A FINITE BENDERS DECOMPOSITION ALGORITHM FOR MIXED INTEGER PROBLEMS. RESOLUTION THROUGH PARAMETRIC BRANCH AND BOUND

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 nonconvex recourse function is approximated via a modified lagrangean relaxation algorithm. The decomposition algorithm is understood as a convexification procedure of the perturbation function that appears when a first stage variable is fixed. Mixed integer subproblems are solved through a parametric branch and bound that simultaneously updates a correct Lagrange multiplier value. Extension to nested decomposition is presented.

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

1. Introduction

Benders decomposition algorithm [2,11] solves a mixed integer linear problem (MILP) bunching 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 considered in 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When the subproblem is non convex, then the recourse function is non convex [3], and the former approach is no longer valid.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 [7,10]. 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 [5], where the subproblem is solved by using LR [4]. However, for non-complete recourse problems, 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 extra development consists of introducing in the subproblem cuts that constrain the region over which the convexification procedure is carried out.

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. Then, a simple example to clarify the previous concepts is presented. Subsequently, the resolution of the subproblem via a parametric branch and bound algorithm is presented, with the purpose of reducing the explored nodes all over the algorithm. Finally, the difficulties encountered when extending this technique to nested case are presented.

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.

Consider a problem of the form

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

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

It is defined the generalized graph G of the problem as

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

so that the problem (P) is reinterpreted as finding a point (r, r_0) in G with minimum ordinate and $r \le 0$, see [6]. G is the image of X under the transformation (g, f). The generalized epigraph of the problem, see [6], is defined as

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

Closely related with this idea is the concept of perturbation function. Consider that the right hand side of problem (P) is being modified obtaining a family of problems whose solutions define a function on the right hand side parameter introduced. This function is known in the literature as *perturbation function* or *value function* [1,9].

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

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 $y = 0$. 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

$$w(\lambda) = \begin{array}{l} \min \lambda g(x) + f(x) \\ x \in X \end{array} \tag{2.6}$$

Assume $w(\lambda)$ has a finite value, then there exists $x^i \in X$ with $\lambda g(x^i) + f(x^i) = w(\lambda)$. This optimal solution determines a level curve $L = \{(r, r_0) / \lambda r + r_0 = \lambda g(x^i) + f(x^i)\}$. So that for $r = 0$ we have the point $(0, \lambda g(x^i) + f(x^i))$ and so it is stated that $\upsilon(0) \geq \lambda g(x^i) + f(x^i) = w(\lambda)$. The *dual problem* traditionally consists of finding the maximum of those minimum values $w(\lambda)$.

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

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

$$w(\lambda) = \min \{f(x^k) + \lambda g(x^k) / k = 1, \dots, K,\ x^k \in \text{extr}(X)\} \tag{2.8}$$

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

$$\begin{array}{l} \max w \\ w \leq f(x^1) + \lambda g(x^1) \\ \dots \\ w \leq f(x^K) + \lambda g(x^K) \\ \lambda \geq 0 \end{array} \tag{2.9}$$

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.

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

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

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.

$$(PR_\lambda) \qquad w(\lambda) = \min \{(\lambda, 1)(r, r_0) = \lambda r + r_0, (r, r_0) \in G\} \tag{2.11}$$

This algorithm is nothing but the *cutting plane method of Kelley* [8] for convex programs. These cuts are called *lagrangean optimality cuts*.

Supposing we ignore unboundness cases, the *lagrangean relaxation algorithm* is summarized on the next steps:

1. Solve problem (MD) and obtain λ
2. Obtain upper bound $\overline{z} = w$
3. Solve problem (PR_λ) and obtain x^k
4. Obtain lower bound $\underline{z} = w(\lambda)$

5. Stop if $\bar{z} - \underline{z} < \text{tol}$, otherwise do $k = k + 1$ and go to 1.

1.1. Constrained perturbation region

When calculating the dual function in previous section the problem (2.5)

$$\begin{aligned} &\min\,(\lambda,1)(r,r_0) = \lambda r + r_0 \\ &r_0 \geq f(x) \\ &r \geq g(x) \\ &x \in X \end{aligned} \tag{2.12}$$

was transformed into the problem

$$\begin{aligned} &\min \lambda g(x) + f(x) \\ &x \in X \end{aligned} \tag{2.13}$$

due to the nonexistence of additional constraints over variables (r, r_0).

If the perturbation function is defined for a constrained set of right hand side values, then the convexification procedure has to take this constrained set into account, and previous transformation is no longer valid. Consider the perturbation function

$$\begin{aligned} \upsilon(y) = \quad &\min f(x) \\ &g(x) \leq y \\ &x \in X \end{aligned} \tag{2.14}$$

be defined for $y \in R$. We consider a constrained generalized graph and epigraph as follows.

$$G = \{(r, r_0) / \exists x \in X \text{ with } r = g(x), r_0 = f(x), r \in R\} \tag{2.15}$$

$$\operatorname{epi} G = \{(r, r_0) / \exists x \in X \text{ with } r \geq g(x), r_0 \geq f(x), r \in R\} \tag{2.16}$$

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

$$(PR_\lambda) \qquad \begin{aligned} w(\lambda) = \quad &\min f(x) + \lambda r \\ &r \geq g(x) \\ &x \in X, r \in R \end{aligned} \tag{2.17}$$

and the dual problem still remains as

$$(D) \qquad \max\{w(\lambda), \lambda \geq 0\} \tag{2.18}$$

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

$$(MD) \qquad \begin{aligned} &\max w \\ &w \leq f(x^i) + \lambda r^i \\ &\lambda \geq 0, i = 1, \ldots, k \end{aligned} \tag{2.19}$$

$$(x^i, r^i) \in \operatorname{extr}\{g(x) \leq r, x \in X, r \in R\}, i : 1, \ldots, k$$

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.

1.2. Phase I of lagrangean relaxation and bounding cuts

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. It is also necessary to obtain a bounding cut that excludes the proposed multiplier value in case problem (PR_λ) turns out to be unbounded. From hereafter it is assumed that the objective function is linear and the constraints are affine. We will exhaustively use the Farkas' law results.

1.1.1. Phase I of lagrangean relaxation

Let problem (P) take the form

$$(P)\qquad \begin{aligned}&\min cx\\ &Ax \le b\\ &x \in X\end{aligned} \tag{2.20}$$

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

$$\begin{aligned}&\min cx\\ &Ax - b \le r\\ &x \in X, r \in R\end{aligned} \tag{2.21}$$

It is clear that system $\{Ax - b \le r, x \in X, r \in R, r = 0\}$ has a solution if and only if system $\{\operatorname{conv}\{Ax - b \le r, x \in X, r \in R\}, r = 0\}$ 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

$$\begin{aligned}&\min r^+ + r^-\\ &r - r^+ + r^- = 0\\ &\operatorname{conv}\begin{Bmatrix}Ax - b \le r\\ x \in X, r \in R\end{Bmatrix}\\ &r^+, r^- \ge 0\end{aligned} \tag{2.22}$$

Feasibility region of the above problem immediately satisfies the *integrality property* [4] that guarantees its optimal value is equivalent to the value obtained by a LR algorithm. Then, for any $\lambda \ge 0$ consider

$$(PR_\lambda)_*\qquad \begin{aligned}w_*(\lambda) = \ &\min r^+ + r^- + \lambda(r - r^+ + r^-)\\ &Ax - b \le r\\ &x \in X, r \in R\\ &r^+, r \ge 0\end{aligned} \tag{2.23}$$

and solve the following dual problem

$$(D)_* \qquad \max\{w_*(\lambda), \lambda \geq 0\} \tag{2.24}$$

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

$$w_*(\lambda) = \begin{cases} -\infty & \lambda > 1 \\ \min \lambda r, (x,r) \in \{Ax - b \leq r, x \in X, r \in R\} & \lambda \leq 1 \end{cases} \tag{2.25}$$

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

$$(D)_* \qquad \max\{w_*(\lambda), 0 \leq \lambda \leq 1\} \tag{2.26}$$

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

$$(MD)_* \qquad \begin{aligned} &\max w \\ &w \leq \lambda r^i \\ &0 \leq \lambda \leq 1, i = 1,\ldots,k \end{aligned} \tag{2.27}$$

$$(x^i, r^i) \in \text{extr}\{Ax - b \leq r, x \in X, r \in R\}, i : 1,\ldots,k$$

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.

$$(MD_{\lambda_0}) \qquad \begin{aligned} &\max w \\ &w \leq \lambda_0 cx^i + \lambda r^i \\ &\lambda \geq 0, i = 1,\ldots,k \end{aligned} \tag{2.28}$$

$$(PR_{\lambda_0\lambda}) \qquad \begin{aligned} w_{\lambda_0}(\lambda) = \quad &\min \lambda_0 cx + \lambda r \\ &(x,r) \in \{Ax - b \leq r, x \in X, r \in R\} \end{aligned} \tag{2.29}$$

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

1.1.2. Bounding cuts

Consider problem (P) of the form

$$(P) \qquad \begin{aligned} &\min cx \\ &Ax \leq b \\ &x \in X \end{aligned} \tag{2.30}$$

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

When solving this problem via LR, in order to obtain the convexification of its associated perturbation function, a multiplier value is proposed through the resolution of a master dual problem (MD) and then we have to solve the problem (PR_λ).

