A FINITE BENDERS DECOMPOSITION ALGORITHM FOR MIXED INTEGER PROBLEMS BY LAGRANGEAN RELAXATION

Santiago Cerisola, Andrés Ramos
INSTITUTO DE INVESTIGACIÓN TECNOLÓGICA
Universidad Pontificia Comillas
Alberto Aguilera 23
28015 Madrid, SPAIN
santiago.cerisola@iit.upco.es

Abstract – This document presents a finite decomposition algorithm to solve mixed integer linear problems. Integer variables appear at the master problem and at the subproblem. The non convex recourse function is approximated via a modified lagrangean relaxation algorithm. This decomposition algorithm is understood as a convexification procedure of the perturbation function that appears when a first stage variable is fixed. Extension to nested Benders decomposition is presented.

Keywords: Benders decomposition, lagrangean relaxation, perturbation function, branch and bound, convexification, mixed integer linear problem, optimality cuts, feasibility cuts, bounding cuts.

1. INTRODUCTION

Benders decomposition algorithm [2,15] solves a mixed integer linear problem (MILP) gathering the integer variables (complicating variables) into a master problem and building a subproblem on the remaining variables. Part of the objective function is explicitly evaluated in the master problem, while the rest constitutes the objective function of the subproblem and it is only introduced into the master problem in an approximate manner. When the subproblem turns out to be linear, its objective function (called *recourse function*) is convex, so that it is immediately approximated at a point building up the tangent with the optimal dual variable. The algorithm proceeds proposing values at the master problem and solving the subproblem to update the approximation of the recourse function.

When the subproblem is non convex, then the recourse function is non convex [3], and the former approach is no longer valid. Then, a way to proceed consists of forming the lower convex envelope of the recourse function [10,14]. This lower convex envelope is traditionally constructed via a *lagrangean relaxation* (LR) procedure (also called *conjugate function* or *Fenchel duality*).

That is the approach presented in Geoffrion's generalized Benders decomposition [6], where the subproblem is solved using LR [7]. However, the simple use of LR to solve a subproblem only yields an approximation of the convex envelope of the recourse function, and additional development is needed to get the exact envelope. This further development consists of introducing into the subproblem cuts that constrain the "tender variables" space over which the convexification procedure is carried out. The "tender variables" are those that connect first and second stage problems. The main issue of the paper consists of constructing the convex envelope of the recourse function restricted to the set of tender variables, which is induced by the first stage constraints. This construction is embedded into the Benders decomposition algorithm so to deal with mixed integer variables in the sub problem.

The paper is organized in the following way. The first part reviews LR and its relation with the powerful concept of perturbation function. The second part presents Benders decomposition within a two-stage problem. S simple example to clarify the previous concepts is presented. The algorithm is extended to nested decomposition and finally computational test over a large scale problem are reported.

2. Perturbation function and lagrangean relaxation algorithm

This section introduces the concept of perturbation function and its relation with LR. Later, the recourse function of a Benders algorithm is interpreted as a perturbation function and solved via LR.

2.1. General problem

Consider a problem (P) of the form

$$(P) \qquad \begin{aligned} &\min f(y) \\ &g(y) \le 0 \\ &y \in Y \end{aligned} \tag{2.1}$$

with Y being the mixed integer solutions of a polyhedron $\overline{Y}$ (i.e., a non convex region). We can assume without loss of generality that region $\overline{Y}$ incorporates the non negativity constraints of variables y , $y \ge 0$.

It is defined the generalized graph G of the problem as

$$G = \{(r, r_0) / \exists y \in Y \text{ with } r = g(y), r_0 = f(y)\} \tag{2.2}$$

so that the problem (P) is reinterpreted as finding a point (r, r_0) in G with minimum ordinate and $r \le 0$, see [9]. G is the image of Y under the transformation (g, f) . The generalized epigraph of the problem, see [9], is defined as

$$\operatorname{epi} G = \{(r, r_0) / \exists y \in Y \text{ with } r \ge g(y), r_0 \ge f(y)\} \tag{2.3}$$

Closely related with this idea is the concept of perturbation function. Consider that the right hand side (RHS) of problem (P) is being modified obtaining a family of problems whose solutions define a function on the RHS parameter introduced. This function is known in the literature as *perturbation function* or *value function* [1,12]

$$\begin{aligned} v(r) = \quad &\min f(y) \\ &g(y) \le r \\ &y \in Y \end{aligned} \tag{2.4}$$

Observe that due to the inequality in problem (P), the perturbation function is non increasing. Problem (P) is understood as finding $v(0)$. It should be clear that finding the convex hull of the generalized epigraph is equivalent to finding the lower convex envelope of the perturbation function.

A LR procedure gives the value of the convexification of the perturbation function at the point $r = 0$. For any $\lambda \ge 0$ define the *dual function* $w(\lambda)$ as

$$w(\lambda) = \min\{(\lambda,1)(r, r_0) = \lambda r + r_0, (r, r_0) \in G\} = \min\{(\lambda,1)(r, r_0) = \lambda r + r_0, (r, r_0) \in \operatorname{epi} G\} \tag{2.5}$$

or equivalently

$$\begin{aligned} w(\lambda) = \quad &\min \lambda g(y) + f(y) \\ &y \in Y \end{aligned} \tag{2.6}$$

Assume $w(\lambda)$ has a finite value, then there exists $y^i \in Y$ with $\lambda g(y^i) + f(y^i) = w(\lambda)$. This optimal solution determines a level curve $L = \{(r, r_0) / \lambda r + r_0 = \lambda g(y^i) + f(y^i)\}$. So that for

$r = 0$ we have the point $\left(0, \lambda g(y^i) + f(y^i)\right)$ and it is stated that $\upsilon(0) \geq \lambda g(y^i) + f(y^i) = w(\lambda)$. The *dual problem* traditionally consists of finding the maximum of those minimum values $w(\lambda)$.

$$(D) \qquad \max\left\{w(\lambda), \lambda \geq 0\right\} \tag{2.7}$$

Assume Y is a polytope and $y^1,...,y^K$ the extreme points of Y. Assume f and g are convex functions (e.g., linear), then we obtain the equivalent expression for the dual function

$$w(\lambda) = \min\left\{f(y^k) + \lambda g(y^k) \,/\, k = 1,...,K,\ y^k \in \mathrm{extr}(Y)\right\} \tag{2.8}$$

which shows concavity of the dual function and allows the dual problem to be formulated as a linear problem

$$\begin{aligned}
&\max w \\
&w \leq f(y^1) + \lambda g(y^1) \\
&\dots \\
&w \leq f(y^K) + \lambda g(y^K) \\
&\lambda \geq 0
\end{aligned} \tag{2.9}$$

Clearly for large scale problems it is not possible to calculate all polytope extreme points, so that the dual function is usually optimized formulating a relaxed problem, denoted *master dual problem* (MD), whose resolution proposes multiplier values.

$$(MD) \qquad \begin{aligned}
&\max w \\
&w \leq f(y^i) + \lambda g(y^i) \quad i = 1,\dots,k \\
&\lambda \geq 0
\end{aligned} \tag{2.10}$$

Evaluation of the dual function at these multiplier values obtains tangential approximations of function $w(\lambda)$, which are incorporated into the master dual problem. These tangential approximations are called *lagrangean optimality cuts.*

Traditional LR algorithm iterates between the master dual problem (MD) and the *lagrangean subproblem* (evaluation of the dual function) (PR_λ) until a certain tolerance is satisfied.

$$(PR_\lambda) \qquad \begin{aligned}
w(\lambda) = \quad &\min \lambda g(y) + f(y) \\
&y \in Y
\end{aligned} \tag{2.11}$$

Assuming only bounded cases, the *lagrangean relaxation algorithm* is summarized on the next steps:

Step 1. Solve problem (MD) and obtain λ
Step 2. Obtain upper bound $\overline{z} = w$
Step 3. Solve problem (PR_λ) and obtain x^k
Step 4. Obtain lower bound $\underline{z} = \max\left\{\underline{z}, w(\lambda)\right\}$
Step 5. Stop if $\overline{z} - \underline{z} < \mathrm{tol}$, otherwise do $k = k + 1$ and go to 1.

2.2. Constrained perturbation region

The problem (2.4) has been transformed into the problem (2.11) due to the nonexistence of additional constraints over variables (r, r_0). If the perturbation function is defined for a constrained set of RHS values, then the convexification procedure has to take this constrained set into account and previous transformation is no longer valid. Consider the perturbation function

$$
\begin{aligned}
v(r) = \quad & \min f(y) \\
& g(y) \le r \qquad r \in R \\
& y \in X
\end{aligned}
\tag{2.12}
$$

be defined for $r \in R$ and assume $0 \in R$. (If $0 \notin R$ then the convexification will not give an approximation of the optimal value $v(0) = \min\{f(y); g(y) \le 0, y \in Y\}$).

We consider a constrained generalized graph and epigraph as follows.

$$G = \{(r, r_0) / \exists y \in Y \text{ with } r = g(y), r_0 = f(y), r \in R\} \tag{2.13}$$

$$\operatorname{epi} G = \{(r, r_0) / \exists y \in Y \text{ with } r \ge g(y), r_0 \ge f(y), r \in R\} \tag{2.14}$$

When calculating the value of the perturbation function convexification at $r = 0$ the dual function $w(\lambda)$ is defined

$$
(PR_\lambda) \qquad
\begin{aligned}
w(\lambda) = \quad & \min f(y) + \lambda r \\
& g(y) \le r \\
& y \in Y, r \in R
\end{aligned}
\tag{2.15}
$$

and the dual problem remains as

$$(D) \qquad \max\{w(\lambda)\} \tag{2.16}$$

From an algorithmic point of view, this dual problem is replaced by a relaxed problem, called master dual problem (MD), which is continuously updated during the LR algorithm

$$
(MD) \qquad
\begin{aligned}
& \max w \\
& w \le f(y^i) + \lambda r^i \qquad i = 1, \ldots, k \\
& \lambda \ge 0
\end{aligned}
\tag{2.17}
$$

with $(y^i, r^i) \in \operatorname{extr}\{g(y) \le r, y \in Y, r \in R\}, i : 1, \ldots, k$

Now it is not possible to eliminate the variable r in the same way as it was eliminated when the perturbation region was the whole euclidean space.

2.3. Phase I of lagrangean relaxation

Previous section implicitly assumed that master dual problem (MD) was a bounded problem, so each resolution would give a new multiplier proposal λ. It did also assume problem (PR_λ) was bounded for each value λ. However, this is not the general situation and a family of cuts is necessary to guarantee master dual problem boundness. Those cuts will appear when solving the minimization of infeasibilities of the subproblem under study. From hereafter it is assumed that the objective function is linear and the constraints are affine. We will exhaustively use the Farkas' law results.

Let problem (P) take the form

$$
(P) \qquad
\begin{aligned}
& \min dy \\
& Wy \le h \\
& y \in Y
\end{aligned}
\tag{2.18}
$$

In the resolution of problem (P) it is necessary to test that the problem is feasible and, if not the case, to provide a minimization of infeasibilities. Its feasibility is equivalent to a non infinite value of the associated perturbation function for $r = 0$, which for a constrained perturbation region R is defined as

$$\begin{aligned}&\min dy\\&Wy-h\le r\\&y\in Y, r\in R\end{aligned}\tag{2.19}$$

It is clear that system $\{Wy-h\le r, y\in Y, r\in R, r=0\}$ has a solution if and only if system $\{\operatorname{conv}\{Wy-h\le r, y\in Y, r\in R\}, r=0\}$ does. Feasibility of this region is tested formulating the minimization of infeasibilities problem. Assuming that infeasibility can only be caused by the constraints $r=0$ this problem takes the form

$$\begin{aligned}&\min r^+ + r^-\\&r-r^+ + r^- = 0\\&\operatorname{conv}\begin{Bmatrix}Wy-h\le r\\ y\in Y, r\in R\end{Bmatrix}\\&r^+, r^-\ge 0\end{aligned}\tag{2.20}$$

Feasibility region of the above problem immediately satisfies the *integrality property* [7] that guarantees its optimal value is equivalent to the value obtained by a LR algorithm. Then, for any λ consider

$$(PR_\lambda)_* \qquad \begin{aligned}w_*(\lambda) = \quad &\min r^+ + r^- + \lambda(r-r^+ + r^-)\\&Wy-h\le r\\&y\in Y, r\in R\\&r^+, r^-\ge 0\end{aligned}\tag{2.21}$$

and solve the following dual problem

$$(D)_* \qquad \max\{w_*(\lambda)\}\tag{2.22}$$

If this problem has positive solution, then primal problem is infeasible due to value $r=0$. Observe that $w_*(\lambda)$ verifies

$$w_*(\lambda)=\begin{cases}-\infty & \lambda<-1, \lambda>1\\ \min \lambda r, (y,r)\in\{Wy-h\le r, y\in Y, r\in R\} & -1\le\lambda\le 1\end{cases}\tag{2.23}$$

So that dual problem $(D)_*$ can be rewritten as

$$(D)_* \qquad \max\{w_*(\lambda), -1\le\lambda\le 1\}\tag{2.24}$$

The resolution of problem $(D)_*$ is carried out formulating a relaxed problem, called master dual problem $(MD)_*$, which is being updated when necessary

$$(MD)_* \qquad \begin{aligned}&\max w\\&w\le \lambda r^i \qquad i=1,\dots,k\\&0\le\lambda\le 1\end{aligned}\tag{2.25}$$

with $(y^i, r^i)\in \operatorname{extr}\{Wy-h\le r, y\in Y, r\in R\}, i:1,\dots,k$

Remark 1. Observe that this cutting plane technique creates a group of planes that correspond to a group of planes of problem (MD) moved to the origin. So in case problem $(MD)_*$ ends with zero solution, problem (MD) has a set of constraints that will guarantee its boundness.

