CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

The Benders Dual Decomposition Method

Ragheb Rahmaniani
Shabbir Ahmed
Teodor Gabriel Crainic
Michel Gendreau
Walter Rei

January 2018

CIRRELT-2018-03

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

The Benders Dual Decomposition Method

Ragheb Rahmaniani[1,2,*], Shabbir Ahmed[3], Teodor Gabriel Crainic[1,4], Michel Gendreau[1,2], Walter Rei[1,4]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Mathematics and Industrial Engineering, Polytechnique Montréal, P.O. Box 6079, Station Centre-Ville, Montréal, Canada H3C 3A7

[3] School of Industrial & Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332, United States

[4] Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

Abstract. Many methods that have been proposed to solve large-scale MILP problems rely on decomposition techniques. These methods exploit either the primal or the dual structure of the problem, yielding the Benders decomposition or Lagrangian dual decomposition methods. We propose a new and high performance approach, called Benders dual decomposition (BDD), which combines the complementary advantages of both methods. The development of BDD is based on a specific reformulation of the Benders subproblem, where local copies of the master variables are introduced in the proposed subproblem formulation and then priced out into the objective function. We show that this method: (i) generates stronger feasibility and optimality cuts compared to the classical Benders method, (ii) can converge to the optimal integer solution at the root node of the Benders master problem and (iii) is capable of generating high quality incumbent solutions at the early iterations of the algorithm. We report encouraging numerical results on various benchmark MILP problems.

Keywords: Benders decomposition, Lagrangian relaxation, dual decomposition, mixed-integer programming.

Acknowledgements. Partial funding for this project has been provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery Grant program and by the Fonds de recherche du Québec through its Team Grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their strategic infrastructure grants. S. Ahmed was supported in part by the National Science Foundation Grant #1633196 and the Office of Naval Research Grants N00014-15-1-2078 and N00014-18-1-2075.

---

* Corresponding author: Ragheb.Rahmaniani@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2018

1. Introduction

Mixed-integer linear programing (MILP) is used to model a wide range of engineering and financial problems (Nemhauser and Wolsey 1988). Owing to the importance and inherent complexities of MILP models, it has been the subject of intense research since the early 1950s (Beale 1965, Jünger et al. 2009, Newman and Weiss 2013). In this article, we consider MILP problems of the following generic form

$$\min_{y,x}\{f^\top y + c^\top x : \ By \geq b, \ Wx + Ty \geq h, \ y \in \mathbb{Z}_+^n, \ x \in \mathbb{R}_+^m\} \tag{1.1}$$

where $f \in \mathbb{R}^n$, $B \in \mathbb{R}^{k\times n}$, $b \in \mathbb{R}^k$, $c \in \mathbb{R}^m$, $W \in \mathbb{R}^{l\times m}$, $h \in \mathbb{R}^l$, $T \in \mathbb{R}^{l\times n}$. We assume that the above problem is feasible and bounded. A prominent general approach to solve problem (1.1) relies on partitioning techniques such as the Benders decomposition method (Benders 1962), especially when part of the input data defined in the program are stochastic (Ruszczyński 1997, Costa 2005, Birge and Louveaux 2011, Rahmaniani et al. 2017a). To solve problem (1.1) with the Benders decomposition (BD) method , we introduce an auxiliary variable θ and the following master problem (MP)

$$MP = \min_{y,\theta}\{f^\top y + \theta : \ By \geq b, \ \theta \geq \bar{\theta}, \ y \in \mathbb{Z}_+^n\} \tag{1.2}$$

where $\bar{\theta}$ is a lower bound on θ to avoid unboundedness of the problem. The MP, the solution values of which define lower bounds for (1.1), is solved in a branch-and-cut method. At each integer node of the branch-and-bound tree, the solution y^* is fixed in the following dual subproblem (DSP)

$$DSP \ = \ \max_{\alpha}\{(h - Ty^*)^\top \alpha : \ W^\top \alpha \leq c, \ \alpha \in \mathbb{R}_+^l\} \tag{1.3}$$

If the above problem is unbounded, a direction of unboundedness r is chosen and the feasibility cut $(h - Ty)^\top r \leq 0$ is added to the MP to exclude all infeasible y solutions satisfying $(h - Ty)^\top r > 0$. Otherwise, a feasible solution to (1.1) is identified (thus allowing the upper bound of the algorithm to be updated), and the optimality cut $(h - Ty)^\top \bar{\alpha} \leq \theta$ is added to the MP. This procedure is repeated until the algorithm converges to an optimal solution. Due to the impact that the quality of the starting lower bound has on the size of the branch-and-bound tree generated by the algorithm, optimality and feasibility cuts can also be generated at fractional nodes of the tree at the beginning of the solution process to rapidly improve the quality of the lower bound. Such a strategy has mainly been applied at the root node, see (McDaniel and Devine 1977, Naoum-Sawaya and Elhedhli 2013, Adulyasak et al. 2015, Gendron et al. 2016, Bodur et al. 2017).

The BD method has been subject of intense research due to its practical importance in handling various MILP problems, e.g., production routing (Adulyasak et al. 2015), power generation (Cerisola et al. 2009), healthcare (Lin et al. 2016), logistics (Cordeau et al. 2006), green wireless network design (Gendron et al. 2016), map labeling (Codato and Fischetti 2006), supply chain management (Santoso et al. 2005), investment planning (Mitra et al. 2016), network design (Rahmaniani et al. 2017b), and so on. It may however perform disappointingly without the inclusion of some (problem-specific) acceleration techniques, see the recent literature reviews by Costa (2005), Rahmaniani et al. (2017a), Fragkogios and Saharidis (2018) and references therein for a complete information. Generally speaking, as highlighted in these reviews, the poor performance of the BD method is due to its shortfalls from both dual and primal perspectives.