$$(MD) \qquad \begin{aligned} &\max w \\ &w \le cx^i + \lambda r^i \\ &\lambda \ge 0, i = 1,\ldots,k \end{aligned} \tag{2.31}$$

$$(PR_\lambda) \qquad \begin{aligned} w(\lambda) = \quad &\min cx + \lambda r \\ &Ax - b \le r \\ &x \in X, r \in R \end{aligned} \tag{2.32}$$

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

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

$$(DPR_\lambda) \qquad \begin{aligned} &\max \pi_1 b + \pi_2 b_1 + \pi_3 r_1 \\ &\pi_1 A + \pi_2 A_{11} = c \\ &-\pi_1 + \pi_3 R_{11} = \lambda \\ &\pi_1, \pi_2, \pi_3 \le 0 \end{aligned} \tag{2.33}$$

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

$$\begin{aligned} c\tilde{x} + \lambda\tilde{r} \le 0, \forall (\tilde{x}, \tilde{r}) / \\ -A\tilde{x} + \tilde{r} \le 0 \\ -A_{11}\tilde{x} \le 0 \\ -R_{11}\tilde{r} \le 0 \end{aligned} \tag{2.34}$$

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

$$B = \begin{Bmatrix} \lambda / c\tilde{x}^j + \lambda\tilde{r}^j \le 0 \\ \forall(\tilde{x}^j, \tilde{r}^j) \text{ extreme ray } \{-A\tilde{x} + \tilde{r} \le 0, -A_{11}\tilde{x} \le 0, -R_{11}\tilde{r} \le 0\} \end{Bmatrix} \tag{2.35}$$

Then, the dual problem (D) takes the form

$$(D) \qquad \max\{w(\lambda), \lambda \ge 0, \lambda \in B\}$$

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 *bounding cut.* Before solving problem (PR_λ) next problem has to be solved

$$
\begin{aligned}
&\max c\tilde{x} + \lambda \tilde{r} \\
&-A\tilde{x} + \tilde{r} \le 0 \\
&-A_{11}\tilde{x} \le 0 \\
&-R_{11}\tilde{r} \le 0 \\
&-1 \le \tilde{x} \le 1, -1 \le \tilde{r} \le 1
\end{aligned} \tag{2.36}
$$

and in case this problem ends with a positive solution, a constraint of the form $c\tilde{x}^j + \lambda \tilde{r}^j \le 0$ is introduced into the master dual problem (MD). The point $(\tilde{x}^j, \tilde{r}^j)$ 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 (2.36) 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 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

$$
(MD) \qquad
\begin{aligned}
&\max w \\
&w \le cx^i + \lambda r^i \\
&0 \le c\tilde{x}^j + \lambda \tilde{r}^j \\
&\lambda \ge 0, i : 1, \ldots, k, j : 1, \ldots, l
\end{aligned} \tag{2.37}
$$

with

$(x^i, r^i) \in \operatorname{extr}\{Ax - b \le r, x \in X, r \in R\}, i : 1, \ldots, k$ and

$(\tilde{x}^j, \tilde{r}^j) \in$ extreme ray $\{-A\tilde{x} + \tilde{r} \le 0, -A_{11}\tilde{x} \le 0, -R_{11}\tilde{r} \le 0\}, j = 1, \ldots, l$

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.

1.1.3. Equality constraints

The perturbation function is no 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 mixed integer linear problems.

Consider problem (P)

$$
(P) \qquad
\begin{aligned}
&\min cx \\
&Ax \le b \\
&Dx = d \\
&x \in X
\end{aligned} \tag{2.38}
$$

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

$$
\begin{aligned}
v(r, r') = \quad &\min cx \\
&Ax - b \le r \\
&Dx - d = r' \\
&x \in X
\end{aligned} \tag{2.39}
$$

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

$$G = \left\{(r, r', r_0) / \exists x \in X \text{ with } r = Ax - b, r' = Dx - d, r_0 = cx, (r, r') \in R\right\} \tag{2.40}$$

$$\operatorname{epi} G = \left\{(r, r', r_0) / \exists x \in X \text{ with } r \geq Ax - b, r' = Dx - d, r_0 \geq cx, (r, r') \in R\right\} \tag{2.41}$$

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

$$(PR_{\lambda_0 \lambda \mu}) \quad \begin{aligned} w_{\lambda_0}(\lambda, \mu) = \quad & \min \lambda_0 cx + \lambda r + \mu(Dx - d) \\ & Ax - b \leq r \\ & x \in X, (r, Dx - d) \in R \end{aligned} \tag{2.42}$$

with $\lambda_0 = 0$ on algorithm phase 1 and $\lambda_0 = 1$ on algorithm phase 2.

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

$$(MD_{\lambda_0}) \quad \begin{aligned} & \max w \\ & w \leq \lambda_0 cx^i + \lambda r^i + \mu(Dx^i - d) \\ & 0 \leq \lambda_0 c\tilde{x}^j + \lambda \tilde{r}^j + \mu D\tilde{x}^j \\ & 0 \leq \lambda \leq 1 \end{aligned} \tag{2.43}$$

with $(x^i, r^i) \in \operatorname{extr}\left\{x \in X, Ax - b \leq r, (r, Dx - d) \in R\right\}$ and $(\tilde{x}^j, \tilde{r}^j)$ are extreme rays for the corresponding region.

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

1. Set $\lambda_0 = 0$
2. Solve problem (MD_{λ_0}) and obtain multiplier values λ and μ
3. Obtain upper bound $\overline{z}_{\lambda_0} = w$
4. If $\lambda_0 = 0$ and $w = 0$ switch to phase 2 set $\lambda_0 = 1$
5. Solve linear relaxation of problem $(PR_{\lambda_0 \lambda \mu})$
6. If linear relaxation is unbounded, then take dual values and form a bounding cut
   Set $l = l + 1$ and go to step 2
7. If linear relaxation is bounded, then continue solving mixed integer linear problem $(PR_{\lambda_0 \lambda \mu})$
   Obtain value (x^k, r^k) and set $k = k + 1$
   Obtain lower bound $\underline{z}_{\lambda_0} = w_{\lambda_0}(\lambda, \mu)$
8. If $\overline{z}_{\lambda_0} - \underline{z}_{\lambda_0} < \text{tol}$ then stop. Otherwise go to 2.
9. If $\lambda_0 = 0$ problem (P) is infeasible
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 that it ends at phase 1, then problem (P) is infeasible, and the final value gives the minimization of infeasibilities due to the complicating constraints $\{Ax \leq b, Dx = d\}$. Infeasibility of problem due to these constraints is identified if a lagrangean subproblem has positive optimal value during phase 1 algorithm. 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 on literature as *deepest cut*.

3. BENDERS DECOMPOSITION

We now face the issue of solving the problem (P) of the form

$$(P) \qquad \begin{aligned} & \min cx + dy \\ & A_1 x + A_2 y \le b \\ & x \in X, y \in Y \end{aligned} \tag{3.1}$$

where feasible regions for first and second stage variables, x and y respectively, incorporate integrality constraints for some variables $X = \{A_{11}x \le b_1, x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}\}$, $Y = \{A_{22}y \le b_2, y \in \mathbb{R}^{n_2} \times \mathbb{Z}^{m_2}\}$. 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

$$(SP_x) \qquad \begin{aligned} \theta(x) = \quad & \min dy \\ & A_2 y \le b - A_1 x \\ & y \in Y \end{aligned} \tag{3.3}$$

1.3. Linear problems

For linear 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) at this optimal value (modifying right hand sides 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 turns 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.

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

$$(SP_{x_0}) \qquad \theta(x_0) = \min\{dy, A_2 y \le b - A_1 x_0, y \in \overline{Y}\} \tag{3.4}$$

where $\overline{Y} = \{y \,/\, A_{22}y \le b_2, y \in \mathbb{R}^{n_2+m_2}\}$.

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

$$(SP_{x_0}) \qquad \begin{aligned} \theta(x_0) = \quad & \max \pi(b - A_1 x_0) + \rho b_2 \\ & \pi A_1 + \rho A_{22} = d \\ & \pi \le 0, \rho \le 0 \end{aligned} \tag{3.5}$$

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

$$\theta(x) \geq \pi^i(b - A_1 x) + \rho^i b_2 \tag{3.6}$$

Linearizing around the point of interest x_0 we have the expression

$$\begin{aligned} &\theta(x) \geq \pi^i(b - A_1 x) + \rho^i b_2 = \pi^i(b - A_1 x_0 + A_1 x_0 - A_1 x) + \rho^i b_2 = \\ &= \pi^i(b - A_1 x_0) + \rho^i b_2 + \pi^i(A_1 x_0 - A_1 x) = \theta^i + \pi^i A_1(x_0 - x) \end{aligned} \tag{3.7}$$

So that expression (3.6) is written as

$$\theta(x) \geq \theta^i + \pi^i A_1(x_0 - x) \tag{3.8}.$$

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

1.1.4. Feasibility cuts

Subproblem (SP_{x_0}) is infeasible if no solution exist for the region $\left\{A_2 y \leq b - A_1 x_0, A_{22} y \leq b_2\right\}$. This is equivalent to assert that there exists no $s_1 \geq 0$, $s_2 \geq 0$ such that

$$\left\{A_2 y + s_1 = b - A_1 x_0,\ A_{22} y + s_2 = b_2\right\} \tag{3.9}$$

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

$$\left\{\tilde{\pi}(b - A_1 x_0) + \tilde{\rho} b_2 \leq 0, \forall \tilde{\pi}, \tilde{\rho} \leq 0 \,/\, \tilde{\pi} A_2 + \tilde{\rho} A_{22} \leq 0\right\} \tag{3.10}$$