Introducing a new parameter λ_0, with value 0 in phase 1 and value 1 in phase 2[1], we formulate lagrangean subproblem $(PR_{\lambda_0\lambda})$ and master dual problem (MD_{λ_0}), which will generalize the LR algorithm

$$(MD_{\lambda_0}) \qquad \begin{aligned} &\max w \\ &w \le \lambda_0 dy^i + \lambda r^i \quad i = 1,\ldots,k \\ &\lambda \ge 0 \end{aligned} \tag{2.26}$$

$$(PR_{\lambda_0\lambda}) \qquad \begin{aligned} w_{\lambda_0}(\lambda) = \ &\min \lambda_0 dy + \lambda r \\ &Wy - h \le r \\ &y \in Y, r \in R \end{aligned} \tag{2.27}$$

where $\lambda_0 = 0$ and $-1 \le \lambda \le 1$ in phase 1 and $\lambda_0 = 1$ in phase 2.

2.4. Bounding cuts

It may be the situation that for a particular multiplier value λ problem (PR_λ) ends with an unbounded solution. In that case, a bounding cut is introduced that eliminates that multiplier proposal. This technique is now described.

Consider problem (P) of equation (2.27) adapted for simplicity to phase 2.

$$(PR_{\lambda}) \qquad \begin{aligned} w(\lambda) = \ &\min dy + \lambda r \\ &Wy - h \le r \\ &y \in Y, r \in R \end{aligned} \tag{2.28}$$

We can assume $Y = \{A_{11}y \le b_1, y \in \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}\}$, $R = \{R_{11}r \le r_1\}$.

Previous problem is unbounded if its linear relaxation is unbounded. On the contrary, a bounded linear relaxation problem implies boundness for the MIP problem. Unboundness of previous linear relaxation problem is equivalent to infeasibility of its dual linear problem (DPR_λ), which takes the form

$$(DPR_{\lambda}) \qquad \begin{aligned} &\max \pi_1 h + \pi_2 b_1 + \pi_3 r_1 \\ &\pi_1 W + \pi_2 A_{11} = d \\ &-\pi_1 + \pi_3 R_{11} = \lambda \\ &\pi_1, \pi_2, \pi_3 \le 0 \end{aligned} \tag{2.29}$$

A direct application of Farkas results assures problem (DPR_λ) is feasible if and only if

$$\begin{aligned} cd + \lambda\tilde{r} &\le 0, \forall(\tilde{y}, \tilde{r}) / \\ -W\tilde{y} + \tilde{r} &\le 0 \\ -A_{11}\tilde{y} &\le 0 \\ -R_{11}\tilde{r} &\le 0 \end{aligned} \tag{2.30}$$

This equation constrains the set of Lagrange multipliers such that problem (PR_λ) is bounded to belong to set B, denoted as *bounding set*. A closed form expression for this set B is then

---

[1] Phase 2 is understood as the algorithm presented at the beginning of section 2 corresponding to problems (2.10) and (2.11).

$$B = \begin{Bmatrix} \lambda \,/\, d\tilde{y}^j + \lambda \tilde{r}^j \leq 0 \\ \forall (\tilde{y}^j, \tilde{r}^j) \text{ extreme ray } \left\{ -W\tilde{y} + \tilde{r} \leq 0, -A_{11}\tilde{y} \leq 0, -R_{11}\tilde{r} \leq 0 \right\} \end{Bmatrix} \tag{2.31}$$

Then, dual problem (D) takes the form

$$(D) \qquad \max \left\{ w(\lambda), \lambda \in B \right\}$$

Calculating all the former extreme rays is an out of question matter, so that when proposing a new multiplier value at the master problem it should be tested if this multiplier value belongs to set B. In the negative case a new constraint will be added to the master dual problem (MD). This new constraint is defined as a *bounding cut*. Before solving problem (PR_λ) next problem has to be solved

$$\begin{aligned} &\max d\tilde{y} + \lambda \tilde{r} \\ &-W\tilde{y} + \tilde{r} \leq 0 \\ &-A_{11}\tilde{y} \leq 0 \\ &-R_{11}\tilde{r} \leq 0 \\ &-1 \leq \tilde{y} \leq 1, -1 \leq \tilde{r} \leq 1 \end{aligned} \tag{2.32}$$

and in case this problem ends with a positive solution, a constraint of the form $d\tilde{y}^j + \lambda \tilde{r}^j \leq 0$ is introduced into the master dual problem (MD). The point $(\tilde{y}^j, \tilde{r}^j)$ represents the optimal solution of previous problem. This constraint eliminates the last multiplier λ from the feasibility set of problem (MD). Observe that the dual problem of (2.32) is precisely the minimization of infeasibilities of problem (DPR_λ).

So in a LR algorithm a master problem is solved to obtain a new multiplier proposal. This master dual problem is built up with constraints that outer approximate the dual function $w(\lambda)$ (lagrangean optimality cuts) and bounding cuts that eliminate multiplier values for which lagrangean subproblem turns out to be unbounded. The master dual problem takes the form

$$(MD) \qquad \begin{aligned} &\max w \\ &w \leq dy^i + \lambda r^i \qquad i:1,\ldots,k \\ &0 \leq d\tilde{y}^j + \lambda \tilde{r}^j \qquad j:1,\ldots,l \\ &\lambda \geq 0 \end{aligned} \tag{2.33}$$

with
$(y^i, r^i) \in \operatorname{extr} \left\{ Wy - h \leq r, y \in Y, r \in R \right\}, i:1,\ldots,k$ and
$(\tilde{y}^j, \tilde{r}^j) \in \text{extreme ray} \left\{ -W\tilde{y} + \tilde{r} \leq 0, -A_{11}\tilde{y} \leq 0, -R_{11}\tilde{r} \leq 0 \right\}, j = 1,\ldots,l$

Remark 2. Observe that the former development of bounding cuts has been done for phase 2, but it should also be done for phase 1. A bounding cut obtained in phase 1 is also valid for phase 2, in the same way as lagrangean optimality cut obtained in phase 1 provides a valid cut for phase 2.

2.5. Detailed Lagrangean Relaxation Algorithm

The perturbation function is non increasing in the case of all the constraints are inequality constraints. The case with equality constraints is quite similar, although concepts need to be redefined. We generalize the above development to the case with inequality and equality constraints and summarize the general relaxation algorithm for MILP problem.

Consider now problem (P)

$$(P) \qquad \begin{aligned} &\min dy \\ &Wy \le h \\ &W'y = h' \\ &y \in Y \end{aligned} \tag{2.34}$$

and consider possible perturbations of the RHS of this problem for $(r, r') \in R$.

$$\begin{aligned} \upsilon(r,r') = \ &\min dy \\ &Wy - h \le r \\ &W'y - h' = r' \\ &y \in Y \end{aligned} \tag{2.35}$$

Graph and epigraph associated with problem (P) are defined as

$$\begin{aligned} G &= \left\{(r,r',r_0) / \exists y \in Y \text{ with } r = Wy - h, r' = W'y - h', r_0 = dy, (r,r') \in R\right\} \\ \operatorname{epi} G &= \left\{(r,r',r_0) / \exists y \in Y \text{ with } r \ge Wy - h, r' = W'y - h', r_0 \ge dy, (r,r') \in R\right\} \end{aligned} \tag{2.36}$$

Assume we want to obtain the lower convex envelope value of the perturbation function at $r = 0$, $r' = 0$. We have to solve a lagrangean subproblem of the form

$$(PR_{\lambda_0\lambda\mu}) \qquad \begin{aligned} w_{\lambda_0}(\lambda,\mu) = \ &\min \lambda_0 dy + \lambda r + \mu(W'y - h') \\ &Wy - h \le r \\ &y \in Y, (r, W'y - h') \in R \end{aligned} \tag{2.37}$$

with $\lambda_0 = 0$ at phase 1 of the algorithm and $\lambda_0 = 1$ at phase 2.

Lagrange multipliers are proposed solving a relaxed master dual problem whose expression is

$$(MD_{\lambda_0}) \qquad \begin{aligned} &\max w \\ &w \le \lambda_0 dy^i + \lambda r^i + \mu(W'y^i - h') \qquad i: 1,...,k \\ &0 \le \lambda_0 d\tilde{y}^j + \lambda \tilde{r}^j + \mu W'\tilde{y}^j \qquad j: 1,...,l \\ &\lambda \ge 0, \end{aligned} \tag{2.38}$$

with $(y^i, r^i) \in \operatorname{extr}\{y \in Y, Wy - h \le r, (r, W'y - h') \in R\}$ and $(\tilde{y}^j, \tilde{r}^j)$ are extreme rays for the corresponding region.

The *lagrangean relaxation algorithm* is summarized on the following steps:

Step 1. Set $\lambda_0 = 0$ and $-1 \le \lambda \le 1$
Step 2. Solve problem (MD_{λ_0}) and obtain multiplier values λ and μ
Step 3. Obtain upper bound $\overline{z}_{\lambda_0} = w$
Step 4. If $\lambda_0 = 0$ and $w = 0$ switch to phase 2 setting $\lambda_0 = 1$
Step 5. Solve linear relaxation of problem $(PR_{\lambda_0\lambda\mu})$
Step 6. If linear relaxation is unbounded, then take dual values and form a bounding cut
Obtain value $(\tilde{y}^k, \tilde{r}^k)$, set $l = l + 1$ and go to step 2
Step 7. If linear relaxation is bounded, then continue solving MILP problem $(PR_{\lambda_0\lambda\mu})$
Obtain value (y^k, r^k) and set $k = k + 1$
Obtain lower bound $\underline{z}_{\lambda_0} = \max\left\{\underline{z}_{\lambda_0}, w_{\lambda_0}(\lambda,\mu)\right\}$
Step 8. If $\overline{z}_{\lambda_0} - \underline{z}_{\lambda_0} < \text{tol}$ then stop. Otherwise go to 2.
Step 9. If $\lambda_0 = 0$ problem (P) is infeasible
Step 10. If $\lambda_0 = 1$ problem (P) is feasible

The LR algorithm ends at phase 2 with the value of the lower convex envelope of problem (P) at $r = 0$, $r' = 0$. In case of ending at phase 1, then problem (P) is infeasible, and the final value gives the minimization of infeasibilities due to the complicating constraints $\{Wy \le h, W'y' = h'\}$. Problem infeasibility due to these constraints is identified if a lagrangean subproblem has positive optimal value during phase 1. In that case the algorithm may stop or may continue to get the minimization of infeasibilities required by most optimization algorithms. This is necessary when incorporating a lagrangean algorithm into a Benders decomposition scheme, producing what is defined in the literature as *deepest cut*.