The performance of the BD method, from a dual perspective, depends on the quality of the cuts chosen to bound the projected term (Holmberg 1990, Crainic et al. 2016, Fischetti et al. 2016). In particular, if the underlying LP relaxation of the problem is weak and/or the subproblems are degenerated, the algorithm performs poorly because the cuts and the root node bound are very weak (Magnanti and Wong 1981, Van Roy 1983, Sahinidis and Grossmann 1991, Geoffrion and Graves 1974, Cordeau et al. 2006, Rahmaniani et al. 2017b). Thus, effective selection of the Benders cuts has been the main focus of several studies, e.g., Magnanti and Wong (1981), Wentges (1996), Codato and Fischetti (2006), Papadakos (2008), Fischetti et al. (2010), Contreras et al. (2011), Sherali and Lunday (2013), among others. In more recent studies, strengthening the Benders cuts has been performed by making use of valid inequalities (VIs), see e.g., Bodur et al. (2017) and Rahmaniani et al. (2017b). Lagrangian techniques have also been used to generate alternative optimality cuts, particularly when the subproblem includes integrality requirements (Cerisola et al. 2009, Zou et al. 2017). In fact, it has been shown that the cuts obtained from the Lagrangian dual subproblems are not only valid for the Benders master problem, they are also generally tighter than the classical ones (Van Roy 1983, Santoso et al. 2005, Mitra et al. 2016).

From a primal point of view, the BD method has no systematic mechanism to generate high quality upper bounds. Indeed, it has oftentimes been observed that the evolution of the upper bound throughout the BD search process stagnates and finding good quality solutions can be quite a challenge, see (Boland et al. 2016). Thus, problem-specific heuristics have been used to generate a pool of high-quality solutions or to improve the quality of the master solutions obtained, see e.g. Poojari and Beasley (2009), Rei et al. (2009), Costa et al. (2012), Pacqueau et al. (2012) and sections 6.3 in Rahmaniani et al. (2017a).

Motivated by the important role that the cuts, the root node bound and the incumbent solution play on the performance of the BD method, we propose a new and high performance decomposition strategy, referred to as Benders dual decomposition (BDD), to overcome the aforementioned primal and dual shortfalls. The development of BDD is based on a specific reformulation of the subproblems where local copies of the master variables are introduced. This reformulation of the subproblems has been used in previous studies to generate generalized Benders cuts (Geoffrion 1972, Flippo and Rinnooy Kan 1993, Grothey et al. 1999, Zaourar and Malick 2014, Fischetti et al. 2016, Zou et al. 2017). In the present case, we apply Lagrangian duality to the proposed subproblem reformulation to price out the coupling constraints that link the local copies to the master variables. This allows us to impose the integrality requirements on the copied variables to obtain MILP subproblems, which are comparable to those defined in *Lagrangian dual decomposition* (LDD) (Ruszczyński 1997, Rush and Collins 2012, Ahmed 2013). As a consequence of obtaining these MILP subproblems, we will show that our proposed strategy efficiently mitigates the primal and dual inefficiencies of the BD method. Also, in contrast to the LDD method, BDD does not require an enumeration scheme (e.g., branch-and-bound) to close the duality gap. Furthermore, our strategy enables a faster convergence for the overall solution process. In summary, the main contributions of this article are the following:

- proposing a family of strengthened optimality and feasibility cuts that dominate the classical Benders cuts at fractional points of the MP,
- showing that the proposed feasibility and optimality cuts can give the convex hull representation of the MP at the root node, i.e., no branching effort being required,
- producing high quality incumbent values while extracting the optimality cuts,
- developing numerically efficient implementation methodologies for the proposed decomposition strategy and presenting encouraging results on a wide range of hard combinatorial optimization problems.

The reminder of this article is organized as follows. In section 2 we present the proposed decomposition strategy and in section 3 we present a toy example to illustrate its benefits. In section 4, the developed implementation methodologies are presented and discussed. The experimental design and computational results are, respectively, provided in sections 5 and 6. The last section includes our conclusions and a discussion of future research directions.

2. The proposed BDD method

The primal Benders subproblem, i.e., the primal form of (1.3), is $\min_x\{c^\top x:\ Wx \geq h - Ty^*,\ x \in \mathbb{R}^m_+\}$. By defining y^* as the current MP solution, without loss of generality, it can be rewritten as

$$\min_{x,z}\{c^\top x:\ Bz \geq b,\ Wx + Tz \geq h,\ z = y^*,\ x \in \mathbb{R}^m_+, z \in \mathbb{R}^n_+\}. \tag{2.1}$$

It should be noted that the constraints $Bz \geq b$ are redundant due to the presence of $z = y^*$. For a given feasible solution y^*, by solving problem (2.1), the following optimality cut can be derived

$$\theta \geq c^\top \bar{x} + (y - \bar{z})^\top \lambda^* \tag{2.2}$$

where $\bar{x}$ and $\bar{z}$, represent the optimal solution of the subproblem (2.1) and λ^* are the dual multipliers associated to the constraints $z = y^*$. If problem (2.1) for y^* is infeasible, then the following feasibility subproblem needs to be solved

$$\min_{x,z,v}\{\mathbb{1}^\top v:\ Bz \geq b,\ Wx + Tz + v \geq h,\ z = y^*,\ x \in \mathbb{R}^m_+, z \in \mathbb{R}^n_+,\ v \in \mathbb{R}^l_+\} \tag{2.3}$$

to generate a feasibility cut of the form

$$0 \geq \mathbb{1}^\top \bar{v} + (y - \bar{z})^\top \beta^* \tag{2.4}$$

where $\bar{v}$ define the optimal values of the v variables, β^* are the dual multipliers associated to the constraints $z = y^*$ and $\mathbb{1}$ is a vector of ones of size l. The optimality and feasibility cuts (2.2) and (2.4) are often referred to as *generalized Benders cuts* (GBC) (Geoffrion 1972, Sahinidis and Grossmann 1991, Grothey et al. 1999, Fischetti et al. 2016). It should be noted that the dual multipliers associated to the equality constraints $z = y^*$ define a subgradient of the objective function.

