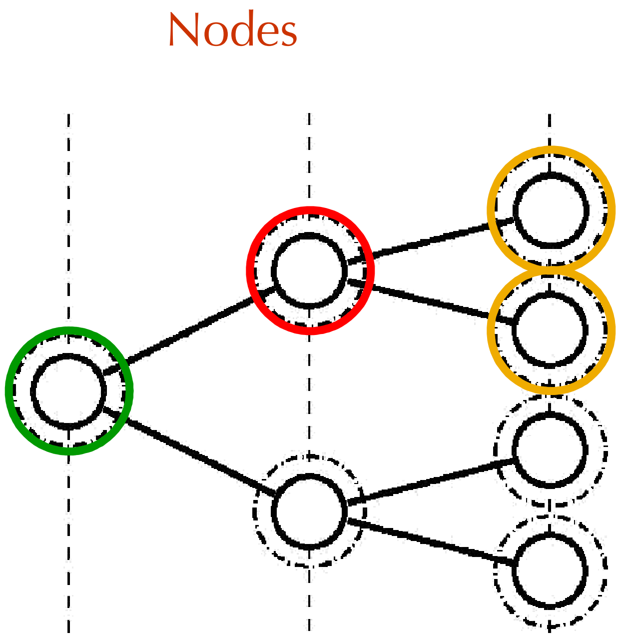
- Formulation and cut aggregation
  - Linear or nonlinear type (linearization around a point)
  - Monocut or multicut
  - More cuts $\Rightarrow$ more information to the master $\Rightarrow$ less iterations. On the other hand, more variables and more constraints in the master

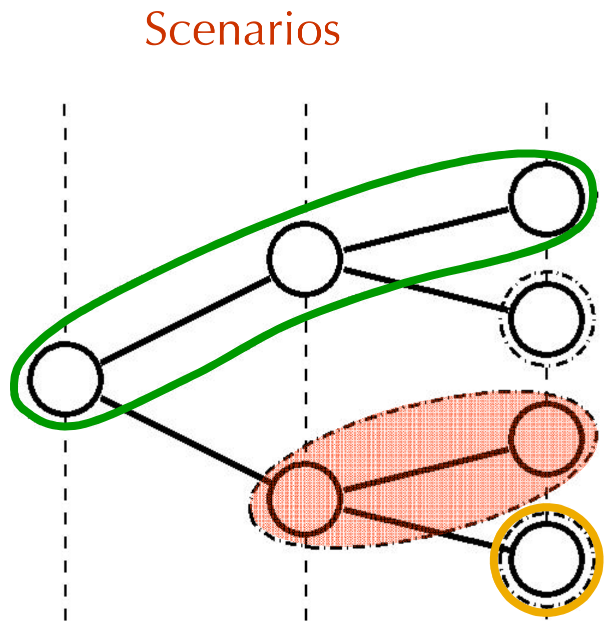
Improvements in decomposition techniques (iv)

- Tree partition or node aggregation (multicoordination)
  - Advantage: reduction in decomposition algorithm iterations
  - Disadvantage: potential increase in problem solution time (interior point method)
  - Methods
    - By nodes
    - By scenarios
    - By subtrees
    - By complete scenarios
    - By graph partition