3. BENDERS DECOMPOSITION

We now face the subject of solving a problem (P) of the form

$$(P) \qquad \begin{aligned} &\min cx + dy \\ &Tx + Wy \le h \\ &x \in X, y \in Y \end{aligned} \tag{3.1}$$

where feasible regions for first and second stage variables, x and y respectively, incorporate integrality constraints for some variables $X = \{A_1 x \le a_1, x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}\}$, $Y = \{A_2 y \le a_2, y \in \mathbb{R}^{n_2} \times \mathbb{Z}^{m_2}\}$. We assume this representation incorporates the non negativity constraints for variables x and y. The resolution of this problem (P) is equivalent to solve the master problem (MP)

$$(MP) \qquad \begin{aligned} &\min cx + \theta(x) \\ &x \in X \end{aligned} \tag{3.2}$$

with the *recourse function* $\theta(x)$ defined as

$$(SP_x) \qquad \begin{aligned} \theta(x) = \quad &\min dy \\ &Wy \le h - Tx \\ &y \in Y \end{aligned} \tag{3.3}$$

3.1. Linear problems

For LP problems, the Benders algorithm [2] proceeds formulating a master problem that incorporates first stage variables and a partial description of the recourse function $\theta(x)$. Resolution of this master problem gives a first stage optimal value x. Evaluation of subproblem (SP_x) at this optimal value (modifying RHS of corresponding equations) gives a supporting hyperplane of the epigraph of the recourse function, named Benders optimality cut. This supporting hyperplane updates master problem, which is solved again. In addition to these supporting planes the algorithm also provides feasibility cuts that eliminate those first stage values of the master problem that turn infeasible the second stage problem. The algorithm continues until a certain tolerance is satisfied. Subsequently, a very briefly review for the linear case is presented.

3.1.1. Optimality cuts

Let us assume integrality constraints are removed from Y so that problem (SP) for a fixed x_0 takes the form

$$(SP_{x_0}) \qquad \theta(x_0) = \min\{dy, Wy \le h - Tx_0, y \in \overline{Y}\} \tag{3.4}$$

where $\overline{Y} = \{y \,/\, A_2 y \le a_2, y \in \mathbb{R}^{n_2+m_2}\}$.

Duality in linear programming immediately derives an equivalent expression for problem (SP_{x_0})

$$(SP_{x_0}) \qquad \begin{aligned} \theta(x_0) = \quad & \max \pi(h - Tx_0) + \rho a_2 \\ & \pi W + \rho A_2 = d \\ & \pi \le 0, \rho \le 0 \end{aligned} \tag{3.5}$$

Resolution of this problem ends with optimal value θ^i, achieved for dual values (π^i, ρ^i). The recourse function then trivially satisfies the next constraint

$$\theta(x) \ge \pi^i(h - Tx) + \rho^i a_2 \tag{3.6}$$

Linearizing around the point of interest x_0 we have the expression

$$\begin{aligned} \theta(x) \ge \pi^i(h - Tx) + \rho^i a_2 = \pi^i(h - Tx_0 + Tx_0 - Tx) + \rho^i a_2 = \\ = \pi^i(h - Tx_0) + \rho^i a_2 + \pi^i(Tx_0 - Tx) = \theta^i + \pi^i T(x_0 - x) \end{aligned} \tag{3.7}$$

So that expression (3.6) is written as

$$\theta(x) \ge \theta^i + \pi^i T(x_0 - x) \tag{3.8}$$

and denoted in the literature as *Benders optimality cut.*

3.1.2. Feasibility cuts

Subproblem (SP_{x_0}) is infeasible if no solution exists for the region $\{Wy \le h - Tx_0, A_2 y \le a_2\}$. This is equivalent to assert that there exists no $s_1 \ge 0$, $s_2 \ge 0$ such that

$$\left\{Wy + s_1 = h - Tx_0,\ A_2 y + s_2 = a_2\right\} \tag{3.9}$$

Direct application of Farkas law implies that a necessary condition for a first stage value x_0 to produce a feasible subproblem is

$$\left\{\tilde{\pi}(h - Tx_0) + \tilde{\rho} a_2 \le 0, \forall \tilde{\pi}, \tilde{\rho} \le 0 \,/\, \tilde{\pi} W + \tilde{\rho} A_2 \le 0\right\} \tag{3.10}$$

This result introduces the feasible set K, as the set of first stage values that guarantee feasibility for second stage problem. A closed form expression for this set is

$$K = \left\{x_0 \,/\, \tilde{\pi}^j(h - Tx_0) + \tilde{\rho}^j a_2 \le 0, \forall \tilde{\pi}^j, \tilde{\rho}^j \text{ extreme ray } \left\{\tilde{\pi} W + \tilde{\rho} A_2 \le 0, \tilde{\pi} \le 0, \tilde{\rho} \le 0\right\}\right\} \tag{3.11}$$

Once a first stage solution x_0 is obtained at the master problem (MP), it is solved the problem

$$\begin{aligned} \theta_*(x_0) = \quad & \max \tilde{\pi}(h - Tx_0) + \tilde{\rho} a_2 \\ & \tilde{\pi} W + \tilde{\rho} A_2 \le 0 \\ & -1 \le \tilde{\pi} \le 0 \\ & -1 \le \tilde{\rho} \le 0 \end{aligned} \tag{3.12}$$

and if the objective function has positive value then a *feasibility cut* is introduced that excludes that first stage value x_0 and has the following form

$$\tilde{\pi}^j(h - Tx) + \tilde{\rho}^j a_2 \le 0 \tag{3.13}$$

Linearizing around the first stage value and letting θ^j be the optimum of problem (3.12) and $\tilde{\pi}^j$ and $\tilde{\rho}^j$ its optimal values we have

$$\begin{aligned} 0 \ge \tilde{\pi}^j(h - Tx) + \tilde{\rho}^j a_2 = \tilde{\pi}^j(h + Tx_0 - Tx_0 - Tx) + \tilde{\rho}^j a_2 = \\ = \theta^j + \tilde{\pi}^j(Tx_0 - Tx) = \theta^j + \tilde{\pi}^j T(x_0 - x) \end{aligned} \tag{3.14}$$

This feasibility cut gets a similar expression to the optimality cut (3.8)

$$0 \geq \theta^{j} + \tilde{\pi}^{j} T(x_0 - x) \tag{3.15}$$

Observe that problem (3.12) is precisely the dual problem of

$$\begin{aligned}
&\min s_1 + s_2 \\
&Wy - s_1 \leq h - Tx_0 \\
&A_2 y - s_2 \leq a_2 \\
&s_1, s_2 \geq 0
\end{aligned} \tag{3.16}$$

that represents the minimization of infeasibilities of problem (SP_{x_0}).

Observe this is the dual situation of the lagrangean decomposition scheme in which a bounding cut excludes a multiplier value if this turns the lagrangean subproblem unbounded.

Benders decomposition proceeds iterating between a linear master problem (MP) and a subproblem (SP) until a certain tolerance is satisfied. The master problem presents an expression

$$(MP) \qquad \begin{aligned}
&\min cx + \theta \\
&0 \geq \theta^{j} + \tilde{\pi}^{j} T(x_0^{j} - x) \quad\quad j = 1, \ldots, l \\
&\theta \geq \theta^{i} + \pi^{i} T(x_0^{i} - x) \quad\quad i = 1, \ldots, k \\
&x \in \overline{X}, \overline{X} = \left\{ A_1 x \leq a_1, x \in \mathbb{R}^{n_1 + m_1} \right\}
\end{aligned} \tag{3.17}$$

3.2. Mixed integer linear problems

The MILP case keeps the same procedure, but face the disadvantage of the non convexity of the recourse function $\theta(x)$. The resolution of problem (P) requires the convexification of this recourse function. Considering the recourse function as the perturbation function of a problem, we want to obtain the convexified expression of the perturbation function

$$\theta(r) = \min \left\{ dy, Wy - h \leq r, y \in Y, r \in R \right\} \tag{3.18}$$

with $R = \left\{ r \,/\, \exists x \in X \,/\, r = -Tx \right\}$. The variables r are the "tender variables" and region R is the domain of the recourse function which is induced by the first-stage constraints.

Following results of section 2, define for any λ the *dual function*

$$(PR_{\lambda}) \qquad \begin{aligned}
w(\lambda) = \quad &\min dy + \lambda r \\
&Wy - h \leq r \\
&y \in Y, r \in R
\end{aligned} \tag{3.19}$$

Its solution determines a level curve of the form $L = \left\{ (r, r_0) \,/\, \lambda r + r_0 = w(\lambda) \right\}$. For $r = -Tx_0$ the resulting point is then $(-Tx_0, w(\lambda) + \lambda T x_0)$. The dual problem consists of finding the maximum of those ordinates

$$(D_x) \qquad \max \left\{ w(\lambda) + \lambda T x_0, \lambda \geq 0 \right\} \tag{3.20}$$

So, in the MILP case the linear resolution of the subproblem is replaced by the LR algorithm that finds the supporting plane of the lower convex envelope of the recourse function at the first stage proposal. Region R is only known through its implicit definition, and will be outer approximated as the algorithm proceeds.

The resolution of dual problem (D_x) ends with an optimal Lagrange multiplier λ^{i} and an optimal value for the dual problem given as $(w(\lambda^{i}) + \lambda^{i} T x_0)$. The epigraph of the perturbation function then immediately satisfies

$$\theta \geq w(\lambda^i) + \lambda^i T x, \quad \forall x \in X \tag{3.21}$$

Denoting $\theta^i = w(\lambda^i) + \lambda^i T x_0$ the optimum value of problem (D_x) and linearizing around the first stage solution we have

$$\theta \geq w(\lambda^i) + \lambda^i T x = w(\lambda^i) + \lambda^i T x_0 - \lambda^i T x_0 + \lambda^i T x = \theta^i + \lambda^i T(-x_0 + x) \tag{3.22}$$

Summarizing

$$\theta \geq \theta^i - \lambda^i T(x_0 - x), \quad \forall x \in X \tag{3.23}$$

This expression recovers the *Benders optimality cut* introduced in the master problem for the LP case and shows the classical result that relates the dual value of a linear problem to the negative of the optimal Lagrange multiplier that maximizes the dual function.

The optimization of the dual function is carried out through a LR algorithm. The end of this algorithm at phase 2 produces an optimal multiplier and an optimal value of the dual function, that are used to form a Benders optimality cut.

3.2.1. Feasibility cuts

In order to check if the first stage proposal turns the subproblem into a feasible one, we have to solve the dual problem of the phase 1 LR algorithm. This dual problem takes the form

$$(D_x)_* \qquad \max\left\{w_*(\lambda) + \lambda T x_0, -1 \leq \lambda \leq 1\right\} \tag{3.24}$$

with

$$(PR_\lambda)_* \qquad \begin{aligned} w_*(\lambda) = \;& \min \lambda r \\ & Wy - h \leq r \\ & y \in Y, r \in R \end{aligned} \tag{3.25}$$

If the phase 1 of LR procedure indicates subproblem is infeasible (positive optimum value) for that first stage proposal then a *feasibility cut* is introduced into the master problem which takes the form

$$0 \geq \theta^j - \lambda^j T(x_0 - x), \quad \forall x \in X \tag{3.26}$$

The master problem on a Benders decomposition algorithm presents the next form

$$(MP) \qquad \begin{aligned} & \min cx + \theta \\ & 0 \geq \theta^j + \tilde{\pi}^j T(x_0^j - x) \quad j = 1,\ldots,l \\ & \theta \geq \theta^i + \pi^i T(x_0^i - x) \quad i = 1,\ldots,k \\ & x \in X, X = \left\{A_1 x \leq a_1, x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}\right\} \end{aligned} \tag{3.27}$$

The values (θ^i, π^i), being $\pi^i = -\lambda^i$, represent the optimal values of LR procedure when this ends at phase 2, and x_0^i is the first stage solution used on that iteration. These values are used to form the outer approximation of the convexified recourse function. The values $(\theta^j, \tilde{\pi}^j)$, being $\tilde{\pi}^j = -\lambda^j$, represent the optimal values of the LR procedure when this ends at phase 1, and x_0^j is the first stage solution used. These values are used to create feasibility cuts to exclude infeasible first stage solutions.

Once the resolution of master problem (MP) produces a first stage value, the associated dual subproblem is optimized by iterating between a relaxed master dual problem $(MD_{x\lambda_0})$ and a subproblem (PR_λ) whose expressions take the form

$$(MD_{x\lambda_0})\qquad \begin{aligned} &\max w + \lambda T x_0 \\ &w \le \lambda_0 d y^i + \lambda r^i \qquad i:1,\dots,k \\ &0 \le \lambda_0 d \tilde{y}^j + \lambda \tilde{r}^j \qquad j:1,\dots,l \\ &\lambda \ge 0 \end{aligned} \tag{3.28}$$

where $\lambda_0 = 0$ and $-1 \le \lambda \le 1$ in phase 1 and $\lambda_0 = 1$ in phase 2.

$$(PR_{\lambda_0\lambda})\qquad \begin{aligned} w_{\lambda_0}(\lambda) = \ &\min \lambda_0 d y + \lambda r \\ &Wy - h \le r \\ &y \in Y, r \in R^i \end{aligned} \tag{3.29}$$

This subproblem resolution gives back the supporting hyperplane of the recourse function when it is interpreted as a perturbation function. This convexification considers R^i as the perturbation region. During the algorithm, region R is being outer approximated with further resolutions of master problem. In the next sections it is commented a way to incorporate cuts to approximate perturbation region R.