2.1. Strengthening the optimality and feasibility cuts

To strengthen the optimality cut (2.2), we price out the constraints $z = y^*$ into the objective function using the dual multipliers λ. By doing so, the following Lagrangian dual problem of (2.1) is obtained

$$\max_{\lambda\in\mathbb{R}^n} \min_{x\in\mathbb{R}^m_+, z\in\mathbb{R}^n_+} \{c^\top x - \lambda^\top (z-y^*) : \; Bz \geq b, \; Wx + Tz \geq h\} \tag{2.5}$$

In the interest of brevity, we henceforth use $X := \{(x,z) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \mid Bz \geq b, \; Wx + Tz \geq h\}$ and $F := \big\{y \in \mathbb{R}^n_+ | \{By \geq b\} \bigcap \{Wx \geq h - Ty : \text{for some } x \in \mathbb{R}^n_+\}\big\}$ to represent the compact form of the feasible region of subproblem (2.1) and the set of all feasible solutions y to the LP relaxation of problem (1.1). By applying this relaxation step, integrality requirements can be imposed on any subset of the variables z, given that they are no longer set equal to y^* (i.e., a master solution that is not guaranteed to be feasible).

In the following proposition, it is shown that problem (2.5) with $z \in \mathbb{Z}^n_+$ can be used to produce a valid optimality cut:

PROPOSITION 1. *For any solution $y^* \in F$, let λ^*, $\bar{x}$ and $\bar{z}$ be optimal solutions obtained by solving the following max-min MILP problem*

$$\max_{\lambda\in\mathbb{R}^n} \min_{(x,z)\in X} \{c^\top x - \lambda^\top (z-y^*) : z \in \mathbb{Z}^n_+\}, \tag{2.6}$$

then

$$\theta \geq c^\top \bar{x} + (y-\bar{z})^\top \lambda^* \tag{2.7}$$

is a valid optimality cut for the Benders master problem.

Proof. See Appendix A.1.

We observe from the previous proof that, when considering an integer solution $y^* \in F \bigcap \mathbb{Z}^n_+$, the optimality cut (2.7) is at most as strong as the classical Benders optimality cut. However, when applied using a fractional solution to the master problem, the cut (2.7) provides an added advantage which is studied in the following Theorem. This is an important observation to make, given that the LP relaxation of the MP is often solved first to quickly obtain valid cuts which enable the Benders method to perform more efficiently (McDaniel and Devine 1977, Naoum-Sawaya and Elhedhli 2013). The next result reports the improvement that is obtained regarding the lower bound provided through the use of the optimality cuts (2.7) when compared to the lower bound provided by (2.2).

THEOREM 1. *Given the dual multiplier $\hat{\lambda} \in \mathbb{R}^n$ obtained from problem (2.1), the optimality cut (2.7) is parallel to optimality cut (2.2) and at least $\sigma \geq 0$ units tighter, where*

$$\sigma = \min_{(x,z)\in X:\ z\in\mathbb{Z}^n_+} \{c^\top x - \hat{\lambda}^\top z\} \quad - \quad \min_{(x,z)\in X} \{c^\top x - \hat{\lambda}^\top z\} \tag{2.8}$$

Proof. See Appendix A.2.

As a direct implication of Theorem 1, the optimality cuts (2.7) are at least as strong than (2.2). If we use the dual multipliers obtained from solving subproblem (2.1), then according to Theorem 1 we can lift the optimality cut by an amount equal to the duality gap of the inner minimization problem. We refer to such cuts as *strengthened Benders cut*. If we optimize the Lagrangian problem (2.6) to generate an optimality cut, then we refer to such cut as *Lagrangian cut* (Zou et al. 2017).

The same strategy can also be used to strengthen the Benders feasibility cuts (2.4). In this case, the following Lagrangian dual for problem (2.3) is used to evaluate $z = y^*$

$$\max_{\beta\in\mathbb{R}^n} \min_{(x,z,v)\in\mathbb{R}^m_+\times\mathbb{R}^n_+\times\mathbb{R}^l_+} \{\mathbb{1}^\top v - \beta^\top(z-y^*) :\ Bz \geq b,\ Wx + Tz + v \geq h\} \tag{2.9}$$

Following Proposition 1, it might appear natural to impose integrality requirements on the z variables. However, given that the constraint set $Bz \geq b$ may not be satisfied, this approach does not guarantee that a valid feasibility cut will be obtained. Therefore, in the following proposition, a modified Lagrangian dual problem is proposed to generate a valid and lifted feasibility cut for the MP.

PROPOSITION 2. *For an arbitrary $y^* \notin F$, let β^*, $\bar{v}$, $\bar{u}$, $\bar{x}$ and $\bar{z}$ be the optimal solution of*

$$\max_{\beta\in\mathbb{R}^n} \min_{(x,z,v,u)\in\mathbb{R}^m_+\times\mathbb{Z}^n_+\times\mathbb{R}^l_+\times\mathbb{R}^k_+} \{\mathbb{1}^\top v + \mathbb{1}^\top u - \beta^\top(z-y^*) :\ Bz + u \geq b,\ Wx + Tz + v \geq h\} \tag{2.10}$$

then,

$$0 \geq \mathbb{1}^\top \bar{v} + \mathbb{1}^\top \bar{u} + (y - \bar{z})^\top \beta^* \tag{2.11}$$

is a violated valid feasibility cut to the master problem.

Proof. See Appendix A.3.

From last part in the proof of Proposition 2, it is clear that the proposed feasibility cut dominates the classical one if the duality gap of the minimization problem is non-zero. Thus, the results of Theorem 1 directly apply to this case as well.

To conclude this section, we study in Theorem 2 the root node bound of the BD method when the proposed optimality and feasibility cuts are used. To ease the exposition of the result, we henceforth use $H := \{(x,z,v,u) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^l_+ \times \mathbb{R}^k_+ | \; Bz + u \geq b, \; Wx + Tz + v \geq h\}$ to represent the feasible region of problem (2.10) and $Y := \{y \in \mathbb{R}^n_+ | \; By \geq b\}$ to indicate the domain of the relaxed y variables.