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

$$K = \left\{x_0 \,/\, \tilde{\pi}^j(b - A_1 x_0) + \tilde{\rho}^j b_2 \leq 0, \forall \tilde{\pi}^j, \tilde{\rho}^j \text{ extreme ray } \left\{\tilde{\pi} A_2 + \tilde{\rho} A_{22} \leq 0, \tilde{\pi} \leq 0, \tilde{\rho} \leq 0\right\}\right\} \tag{3.11}$$

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

$$\begin{aligned} \theta_*(x_0) = \quad & \max \tilde{\pi}(b - A_1 x_0) + \tilde{\rho} b_2 \\ & \tilde{\pi} A_2 + \tilde{\rho} A_{22} \leq 0 \\ & -1 \leq \tilde{\pi} \leq 0 \\ & -1 \leq \tilde{\rho} \leq 0 \end{aligned} \tag{3.12}$$

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

$$\tilde{\pi}^j(b - A_1 x) + \tilde{\rho}^j b_2 \leq 0 \tag{3.13}$$

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

$$\begin{aligned} &0 \geq \tilde{\pi}^j(b - A_1 x) + \tilde{\rho}^j b_2 = \\ &= \tilde{\pi}^j(b - A_1 x + A_1 x_0 - A_1 x_0) + \tilde{\rho}^j b_2 = \\ &= \tilde{\pi}^j(-A_1 x + A_1 x_0) + \theta^j = \tilde{\pi}^j A_1(x_0 - x) + \theta^j \end{aligned} \tag{3.14}$$

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

$$0 \geq \theta^j + \tilde{\pi}^j A_1(x_0 - x) \tag{3.15}$$

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

$$\begin{aligned} &\min s_1 + s_2 \\ &A_2 y - s_1 \leq b - A_1 x_0 \\ &A_{22} y - s_2 \leq b_2 \\ &s_1, s_2 \geq 0 \end{aligned} \tag{3.16}$$

that represents the minimization of infeasibilities of problem (SP).

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

$$(MP) \qquad \begin{aligned} &\min cx + \theta \\ &0 \geq \theta^j + \tilde{\pi}^j A_1(x_0^j - x) \quad j = 1, \ldots, l \\ &\theta \geq \theta^i + \pi^i A_1(x_0^i - x) \quad i = 1, \ldots, k \\ &x \in \overline{X}, \overline{X} = \left\{ A_{11} x \leq b_1, x \in \mathbb{R}^{n_1 + m_1} \right\} \end{aligned} \tag{3.17}$$

1.4. Mixed integer linear problems

The mixed integer case keeps the same procedure, but face the disadvantage of the no 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

$$\theta(r) = \min \left\{ dy, A_2 y - b \leq r, y \in Y, r \in R \right\} \tag{3.18}$$

with $R = \left\{ r \,/\, \exists x \in X \,/\, r = -A_1 x \right\}$

Following results of section 2, define for any $\lambda \geq 0$ the *dual function* by

$$(PR_\lambda) \qquad \begin{aligned} w(\lambda) = \quad &\min dy + \lambda r \\ &A_2 y - b \leq r \\ &y \in Y, r \in R \end{aligned} \tag{3.19}$$

Its solution determines a level curve of the form $L = \left\{ (r, r_0) \,/\, \lambda r + r_0 = w(\lambda) \right\}$. For $r = -A_1 x$ the resulting point is then $(-A_1 x, w(\lambda) + \lambda A_1 x)$. The dual problem consists of finding the maximum of those ordinates points

$$(D_x) \qquad \max \left\{ w(\lambda) + \lambda A_1 x, \lambda \geq 0 \right\} \tag{3.20}$$

So, in the mixed integer 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 $w(\lambda^i)+\lambda^i A_1 x$. The epigraph of the perturbation function then immediately satisfies

$$\theta \geq w(\lambda^i)+\lambda^i A_1 x, \quad \forall x \in X \tag{3.21}$$

Denoting $\theta^i = w(\lambda^i)+\lambda^i A_1 x$ the optimum value of problem (D_x) and linearizing around the first stage solution we have

$$\begin{aligned} &\theta \geq w(\lambda^i)+\lambda^i A_1 x = \\ &= w(\lambda^i)+\lambda^i A_1 x_0 - \lambda^i A_1 x_0 + \lambda^i A_1 x = \\ &= \theta^i + \lambda^i A_1(-x_0 + x), \quad \forall x \in X \end{aligned} \tag{3.22}$$

Summarizing

$$\theta \geq \theta^i - \lambda^i A_1(x_0 - x), \quad \forall x \in X \tag{3.23}$$

This expression recovers the *Benders optimality cut* introduced in the master problem for the linear 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 on phase 2 produces an optimal multiplier and an optimal value of the dual function, that are used to form a *Benders optimality cut*.

1.1.5. 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

$$(D_x)_* \qquad \max\left\{w_*(\lambda)+\lambda A_1 x, 0 \leq \lambda \leq 1\right\} \tag{3.24}$$

with

$$(PR_\lambda)_* \qquad \begin{aligned} w_*(\lambda) = \quad & \min \lambda r \\ & A_2 y - b \leq r \\ & y \in Y, r \in R \end{aligned} \tag{3.25}$$

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

$$0 \geq \theta^j - \lambda^j A_1(x_0 - x), \quad \forall x \in X \tag{3.26}$$

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

$$(MP) \qquad \begin{aligned} & \min cx + \theta \\ & 0 \geq \theta^j + \tilde{\pi}^j A_1(x_0^j - x) \quad\quad j = 1,\ldots,l \\ & \theta \geq \theta^i + \pi^i A_1(x_0^i - x) \quad\quad i = 1,\ldots,k \\ & x \in X, X = \left\{A_{11}x \leq b_1, x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{m_1}\right\} \end{aligned} \tag{3.27}$$

where values (θ^i, π^i), being $\pi^i = -\lambda^i$, represent the optimal values of LR procedure when this ends on 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. Values $(\theta^j, \tilde{\pi}^j)$, being $\tilde{\pi}^j = -\lambda^j$, represent the optimal values of the LR procedure when this ends on phase 1, with 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

$$(MD_{x\lambda_0}) \qquad \begin{aligned} &\max w + \lambda A_1 x \\ &w \le \lambda_0 dy^i + \lambda r^i \\ &0 \le \lambda_0 d\tilde{y}^j + \lambda \tilde{r}^j \\ &\lambda \ge 0, i:1,...,k, j:1,...,l \end{aligned} \tag{3.28}$$

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

$$(PR_{\lambda_0\lambda}) \qquad \begin{aligned} w_{\lambda_0}(\lambda) = \ &\min \lambda_0 dy + \lambda r \\ &A_2 y - b \le r \\ &y \in Y, r \in R^i \end{aligned} \tag{3.29}$$

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 to be 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.

1.5. Perturbation cuts

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

$$(P) \qquad \begin{aligned} &\min c_1 x_1 + c_2 x_2 + dy \\ &A_1 x_2 + A_2 y \le b \\ &(x_1, x_2) \in X, y \in Y \end{aligned} \tag{3.30}$$

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

$$R = \{r / \exists x_2 \in S / r = -A_1 x_2\} \tag{3.31}$$

So we are interested in finding $S = \{x_2 / \exists x_1 / (x_1, x_2) \in X\}$. 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.

1.1.6. 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

$$(MP)\qquad \begin{aligned} &\min c_1x_1 + c_2x_2 + \theta \\ &0 \ge \theta^j + \tilde{\pi}^j A_1(x_0^j - x_2) \quad j = 1,\dots,l \\ &\theta \ge \theta^i + \pi^i A_1(x_0^i - x_2) \quad i = 1,\dots,k \\ &x \in \bar{X}, \bar{X} = \left\{A_{11}x \le b_1, x \in \mathbb{R}^{n_1+m_1}\right\} \end{aligned} \qquad (3.32)$$

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 $(d_1, d_2, ..., d_{n_1+m_1}, d_{n_1+m_1+1})$ be the edges at that extreme point. Let $(\widehat{d}_1, \widehat{d}_2, ..., \widehat{d}_{n_1+m_1}, \widehat{d}_{n_1+m_1+1})$ 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.

1.1.7. 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 $\left\{X_n, n = 1,...,N\right\}$ this partition. Let C_n^i the cone obtained at each optimum point of the each partition. Then the convex sum of these cones constrains region S. Observe that in case the partition $\left\{X_n, n = 1,...,N\right\}$ 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.

1.6. 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 this point's projection is as extreme point of S. In that case then 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 it is enough to check the last constraint introduced to 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 lagrangean relaxation phase 2. Let λ^i and $w(\lambda^i) + \lambda^i A_1 x^p$ the optimal solution. Calculate upper bound $\overline{z} = cx^p + w(\lambda^i) + \lambda^i A_1 x^p$. 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.

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

Let $\theta_{R^p}(x)$ the convexification of $\theta(x)$ when the perturbation region is R^p. $\theta(x) \geq \theta_{R^p}(x)$, $\forall x \in X$.

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

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 have these cases.

Case a. If $\theta_{R^p}(x^p) > \theta^p$, then a new cut is generated. The algorithm proceeds formulating an augmented 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.

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

Assume $\theta_{R^p}(x^p) = \theta^p$ and assume that x^p is not an optimal solution for problem $\min\left\{cx + \theta(x), x \in X\right\}$. Then

$$cx + \theta(x) \geq m > cx^p + \theta^p, \forall x \in X \tag{3.35}$$

differentiating the following two cases.