3.3. Perturbation cuts

In the general case not all the first stage variables will modify the RHS parameters of the second stage subproblem, only a group of first stage variables are tied with a group of second stage variables. Consider a problem (P) of the form

$$(P)\qquad \begin{aligned} &\min c_1 x_1 + c_2 x_2 + dy \\ &Tx_2 + Wy \le h \\ &(x_1, x_2) \in X, y \in Y \end{aligned} \tag{3.30}$$

We define the x_2 space to be the space of coupling variables between first and second stages.

Define the shadow S of region X over the coupling variable space as

$$S = \{x_2 / \exists x_1 / (x_1, x_2) \in X\}$$

and define the perturbation region R as

$$R = \{r / \exists x_2 \in S / r = -Tx_2\} \tag{3.31}$$

So we are interested in finding S. This is the projection of set X over the euclidean space of the coupling variables x_2. Let $P(x_1, x_2) = x_2$ this projection. We differentiate between the case of a linear master problem and a mixed integer one. The second case extends the first one.

3.3.1. Linear master problem

The idea for obtaining a constraint for region S comes from inspection of an optimal solution of the master problem. Let the master problem at an iteration of the algorithm be given as

$$(MP)\qquad \begin{aligned} &\min c_1 x_1 + c_2 x_2 + \theta \\ &0 \ge \theta^j + \tilde{\pi}^j T(x_0^j - x_2) \qquad j = 1,\dots,l \\ &\theta \ge \theta^i + \pi^i T(x_0^i - x_2) \qquad i = 1,\dots,k \\ &x \in \overline{X}, \overline{X} = \{A_1 x \le a_1, x \in \mathbb{R}^{n_1+m_1}\} \end{aligned} \tag{3.32}$$

and let (x_1^0, x_2^0, θ_0) the optimal solution. As $x \in \mathbb{R}^{n_1+m_1}$ and $\theta \in \mathbb{R}$ then the optimal point is given as the intersection of $n_1 + m_1 + 1$ planes. Let $(d_1, d_2, \dots, d_{n_1+m_1}, d_{n_1+m_1+1})$ be the edges at that extreme point. Let $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_{n_1+m_1}, \hat{d}_{n_1+m_1+1})$ be the projection of these edges over the euclidean space of the coupling variables x_2. Then, we eliminate those projected edges that can be obtained as a positive linear combination of the remaining ones. If the positive cone generated by those remaining projected edges generates the coupling variable space, then x_2^0 is not an extreme point of region S. On the contrary, those remaining projected edges generate a cone C^i such that $S \subset C^i$. See Appendix A for a detailed description.

3.3.2. Mixed integer master problem

In this case the problem will be solved with a B&B algorithm. The B&B algorithm ends with a partition of the master problem feasible region. Let $\{X_n, n = 1,...,N\}$ this partition. Let C_n^i the cone obtained at the optimum point of each partition. Then the convex sum of these cones constrains region S. Observe that in case the partition $\{X_n, n = 1,...,N\}$ reduces to a point for each set, i.e., Pure Integer Problem (PIP), then the region S is obtained as the convex hull of the set of extreme points of S. More details about constraining the shadow region S can be found in Appendix A.

3.4. Outline of the Benders algorithm

Assume we are at iteration p of the algorithm. At this moment we have a family of optimality cuts and a family of feasibility cuts for the master problem. We do also have a collection of cuts for the perturbation region R, so we have an approximation R^{p-1} $(R \subset R^{p-1})$ of this perturbation region.

Step 1. Solve master problem (MP) and obtain (x^p, θ^p). Check if the projection of this point is an extreme point of S. In that case generate a group of constraints for region R. Then we have the new region approximation R^p $(R \subset R^p \subset R^{p-1})$. Calculate lower bound $\underline{z} = cx^p + \theta^p$. Pass value x^p to the subproblem. Go to step 2.

Step 2. Solve subproblem (SP_{x^p}). Eliminate those extreme points (y^i, r^i) $i \in I$ out of the master dual problem $(MD_{x^p\lambda_0})$ (that until this moment have been used to form approximations of the dual function) such that $r^i \notin R^p$. Observe that it is enough to check the last constraint introduced into the perturbation region.
Solve the LR phase 1 to check feasibility of the master proposal. Let λ^j and $w_*(\lambda^j)$ the optimal solution. If $w_*(\lambda^j) > 0$ then generate a feasibility cut for the master problem with the dual value $\pi^j = -\lambda^j$. Go to step 1.
If $w_*(\lambda^j) = 0$, then solve the LR phase 2. Let λ^i and $w(\lambda^i) + \lambda^i T x^p$ the optimal solution. Calculate upper bound $\overline{z} = cx^p + w(\lambda^i) + \lambda^i T x^p$. Stop if difference between bounds is close enough. Otherwise generate an optimality cut for the master problem with the multiplier $\pi^i = -\lambda^i$. Go to step 1.

3.5. Convergence proof

Proposition. The Benders algorithm as proposed on section 3.4 is finite and ends with an optimal solution of original problem (P).

Proof.

Let R^p the outer approximation of the perturbation region available at iteration p of the algorithm. Let $\theta_{R^p}(x)$ the convexification of $\theta(x)$ when the perturbation region is R^p. Immediately $\theta(x) \geq \theta_{R^p}(x)$, $\forall x \in X$.

Let $\hat{\theta}^p(x)$ the outer approximation of $\theta(x)$ available at iteration p. Immediately $\theta(x) \geq \hat{\theta}^p(x)$.

By algorithm construction we also have that $\theta(x) \geq \theta_{R^p}(x) \geq \hat{\theta}^p(x)$ $\forall x \in X$.

Let (x^p, θ^p) the solution obtained at the master problem. Then $\theta^p = \hat{\theta}^p(x^p)$.

Let $\theta_{R^p}(x^p)$ the solution obtained at the subproblem. This value represents the value of the convexification of the recourse function when the perturbation region is R^p. Then we may have these cases.

Case a. If $\theta_{R^p}(x^p) > \theta^p$, then a new cut is generated. The algorithm proceeds formulating a relaxed master problem (MP) and obtaining new first stage values.
Case b. If $\theta_{R^p}(x^p) = \theta^p$, then x^p is an optimal solution for problem $\min\left\{cx + \theta_{R^p}(x), x \in X\right\}$. In this case we observe that

$$\begin{aligned} &\min\left\{cx + \hat{\theta}^p(x), x \in X\right\} = cx^p + \hat{\theta}^p(x^p) = cx^p + \theta^p = \\ &= cx^p + \theta_{R^p}(x^p) \geq \min\left\{cx + \theta_{R^p}(x), x \in X\right\} \geq \min\left\{cx + \hat{\theta}^p(x), x \in X\right\} \end{aligned} \tag{3.33}$$

and immediately

$$cx^p + \theta^p = \min\left\{cx + \theta_{R^p}(x), x \in X\right\} \tag{3.34}$$

so that this means that x^p is an optimal solution for problem $\min\left\{cx + \theta_{R^p}(x), x \in X\right\}$.

We are also interested in proving that x^p is an optimal solution for problem
$\min\left\{cx + \theta(x), x \in X\right\}$. Assume the contrary, then

$$cx + \theta(x) > m \quad \forall x \in X$$

but

$$\min\left\{cx + \theta(x), x \in R^p\right\} = m$$

which implies $x^p \in R^p \setminus X$. This is a contradiction because $x^p \in X$, due it has been obtained as a solution of the master problem.

It only remains to proof that there can only be a finite number of iterations in which the solution of master problem and the solution of the subproblem fill to satisfy situation presented in case a.

Assume that no cuts for the perturbation region are generated. The function $\theta_{R^p}(x)$ is a piecewise convex linear function. This means there is only necessary a finite number of cuts (perhaps a high number) to build it up as the maximum of linear functions. If the same point at the master problem is obtained, we can assure that the point is an optimal solution because repetition of the subproblem ended with the situation described in case b. So that repetitions of master problem always give back different first stage proposals.

In case $\theta_{R^p}(x^p) > \theta^p$, then a new cut is generated. So that the number of iterations with no added constraints for the perturbation region is finite. A new perturbation cut is introduced in case a first stage solution results to be an extreme point of shadow S. Region S has a finite number of extreme points, so that there are only a finite number of iterations in which the perturbation region is updated due to the non possibility of repetitions of first stage solutions. After that finite number of iterations situation presented in case b will then appear after a finite number of iterations. Considering that the lagrangean relaxation algorithm is a finite algorithm and then the proposed algorithm is also finite.

4. EXAMPLE

Consider the problem

$$\begin{aligned}
&\min -0.3x - 1.5y - z \\
&0 \le x \le 5 \\
&x + y \le 3.7 \\
&y + z \le 5.2 \\
&y \ge 0, z \ge 0 \\
&y \in \mathbb{Z}, z \in \mathbb{Z}
\end{aligned}$$

whose optimal solution is -6.71 achieved at $x = 0.7$, $y = 3$, $z = 2$.

Solving this program by Benders decomposition we formulate this master problem (MP)

$$\begin{aligned}
&\min -0.3x + \theta(x) \\
&0 \le x \le 5
\end{aligned}$$

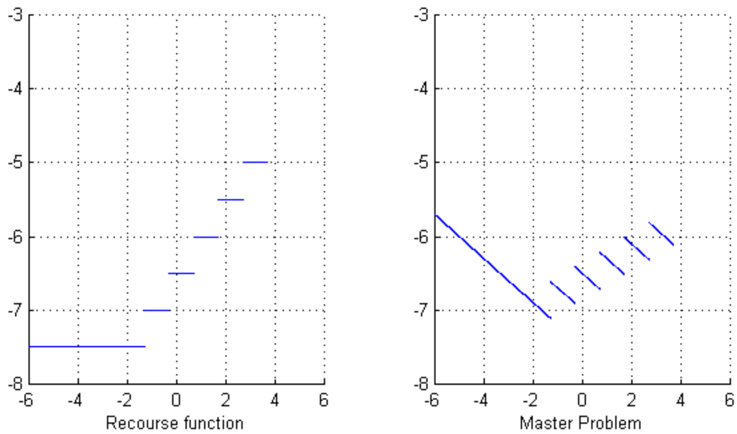
and this subproblem (SP)

$$\begin{aligned}
\theta(x) = &\min -1.5y - z \\
&y \le 3.7 - x \\
&y + z \le 5.2 \\
&y \ge 0, z \ge 0 \\
&y \in \mathbb{Z}, z \in \mathbb{Z}
\end{aligned}$$

The expression of recourse function $\theta(x)$, depicted in the next figure, is

$$\begin{array}{lllllll}
x > 3.7 & & & \Rightarrow & \theta(x) = \infty & & \\
x \in (2.7, 3.7] & \Rightarrow & y < 1 & \Rightarrow & y = 0 & z = 5 & \theta(x) = -5 \\
x \in (1.7, 2.7] & \Rightarrow & y < 2 & \Rightarrow & y = 1 & z = 4 & \theta(x) = -5.5 \\
x \in (0.7, 1.7] & \Rightarrow & y < 3 & \Rightarrow & y = 2 & z = 3 & \theta(x) = -6 \\
x \in (-0.3, 0.7] & \Rightarrow & y < 4 & \Rightarrow & y = 3 & z = 2 & \theta(x) = -6.5 \\
x \in (-1.3, -0.3] & \Rightarrow & y < 5 & \Rightarrow & y = 4 & z = 1 & \theta(x) = -7 \\
x \le -1.3 & & & \Rightarrow & y = 5 & z = 0 & \theta(x) = -7.5
\end{array}$$

Figure 1. Recourse function and Master Problem



This two stage problem presents a suitable structure to be solved with the proposed algorithm. The use of the proposed method improves the solution obtained when solving the problem relaxing the subproblem integrality conditions (traditional Benders decomposition algorithm) or using the LR algorithm as the method to solve the subproblem.

In the first case, the Benders algorithm behaves proposing as primal values $x = 5$, $x = 3.7$ and $x = 0$. For each proposal, the Benders cuts built are $x \le 3.7$, $\theta \ge 0.5x - 7.05$. For $x = 0$ the stopping rule is satisfied, and the algorithm finishes.