THEOREM 2. *Let Z_{IP} and Z^{LP}_{BD} be, respectively, the optimal objective value of problem (1.1) and the root node bound of the Benders MP with the proposed Lagrangian optimality and feasibility cuts, then $Z_{IP} = Z^{LP}_{BD}$.*

Proof. See Appendix A.4.

From the proof of Theorem 2, one observes that our method is closely related to the LDD method (see Appendix B for more details on this method). This decomposition approach has been successfully applied in the context of solving stochastic programs (Ahmed 2013), where it is referred to as *scenario decomposition*. As shown in Ahmed (2013), Lagrangian dual decomposition enables a stochastic program to be separated via the scenarios, allowing lower bounds to be computed more efficiently within a general search process. In the present case, as demonstrated in Theorem 2, when applied in the context of the Benders algorithm, this decomposition strategy strengthens the cuts generated while solving the LP in such a way as to close the gap at the root node. While solving (2.6) and (2.10) each time to optimality to generate the associated cuts may not be computationally efficient, the previous theoretical results nevertheless provide clear guidelines defining how Benders cuts can be lifted to improve the quality of the lower bound generated at each iteration of the algorithm. Furthermore, it should be noted that the proposed cut generation strategy is applicable to a wider range of problems. Specifically, from the definitions of problems (2.6) and (2.10), it is clear that the generation of optimality and feasibility cuts is independent of the specific structure of set X. Thus, set X could include nonlinear constraints and integer requirements on the variables.

3. Example

To illustrate the benefits of the proposed method, consider following toy problem:

$$\min_{y \in \{0,1\},\; x \geq 0} \{x : x + 15y \geq 8, 3x + 10y \geq 13, x + 10y \geq 7, 2x - 10y \geq -1, 2x - 70y \geq -49\} \tag{3.1}$$

The optimal solution to this problem is $y = 0$, which has a cost of 8 units. We solve the LP relaxation of this problem using the (i) classical, (ii) strengthened and (iii) Lagrangian cuts. The detailed results of each iterative procedure can be found in Appendix C. In Figure 1, we have graphically depicted the cuts generated at each iteration of these methods in the (y, θ)-space.

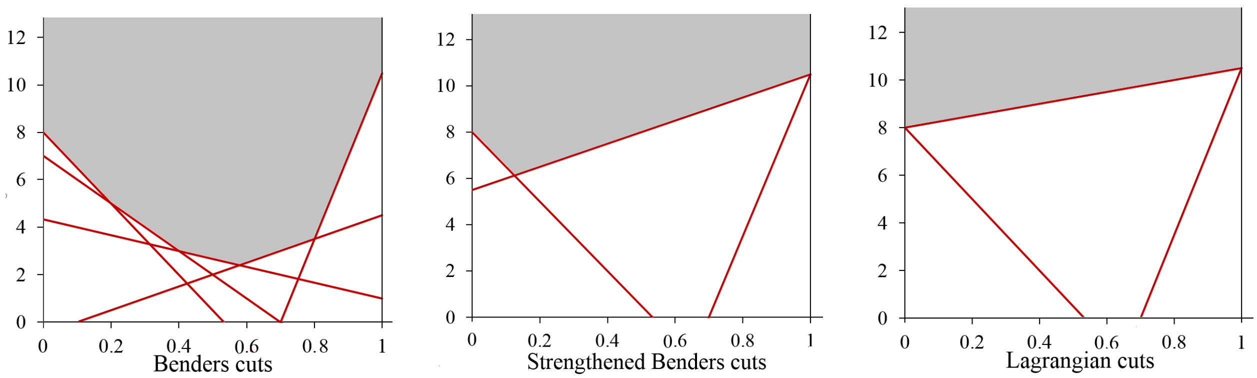

Figure 1 Performance of different cuts in cutting the solution space of LP master problem

In solving the LP relaxation of the master problem with the classical Benders cuts, after 5 iterations the method converges to the optimal LP solution $y = 0.58$ and its objective value of 2.4 units. If the strengthened Benders cuts are used, then the first two iterations of the algorithm generate the same cuts as the classical ones. The reason for this is that the y solutions for these two iterations are integer and as a result the Benders cuts are the strongest. At the third iteration, solving the MP produces the solution $y = 0.65$. For this solution the associated strengthened Benders cut is parallel to the classical Benders cuts but tighter by 6.125 units. At this point, the proposed method with the strengthened cuts converges. For the master solution $y = 0.65$, the Lagrangian cut provides the convex hull representation of the MP and, consequently, the algorithm converges to the optimal integer solution. As demonstrated in the previous section, in this case, while solving the LP relaxation of the MP, the optimal integer solution to the problem (3.1) is obtained.

4. Implementation details

As clearly shown in section 2 and numerically illustrated on a simple example in section 3, the proposed method reduces the number of required cuts for convergence and significantly tightens the root node bound. However, this is achieved at the cost of solving one (or several) MILP subproblem(s), which may create a numerical bottleneck for the Benders algorithm (especially when solving large-scale optimization problems). Therefore, in this section, we develop a series of strategies that aim to apply the proposed method in a computationally efficient manner.

4.1. Multi phase implementation

To alleviate the computational burden of optimizing the Lagrangian dual problem (2.6) for numerous iterations, we propose to implement a three-phase strategy to generate the Benders cuts. This is motivated by the fact that, at the initial iterations of the Benders algorithm, the master solutions are usually of very low quality. At this point, the derived cuts provide a poor approximation of the value function at the optimal region of the problem. In turn, it may not be worthwhile to invest the effort of lifting these cuts. However, as the Benders algorithm progresses and more and more cuts are added to the MP (thus improving the quality of the lower bound provided by the MP and guiding the search process towards more promising areas of the feasible region), then a lifting strategy may be applied to accelerate the convergence of the algorithm. The proposed multi phase implementation works as follows.

Phase 1. To quickly derive valid cuts, we first solve the LP relaxation of the MP with the classical cuts (2.2) and (2.4), which is a strategy that was originally proposed by McDaniel and Devine (1977). Given that this strategy has become one of the main methods used to efficiently apply the Benders algorithm on numerous applications, see (Rahmaniani et al. 2017a), we thus apply it in this phase of our implementation.