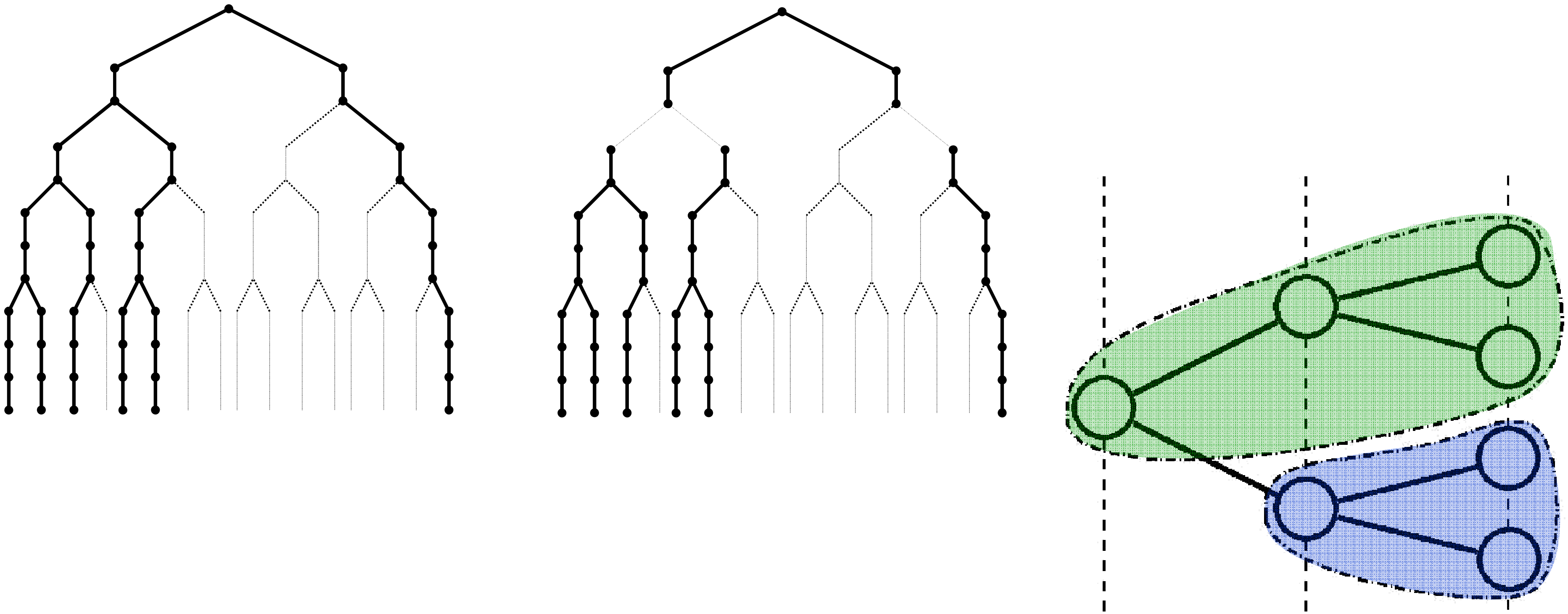
Node and scenario partition





Subtree partition

- Ascendant node aggregation (from the leaves to the root) is the one with the best performance



S. Cerisola, A. Ramos *Node Aggregation in Stochastic Nested Benders Decomposition Applied to Hydrothermal Coordination* 6th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS) RUM-102. Madeira, Portugal September 2000

Decomposition in grid computing

- Distributed computing
  - J.M. Latorre *Resolución distribuida de problemas de Optimización estocástica. Aplicación al problema de coordinación hidrotérmica* Doctoral thesis. Universidad Pontificia Comillas. November 2007
  - J.M. Latorre, S. Cerisola, A. Ramos, R. Palacios *Analysis of Stochastic Problem Decomposition Algorithms in Computational Grids* Annals of Operations Research 166 (1): 355-373 Feb 2009
- GAMS grid
  - Use of multiple cores of a computer

Multistage stochastic integer programming PLE-P

- Decomposition for multistage problems with integer variables in each stage
  - S. Cerisola *Benders' decomposition for mixed integer problems. Application to a medium term hydrothermal coordination problem* Doctoral thesis. Universidad Pontificia Comillas. April 2004
  - S. Cerisola, A. Baillo, J.M. Fernandez-Lopez, A. Ramos, R. Gollmer *Stochastic Power Generation Unit Commitment in Electricity Markets: A Novel Formulation and A Comparison of Solution Methods* Operations Research 57 (1): 32-46 Jan-Feb 2009
  - S. Cerisola, J.M. Latorre, A. Ramos *Stochastic Dual Dynamic Programming Applied to Nonconvex Hydrothermal Models* European Journal of Operational Research 218 (2012) 687–697 10.1016/j.ejor.2011.11.040

Simulation in stochastic optimization

Why do we need simulation?