In case of solving the subproblem via a LR algorithm, the Benders algorithm behaves in a similar way, proposing as primal values $x = 5$, $x = 3.7$ and $x = 0$. For each proposal, the Benders cuts built are $x \le 3.7$, $\theta \ge 0.5x - 6.85$. For $x = 0$ the stopping rule is satisfied, and the algorithm finishes.

When using the proposed method, the primal values proposed are also $x = 5$, $x = 3.7$ and $x = 0$. There are also obtained the Benders cuts $x \le 3.7$ and $\theta \ge 0.5x - 6.85$. However, the outer approximation of the recourse function at $x = 0$ is improved when incorporating the perturbation cuts that constrain the domain of the recourse function. The domain $0 \le x \le 3.7$ switches to $-3.7 \le r \le 0$, being r the variable that modifies the right hand side.

Let us solve the subproblem for $x = 0$. We avoid testing feasibility of this point. The dual function is now

$$\begin{aligned} w(\lambda) = \min &- 1.5y - z + \lambda r \\ & y - 3.7 \le r \\ & y + z \le 5.2 \\ & y \ge 0, z \ge 0 \\ & -3.7 \le r \le 0 \\ & y \in \mathbb{Z}, z \in \mathbb{Z} \end{aligned}$$

obtaining

$$w(\lambda) = \begin{cases} -6.5 - 0.7\lambda & 0 \le \lambda \le 0.5 \\ -5 - 3.7\lambda & \lambda \ge 0.5 \end{cases}$$

The dual problem now is to $\max\{w(\lambda) + \lambda Tx, \lambda \ge 0\} = \max\{w(\lambda), \lambda \ge 0\} = -6.5$ for $\lambda = 0$.

This solution generates the Benders optimality cut $\theta \ge -6.5$, which is introduced into a new resolution of the master problem.

Now solve the relaxed master problem

$$\begin{aligned} &\min - 0.3x + \theta \\ &0 \le x \le 5 \\ &x \le 3.7 \\ &\theta \ge 0.5x - 6.85 \\ &\theta \ge -6.5 \end{aligned}$$

with solution $x = 0.7$, $\theta = -6.5$.

In order to generate a new Benders cut, or check the optimality of the solution proposed, it is solved the dual problem. Observe that no perturbation cut is introduced now because the point $x = 0.7$ belongs to the interior of the region S. The dual problem has to maximize the function

$$w(\lambda) + 0.7\lambda = \begin{cases} -6.5 & 0 \le \lambda \le 0.5 \\ -5 - 3\lambda & \lambda \ge 0.5 \end{cases}$$

whose optimum is achieved at -6.5. Observe that the stopping rule is now satisfied.

The solution given by the proposed method finally is $x = 0.7$, $\theta = -6.5$, $y = 3$, $z = 2$.

Summary of results

We summarize the example results on the following table and figure for the three alternatives.

	Optimal value	**Optimal Solution**		**Duality gap**
Resolution of sub problem linear relaxation	$z = -7.05$	$x = 0$	$\theta = -7.05$	0.34
Resolution of MIP subproblem by lagrangean relaxation	$z = -6.85$	$x = 0$	$\theta = -6.85$	0.14
Resolution of MIP subproblem by the proposed method	$z = -6.71$	$x = 0.7$	$\theta = -6.5$	0.00

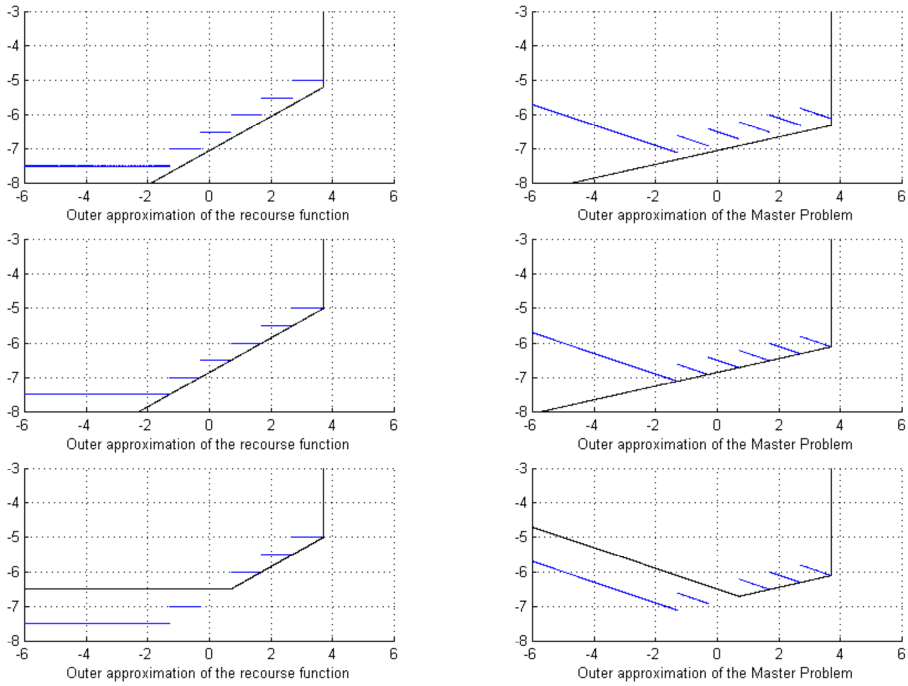


Figure 2. Comparison of generated cuts

The above figures represent the Benders optimality and feasibility cuts obtained when applying the decomposition algorithm to the numerical example of the beginning of the section. Figures on the first column plot the recourse function and the cuts obtained when solving the subproblem linear relaxation, the subproblem via the lagrangean relaxation algorithm and the subproblem via the proposed method. Observe that the proposed method finds the exact convex envelope of the recourse function over its domain $0 \leq x \leq 5$. That is not the case of the previous methods. On the right column it is plotted the outer approximation of the master problem. Recall that within a decomposition scheme the resolution of the master problem is equivalent to the resolution of the complete problem. Those outer approximations depend on the methodology chosen to solve the Benders subproblem. The correct convexification given by the third method provides an exact approximation of the mater problem and consequently its optimization gives the correct optimal point for the problem.

5. Nested Benders decomposition

Nested situations appear when the second stage (or the subproblem) of a two-stage problem is solved with decomposition. This situation creates a chain of problems that are solved proposing primal solutions for the subproblem and giving back dual values to create an outer approximation of the associated recourse functions.

In a two-stage scheme, when introducing integer variables on the process, dual values have to be calculated with a LR procedure. In a nested case, this scheme is also maintained. The feasible regions for subproblems are modified as new *perturbation cuts* are introduced from master problems, and *feasibility* and *Benders optimality cuts* are introduced from subproblems.

Consider a multi-stage problem of the form [5]

$$(TP)\qquad \begin{aligned} &\min c_1x_1 + \min(c_2x_2 + \cdots + \min c_Tx_T) \\ &T_1x_1 + W_2x_2 \qquad\qquad\qquad \le h_2 \\ &\qquad T_2x_2 + W_3x_3 \qquad\qquad \le h_3 \\ &\vdots \\ &\qquad\qquad T_{T-1}x_{T-1} + W_Tx_T \le h_T \\ &x_t \in X_t, t = 1,\dots,T \end{aligned} \tag{5.1}$$

where X_t represents the feasible region of stage t variables and may contain the non negativity constraints for some or all of the variables appearing in the problem together with integrality constraints for some of the variables.

The classical L-shaped method [15] is extended to multistage problems formulating the decision problem for stage t variables as

$$(P_t)\qquad \begin{aligned} &\theta(x_{t-1}) = \min c_tx_t + \theta(x_t) \\ &W_tx_t \le h_t - T_{t-1}x_{t-1} \\ &x_t \in X_t \end{aligned} \tag{5.2}$$

and proposing an extension of the two-stage algorithm, that outer approximates the above problem and solves sequentially a family of relaxed master problems (RMP_t) written as

$$(RMP_t)\qquad \begin{aligned} &\theta(x_{t-1}) = \min c_tx_t + \theta_t \\ &W_tx_t \le h_t - T_{t-1}x_{t-1} \\ &0 \ge \theta_t^j + \tilde{\pi}^jT_t(x_t^j - x_t) \quad j = 1,\dots,l \\ &\theta_t \ge \theta_t^i + \pi^iT_t(x_t^i - x_t) \quad i = 1,\dots,k \\ &x_t \in X_t \end{aligned} \tag{5.3}$$

On expression (5.3) x_t^j and x_t^i represents the different proposals that problem (RMP_t) has passed to the next problem. An equation like $0 \ge \theta_t^j + \tilde{\pi}^jT_t(x_t^j - x_t)$ is generated when problem (RMP_{t+1}) turns out to be infeasible for x_t^j primal value and an equation like $\theta_t \ge \theta_t^i + \pi^iT_t(x_t^i - x_t)$ is generated when problem (RMP_{t+1}) turn out to be feasible. These are denoted as within the two stage situation *infeasibility* and *optimality cuts*.

Typically nested algorithms iterate a forward pass from stage 1 to stage T solving problem (RMP_t) and obtaining primal values that modify the right hand side value of problem (RMP_{t+1}). A backward pass iterates from stage T to stage 1 solving problem (RMP_t) and providing a dual value that augment problem (RMP_{t-1}) with an optimality cut or a feasibility cut when necessary. An upper bound is obtained when a complete forward pass is performed by means of evaluating optimal stage t values in problem (TP). A lower bound is obtained at each resolution of stage 1 subproblem. The nested algorithm stops when the relative difference of those bounds is less than a specified tolerance.

On the resolution of problem (RMP_t) with LR, once a multiplier value λ is proposed, the following problem that computes the dual function $w_t(\lambda)$ must be solved

$$
(PR_{t,\lambda}) \qquad
\begin{aligned}
& w_t(\lambda) = \min c_t x_t + \theta_t + \lambda r \\
& W_t x_t - h_t \le r \\
& 0 \ge \theta_t^j + \tilde{\pi}^j T_t (x_t^j - x_t) \quad j = 1,\dots,l \\
& \theta_t \ge \theta_t^i + \pi^i T_t (x_t^i - x_t) \quad i = 1,\dots,k \\
& x_t \in X_t, r \in R_t^n
\end{aligned}
\tag{5.4}
$$

and a dual problem is optimized

$$
(D_{t,x_{t-1}}) \qquad \max\left\{ w_t(\lambda) + \lambda T_{t-1} x_{t-1} \right\}
$$

Here R_t^n indicates the perturbation region of stage t available at iteration n of the nested decomposition algorithm. It is the outer approximation of the tender variable region R_t, which is defined as

$$
R_t = \left\{ r \,/\, \exists x_{t-1} \in X_{t-1} \,/\, r = -T_{t-1} x_{t-1} \right\}
$$

The optimization of the dual problem is replaced by the iterative resolution of the lagrangean subproblem $(PR_{t,\lambda})$ and a master dual problem $(MD_{t,x_{t-1}})$ which is given as

$$
(MD_{t,x_{t-1}}) \qquad
\begin{aligned}
& \max w + \lambda T_{t-1} x_{t-1} \\
& w \le c_t x_t^k + \theta_t^k + \lambda r^k \quad i = 1,\dots,K
\end{aligned}
\tag{5.5}
$$

Remarks. Above description of the relaxed dual problem does not incorporate phase 1 neither bounding cuts that avoid unbounded lagrangean subproblems. Of course it is possible that a primal proposal turns problem (RMP_t) into an infeasible one, and in that situation termination of the lagrangean algorithm at phase 1 will provide a feasibility cut for problem (RMP_{t-1}). The notation has been reduced for simplicity of the exposition.

Otherwise, on repetitive resolutions of the LR algorithm that solves problem (RMP_t), it is possible that previously calculated lagrangean cuts remain valid in a new resolution of the LR method. Whether these cuts are valid or not may be checked before the performing the LR algorithm. A code that performs this algorithm may improve its efficiency incorporating the valid cuts in new resolutions [4].

Another point that needs to be commented is the calculation of the shadow (and consequently the perturbation region) for nested situations. This presents an additional complexity for third and future stages on a nested scheme. A description of the computation of these cuts can be found in Appendix B.

Next section presents the application of the proposed decomposition algorithm to a large scale optimal allocation problem extracted from a large problem of optimal railway electrification [13]. The results are focused to compare the direct resolution of the problem and the resolution by used of the proposed algorithm.