Case 1. $P(x^p) \in \mathrm{int}(S)$. Then there is a neighborhood U ($P(x^p) \in U \subset S$) such that $cx + \theta(x) \geq m$, $\forall x \in U$. The function $\theta_{R^p}(x)$ then the immediately satisfies

$$
\begin{aligned}
cx + \theta_{R^p}(x) &\geq cx + \hat{\theta}^p(x) \qquad && \forall x \in X \\
cx + \theta_{R^p}(x) &\geq m && \forall x \,/\, P(x) \in U
\end{aligned}
\tag{3.36}
$$

then

$$
cx + \theta_{R^p}(x) \geq \max(cx + \hat{\theta}^p(x), m) \quad \forall x \in X \tag{3.37}
$$

In particular, for $x = x^p$ observe that $cx^p + \theta_{R^p}(x^p) \geq m > cx^p + \theta^p$ and then $\theta_{R^p}(x^p) > \theta^p$ which is a contradiction with the initial assumptions.

Case 2. $P(x^p) \in \mathrm{fr}(S)$. In that case the algorithm determines a cone C^p that constrains the perturbation region R. The algorithm construction then implies that

$$
\begin{aligned}
\theta_{R^p}(x^p + \lambda d) &= \infty \qquad && d \notin C^p \\
\theta_{R^p}(x^p + \lambda d) &\geq \hat{\theta}(x^p + \lambda d) && d \in C^p
\end{aligned}
\tag{3.38}
$$

and that the function $\theta_{R^p}(x)$ satisfies

$$
\begin{aligned}
c(x^p + \lambda d) + \theta_{R^p}(x^p + \lambda d) &= \infty && d \notin C^p \\
c(x^p + \lambda d) + \theta_{R^p}(x^p + \lambda d) &\geq \max(c(x^p + \lambda d) + \hat{\theta}(x^p + \lambda d), m) \quad && d \in C^p
\end{aligned}
\tag{3.39}
$$

In particular, for $x = x^p$ observe that $cx^p + \theta_{R^p}(x^p) \geq m > cx^p + \theta^p$ and then the contradiction appears.

Observe that the argument in case 2 cannot be carried out in the case of simple use of lagrangean relaxation (i.e., without the introduction of the perturbation region). The convexified function obtained is not forced to satisfy the last constraints.

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 only is 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, then the proposed algorithm is also finite.

4. EXAMPLE

Consider the problem

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

The optimal solution of this problem is -6.71 achieved at $x = 0.7$, $y = 3$, $z = 2$.
Solving this program by Benders decomposition we formulate this master problem (MP)

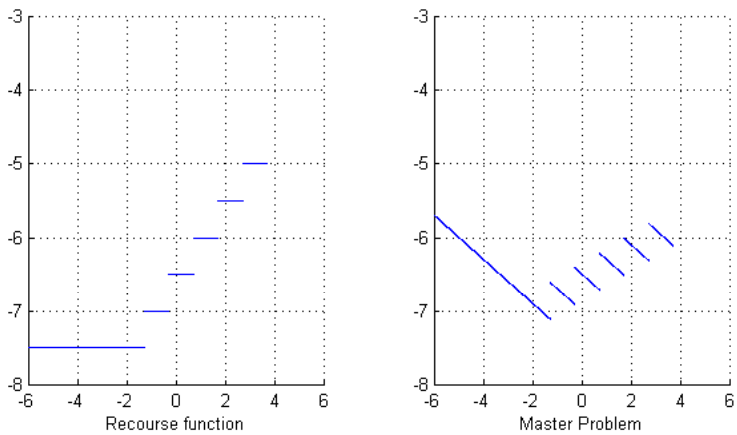
$$\begin{aligned}&\min -0.3x+\theta(x)\\&0\le x\le 5\end{aligned}$$

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

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



First it is solved relaxing the subproblem integrality conditions. Then, the subproblem is solved directly by LR, and finally its is solved with the above described approach. For simplicity of the exposition, in the LR we avoid obtaining the dual function interactively. The dual function is analytically calculated and optimized.

1.8. Relaxation of subproblem integrality conditions

Iteration 1: solve master problem

$$\min -0.3x$$
$$0 \le x \le 5$$

with solution $x = 5$.

Solve subproblem for $x = 5$

$$\begin{aligned} \theta(5) = \min -1.5y - z \\ y \le 3.7 - 5 \\ y + z \le 5.2 \\ y \ge 0, z \ge 0 \end{aligned}$$

obtaining a *feasibility cut* $x \le 3.7$

Iteration 2: solve augmented master problem

$$\begin{aligned} &\min -0.3x \\ &0 \le x \le 5 \\ &x \le 3.7 \end{aligned}$$

with solution $x = 3.7$

Solve subproblem for $x = 3.7$

$$\begin{aligned} \theta(3.7) = \min -1.5y - z \quad &\text{whose dual take the form of } \theta(3.7) = \max 5.2\pi_2 \\ y \le 3.7 - 3.7 \quad & \quad \pi_1 + \pi_2 \le -1.5 \\ y + z \le 5.2 \quad & \quad \pi_2 \le -1 \\ y \ge 0, z \ge 0 \quad & \quad \pi_1, \pi_2 \le 0 \end{aligned}$$

with solution $\pi_1 = -0.5$, $\pi_2 = -1$, $\theta(3.7) = -5.2$ that generates the following *Benders optimality cut*

$$\theta \ge \theta(3.7) + \pi_1(3.7 - x)$$

$$\theta \ge 0.5x - 7.05$$

Iteration 3: solve augmented master problem

$$\begin{aligned} &\min -0.3x + \theta \\ &0 \le x \le 5 \\ &x \le 3.7 \\ &\theta \ge 0.5x - 7.05 \end{aligned}$$

with solution $x = 0$, $\theta = -7.05$

Solve subproblem for $x = 0$

$$\theta(0) = \min -1.5y - z \quad \text{whose dual take the form of} \quad \theta(0) = \max 5.2\pi_2$$

$$\begin{array}{ll} y \le 3.7 - 0 & \pi_1 + \pi_2 \le -1.5 \\ y + z \le 5.2 & \pi_2 \le -1 \\ y \ge 0, z \ge 0 & \pi_1, \pi_2 \le 0 \end{array}$$

with solution $\pi_1 = -0.5$, $\pi_2 = -1$, $\theta(0) = -7.05$. The stopping rule is verified, so the optimal solution is –7.05.

1.9. Subproblem resolution by lagrangean relaxation.

Iteration 1: solving the master problem determines the optimal solution $x = 5$.

Solve subproblem for $x = 5$. We firstly solve the minimization of infeasibilities problem

$$\begin{aligned} \theta_*(5) = \min\, & s^+ + s^- \\ & y + s^+ - s^- \le 3.7 - 5 \\ & y + z \le 5.2 \\ & y \ge 0, z \ge 0 \\ & y \in \mathbb{Z}, z \in \mathbb{Z} \end{aligned}$$

by LR formulating the lagrangean

$$L(y, z, \lambda) = s^+ + s^- + \lambda(y + s^+ - s^- - 3.7 + 5)$$

or equivalently

$$L(y, z, \lambda) = (1 + \lambda)s^+ + (1 - \lambda)s^- + \lambda(y - 3.7 + 5)$$

If $\lambda > 1$ then the dual function $w_*(\lambda) = -\infty$, so the dual function for $0 \le \lambda \le 1$ takes the form

$$\begin{aligned} w_*(\lambda) = \min\, & \lambda(y - 3.7 + 5) \\ & y + z \le 5.2 \\ & y \ge 0, z \ge 0 \\ & y \in \mathbb{Z}, z \in \mathbb{Z} \end{aligned}$$

that has solution $y = 0$ for every $\lambda \in [0,1]$. Then , $0 \le \lambda \le 1$.

Dual problem $\max\{w_*(\lambda), 0 \le \lambda \le 1\}$ ends with solution $\lambda = 1$, $w_*(1) = 1.3 > 0$, so a *feasibility cut* is generated determined as

$$w_*(1) + \tilde{\pi}(5 - x) \le 0$$

with $\tilde{\pi} = -\lambda = -1$

$$\begin{aligned} & 1.3 + (x - 5) \le 0 \\ & x \le 3.7 \end{aligned}$$

It is the same feasibility cut obtained in case of relaxing subproblem integrality conditions.

Iteration 2: the master problem is solved ending with optimal solution $x = 3.7$ and solve the subproblem for this value.

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

The minimization of infeasibilities problem leads to the resolution of the first stage dual function

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

with $w_*(\lambda) \equiv 0$. The subproblem is then feasible and a *bounding cut* is obtained for the second stage $w \le 0$. The subproblem is now solved via LR

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

obtaining

$$
w(\lambda) = \begin{cases} -7.5 + 5\lambda & 0 \le \lambda \le 0.5 \\ -5 & \lambda \ge 0.5 \end{cases}
$$

Dual problem $\max\{w(\lambda), \lambda \ge 0\}$ presents multiple solutions due to the no differentiability of the recourse function at the point $x = 3.7$. Take as solution $\lambda = 0.5$, $w(0.5) = -5$. We formulate then an *optimality cut* with $\pi = -\lambda = -0.5$ and $w(0.5) = -5$.

$$
\theta \ge w(0.5) + \pi(3.7 - x)
$$

This is

$$
\begin{aligned}
& \theta - 0.5x \ge -5 - 0.5 \cdot 3.7 \\
& \theta \ge 0.5x - 6.85
\end{aligned}
$$

Iteration 3: The master problem is now

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

with solution $x = 0$, $\theta = -6.85$.