Phase 2. We then generate the strengthened optimality Benders cuts by first fixing the dual multipliers λ to the optimal values obtained by solving the problem (2.1). Using the obtained λ values, the inner minimization problem in (2.6) (i.e., the Lagrangean dual subproblem) is then solved to find the values of $\bar{x}$ and $\bar{z}$, which are then used to generate a strengthened optimality Benders cuts. Similarly, if the MP solution is infeasible at this point, a strengthened feasibility cut is generated by first fixing the dual multipliers β to the optimal values obtained by solving the problem (2.3). Using values β, the inner minimization problem in (2.10) is then solved to find the values of $\bar{v}$, $\bar{u}$ and $\bar{z}$, which are then applied to produce a strengthened feasibility cut. Overall, this second phase starts the lifting process of the proposed cut generation strategy.

Phase 3. In this last phase, Lagrangian cuts are generated. To do so, we heuristically solve the Lagrangian dual problem (2.6). Therefore, to generate an optimality cut and assuming that a series of values $\bar{x}^v$ and $\bar{z}^v$, for $v = 1, 2, \ldots, t-1$, have been obtained via the process by which a strengthened cut is generated (i.e., solving the inner minimization problem (2.6) for a series of values λ^v, for $v = 1, 2, \ldots, t-1$), the following regularized problem is solved to update the dual multipliers λ:

$$\max_{\lambda \in \mathbb{R}^n, \eta \in \mathbb{R}^1} \eta - \frac{\delta}{2}||\lambda^{t-1} - \lambda||_2^2 \tag{4.1}$$

$$\eta \leq c^\top \bar{x}^v + (y - \bar{z}^v)^\top \lambda \qquad \forall v = 1, 2, \ldots, t-1. \tag{4.2}$$

The objective function (4.1) seeks to maximize the value by which the cut is lifted (i.e., value η) minus the distance value $\frac{\delta}{2}||\lambda^{t-1} - \lambda||_2^2$. The latter component of the objective function, which defines a trust-region for the Lagrangean multipliers, is included to stabilize the updating process for a given prefixed value $\delta \in \mathbb{R}^1_+$ that represents the stabilization parameter. It thus enables a new vector λ^t of values to be found that is close to the one obtained in the previous iteration (i.e., λ^{t-1}). As for the constraints (4.2), they provide an inner approximation of the Lagrangian dual subproblem. Given the new Lagrangian multiplier values λ^t, the inner minimization problem in (2.6) is instantiated and then solved to obtain $\bar{x}^t$ and $\bar{z}^t$. At this point, the constraint $\eta \leq c^\top \bar{x}^t + (y - \bar{z}^t)^\top \lambda$ is added to (4.2) and the updated regularized program is solved. This process is repeated until either the new multiplier values fail to lift the cut by at least $\gamma\%$ when compared to the previous iteration, or, a maximum number of iterations (i.e., defined as parameter κ) is reached. Finally, it should be noted that the same procedure is applied to generate Lagrangian feasibility cuts by simply interchanging the appropriate programs.

4.2. Partially relaxed subproblems

To further reduce the computational burden of solving one or several MILP subproblems to generate a single cut, we suggest in this section two relaxation strategies that can be applied. The first strategy applies the integrality requirements only on a subset of the z variables following the relaxation of the constraint set $z = y^*$. We thus partition the variables as follows: $z^\top = [z_I, z_J]^\top$ using two sets I and J such that $I \subset \{1, \ldots, n\}$, $I \subset \{1, \ldots, n\}$, $I \cap J = \emptyset$ and $I \cup J = \{1, \ldots, n\}$, where z_I and z_J represent the subvectors of z that include the variables whose indexes are included in the subsets I and J, respectively. The integrality requirements are then imposed solely on the variables of the subvector z_I. Therefore, the formulation of the cut generation problem becomes: $\max_{\lambda\in\mathbb{R}^n}\left\{\lambda^\top y^* + \min_{(x,z)\in X}\{c^\top x - \lambda_I^\top z_I - \lambda_J^\top z_J : z_I \in \mathbb{Z}_+^{|I|}\}\right\}$. Given that the integrality constraints on the variables z_J are relaxed, this problem provides an optimality cut that is weaker than the Lagrangian cut. However, it remains stronger when compared to the Benders cut. At each iteration of the Benders algorithm, to define the subsets I and J we apply the following procedure: we initially set $I=\emptyset$ and $J=\{1,\ldots,n\}$; then if a master variable y_a (where $a\in\{1,\ldots,n\}$ is the index associated with a variable in the y vector) takes a fractional value such that $\varepsilon_0+\lfloor y_a^*\rfloor \le y_a^* \le \lceil y_a^*\rceil-\varepsilon_1$, where y^* is the current master solution and the non-negative values ε_0 and ε_1 are two preset parameters such that $\varepsilon_0+\varepsilon_1<1$, then a is removed from J and added to I. Therefore, parameters ε_0 and ε_1 are used to adjust the width of the fractional interval that is used to identify the variables on which the integrality restrictions will be imposed.

As a follow-up to the previous approach, a second relaxation strategy is proposed. This strategy, which is problem specific, is based on the idea of fixing variables in the Lagrangian subproblems. To apply it, we need to assume that fixing a z variable to its upper limit in the inner minimization problem of (2.6) widens the solution space, i.e., $\min_{(x,z)\in X}\{c^\top x : z=y^*\}\ge \min_{(x,z)\in X}\{c^\top x : z\ge y^*\}$. Such an assumption can be made in a wide gamut of applications, particularly in cases where capacity constraints are imposed. When this assumption is observed, then the following result can be applied:

PROPOSITION 3. *Let y^* be the current master solution and $\hat{\lambda}$ be the corresponding dual multiplier obtained from linear program (2.1). Furthermore, let I' be a subset of I such that $y_a^*\ge ub_a-\tilde{\varepsilon}$ and $\hat{\lambda}_a\ge \upsilon$ for every $a\in I'$ and for given values $\tilde{\varepsilon}\ge 0$ and $\upsilon\ge 0$, where ub_a is the upper bound on variable y_a. Then, the following restricted Lagrangian dual problem*

$$\max_{\lambda\in\mathbb{R}^n}\Big\{\min_{(x,z)\in X: z_I\in\mathbb{Z}_+^{|I|}}\{c^\top x-\lambda^\top(z-y^*):\ z_{I'}=ub_{I'}\}:\lambda_{I'}=\hat{\lambda}_{I'}\Big\},\tag{4.3}$$

can be used in replacement of (2.6) to generate approximate Lagrangian cuts, where $ub_{I'}$ is the vector of upper bounds associated with the variables whose indexes are in I'.