6. A LARGE SCALE PROBLEM

The proposed algorithm has been codified in C\C++ using Concert Technology [8], which enables to solve mathematical programming problems by use of the commercial optimizer Cplex [8]. The code performs the Benders decomposition solving the proposed LR method when a new dual variable is needed to outer approximate any recourse function. The perturbation cuts introduced on the code are those that incorporate the natural bounds for the tender variables. The remaining of the perturbations cuts, as described on this paper, have not been introduced. This induces a gap into the final solution that measures the accuracy of the solution obtained. That accuracy is compared with that obtained by the direct resolution of the problem with Cplex.

6.1. Problem description

This problem consists of determining the optimal division into sectors of a whole line that needs to be electrified. Apart from deciding the number of sectors and the length for each sector, it is also necessary to select the proper overhead power cable that will be used to cover that sector. On the division between sectors is allocated a substation, which provides the necessary power for the train consumptions. The problem introduces very simplified security constraints, implying that a substation must be able to feed the sector on its left as well as the sector on its right, in case another substation goes down. There is a fixed cost for each sector that represents the installation cost for a new substation and a variable cost of the type of catenary used. The problem is solved modelling it with a stair case structure that makes it suitable to be solved using the decomposition algorithm.

The problem has been modelled focusing of the individual subproblem that appears on each sector. The modelling is focused so to have the minimum number of coupling constraints between sectors. In the modelling presented, this number of coupling constraints is limited to one, the total length coverage constraints. Let

L	length of the line
I	collection of possible sectors
T	set of trains
SC	set of scenarios
C	set of catenaries

the collection of sets for the problem and

$TrainPos_{t,sc}$	Position of train t on the scenario sc
$I_{t,sc}$	Intensity of train t on the scenario sc
Z_c	Impedance of catenary c
$Fixed_i$	Fixed cost of employing sector i
$CatCost_c$	Variable cost of catenary c

the collection of data necessary to solve it. In order to create the mathematical programming problem that solves it, the next constraints are necessary.

Total length coverage

This constraint it the one that connect sectors, implying that the beginning of a sector must coincides with the end of previous sector.

$$Length_i = Pos_i - Pos_{i-1} \tag{6.1}$$

where

$Length_i$	Length of sector i

Pos_i Position of sector i.
$Pos_{i-1} = 0$ if i represents the first sector and
$Pos_i = L$ if i represents the last sector.

Existence of sector

Initially, the number of possible sectors is greater than the final number of sectors that the optimization algorithm decides to use. There is a cost associated with each sector that represents the fixed cost of the substation at the end of the sector. That cost is later introduced into the objective function through a binary variable.

$$L * u_i > Length_i \tag{6.2}$$

where

u_i Boolean variable that indicates the existence of sector i

Partial length coverage

Indicates the length of a sector may be covered by at least one catenary

$$Length_i \leq \sum_{c \in C} LengthCat_{i,c} \tag{6.3}$$

where

$LengthCat_{i,c}$ Length of sector i covered with catenary c

Relative Positions

The idea of the modeling is to have a pair of variables that determine, for each train and each scenario, the distance to the left extreme and the distance to the right extreme of its current sector. The current sector is understood as the sector where the train is. These variables are obtained after the logical manipulations of coming lines. Let

$$LeftDist_{t,sc,i} = TrainPos_{t,sc} - Pos_i + Length_i \tag{6.4}$$

$$RightDist_{t,sc,i} = Length_i - RightDist_{t,sc,i} \tag{6.5}$$

where

$LeftDist_{t,sc,i}$ indicates the distance of each train t at scenario sc to the left extreme of each sector i. Observe that this variable is positive for each train current sector and for the previous sectors. The distance is negative for the sectors that comes after the sector i. Symmetrically, $RightDist_{t,sc,i}$ indicates the distance of each train t at scenario sc to the right extreme of each sector i. This distance is positive for each train current sector and sectors that comes after it. The distance is negative for the sectors that precede sector i.

Distance $LeftDist_{t,sc,i}$ to identify a binary variable that indicates the current sector and preceding sectors of each train t of scenario sc.

$$0 \leq L * v_{t,sc,i} - LeftDist_{t,sc,i} \leq L \tag{6.6}$$

Observe that

$$\begin{cases} \text{if} \quad LeftDist_{t,sc,i} < 0 \quad \text{then} \quad v_{t,sc,i} = 0 \\ \text{if} \quad LeftDist_{t,sc,i} > 0 \quad \text{then} \quad v_{t,sc,i} = 1 \end{cases}$$

That binary variable $v_{t,sc,i}$ is used to determine a positive variable $DefRightDist_{t,sc,i}$ (*Definitive Right Distance*), that computes the distance of each train to the right extreme of its current sector and gets a null value for the remaining sectors. Another intermediate positive variable $RightDist^{+}{}_{t,sc,i}$ is necessary for the logical getting of the definitive distance. The next constraints are then included in the modeling.

$$RightDist^{+}{}_{t,sc,i} \geq RightDist_{t,sc,i} \tag{6.7}$$

$$0 \leq RightDist^{+}{}_{t,sc,i} - DefRightDist_{t,sc,i} \leq (1 - v_{t,sc,i}) * L \tag{6.8}$$

Observe that above group of constraints do what it is required from them. The optimization problem that will be constructed penalizes the variables that represent the definitive distances, so that,

If the train is its current sector, then $v_{t,sc,i} = 1$, and consequently

$$DefRightDist_{t,sc,i} = RightDist^{+}{}_{t,sc,i} = RightDist_{t,sc,i}$$

For a sector i preceding the current sector of a train, $v_{t,sc,i} = 1$, and consequently

$$DefRightDist_{t,sc,i} = RightDist^{+}{}_{t,sc,i} \equiv 0$$

For a sector i following the current sector of a train, $v_{t,sc,i} = 0$, and constraint (6.8) imposes no condition over the variable $DefRightDist_{t,sc,i}$, so that the optimization problem will take the variable to it lower bound, and consequently,

$$DefRightDist_{t,sc,i} \equiv 0$$

The procedure to obtain the symmetrical variable to $DefRightDist_{t,sc,i}$ may follows the symmetrical steps followed to obtain the definitive right distance variable. An alternative that uses previously logical constraints is presented.

Consider a binary variable $w_{t,sc,i} = 0$ that indicates the current sector of a train through next expression

$$DefRightDist_{t,sc,i} \leq L * w_{t,sc,i} \tag{6.9}$$

Expression that implies

$$\{\text{If} \;\; DefRightDist_{t,sc,i} > 0 \;\; \text{then} \; w_{t,sc,i} = 1$$

Now, the binary variable $w_{t,sc,i}$ is used to determine the positive variable $DefLeftDist_{t,sc,i}$. An intermediate positive variable $LeftDist^{+}{}_{t,sc,i}$ is also necessary.

$$LeftDist^{+}{}_{t,sc,i} \geq LeftDist_{t,sc,i} \tag{6.10}$$

$$0 \leq LeftDist^{+}{}_{t,sc,i} - DefLeftDist_{t,sc,i} \leq (1 - w_{t,sc,i}) * L \tag{6.11}$$

Equations (6.10) and (6.11) do what they are required to do,

If the train is in its current sector, then $w_{t,sc,i} = 1$, and consequently

$$DefLeftDist_{t,sc,i} = LeftDist^{+}{}_{t,sc,i} = LeftDist_{t,sc,i}$$

In other case $w_{t,sc,i}$ is not commanded to get any value, but a null value of this binary variable will allow the variable $DefLeftDist_{t,sc,i}$ to be free to reach its lower bound, so that

$$DefLeftDist_{t,sc,i} \equiv 0$$

Voltage Drop

This value depends on the catenary chosen to cover each sector. In order to avoid a nonlinear modeling of the problem, the extra positive variables $CatLeftDist_{t,sc,i,c} \equiv 0$ and $CatRightDist_{t,sc,i,c} \equiv 0$ are introduced and related to previous positive variables through next expressions.

$$DefLeftDist_{t,sc,i,c} = \sum_{c \in C} CatLeftDist_{t,sc,i,c} \tag{6.12}$$

$$DefRightDist_{t,sc,i,c} = \sum_{c \in C} CatRightDist_{t,sc,i,c} \tag{6.13}$$

Now, the security criterion imposed on the model requires the voltage drop to the left extreme and the voltage drop to the right extreme not to exceed a determined value $V_{\max}$.

$$\sum_{t \in T} \sum_{c \in C} Z_c * CatRightDist_{t,sc,i,c} \leq V_{\max} \tag{6.14}$$

$$\sum_{t \in T} \sum_{c \in C} Z_c * CatLeftDist_{t,sc,i,c} \leq V_{\max} \tag{6.15}$$

Use of Catenary

A collection of constraints must be imposed that imply the use of a catenary.

Let $Cat_{i,c}$ a binary variable that indicates if catenary c has been used on sector i. Then, the next constraints are necessary.

$$CatLeftDist_{t,sc,i,c} \leq L * Cat_{i,c} \tag{6.16}$$

$$CatRightDist_{t,sc,i,c} \leq L * Cat_{i,c} \tag{6.17}$$

$$LengthCat_{i,c} \leq L * Cat_{i,c} \tag{6.18}$$

$$\sum_{c \in C} Cat_{i,c} = 1 \tag{6.19}$$

Objective function

The objective function incorporates the cost derived from the fixed cost of each sector and the cost associated with the catenaries employed.

$$\sum_{i \in I} \left(Fixed_i * u_i + \sum_{c \in C} CatCost_c LengthCat_{i,c} \right)$$

The modeling of this problem has been done so that variables and constraints attached to a sector are independent from the collection of variables and constraints for another sector. With this modeling, the constraints matrix of this problem presents a stair case structure. As an example consider figure 3, that represents the constraints matrix for the presented problem for a situation of five sectors.

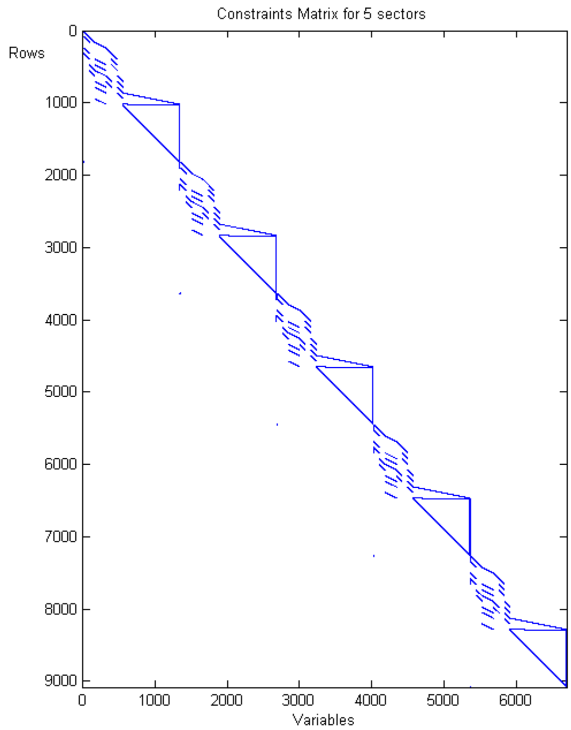


Figure 3 . Constraints Matrix for a five sectors electrification problem

6.2. Numerical Results

This section presents numerical results about the application of the decomposition algorithm to the problem above presented. The numerical tests have been carried out over lines with 200, 250 and 300 kilometers. A collection of scenarios is considered that represents the train positions along the line on different time intervals. Scenarios are a collection of possibilities that may be considered simultaneously so that the physical constraints of the voltage across the line is maintained in its security levels. 6 scenarios are considered that represent the trains separated by 24 kilometers, with the first train starting on kilometer 0, 4, 8, 12, 16 and 20 respectively.

The problem for different lengths has been solved with different number of sectors. In any case the number of scenarios has been fixed to 6. The more sectors, the bigger the size of the problem that has to be solved. The size of the problem also increases with the total length of the line, because of the number of variables that come associated with the number of trains, that naturally increases.

The problems have also been solved directly with Cplex 7.5, and comparison about the direct resolution and the resolution with the decomposition algorithm is presented. When using the decomposition algorithm, it has been considered one subproblem for each of the sectors implied.

The summary of the computational results are presented on next table, al later commented.