Solve subproblem for $x = 0$

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

We skip if $x = 0$ is feasible. The dual function at phase 2 takes the form

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

or

$$w(\lambda) = \begin{cases} -7.5 + 1.3\lambda & 0 \le \lambda \le 0.5 \\ -5 - 3.7\lambda & \lambda \ge 0.5 \end{cases}$$

and the $\max\{\omega(\lambda), \lambda \ge 0\}$ is achieved at $\lambda = 0.5$, $w(0.5) = -6.85$. This solution ends the algorithm because the stopping rule is satisfied.

1.10. Subproblem resolution by the proposed method

Iteration 1: solving the master problem determines the optimal solution $x = 5$

Solve subproblem for $x = 5$. We have the following constraint $x \le 5$ for the projection S. So that we have the following constraint for the perturbation region $r \ge -5$. Phase 1 has to solve the problem

$$\begin{aligned} w_*(\lambda) = \min & \lambda r \\ & y - 3.7 \le r \\ & y + z \le 5.2 \\ & y \ge 0, z \ge 0 \\ & r \ge -5 \\ & y \in \mathbb{Z}, z \in \mathbb{Z} \end{aligned}$$

The dual problem is stated then as $\max\{w_*(\lambda) + \lambda A_1 x, 0 \le \lambda \le 1\} = \max\{w_*(\lambda) + 5\lambda, 0 \le \lambda \le 1\} = \max\{1.3\lambda, 0 \le \lambda \le 1\} = 1.3$ at $\lambda = 1$.

A *feasibility cut* is obtained with expression $x \le 3.7$. Observe this is precisely the cut obtained when solving subproblem relaxing integrality constraints. This feasibility cut induces a cut for the perturbation function that is given as $r \ge -3.7$.

Iteration 2: once solved the augmented master problem we obtain $x = 3.7$ as a solution.

Solve subproblem for $x = 3.7$. Phase 1 of the subproblem is now $\max\{w_*(\lambda) + \lambda A_1 x, 0 \le \lambda \le 1\} = \max\{w_*(\lambda) + 3.7\lambda, 0 \le \lambda \le 1\} = \max\{0, 0 \le \lambda \le 1\} = 0$. Then the subproblem is feasible. This solution is obtained for $y = 0$, $r = -3.7$, $z \in [0,5] \cap \mathbb{Z}$. Multiplicity of solutions is due to the no differentiability of the recourse function at $x = 3.7$. This solution generates a cut for future phase 2 of LR, preventing the dual function to be unbounded. This cut is given as $w \le 0$. Now we face the solution of

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

Check that

$$
w(\lambda) = \begin{cases} -7.5 + 1.3\lambda & 0 \le \lambda \le 0.5 \\ -5 - 3.7\lambda & \lambda \ge 0.5 \end{cases}
$$

so that we have

$$
w(\lambda) + 3.7\lambda = \begin{cases} -7.5 + 5\lambda & 0 \le \lambda \le 0.5 \\ -5 & \lambda \ge 0.5 \end{cases}
$$

Dual problem now solves the problem $\max\{w(\lambda) + \lambda A_1 x, \lambda \ge 0\} = \max\{w(\lambda) + 3.7\lambda, \lambda \ge 0\} = -5$. The optimal solution is obtained for $\lambda \ge 0.5$. Again multiple solutions appear because of the no differentiability of the recourse function. Take $\lambda = 0.5$ and create the Benders optimality cut $\theta \ge 0.5x - 6.85$. Observe this is the same cut obtained when solving the subproblem directly by LR.

Iteration 3: the resolution of the augmented master problem ends with $x = 0$, $\theta = -6.85$. The region S is now updated with $x \ge 0$, and then the perturbation region with the constraint $r \le 0$.

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

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

or

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

The dual problem now $\max\{w(\lambda) + \lambda A_1 x, \lambda \ge 0\} = \max\{w(\lambda), \lambda \ge 0\} = -6.5$ for $\lambda = 0$.

This solution generates the Benders optimality cut $\theta \ge -6.5$.

Iteration 4: now solve the augmented Master Problem

$$\min -0.3x+\theta$$
$$0 \le x \le 5$$
$$x \le 3.7$$
$$\theta \ge 0.5x - 6.85$$
$$\theta \ge -6.5$$

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

Now solve 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 . Dual problem has to maximize the function

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

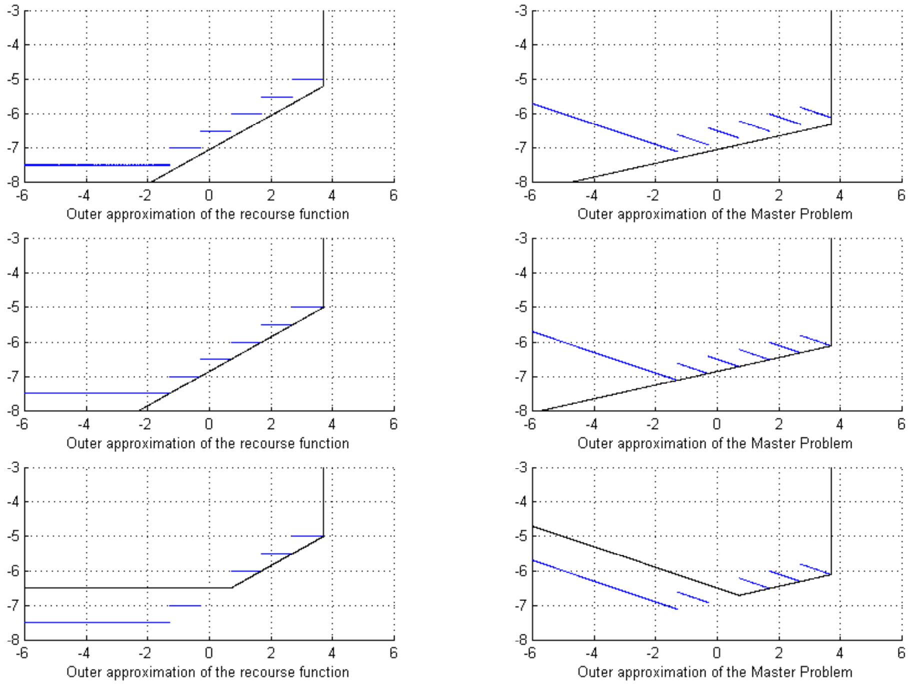
The dual problem now $\max\{w(\lambda)+\lambda A_1 x, \lambda \ge 0\} = \max\{w(\lambda)+0.7\lambda, \lambda \ge 0\} = -6.5$ that gives the optimal solution -6.5 , with multiplicity of solutions due to the no differentiability of the recourse function at point $x = 0.7$. Observe that the stopping rule is now satisfied. For $\lambda = 0$ the optimum value of the dual function is achieved at $y = 3$, $z = 2$, $r = -0.7$.

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

1.11. Summary of results

We summarize the example results on the following table and figure fore the three algorithms.

Optimal solution				Duality gap
A. Relaxation of subproblem integrality conditions	$z = -7.05$	$x = 0$	$\theta = -7.05$	0.34
B. Subproblem resolution by LR	$z = -6.85$	$x = 0$	$\theta = -6.85$	0.14
C. Subproblem resolution by the proposed method	$z = -6.71$	$x = 0.7$	$\theta = -6.5$	0.00



5. Parametric branch and bound

When applying LR to solve a subproblem, or the approach presented in which a partial description of the perturbation region is introduced, once a multiplier is proposed from the relaxed dual problem, a MIP problem must be solved. This resolution of the dual problem makes the calculation of Benders cut very cumbersome. A natural solution then consists of continuously updating the Lagrange multiplier as the branching step of the branch and bound (B&B) procedure goes on. These ideas are presented on this section with final commentaries about its adequacy within the Benders decomposition framework.

Let Y_1, Y_2 be two closed sets and assume $Y_1 \subset Y_2$. Suppose we want to solve

$$\begin{array}{lcl} \min z_1 = cy & & \min z_2 = cy \\ Ay \le b & \text{and} & Ay \le b \\ y \in Y_1 & & y \in Y_2 \end{array} \tag{5.1}$$

Then it is clear that $z_1 \ge z_2$. If we formulate the lagrangean dual function for each problem then $w_2(\lambda) \le w_1(\lambda)$. This is the case with a B&B procedure, in which a decreasing sequence of close sets Y_n, $n = 1,...,N$ is obtained, such that

$$Y = \bigcap_{n=1}^{N} Y_n, \quad Y \subset Y_n \subset \overline{Y}, \quad n = 1,...,N \tag{5.2}$$

where $\overline{Y}$ represents the region in which integrality constraints are relaxed. Thus, when considering the dual function for each of the previous regions we have that

$$\overline{w}(\lambda) \le w_n(\lambda) \le w(\lambda) \quad n = 1,...,N \tag{5.3}$$

and $w_n(\lambda) \to w(\lambda)$ as the B&B algorithm proceeds. Moreover, the previous sequence is a finite sequence, so the last approximate dual function is the function we are looking for.

This observation enables us to formulate a B&B algorithm that includes the updating of the relaxed constraint multiplier by optimizing an approximation of the dual function. This approximation is continuously updated as new nodes are explored, so that new cuts are added and previous cuts are deleted from the branching node. It should be taken into account that the approximation built is neither a lower nor an upper approximation of the final dual function. This is because at each iteration of the B&B algorithm we do not have all extreme points of the set Y_n, but just a few, corresponding to the number of different multiplier evaluations at that node.