Proof. See Appendix A.5

Given that the Lagrangian dual problem needs to be solved several times per iteration, we expect that these strategies will noticeably enhance the overall numerical performance of the Benders algorithm. However, relaxing the subproblems will weaken the optimality cuts derived. Therefore, careful consideration must be given to the specific values to which the different parameters defining the strategies will be fixed. In the case of the second strategy, it should be noted that the parameters ε and υ define the conditions by which the variables whose indexes will be included in subset I' are chosen. Subset I' is then used to instantiate the restricted Lagrangian dual problem (4.3), which is solved to generate the approximate Lagrangian cuts. The more y variable indexes are added to subset I' and the more problem (4.3) is restricted, thus reducing the numerical burden to produce a cut. However, the larger the restriction that is applied and the weaker the cut that is derived. Clearly, a trade-off between numerical efficiency and cut strength needs to be established when applying the strategy on a specific application.

4.3. ε-optimal cuts

In many applications finding optimal solutions for the Lagrangian dual subproblems, even when applying the aforementioned strategies, may not be easy. In our pilot numerical experiments, we observed that closing the optimality gap from, for example, 1% to 0% is often the most time consuming part of solving a subproblem. Therefore, in this section, we propose to use ε-optimal solutions of the Lagrangian dual subproblems to generate approximate cuts. The following results shows the validity of this approach:

PROPOSITION 4. *Let $y \in Y$ and $(\bar{x}(\varepsilon), \bar{z}(\varepsilon)) \in X$ be an ε-optimal solution of the minimization problem of the Lagrangian dual. Then, the following cut can be derived:*

$$\theta \geq c^\top \bar{x}(\varepsilon) + (y - \bar{z}(\varepsilon))^\top \lambda^* - \varepsilon \qquad \forall y \in Y \tag{4.4}$$

Proof. See Appendix A.6.

4.4. Upper bounding

In this last section, we detail how the cut generation strategy can also be used to improve the upper bound that is generated throughout the Benders solution process. We first highlight the fact that solving the inner minimization problem in (2.6) provides an integer $\bar{z}$ solution, which is feasible for the original problem (1.1) given the copied constraints in the proposed reformulation (2.5). Therefore, the cut generation process produces solutions that can be used as incumbents for the overall problem. Thus, the upper bound can be easily updated by evaluating the objective value of these solutions, i.e., $f^\top \bar{z} + c^\top \bar{x}$. If this value is lower than the current incumbent, then the associated solution is used to generate a classical optimality cut as well.

It should be noted that for cases where there are multiple subproblems (i.e., stochastic models), then the solution from one subproblem may be infeasible for the overall problem given that it may not satisfy all the constraints in the other subproblems. Moreover, to obtain the objective value of a solution in this case, all subproblems associated with the solution need to be evaluated. We thus first propose to record the extracted $\bar{z}$ solutions in a pool. From this pool we then merely evaluate those solutions which are generated at least τ times. This is motivated by the heuristic notion that a solution with a higher level of saturation might correspond to an optimal one. Finally, if the objective value of the solution is already higher than the current incumbent value, then it can be simply discarded considering that no improvement is possible.

5. Experimental Design

To study whether our method is numerically beneficial, we complement the theoretical analysis presented in previous sections with an extensive computational study. In this section, we describe the MILP problems used to conduct our numerical analyses, the general characteristics of the test instances and how the algorithms were implemented.

5.1. Problems studied and test instances

We test our method on benchmark instances of three different optimization problems from the scientific literature. Detailed descriptions of these problems can be found in Appendix D. Here, we provide a high level summary of these problems and instances.

Our first test instances are related to the *fixed-charge multicommodity capacitated network design* (FMCND) problem with stochastic demands. This problem naturally appears in many practical applications (Klibi and Martel 2012) and it has been numerically shown to be notoriously hard to solve (Crainic et al. 2011). In addition, it lacks the complete recourse property, which entails that the generation of feasibility cuts is necessary to ensure the convergence of the BD method. We have considered 7 classes of instances (r04 to r10) from the **R** set, as developed in Crainic et al. (2001). Each class includes 5 instances with varying cost and capacity ratios.

Our second test instances are related to the *capacitatied facility location* (CFL) problem, which was introduced by Louveaux (1986) and addressed in Bodur et al. (2017), Fischetti et al. (2016) and Boland et al. (2016) among others. To avoid the generation of feasibility cuts, which do not contribute towards the improvement of the lower bound generated by the BD method, the complete recourse property can be enforced via the inclusion of a constraint in the MP. As for the instances, we use the deterministic **CAP** instances (101 to 134), which are part of the OR-Library. These instances include 50 customers with 25 to 50 potential facilities. For the stochastic variant, we have used the scenarios generated by Bodur et al. (2017) where each scenario includes 250 scenarios. It should be noted that the deterministic instances of this problem are referred to as "CFL" and the stochastic ones as "CFL-S".

Fianlly, our third set of benchmark instances are associated to the *stochastic network interdiction* (SNI) problem proposed by Pan and Morton (2008). It is important to note that this problem is structurally different from the previous ones, in the sense that there are no fixed costs associated to the master variables in the objective function. Moreover, due to the presence of a budget constraint, the variable fixing strategy as detailed in section 4.2 cannot be applied. Regarding the instances, we have considered those which have been described and used by Pan and Morton (2008), Bodur et al. (2017) and Boland et al. (2016). All instances have 456 scenarios and 320 binary master variables associated to the same network of 783 nodes and 2586 arcs. We specifically consider those instances which are part of the classes referred to as "snipno" 3 and 4 (see Pan and Morton (2008) Tables 3 and 4). Each class includes 5 different instances and for each instance we have considered varying budget limits (i.e., 30, 40, 50, 60, 70, 80 and 90 units).