- It is used when the number of states of random parameters too high
- Computation of expectation in the recourse function (multicut) or expectation in the cut terms (monocut)
- Equivalent to integrate or sample in the random parameter hyperspace with known probability density function. A sample is a combination of random parameter values
- Each sample is computationally cumbersome (solving an LP problem)
- Simulate is equivalent to integrate or sample in hyperspace of random parameters with a known probability density function

Types of sampling

- External sampling
  - We take samples to reduce the problem size and then we solve the stochastic optimization problem
    - In SDDP we take samples in the forward pass
- Internal sampling
  - We take samples at the same time that we solve the stochastic optimization problem
    - In a two-stage planning problem with the expected value for the second state substituted by the sample mean of the second stage

12

Stochastic dual dynamic programming

Stochastic Dual Dynamic Progamming SDDP

- Nested Benders' decomposition with
  - Forward pass: node sampling instead of solving all the nodes
  - Backward pass: solution of all the nodes of the recombining tree. It approximates for each scenario the recourse function for the sampled values obtained in the forward iteration
- Therefore we have stochastic convergence
  - Lower bound is deterministic while upper bound is stochastic
  - Stopping criterion: lower bound enters the confidence interval of the upper bound
  - I.e., if the stopping criterion is 1% for a confidence level of 95%. The algorithm stops with we are sure with a 95% that the relative difference between upper and lower bounds is lower than 1%

Stochastic Dual Dynamic Progamming SDDP

Step 0 Set $I_t^{\xi_t} = J_t^{\xi_t} = 0$. Set $\theta_t^{\xi_t} \equiv 0$ at the initial iteration

Step 1 Simulate N scenarios $(h_t^{\xi_t})^n$, $n: 1,\ldots,N$, $t = 1,\ldots,T$

Forward pass:

Repeat for $n: 1,\ldots,N$

Repeat for $t = 1,\ldots,T$

Solve $(RP_t^{\xi_t})$ with r. hand side value $(h_t^{\xi_t})^n$ and obtain solution $(x_t^{\xi_t})^n$

If $t = 1$ obtain lower bound $\underline{z} = v(RP_1^{\xi_1})$

If infeasible: stop forward pass for simulation n

Step 2 Upper bound computation:

Evaluate objective function of the complete (deterministic) problem for each of the primal solutions so far obtained. $\overline{z} = \frac{1}{N}\sum_{t=1}^{T} c_t \left(x_t^{\xi_t}\right)^n$

Step 3 (stopping rule)

If $\overline{z} - \underline{z} < tol$ stop, $x_1^{\xi_1}$ is optimal solution, else go to Step 4

Step 4 Backward pass

Repeat for $t = T,\ldots,1$

Repeat for each node ξ_t of stage t

Repeat for each proposal obtained in forward pass, modifying the right hand side value of subproblem $(RP_t^{\xi_t})$

Solve $(RP_t^{\xi_t})$

If feasible: obtain objective $\theta_t^{\xi_t,i} = v(RP_t^{\xi_t})$ and dual values $\pi_t^{\xi_t,i}$