Another interesting situation is the possibility of having a family of cuts of the (MD) problem for a determined iteration n of the B&B algorithm that guarantees that the approximated master dual problem is bounded, and finding that branching at a node (deleting cuts obtained at the branching node) turns the next approximation of the dual problem unbounded. This is actually a possibility because there can be infeasible MIP problems whose linear relaxation is feasible.

The parametric B&B is presented to obtain the convexification of the perturbation function at the point of interest. However, this technique may independently be used to solve a problem via LR. When facing the resolution of the subproblem a partial description of the perturbation region is at hand. In case this partial description is not available, the following algorithm solves the LR problem through the parametric B&B. At further resolutions of the subproblem, information about the branching tree should be used to accelerate the convergence towards the optimum.

1.12. Outline of the method

Consider the following subproblem $(PR_{\lambda_0\lambda})$ and relaxed master dual problem $(MD_{x\lambda_0})$, which will take part in the development of the parametric B&B algorithm

$$(PR_{Y\lambda_0\lambda}) \qquad \begin{aligned} w_{\lambda_0}(\lambda) = \quad & \min \lambda_0 dy + \lambda r \\ & A_2 y - b \le r \\ & y \in Y, r \in R^i \end{aligned} \tag{5.4}$$

$$(MD_{x\lambda_0}) \qquad \begin{aligned} & \max w + \lambda A_1 x \\ & w \le \lambda_0 dy^i + \lambda r^i \\ & 0 \le \lambda_0 d\tilde{y}^j + \lambda \tilde{r}^j \\ & \lambda \ge 0, i = 1,\ldots,k, j = 1,\ldots,l \end{aligned} \tag{5.5}$$

where Y is the branching region available in an iteration of the B&B algorithm.

Step 1. (Solve root node). Set $\lambda_0 = 0$. Select $\lambda \in [0,1]$. Solve subproblem $(PR_{\bar{Y}\lambda_0\lambda})$.

If problem $(PR_{\bar{Y}\lambda_0\lambda})$ is unbounded, then obtain a bounding cut and solve $(MD_{\lambda_0 x})$. Go to step 1. Continue in other case.

If $w_{\lambda_0}(\lambda) > 0$ then problem (P) is infeasible. The process stops or continues in order to find the sum of infeasibilities required at each primal optimization algorithm. In that case evaluations with $\lambda_0 = 1$ are avoided hereafter.

The problem solved represents the root node n_1 of the B&B algorithm and determines a first (y^1, r^1) point that approximates the dual function through the lagrangean. Identify if (y^1, r^1) satisfy integrality conditions or not. Evaluate the bounding functions $L_{\lambda_0}(n_1) = (L_0(n_1), L_1(n_1))$ where $L_0 = \lambda r^1$ for $\lambda_0 = 0$ and $L_1 = dy^1 + \lambda r^1$ for $\lambda_0 = 1$.

At this point we assume we have explored a set N of tree nodes (evaluated once or more times for different multiplier values) and that we have an approximation of the dual function where we can identify those cuts obtained with integer feasible points. The region that branching defines at this moment is denoted Y_N. We also assume we have a set T of terminal nodes $T \subset N$ and that all approximating cuts of the dual function approximation come from these terminal nodes, except those cuts obtained from previous integer feasible solutions encountered. For each terminal node $t \in T$ we have a lower bounding function $L_{\lambda_0}(t) = (L_0(t), L_1(t))$ that represents the objective functions evaluated for the latest multiplier proposed. For each region N we have the upper bounding function $U_{\lambda_0}(N)$ of the relaxed dual problem for region Y_N. We also have a lower bounding function $L_{\lambda_0}(N)$ defined as $L_{\lambda_0}(N) = \min\left\{L_{\lambda_0}(t), t \text{ terminal nodes of } N\right\}$.

Step 2. (Solve dual problem and obtain new multiplier). Set $\lambda_0 = 0$. Solve problem (MD_{x,λ_0}). Let $U_{\lambda_0}(N) = w + \lambda A_1 x$. If $w > 0$ then the dual problem is unbounded, so that we do not have enough cuts to assure boundness for the dual problem. The primal problem (P) could be infeasible. If $w = 0$ and all active constraints are cuts coming from integer solutions of the subproblem, then problem (P) is feasible and we will never evaluate dual problem or subproblem for $\lambda_0 = 0$. If $w = 0$ do $\lambda_0 = 1$ and solve problem (MD_{x,λ_0}). Let $U_{\lambda_0}(N) = w + \lambda A_1 x$. At this point there appear different strategies for the parametric B&B algorithm. These strategies come from the mixture of LR and classical branching strategies of the B&B algorithm.

Step 3. (Reevaluate terminal nodes for the new multiplier λ proposed). We assume this value is different from the latest multiplier value used to evaluate the terminal nodes so far. Then we obtain new Y_N extreme points that are appended (on the form of *lagrangean optimality cuts*) to the dual function approximation. We could also obtain new values to form *bounding cuts*. Update $L_{\lambda_0}(t)$, $t \in T$ for the proper λ_0 value. If all nodes remain bounded and no new Y_N extreme points are found, then necessarily $U_{\lambda_0}(N) = L_{\lambda_0}(N)$ and go to step 4. In other case go to step 2 and obtain a new multiplier or go to step 4.
This step could be skipped in the algorithm or just passed through it on certain predefined iterations. Multiple possible ways of reevaluating the terminal nodes are possible but we are not trying to face that problem. If the algorithm does not give the possibility of branching at this moment (all terminal nodes are integer feasible), then this step is unavoidable. Iterations between step 3 and step 2 represent the classical LR iterations when the subproblem feasible region has been frozen to Y_N.

Step 4. (Selection of a branching node). For the outline of the algorithm we assume we use a best-bound strategy. This method chooses the terminal node with the lowest bounding function, expecting the solution to be among the descendent nodes of that terminal node. So, select the terminal node as the one with the lowest bounding function $L_{\lambda_0}(N)$ that depends on the λ_0 value at this moment. The branching procedure will be done based on the last optimal solution encountered at that node. In the traditional B&B, if that solution is integer feasible, then the algorithm stops with that solution as the optimal one. In this parametric B&B version, the algorithm does not stop, without having been proved before that the current multiplier λ is the optimal multiplier. If the last solution is integer feasible, then go to step 2. If the last solution is integer feasible and $U_{\lambda_0}(N) = L_{\lambda_0}(N)$, then the parametric B&B algorithm stops.

Step 5. (Branch and solve). Select a branching variable and create a set of subsequent subproblems adding constraints that represent new bounds for certain variables. That branching defines a region $Y_{N+1} \subset Y_N$. Solve the subproblems $(PR_{Y_{N+1}\lambda_0\lambda})$ associated with those new descendent subproblems of the branching node. Delete *lagrangean optimality cuts* at the master dual problem that where formed with non-integer points obtained at the branching node[2]. Augment problem $(MD_{x\lambda_0})$ with those cuts that define the optimal solutions of the descendent subproblems (*lagrangean optimality or bounding cuts*). For each new terminal node n (the new subproblems) update its lower bounding function $L_{\lambda_0}(n)$. In case a subproblem is infeasible then prune it and consider $L_{\lambda_0}(n) = \infty$. That node will never be evaluated again. Go to step 2.

This step is subject to all branching strategies developed for B&B algorithm. The reason of going to step 2 at this moment lays on the expected low computational cost of solving problem $(MD_{x\lambda_0})$, which is an augmented problem of the previous $(MD_{x\lambda_0})$ solved, and is suitable to be solved with a dual simplex algorithm.

1.13. Convergence proof

Lemma 1. Consider situation presented at step 3. If the reevaluation of terminal nodes does not produce new Y_N extreme points, then $U_{\lambda_0}(N) = L_{\lambda_0}(N)$ and the parametric B&B algorithm must go to step 4.

Proof. Consider we have the Y_N region at this point of the algorithm. We have then that $Y_N = \bigcup_{t\in T} Y_{Nt}$, where Y_{Nt} represents each one of the convex regions determined at each terminal node of the B&B algorithm. We suppose we have a specific λ_0 value so we can avoid carrying it as a subindex all through the proof.

Let $w_N(\lambda)$ the dual function obtained considering just the Y_N region. We have an outer approximation of this function $\tilde{w}_N(\lambda)$ so that

$$w_N(\lambda) \le \tilde{w}_N(\lambda) \tag{5.6}$$

Then $U_{\lambda_0}(N) = \max\{\tilde{w}_N(\lambda) + \lambda A_1 x, \lambda \ge 0\} = \tilde{w}_N(\hat{\lambda}) + \hat{\lambda} A_1 x$.

[2] When finding an unbounded problem at a node of the branching tree a bounding cut is introduced into the master dual problem (MD). That unbounded problem implies problem at the root node is unbounded. That bounding cut eliminates the multiplier value that turns the problem unbounded, and so it is not necessary to be deleted from problem (MD) on future branching of that node.

It is stated that if reevaluation of terminal nodes remains with all nodes bounded, and does not produce new Y_N extreme points, then $U_{\lambda_0}(N) = w_N(\hat{\lambda}) + \hat{\lambda} A_1 x$. By the contrary, we have that

Case 1.

$$\begin{aligned} U_{\lambda_0}(N) > w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = \min\left\{\lambda_0 dy + \hat{\lambda}(r + A_1 x) / Ay - b \le r, y \in Y_N, r \in R^i\right\} = \\ = \lambda_0 dy^i + \hat{\lambda}(r^i + A_1 x) / (y^i, r^i) \in \left\{Ay - b \le r, y \in Y_N, r \in R^i\right\} \end{aligned} \tag{5.7}$$

So that the point (y^i, r^i) does not belong to the set of points defining the approximation $\tilde{w}_N(\hat{\lambda})$, opposed to the initial assumption.