5.2. Parameter settings and implementation details

All algorithms (both the BDD and Benders methods) are implemented in a branch-and-cut framework, i.e., a branch-and-bound tree is built and Benders cuts are generated only at the root node and whenever a potential incumbent solution is found. We apply our proposed strategy exclusively at the root node of the branch-and-bound tree. To optimize the Lagrangian dual in the LDD method, we use the same technique as the one previously discussed in section 4.1.

In order to make the implementations simple and easily replicable, we do not employ any specialized codes or algorithms. Thus, we solve all the derived problems with a general-purpose solver. Accordingly, we also avoid incorporating any acceleration technique in our algorithm in order to perform a fair assessment of our proposed decomposition method versus the classical one.

In all methods, cuts (both feasibility and optimality) are generated by solving each subproblem within an optimality gap of 0.5% (i.e., $\varepsilon = 0.5\%$). Moreover, to generate the Lagrangian cuts for the FMCND and CFL instances, we apply the variable fixing strategies defined in section 4.2. To provide a thorough numerical assessment of our proposed decomposition method, we have implemented the following four variants of the strategy:

- BDD_1**:** uses the strengthened Benders cuts by imposing the integrality requirements on all the copied variables in the subproblem, i.e., $J = \emptyset$ and $I = \{1,..,n\}$,
- BDD_2**:** uses the strengthened Benders cuts by imposing the integrality requirements on a subset of the copied variables in the subproblem according to the strategy detailed in section 4.2
- BDD_3**:** similar to BDD_1 but also generates Lagrangian cuts,
- BDD_4**:** similar to BDD_2 but also generates Lagrangian cuts.

To set the values for the different search parameters described in section 4, we conducted a series of preliminary experiments over a small subset of the instances and we have chosen the following values for the parameters in our algorithms: $\delta = \frac{1}{t+1}10^{-2}$, $\kappa = 10$, $\gamma = 10^{-1}$, $\varepsilon_0 = 10^{-2}$, $\varepsilon_1 = 10^{-1}$, $\tilde{\varepsilon} = 10^{-2}$, $\tau = 3$, $\upsilon = 10^2$, where t is the iteration counter. As for the stopping criteria, an amount of 10 hours was allotted for the overall solution of the considered problems. It should be noted that this amount included a time limit of 1 hour to optimize and lift the root node bound. The overall optimality gap to terminate the algorithms was fixed at 1%.

Finally, all programs were coded in C++, using the CPLEX version 12.7 as the optimization solver. The codes were compiled with g++ 4.8.1 and executed on Intel Xeon E7-8837 CPUs running at 2.67GHz. with 32GB RAM under a Linux operating system and in a single-thread mode. The branch-and-cut algorithm was also implemented using the CPLEX callable libraries.

6. Computational results

In this section, we quantify the computational benefits of the proposed decomposition methodology when solving the instances considered. We first analyze the performance and behavior of the various variants of our method (i.e., BDD_1, BDD_2, BDD_3 and BDD_4). To do so, we assess the quality of the lower and upper bounds obtained by our variants at the root node when compared with the classical decomposition methods. We will then evaluate the convergence behavior of our approach and compare its performance with the state-of-the-art optimization solver CPLEX.

6.1. Computational results at the root node

In Table 1, both the time requirements and the percentages of the root gap with respect to the optimal solutions are reported for the BD, LDD and the four variants of our method when these strategies are used to solve all considered benchmark instances. Recall that these experiments were obtained by allotting a maximum of 1 hour of running time.

We observe that BDD_1 and BDD_2 tighten the root node bound by more than 4% for FMCND and less than 1% for the CFL-S and CFL instances. However, they fail to improve this bound for the SNI instances. This observation can be explained by the fact that, in SNI the duality gap of the subproblems is very small. Thus, according to Theorem 1, the strengthened optimality cuts are close to the classical Benders cuts.

This being said, we observe that the Lagrangian cuts are very effective in tightening the root node bound. The most significant improvements are attained for the CFL-S instances, where the average gap at the root node is reduced to less than 1.5% from 17.81% by using the Lagrangian cuts (see the results of BDD_3 and BDD_4 methods). The same type of significant improvements are also observed for the CFL and FMCND instances. Even in the case of the SNI instances, where the obtained improvements are less significant, the root node bound is still tightened by approximately 9%. It should be noted that these stronger root node bounds are achieved at the cost of higher running times. For example, on average, the time requirement of BDD_3 is more than 30 times higher than the BD method for the CFL-S instances. We thus investigate in section 6.2 if this additional effort at the root node pays off in the overall performance of the algorithm.

Our method, when applied using the Lagrangian cuts, also outperforms the LDD algorithm considering that it generates much tighter root node bounds in shorter times. These results can be explained by the fact that our approach effectively integrates the complementary advantages of both the BD and LDD methods, i.e., an effective search mechanism coupled with the generation of strong cuts. Regarding these results, it

Table 1 Average Percentage Gap of the Lower Bound at the Root Node from the Optimal Solution in Different Methods