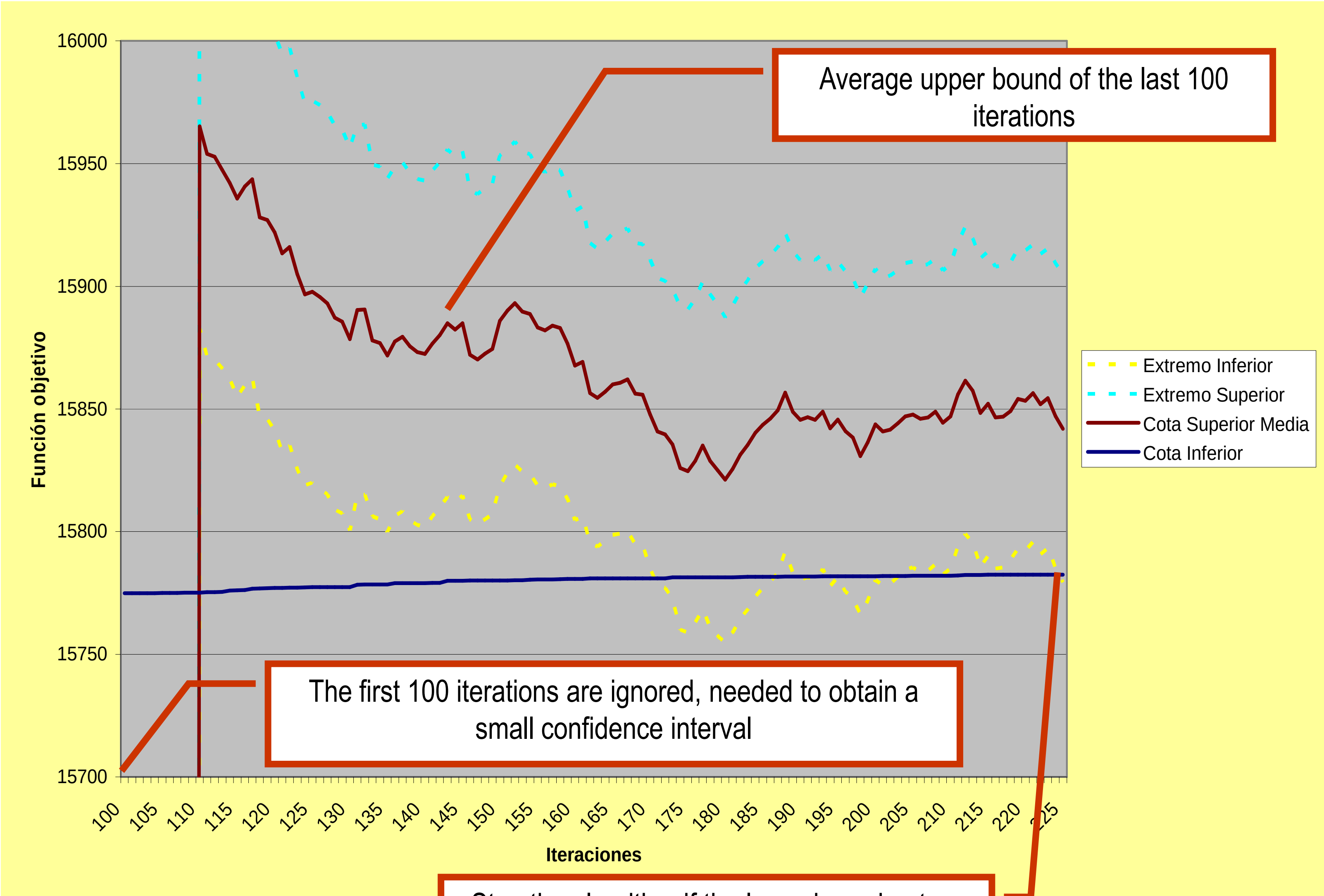
Augment $I_t^{\xi_t} = I_t^{\xi_t} + 1$

If infeasible: obtain sum of infeasibilities $\tilde{\theta}_t^{\xi_t,j}$ and dual values $\tilde{\pi}_t^{\xi_t,j}$