200 Kilometers and 3 Sectors		**Cplex**		**Benders**		
Rows	3792				Lower Bound	Upper Bound
Columns	2797	Solution	8227.2	Iteration 7	8215.68	8227.2
Non Zeros	11006	Accuracy	0.0%	Accuracy	0.14 %	
Binaries	342	Time	5.4 secs	Time	43.4 secs	

200 Kilometers and 4 Sectors		**Cplex**		**Benders**		

Rows	5056				Lower Bound	Upper Bound
Columns	3729	Solution	8227.2	Iteration 9	8190.9	8227.2
Non Zeros	14711	Accuracy	0.0%	Accuracy	0.44 %	
Binaries	456	Time	40 secs	Time	163.5 secs	

250 Kilometers and 4 Sectors		**Cplex**			**Benders**	
Rows	6160				Lower Bound	Upper Bound
Columns	4545	Solution	10534	Iteration 10	10278.2	10534
Non Zeros	17951	Accuracy	0.0%	Accuracy	2.42 %	
Binaries	552	Time	266.5 secs	Time	236.6 secs	

250 Kilometers and 5 Sectors		**Cplex**			**Benders**	
Rows	7700				Lower Bound	Upper Bound
Columns	5681	Solution	10534	Iteration 13	10252	10534
Non Zeros	22472	Accuracy	2.66 %	Accuracy	2.67 %	
Binaries	690	Time	2247 secs	Time	905 secs	

300 Kilometers and 4 Sectors		**Cplex**			**Benders**	
Rows	7264				Lower Bound	Upper Bound
Columns	5361	Solution	12872.6	Iteration 9	12556.9	12874.7
Non Zeros	21191	Accuracy	2.45 %	Accuracy	2.46 %	
Binaries	648	Time	203 secs	Time	276 secs	

300 Kilometers and 5 Sectors		**Cplex**			**Benders**	
Rows	9080				Lower Bound	Upper Bound
Columns	6701	Solution	12872.6	Iteration 10	12296.3	12840.8
Non Zeros	26528	Accuracy	3.85 %	Accuracy	4.24 %	
Binaries	810	Time	10584 secs	Time	830 secs	

300 Kilometers and 6 Sectors		**Cplex**			**Benders**	
Rows	10896				Lower Bound	Upper Bound
Columns	8041	Solution	12840.8	Iteration 9	12273.1	12840.8
Non Zeros	31865	Accuracy	4.41%	Accuracy	4. 42 %	
Binaries	972	Time	9300 secs	Time	949.6 secs	

Table 2. Summary of Optimal Electrification Problem results

Discussion about the results.

Table 2 describes, on its left part, the problem size for the different cases tested. On the right part of the table appears information about the resolution of the problems with the Benders decomposition algorithm. It is presented the computation time as well as the accuracy of the solution obtained. The upper bound column gives the objective value that the algorithm achieves. On the central part of the table it is presented information about the resolution of the problems directly. The branch and bound algorithm is stopped as soon as the accuracy of the Benders method is achieved, in order to get comparable computation times. The solution given is that of the objective function at the moment of stopping the algorithm.

On small problems, the objective function achieved by the decomposition algorithm is that of the direct resolution of the problem. However, the solution is not the same. This problem is highly degenerate so that multiple alternative solution exists. For example, for the case of 250 kilometers and 4 sectors considered, the final position for the sectors given by the direct resolution of the problem is $[48, 116, 184, 250]$ while with the Benders algorithm the solution is $[64, 132, 200, 250]$.

When the number of introduced sectors is larger than the necessary, a sector will be assigned a length of zero. This is for example the situation with the case of 300 kilometers. The solution obtained when selecting just four sectors is worse than the one obtained when selecting five sectors. However, the solution with six sectors does not improve the solution of the five sectors case. In this case, the first sector out of six is assigned length zero and the remaining decision variables are exactly the same. The solution for the five sectors problems is $[64, 132, 200, 236, 300]$ and the solution for the six sectors problems is $[0, 64, 132, 200, 236, 300]$.

It is necessary to comment that the larger the problem, the more efficient the decomposition algorithm is when applied to this problem. In the problem with 300 kilometers and 5 scenarios, not only the computation time outperforms that of the Cplex, but the solution obtained presents minor cost. This comparison of the computation time in terms of the size of the problem is roughly presented on figure 4. The horizontal axe present the number of binary variables and the vertical axe presents computation times in seconds.

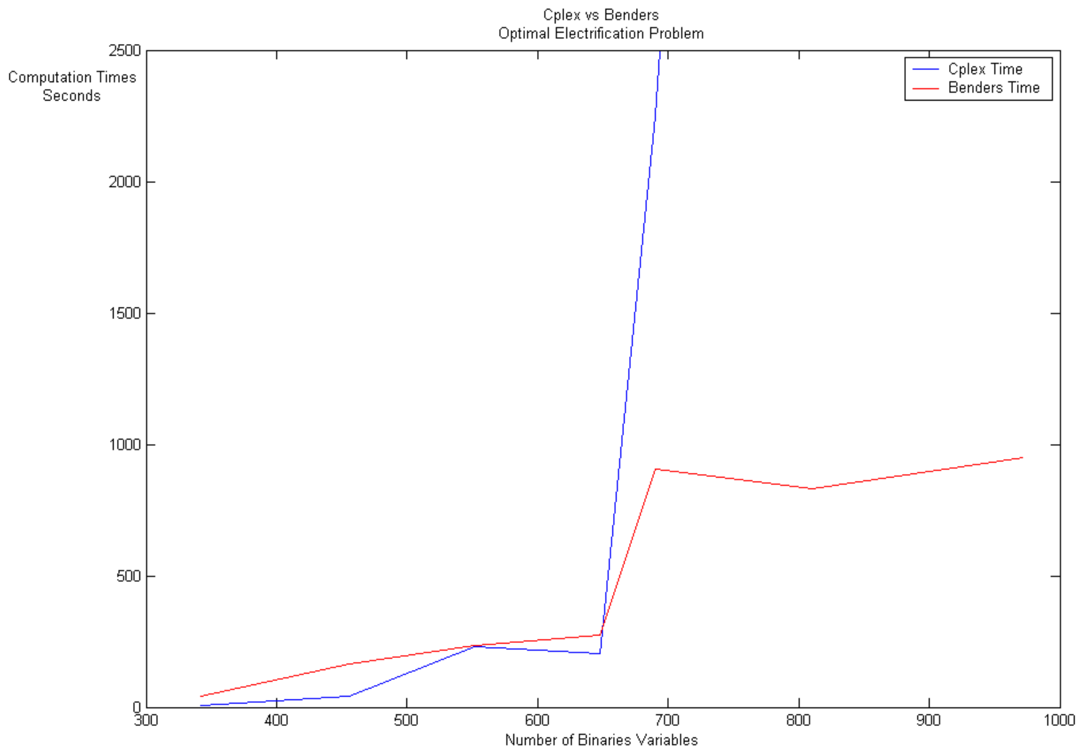


Figure 4. Comparison of computation times

Finally, this presentation of the results is finished with a figure of the outer approximation of the recourse functions that the generated cuts give. Figure 5 presents those results for the case of 300 kilometers and 5 sectors. In this situation there exist 4 recourse functions that indicate the cost of the electrification from the point where the sector ends. For instance, the upper subfigure of figure 5 indicates the cost of the electrification depending of the final position of sector 1. The same represents the remaining subfigures, but introducing in the plot those infeasibility cuts that where obtained in the process of the decomposition algorithm. These infeasibility cuts obtained for this particular case had next form

$$Pos_2 \geq 16$$
$$Pos_3 \geq 100$$
$$Pos_4 \geq 216$$

limiting the final positions of the sectors along the line.

These infeasibility cuts are obtained in the algorithm when a subproblem turns out to be infeasible. A particular feature of this electrification problem is the appearance of feasible LP problems that turns into infeasible MIP problems. In that case the LR phase 1 provides the necessary feasibility cuts that the algorithm requires. In the particular case of 300 kilometers and 5 sectors, the expression $Pos_4 \geq 216$ is a feasibility cut obtained via LR relaxation phase 1. Previously was obtained the cut $Pos_4 \geq 200$ on previous iterations of the Benders method. Feasibility cut that turn insufficient for the case.

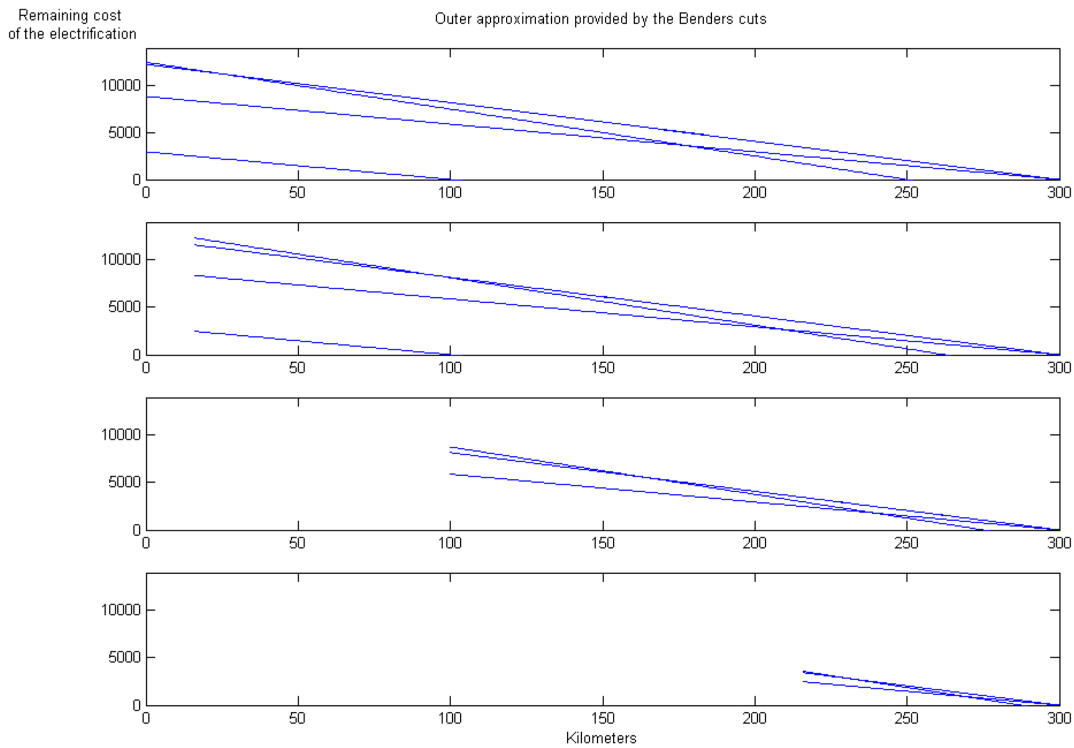


Figure 5. Outer approximation provided by the Benders cuts

7. Conclusions

This paper has presented a finite Benders decomposition algorithm for mixed integer linear programs. Following traditional lines about nonlinear duality theory, the non convex recourse function is convexified formulating a LR problem whose resolution produces correct dual values that outer approximate the non convex recourse function.

In the Benders algorithm, the recourse function is understood as the perturbation function of the subproblem when the RHS of coupling constraints is modified. For the LR procedure, a family of cuts denoted as perturbation cuts is introduced that constrains the perturbation region. This perturbation function domain is precisely the projection or shadow of the first stage feasible region over the first stage coupling variables space and is continuously updated as Benders algorithm proceeds and new perturbation cuts are found.

The algorithm converges to the optimal value of the problem, and at the optimal solution there is no duality gap between the primal solution and the resolution through the LR.

The situation is generalized to nested decomposition, with the added difficulty of calculating the perturbation region for third and further stages.

The paper has finally presented an application of the algorithm to a large scale problem, where the advantage of applying it to large scale instances of it is appreciated.