Case 2.

$$U_{\lambda_0}(N) > w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = \min\left\{\lambda_0 dy + \hat{\lambda}(r + A_1 x) / Ay - b \le r, y \in Y_N, r \in R^i\right\} = -\infty$$

So that a region Y_{Nt} is unbounded. This contradicts initial assumptions.

Then we have that

$$\begin{aligned} U_{\lambda_0}(N) = \max\left\{\tilde{w}_N(\lambda) + \lambda A_1 x, \lambda \ge 0\right\} \ge \\ \ge \max\left\{w_N(\lambda) + \lambda A_1 x, \lambda \ge 0\right\} \ge w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = U_{\lambda_0}(N) \end{aligned} \tag{5.8}$$

So that $\hat{\lambda}$ maximizes $w_N(\lambda) + \lambda A_1 x$ and

$$U_{\lambda_0}(N) = w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = L_{\lambda_0}(N) \tag{5.9}$$

We can even relax the proposition and obtain the following immediately corollary.

Corollary. If reevaluation of terminal nodes ends with a lowest terminal node such that its extreme point is not new, then $U_{\lambda_0}(N) = L_{\lambda_0}(N)$ and the parametric B&B algorithm must go to step 4.

Remark. Observe that the corollary implicitly assumed that all associated problems remain bounded.

Lemma 2. Consider the situation at step 4. If the lowest terminal node turns out to have integer feasible solution and $U_{\lambda_0}(N) = L_{\lambda_0}(N)$ then the actual multiplier is the optimal multiplier for the LR procedure at region Y_N. Even more, the actual multiplier is the optimal multiplier for the LR procedure at region Y. Then, the parametric B&B algorithm stops.

Proof of the final remark. We always have that $w(\lambda) \ge w_N(\lambda)$

If $\hat{\lambda}$ is optimal multiplier $U_{\lambda_0}(N) = L_{\lambda_0}(N)$, then $\hat{\lambda} = \arg\max\left\{w_N(\lambda) + \lambda A_1 x, \lambda \ge 0\right\}$. Then

$\max\left\{w(\lambda) + \lambda A_1 x, \lambda \ge 0\right\} \ge \max\left\{w_N(\lambda) + \lambda A_1 x, \lambda \ge 0\right\} = w_N(\hat{\lambda}) + \hat{\lambda} A_1 x$

Now $w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = \min\left\{\lambda_0 dy + \hat{\lambda}(r + A_1 x), Ay - b \le r, y \in Y_N, r \in R^i\right\} = \lambda_0 dy^i + \hat{\lambda}(r^i + A_1 x)$ with $(y^i, r^i) \in \left\{Ay - b \le r, y \in Y, r \in R^i\right\}$, because the terminal node satisfies integrality constraints.

So we assure then that

$w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = \min\left\{\lambda_0 cy + \hat{\lambda}(r + A_1 x), Ay - b \le r, r \in R^i, y \in Y_N\right\} =$

$= \min\left\{\lambda_0 cy + \hat{\lambda}(r + A_1 x), Ay - b \le r, r \in R^i, y \in Y\right\} = w(\hat{\lambda}) + \hat{\lambda} A_1 x$

and finally $\max\left\{w(\lambda) + \hat{\lambda} A_1 x, \lambda \ge 0\right\} \ge \max\left\{w_N(\lambda) + \hat{\lambda} A_1 x, \lambda \ge 0\right\} = w_N(\hat{\lambda}) + \hat{\lambda} A_1 x = w(\hat{\lambda}) + \hat{\lambda} A_1 x$

The optimal multiplier at region Y_N is the optimal multiplier that the algorithm seeks.

1.14. Parametric B&B within the Benders algorithm

A few features ought to be commented when using the parametric B&B to solve the subproblem repetitively in a Benders algorithm.

Remark 1. Observe that dual function $w(\lambda)$ is independent of the first stage value proposed at the master problem (MP). So outer approximations obtained at a certain iteration of the Benders algorithm are valid on further subproblem resolutions, in case no new *perturbation cuts* are introduced.

Remark 2. Assume a subproblem resolution has ended with an approximation of the dual problem given as

$$(MD_{\lambda_0 x}) \qquad \begin{aligned} &\max w + \lambda A_1 x \\ &w \le \lambda_0 d y^i + \lambda r^i \\ &0 \le \lambda_0 d\tilde{y}^j + \lambda \tilde{r}^j \\ &\lambda \ge 0, i = 1,\ldots,k, j = 1,\ldots,l \end{aligned} \tag{5.10}$$

with $(y^i, r^i) \in \{Ay - b \le r, y \in Y, r \in R^i\}$, $i = 1,\ldots,k$.

Assume that the new resolution of the master problem ends with (x^p, θ^p) as a solution, with the projection of x^p defining an extreme point for the shadow S. So the approximated perturbation region R^i is constrained with new cuts defining a new perturbation region R^{i+1}. The dual problem is now redefined deleting those cuts such that $r^i \notin R^{i+1}$. These new constraints turn infeasible a set of terminal nodes of the branching tree. These nodes will never be evaluated again.

Remark 3. (Premature end of the parametric B&B algorithm).

Assume that a new resolution of the master problem (MP) ends with (x^p, θ^p) as a solution. If x^p is not an optimal solution, then the resolution of the subproblem ends with value $\theta_{R^i}(x^p)$, and $\theta_{R^i}(x^p) > \theta^p$. Let $(U_{\lambda_0}(N), L_{\lambda_0}(N))$ the upper and lower bounds that the parametric B&B algorithm is obtaining as the branching proceeds. Let λ be a proposed multiplier. Solve subproblem

$$\begin{aligned} w_{\lambda_0,N}(\lambda) = \quad &\min \lambda_0 dy + \lambda r \\ &A_2 y - b \le r \\ &y \in Y_N, r \in R^i \end{aligned} \tag{5.11}$$

and let (y^i, r^i) its optimal point. In this case the epigraph of the problem (denoted $\operatorname{epi} G_N$) is contained in the region $\{(r_0, r) / \lambda_0 r_0 + \lambda r \ge \lambda_0 d y^i + \lambda r^i = w_{\lambda_0,N}(\lambda)\}$. If $w_{\lambda_0,N}(\lambda) + \lambda A_1 x > \theta_p$, then $\pi = -\lambda$ and $w_{\lambda_0,N}(\lambda) + \lambda A_1 x$ determine a valid cut for the master problem, that eliminates the previous master problem solution (x^p, θ^p). So that the parametric B&B could be stopped (called *premature end*) and the master problem could be solved again with that Benders optimality cut. If the parametric B&B is not stopped until optimality, the Benders cut obtained is known in the literature as *deepest cut*. The same situation is applied when the master problem solution turns the subproblem infeasible and a *feasibility cut* is generated. In this case the eliminated first stage solution will not appear on further master problems resolutions. Considering non deepest cuts could speed down the Benders algorithm convergence towards the optimum.

6. 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.

When introducing integer variables on the process, dual values have to be calculated with a LR procedure. This procedure approximates the dual function $w(\lambda)$ with cuts $\left\{w \leq dy^i + \lambda r^i\right\}$ formed with points belonging to the feasible region $\left\{A_2 y - b \leq r, y \in Y, r \in R^i\right\}$. If this feasible region is not completely known a priori, then cuts that approximate the dual function are eliminated as the corresponding points are found not to belong to the feasible region. This is the situation presented on section 3.4, where new approximations of the perturbation region eliminated cuts of the relaxed master dual problem (MD).

In nested case, this scheme is 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 three-stage problem of the form

$$(P) \qquad \begin{aligned} &\min c_1 x + c_2 y + c_3 z \\ &A_1 x + A_2 y \leq b_1 \\ &A_3 y + A_4 z \leq b_2 \\ &x \in X, y \in Y, z \in Z \end{aligned} \qquad (6.1)$$

Second stage problem then solves

$$\begin{aligned} &\min c_2 y + \theta_2(y) \\ &A_2 y \leq b_1 - A_1 x \\ &y \in Y \end{aligned} \qquad (6.2)$$

In any algorithm iteration that problem is replaced by a relaxed problem whose expression is

$$(MP) \qquad \begin{aligned} &\min c_2 y + \theta_2 \\ &A_2 y \leq b_1 - A_1 x \\ &\theta_2 \geq \theta_2^p(y) \\ &y \in Y \end{aligned} \qquad (6.3)$$

with $\theta_2 \geq \theta_2^p(y)$ representing the collection of *Benders* (*optimality* or *feasibility*) *cuts* that outer approximate the recourse function at that algorithm iteration. On the resolution of this problem with LR, once a multiplier value λ is proposed, the following problem must be solved

$$(PR_\lambda) \qquad \begin{aligned} &\min c_2 y + \theta_2 + \lambda r \\ &A_2 y - b_1 \leq r \\ &\theta_2 \geq \theta_2^p(y) \\ &y \in Y, r \in R^i \end{aligned} \qquad (6.4)$$

with R^i representing the approximated perturbation region at that algorithm iteration. Repetitions of previous problem create an approximation of the dual function $w(\lambda)$ and an approximation of the dual problem given as

$$(MD) \quad \begin{aligned} & \max w + \lambda A_1 x \\ & w \le c_2 y^1 + \theta_2^1 + \lambda r^1 \\ & \dots \\ & w \le c_2 y^k + \theta_2^k + \lambda r^k \end{aligned} \tag{6.5}$$

with $(y^k, \theta_2^k, r^k) \in \{y \in Y, \theta_2 \ge \theta_2^p(y), r \in R^i\}$ considering only phase 2 and lagrangean optimality cuts. The successive resolutions of its master problem and its subproblems generate constraints for the perturbation region R^{i+1} (*perturbation cuts*) and outer approximations of the recourse function $\theta_2 \ge \theta_2^{p+1}(y)$ (*Benders optimality cuts* and *feasibility cuts*). Then when facing the resolution of the second stage subproblem, there must be removed those cuts such that their associated points (y^k, θ_2^k, r^k) do not satisfy the new constraints. (i.e. $(y^k, \theta_2^k, r^k) \notin \{y \in Y, \theta_2 \ge \theta_2^{p+1}(y), r \in R^{i+1}\}$).