Augment $J_t^{\xi_t} = J_t^{\xi_t} + 1$

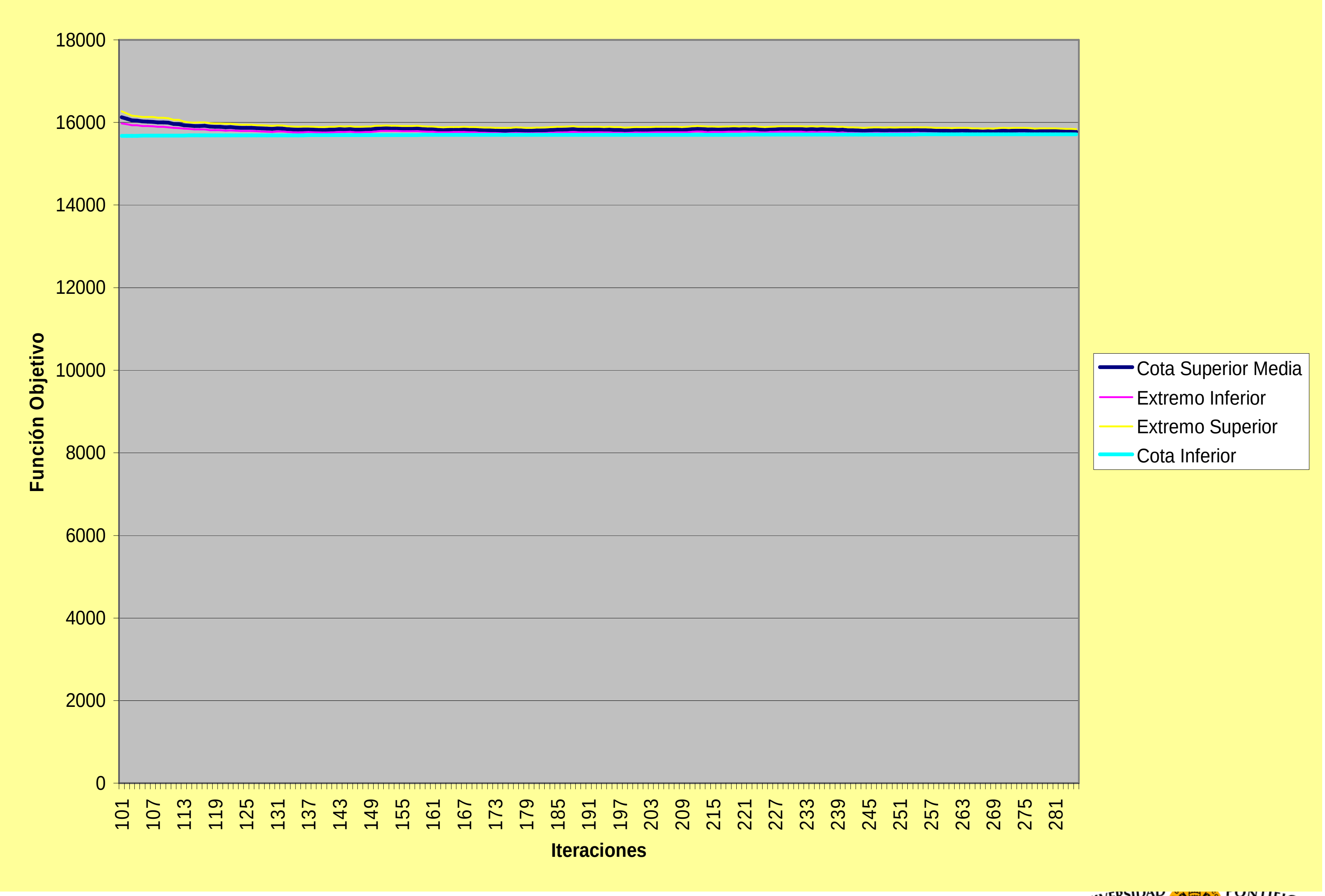
Go to step 1

Stochastic convergence in SDDP (i)