8. References

1. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.(1993): *Nonlinear programming, theory and algorithms*, Wiley
2. Benders, J.F.(1962): *Partitioning procedures for solving mixed-variables programming problems*, Numerische Mathematik 4, 238-252
3. Birge, J.R., Louveaux, F.V. (1997): *Introduction to stochastic programming*, Springer
4. Cerisola, S., Ramos, A. (2003): *An Algorithm for Multistage Integer Programs.* Instituto Investigación Tecnológica, Universidad Pontificia Comillas (IIT-03-083A), Madrid
5. Ermoliev, Y., Wets, R. (1988): *Numerical Techniques for Stochastic Optimization.* Springer-Verlag , Berlin
6. Geoffrion, A. M. (1972): *Generalized Benders Decomposition* J. Optim. Theory Appl. 10, 4, 237-260
7. Geoffrion, A. M. (1974): *Lagrangean relaxation for integer programming* Mathematical Programming Study 2. Vol 2, pp 82-114
8. ILOG (2003): http://www.ilog.com
9. Lemaréchal, C., Renaud, A. (2001) : *A geometric study of duality gaps, with applications* Mathematical programming, Serie A. 90:399-427
10. Louveaux, F.V., Van der Vlerk, M.H. (1993): *Stochastic programming with simple integer recourse*. Mathematical programming, 61:301-325
11. Minoux, M. (1986): *Mathematical Programming, theory and algorithms*, Wiley
12. Nemhauser, G.L., Wolsey, L.A. (1988): *Integer and combinatorial optimization*, Wiley, New York
13. Pilo, E. (2003): Optimizacion del Diseño de la Electrificacion de Trenes de Alta Velocidad. PhD thesis, Universidad Pontificia Comillas, Madrid
14. Van der Vlerk, M.H. (1995): *Stochastic programming with integer recourse*. PhD thesis, University of Groningen, The Netherlands
15. Van Slyke, R.M., Wets, R. (1969): *L-Shaped linear programs with applications to optimal control and stochastic programming*, SIAM Journal of Applied Mathematics, 17, 4, 638-663

Appendix A

A.1 Linear problem

Consider the problem

$$(MP) \qquad \begin{aligned} & \min c_1 x_1 + c_2 x_2 + \theta \\ & 0 \geq \theta^j + \tilde{\pi}^j T(x_0^j - x_2) \quad j = 1,\dots,l \\ & \theta \geq \theta^i + \pi^i T(x_0^i - x_2) \quad i = 1,\dots,k \\ & x \in \overline{X}, \overline{X} = \left\{ A_1 x \leq a_1, x \in \mathbb{R}^{n_1+m_1} \right\} \end{aligned} \qquad \text{(A.1)}$$

Let (x_1, x_2, θ) be the optimal solution of this problem and let $x_2 = P(x_1, x_2, \theta)$ the projection of this point over the coupling variable space $\mathbb{R}^{n_2'}$. Let $(d_1, d_2, \dots, d_{n_1+m_1}, d_{n_1+m_1+1})$ the edges of the feasible region at point (x_1, x_2, θ). Let $(\widehat{d}_1, \widehat{d}_2, \dots, \widehat{d}_{n_1+m_1}, \widehat{d}_{n_1+m_1+1})$ the projection of these edges over the coupling variable space. We have then that x_2 is an extreme point of the shadow S if and only if the coupling variable space cannot be expressed as a positive linear combination of the vectors $(\widehat{d}_1,\widehat{d}_2,...,\widehat{d}_{n_1+m_1},\widehat{d}_{n_1+m_1+1})$. This situation leads to the following criterion based on Farkas' law.

Criterion 1.

Let $e_1=(1,0,0,...,0)$, $e_2=(0,1,0,...,0)$, ... , $e_{n_2'}=(0,0,0,...,1)$, $e_0=(-1,-1,-1,...,-1)$. It is immediately that every $v\in\mathbb{R}^{n_2'}$ can be expressed as a positive linear combination of previous elements.

Consider the family of problems

$$(P_i)\qquad \begin{array}{l}\min e_i x\\ Dx\le 0\\ -1\le x\le 1\end{array}\qquad\text{(A.2)}$$

where D represents a matrix whose rows are the elements $(\widehat{d}_1,\widehat{d}_2,...,\widehat{d}_{n_1+m_1},\widehat{d}_{n_1+m_1+1})$.

Positive solution of this problem (P_i) implies e_i cannot be expressed as a positive linear combination of vectors $(\widehat{d}_1,\widehat{d}_2,...,\widehat{d}_{n_1+m_1},\widehat{d}_{n_1+m_1+1})$. Consequently, positive solution of at least one of those problems implies x_2 is an extreme point of the shadow S. Denote $C(x_2)$ the cone defined at point x_2 by directions $(\widehat{d}_1,\widehat{d}_2,...,\widehat{d}_{n_1+m_1},\widehat{d}_{n_1+m_1+1})$.

Criterion 2.

This criterion is modified so as to find the minimum set of projected edges that define the same positive cone $C(x_2)$.

For $k=1$ to $K=n_1+m_1+1$ consider the problem

$$(P_k)\qquad \begin{array}{l}\min \widehat{d}_k x\\ \widehat{D}_k x\le 0\\ -1\le x\le 1\end{array}\qquad\text{(A.3)}$$

where $\widehat{D}_k$ represents matrix D with the k-row removed.

Positive solution of this problem implies $\widehat{d}_k$ is not a positive linear combination of the remaining vectors. So it is an extreme direction of the cone at point x_2. On the contrary, negative or null solution of this problem implies $\widehat{d}_k$ can be represented as a positive linear combination of the remaining vectors. So it is not necessary any more and it is deleted from the family $(\widehat{d}_1,\widehat{d}_2,...,\widehat{d}_{n_1+m_1},\widehat{d}_{n_1+m_1+1})$ and matrix D updated with the k-row removed. This algorithm ends with a minimal set of extreme directions that generates the positive cone at point x_2, $C(x_2)$.

These both criteria can be interchanged so as to firstly obtain a minimal set of extreme directions and later to check if the positive cone generated (by the remaining vectors) is the whole space or not.

A.2 Mixed integer problem

In this case the solution of problem

$$(MP)\qquad \begin{array}{ll}\min c_1x_1+c_2x_2+\theta & \\ 0\ge\theta^j+\tilde{\pi}^jT(x_0^j-x_2) & j=1,\dots,l\\ \theta\ge\theta^i+\pi^iT(x_0^i-x_2) & i=1,\dots,k\\ x\in X & \end{array}\qquad\text{(A.4)}$$

ends with a set of terminal nodes N, together with a solution $x_n=(x_1^n,x_2^n)$ for each node and a family of edges $(d_1^n,d_2^n,...,d_{n_1+m_1}^n,d_{n_1+m_1+1}^n)$ for each node. Consider $x_0=(x_1^0,x_2^0)$ the optimal solution of problem (MP). Then we consider the positive cone generated by all projected edges together with vectors connecting all the projected solutions. So we consider the positive cone generated by $(\widehat{d}_1^n, \widehat{d}_2^n, ..., \widehat{d}_{n_1+m_1}^n, \widehat{d}_{n_1+m_1+1}^n)$ from $n = 1, \dots, N$ and the vectors $(\overrightarrow{x_2^0 x_2^1}, \overrightarrow{x_2^0 x_2^2}, \dots, \overrightarrow{x_2^0 x_2^N})$.

We apply now criterion 2 for considering a minimal set of extreme directions and later criterion 1 (over this minimal set) to check whether x_2^0 is an extreme point for S or not.

APPENDIX B

B.1 Inequality constraints

Consider a nested situation as the one presented on section 6 and let and let the problem (P) take the form

$$(P) \qquad \begin{aligned} &\min c_{11}x_1 + c_{12}x_2 + c_{21}y_1 + c_{22}y_2 + c_3 z \\ &T_1 x_2 + W_2 y \qquad \le h_1 \\ &\qquad T_2 y_2 + W_3 z \le h_2 \\ &x = (x_1, x_2) \in X, y = (y_1, y_2) \in Y, z \in Z \end{aligned} \tag{B.1}$$

It is defined the shadow S_1 of the region X over the coupling variable space as

$$S_1 = \{x_2 / \exists x_1 / (x_1, x_2) \in X\} \tag{B.2}$$

so that its associated perturbation region keeps the form

$$R_1 = \{r / \exists x_2 \in S_1 / r = -T_1 x_2\} \tag{B.3}$$

The interest on further stages lays on the next projection (particularized for three stages in the appendix)

$$S_2 = \{y_2 / \exists x = (x_1, x_2) \in X, \exists y_1 / x \in X, y = (y_1, y_2) \in Y, A_1 x_2 + A_2 y \le b_1\} \tag{B.4}$$

so that its associated perturbation region keeps the form

$$R_2 = \{r / \exists y_2 \in S_2 / r = -T_2 y_2\} \tag{B.5}$$

In a nested decomposition procedure, we are interested in finding S_1 and S_2. Constraints for region S_1 (and consequently for R_1) are found after solution of a first stage master problem and application of criterion 1 and 2. A few comments are necessary when solving the second stage problem. Let the master problem solved in a second stage take the form

$$(MP) \qquad \begin{aligned} &\min c_{21}y_1 + c_{22}y_2 + \theta_2 \\ &W_2 y \le h_1 - T_1 x_2 \\ &\theta_2 \ge \theta_2^p(y_2) \\ &y \in Y \end{aligned} \tag{B.6}$$

Let (y_1, y_2, θ_2) be the optimal solution. The problem now is to check if $y_2 = P(y_1, y_2, \theta_2)$ is an extreme point of the region S_2 and, in that case, to calculate the associated positive cone to constrain the perturbation region. It might be pointed out that S_2 is the projection of the feasible region

$$\begin{aligned} &T_1 x_2 + W_2 y \le h_1 \\ &\theta_2 \ge \theta_2^p(y) \\ &x \in X, y \in Y \end{aligned} \tag{B.7}$$

over the coupling variable space y_2, defined by variables connecting second and third stage.

Definition. y_2 is a 2-extreme point of S_2 if it satisfies criterion 1 when considering $y_2 = P(y_1, y_2, \theta_2)$, with (y_1, y_2, θ_2) optimal solution of second stage problem.

We immediately obtain the following results.

Fact. If y_2 is not a 2-extreme point of S_2 then y_2 is not an extreme point of S_2.

Assume then y_2 is a 2-extreme point of S_2. We have the following fact.

Fact. If y_2 is a 2-extreme point and $W_2 y \leq h_1 - T_1 x_2$ is not active, then y_2 is an extreme point.

In case $W_2 y \leq h_1 - T_1 x_2$ is an active constraint, then y_2 will be an extreme point if an only if the region $\{y, \theta_2 \geq \theta_2^p(y), W_2 y \leq b_1 - T_1 x_2\}$ is as biggest as possible, i.e., there is no $x' = (x_1', x_2')$ such that the region $\{y, \theta_2 \geq \theta_2^p(y), W_2 y \leq h_1 - T_1 x_2\}$ is included in the region $\{y, \theta_2 \geq \theta_2^p(y), W_2 y \leq h_1 - T_1 x_2'\}$.

Then consider the point $x = (x_1, x_2)$ obtained at first stage problem. In case x_2 is not an extreme point for the region S_1 (projection of first stage feasible region over the coupling variable space), then x_2 can freely move alongside any direction, so that there is a vector d with $T_1 x_2 \geq T_1(x_2 + d)$. We are assuming without lost of generality T_1 is a nonsingular matrix.

In case x is an extreme point for its region S_1 then consider the positive cone $C(x_2)$ to be given by the extreme directions $(\widehat{d}_1, \widehat{d}_2, ..., \widehat{d}_k)$. These extreme directions are found as a result of applying criterion 2 to the first stage solution.

Then, if $T_1 \widehat{d}_k \geq 0$, $\forall k: 1, ..., K$ we can assure y_2 is an extreme point for region S_2.

The previous results are summarized on the following proposition.

Proposition. Let $y_2 = P(y_1, y_2, \theta_2)$ the projection of the optimal second stage solution onto the space of coupling variables. Then y_2 is an extreme point of S_2 if and only if:

1. y_2 is a 2-extreme point and $W_2 y \leq h_1 - T_1 x_2$ is not an active constraint.
2. y_2 is a 2-extreme point, x_2 is an extreme point of S_1 and $T_1 \widehat{d}_k \geq 0$, $\forall \widehat{d}_k$ extreme direction of $C(x_2)$.

B.2 Equality constraints

Now the situation is slightly different. Consider the problem

$$(MP) \qquad \begin{aligned} & \min c_{21} y_1 + c_{22} y_2 + \theta_2 \\ & W_2 y = h_1 - T_1 x_2 \\ & \theta_2 \geq \theta_2^p(y_2) \\ & y \in Y \end{aligned} \tag{B.8}$$

and assume its feasibility. In other case a feasibility cut will be generated and we will not be worried about the optimal solution. In case feasibility its resolution is equivalent to the problem

$$\begin{aligned} & \min c_{21} y_1 + c_{22} y_2 + \theta_2 + M y^+ \\ & W_2 y - y^+ \leq h_1 - T_1 x_2 \\ & \theta_2 \geq \theta_2^p(y_2) \\ & y \in Y, y^+ \geq 0 \end{aligned} \tag{B.9}$$

with M big enough. Then we transform it to the situation presented on section B.1.