A few issues must also be commented about nested situations. When second stage is solved with the purpose of obtaining a proposal for the third stage problem (*forward pass*), then the problem that has to be solved presents the feasible region

$$\begin{aligned} & A_2 y - b_1 \le -A_1 x \\ & \theta_2 \ge \theta_2^p(y) \\ & y \in Y \end{aligned} \tag{6.6}$$

On the other hand, when trying to obtain a dual value (*backward pass*), then the feasible region of the parametric branch and bound is

$$\begin{aligned} & A_2 y - b_1 \le r \\ & \theta_2 \ge \theta_2^p(y) \\ & y \in Y, r \in R^i \end{aligned} \tag{6.7}$$

The second region contains first region. An efficient implementation would try to take advantage of these similarities, for example reformulating the region for the first situation as

$$\begin{aligned} & A_2 y - b_1 \le r \\ & r = -A_1 x \\ & \theta_2 \ge \theta_2^p(y) \\ & y \in Y, r \in R^i \end{aligned} \tag{6.8}$$

The calculation of the shadow (and consequently the perturbation region) for third and future stages can be found in Appendix B.

7. CONCLUSIONS

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

In the Benders algorithm, the recourse function is understood as the perturbation function of the subproblem when the right hand side 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 LR procedure is introduced within a B&B algorithm, developing a parametric B&B algorithm suitable for this situation. This parametric B&B gives back the minimization of infeasibilities (due to the relaxed constraints) in case the subproblem turns out to be infeasible. In the feasible case, it gives back the value of the lower convex envelope of the recourse function at the point proposed at the master problem.

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

8. References

[1] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear programming, theory and algorithms*, Wiley (1993).

[2] J.F. Benders, *Partitioning procedures for solving mixed-variables programming problems*, Numerische Mathematik 4, 238-252 (1962).

[3] J.R. Birge and F. Loveaux, *Introduction to stochastic programming*, Springer (1997).

[4] Geoffrion, A. M. *Lagrangean relaxation for integer programming* Mathematical Programming Study 2. Vol 2, pp 82-114. 1974.

[5] Geoffrion, A. M. *Generalized Benders Decomposition* J. Optim. Theory Appl. 10 237-259.

[6] C. Lemaréchal and A. Renaud *A geometric study of duality gaps, with applications* Mathematical programming, Serie A. 90:399-427, 2001.

[7] F.V. Loveaux and M.H. van der Vlerk. *Stochastic programming with simple integer recourse*. Mathematical programming, 61:301-325, 1993.

[8] M. Minoux, *Mathematical Programming*, theory and algorithms, Wiley (1986).

[9] G.L. Nemhauser and L.A. Wolsey, *Integer and combinatorial optimization*, Wiley, New York (1988).

[10] M.H. van der Vlerk. *Stochastic programming with integer recourse*. PhD thesis, University of Groningen, The Netherlands, 1995.

[11] R.M. Van Slyke and R. Wets, *L-Shaped linear programs with applications to optimal control and stochastic programming*, SIAM Journal of Applied Mathematics

1. Appendix A

1.1. Linear problem

Consider the problem

$$(MP) \qquad \begin{aligned} &\min c_1 x_1 + c_2 x_2 + \theta \\ &0 \geq \theta^j + \tilde{\pi}^j A_1 (x_0^j - x_2) \quad j = 1,\dots,l \\ &\theta \geq \theta^i + \pi^i A_1 (x_0^i - x_2) \quad i = 1,\dots,k \\ &x \in \overline{X}, \overline{X} = \left\{ A_{11} x \leq b_1, x_1 \in \mathbb{R}^{n_1'}, x_2 \in \mathbb{R}^{n_2'} \right\} \end{aligned} \tag{A.1}$$

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 $(d_1, d_2, ..., d_{n_1+m_1}, d_{n_1+m_1+1})$ the edges of the feasible region at point (x_1, x_2, θ). Let $(\widehat{d}_1, \widehat{d}_2, ..., \widehat{d}_{n_1+m_1}, \widehat{d}_{n_1+m_1+1})$ 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

$$(P_i) \qquad \begin{aligned} &\min e_i x \\ &Dx \leq 0 \\ &-1 \leq x \leq 1 \end{aligned} \tag{A.2}$$

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

$$(P_k) \qquad \begin{aligned} &\min \widehat{d}_k x \\ &\widehat{D}_k x \leq 0 \\ &-1 \leq x \leq 1 \end{aligned} \tag{A.3}$$

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.

1.2. Mixed integer problem

In this case the solution of problem

$$(MP) \qquad \begin{aligned} & \min c_1 x_1 + c_2 x_2 + \theta \\ & 0 \geq \theta^j + \tilde{\pi}^j A_1 (x_0^j - x_2) \quad j = 1,\dots,l \\ & \theta \geq \theta^i + \pi^i A_1 (x_0^i - x_2) \quad i = 1,\dots,k \\ & x \in X \end{aligned} \tag{A.4}$$

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 $(d_1^n, d_2^n, \dots, d_{n_1+m_1}^n, d_{n_1+m_1+1}^n)$ 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 $(\widehat{d}_1^n, \widehat{d}_2^n, \dots, \widehat{d}_{n_1+m_1}^n, \widehat{d}_{n_1+m_1+1}^n)$ 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

1.3. Inequality constraints

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

$$(P) \qquad \begin{aligned} & \min c_{11} x_1 + c_{12} x_2 + c_{21} y_1 + c_{22} y_2 + c_3 z \\ & A_1 x_2 + A_2 y \leq b_1 \\ & A_3 y_2 + A_4 z \leq b_2 \\ & x = (x_1, x_2) \in X, y = (y_1, y_2) \in Y, z \in Z \end{aligned} \tag{B.1}$$

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

$$R_1 = \{ r \,/\, \exists x_2 \in S_1 \,/\, r = -A_1 x_2 \} \tag{B.3}$$

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

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

so that its associated perturbation region keeps the form

$$R_2 = \{ r \,/\, \exists y_2 \in S_2 \,/\, r = -A_3 y_2 \} \tag{B.5}$$

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

$$(MP) \qquad \begin{aligned} &\min c_{21}y_1 + c_{22}y_2 + \theta_2 \\ &A_2 y \le b_1 - A_1 x_2 \\ &\theta_2 \ge \theta_2^p(y_2) \\ &y \in Y \end{aligned} \qquad \text{(B.6)}$$

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

$$\begin{aligned} &A_1 x_2 + A_2 y \le b_1 \\ &\theta_2 \ge \theta_2^p(y) \\ &x \in X, y \in Y \end{aligned} \qquad \text{(B.7)}$$

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 $A_2 y \le b_1 - A_1 x_2$ is not active, then y_2 is an extreme point.

In case $A_2 y \le b_1 - A_1 x_2$ is an active constraint, then y_2 will be an extreme point if an only if the region $\{y, \theta_2 \ge \theta_2^p(y), A_2 y \le b_1 - A_1 x_2\}$ is as biggest as possible, i.e., there is no $x' = (x_1', x_2')$ such that the region $\{y, \theta_2 \ge \theta_2^p(y), A_2 y \le b_1 - A_1 x_2\}$ is included in the region $\{y, \theta_2 \ge \theta_2^p(y), A_2 y \le b_1 - A_1 x_2'\}$.

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 $A_1 x_2 \ge A_1(x_2 + d)$. We are assuming without lost of generality A_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 $(\hat{d}_1, \hat{d}_2, ..., \hat{d}_k)$. These extreme directions are found as a result of applying criterion 2 to the first stage solution.

Then, if $A_1 \hat{d}_k \ge 0$, $\forall k : 1, ..., K$ 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 $A_2 y \le b_1 - A_1 x_2$ is not an active constraint.
2. y_2 is a 2-extreme point, x_2 is an extreme point of S_1 and $A_1 \hat{d}_k \ge 0$, $\forall \hat{d}_k$ extreme direction of $C(x_2)$.

1.4. Equality constraints

Now the situation is slightly different. Consider the problem

$$(MP) \qquad \begin{aligned} &\min c_{21}y_1 + c_{22}y_2 + \theta_2 \\ &A_2 y = b_1 - A_1 x_2 \\ &\theta_2 \geq \theta_2^p(y_2) \\ &y \in Y \end{aligned} \tag{B.8}$$

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

$$\begin{aligned} &\min c_{21}y_1 + c_{22}y_2 + \theta_2 + My^+ \\ &A_2 y - y^+ \leq b_1 - A_1 x_2 \\ &\theta_2 \geq \theta_2^p(y_2) \\ &y \in Y, y^+ \geq 0 \end{aligned} \tag{B.9}$$

with M big enough. Then we transform it to the situation presented on section 10.1.