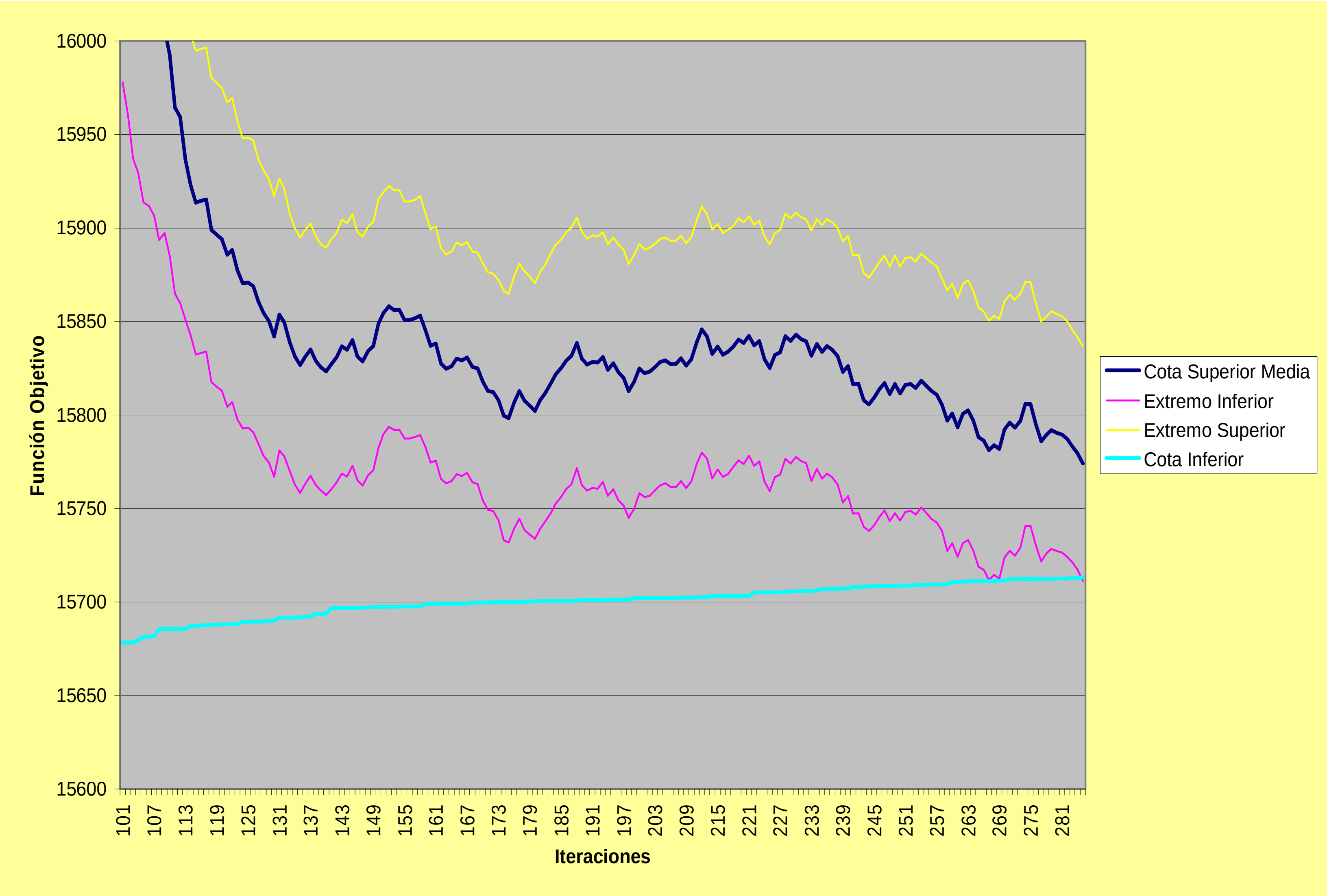


Stop the algorithm if the lower bound enters into the confidence interval of the upper bound for a confidence level of 95 %

Stochastic convergence in SDDP (ii)



Stochastic convergence in SDDP (ii) (detail)



Comparison of decomposition methods

Benders	SDDP
Suitable for stochastic problems with tree structure	Suitable for stochastic problems where stochasticity is introduced independently by scenarios
Scenario tree with no fixed structure (symmetrical or non symmetrical)	Recombining scenario tree (dependence of one scenario with respect to other is modeled by transition probabilities)
Solution of the deterministic equivalent problem	Impossible to solve the deterministic equivalent problem
Multicut	Multicut
Flexible node aggregation (tree partition)	Rigid node aggregation (tree partition) conditioned by the branching periods
In forward pass all the scenarios are solved	In forward pass only one scenario is solved
Deterministic stopping criterion	Stochastic stopping criterion
Exponential time increase with number of scenarios	Linear time increase with number of scenarios

Prof. Andres Ramos

http://www.iit.upcomillas.es/aramos/

Andres.Ramos@upcomillas.es

arght@mit.edu