Prob.	Inst.	Benders decomposition		Dual decomposition		The proposed BDD method							
						BDD_1		BDD_2		BDD_3		BDD_4	
		Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)
FMCND	r04	22.48	6.27	5.89	2479.79	19.52	49.46	19.89	43.95	1.64	900.98	2.25	773.49
	r05	17.95	46.68	9.82	2859.31	13.17	239.90	13.78	232.09	1.14	2160.62	1.26	2105.21
	r06	20.96	336.02	10.73	3349.31	14.63	1026.42	14.86	960.73	9.72	2977.36	8.86	2965.59
	r07	18.56	16.19	5.39	3146.00	16.50	70.10	16.82	61.48	3.58	1048.47	4.62	922.98
	r08	20.57	87.14	9.38	3067.86	17.35	376.39	17.44	331.62	7.12	2673.75	6.22	2051.87
	r09	23.43	596.56	12.85	3876.46	16.99	1683.00	16.93	1832.72	12.81	3681.55	11.80	3571.77
	r10	(45.22)	(3600.56)	(19.37)	(3759.72)	-	-	-	-	-	-	-	-
	Mean	**20.66**	**181.48**	**9.01**	**3129.79**	**16.36**	**574.21**	**16.62**	**577.10**	**6.00**	**2240.46**	**5.83**	**2065.15**
CFL-S	101-104	22.63	16.31	5.17	3588.80	21.38	63.49	21.44	53.15	1.03	696.72	0.76	475.54
	111-114	8.45	77.79	5.74	3806.09	7.73	310.84	7.73	283.79	1.05	1582.18	0.97	1955.86
	121-124	19.23	77.69	12.50	3650.62	18.38	276.00	18.36	231.65	1.51	3003.90	1.38	2762.77
	131-134	24.14	69.67	17.77	3674.22	23.76	172.27	23.75	173.34	2.32	2229.10	1.83	3342.80
	Mean	**18.61**	**60.37**	**10.30**	**3679.93**	**17.81**	**205.65**	**17.82**	**185.48**	**1.47**	**1877.97**	**1.23**	**2134.24**
CFL	101-104	23.80	0.44	0.00	0.03	23.10	0.98	23.49	0.83	2.89	16.48	9.80	14.73
	111-114	11.22	3.11	0.35	0.16	10.57	6.15	10.55	4.83	2.20	174.14	2.20	226.24
	121-124	21.29	2.35	0.00	0.07	21.29	1.18	21.07	1.57	3.41	146.26	2.99	154.51
	131-134	24.36	1.29	0.00	0.04	24.36	0.83	24.15	1.24	4.31	112.69	6.12	199.56
	Mean	**20.17**	**1.80**	**0.09**	**0.08**	**19.83**	**2.28**	**19.82**	**2.12**	**3.20**	**112.39**	**5.28**	**148.76**
SNI	3-30	22.47	128.20	19.59	3751.47	22.42	162.82	22.42	149.65	14.37	1065.85	14.35	1221.82
	3-40	26.21	143.18	23.10	3837.34	26.21	182.12	26.21	170.39	17.35	1362.20	17.35	1377.63
	3-50	27.53	155.60	23.46	3827.19	27.53	193.94	27.53	177.32	19.20	1303.32	19.30	1367.68
	3-60	28.17	164.83	24.42	3890.09	28.17	204.55	28.17	177.51	22.14	1095.66	22.18	1100.10
	3-70	28.88	167.15	25.30	3877.71	28.88	204.70	28.88	173.34	21.90	979.65	22.15	964.42
	3-80	30.87	179.76	26.95	3786.87	30.85	212.44	30.85	178.77	20.53	1316.73	20.81	1294.92
	3-90	33.13	198.12	28.66	3841.09	33.13	226.75	33.13	173.33	23.95	1056.19	24.24	1100.43
	4-30	25.15	82.10	21.80	3815.90	25.14	130.79	25.14	129.57	14.49	1089.79	14.88	1047.74
	4-40	28.45	90.05	26.99	4015.53	28.45	144.97	28.45	143.55	18.47	1254.47	18.46	1274.26
	4-50	30.54	98.57	27.47	3730.99	30.54	161.58	30.54	147.27	21.29	1195.25	22.11	1054.68
	4-60	32.07	97.85	30.44	3790.82	32.07	160.11	32.07	152.50	25.30	1007.63	25.06	979.38
	4-70	32.61	102.49	32.50	3808.80	32.61	158.54	32.61	142.50	25.24	830.67	24.94	919.71
	4-80	33.27	105.02	33.03	3846.78	33.27	169.71	33.27	143.75	22.76	916.80	22.50	916.77
	4-90	36.17	108.51	35.90	3832.19	36.16	163.54	36.16	133.21	22.77	1082.82	21.53	1268.45
	Mean	**29.68**	**130.10**	**27.12**	**3832.34**	**29.67**	**176.90**	**29.67**	**156.62**	**20.70**	**1111.22**	**20.70**	**1134.86**

should be noted that the LDD method solves the deterministic variants of the facility location problem (i.e., CFL) very quickly. This is explained by the fact that this specific decomposition strategy when applied on deterministic problems is equivalent to solving the original problem with CPLEX. Considering that the CFL instances are of small size, they can be solved in a few seconds. Thus, we exclude the CFL instances from the numerical analyses conducted in the next sections so as to focus on more challenging problems.

Finally, considering the results obtained on the $r10$ instances, the LDD method reaches an optimality gap of 19.37% which is much tighter than that obtained by the BD method (45.22%) or our algorithms (that fail to produce an optimality gap). For these instances, the four proposed BDD variants require more that 1 hour of computation time to solve the LP relaxation. Thus, our algorithm terminates before generating any strengthened or Lagrangian cuts. However, if the time limit to solve the root node is increased, then we observe that the proposed BDD_3 and BDD_4 variants reach a much tighter bound than the LDD method, as reported in Table 4. We next specifically analyze the impact of generating strengthened and Lagrangian feasibility cuts and consider how the upper bound obtained at the root node is affected by the use of the proposed BDD method.

6.1.1. Impact of the proposed feasibility cuts The results reported for FMCND in Table 1 were obtained using the classical feasibility cuts. We now assess how the performance of our method is affected when the proposed strengthened and Lagrangian feasibility cuts are generated in the second and third phase of the algorithm. These numerical results are presented in Table 2. It should be noted that we have only presented the results for the BDD_1 and BDD_3 variants (the results obtained for BDD_2 and BDD_4 being very similar to them, respectively).

When analyzing the results in Table 2, it is clear that generating strengthened and Lagrangian feasibility cuts has less impact on tightening the root node bound, when compared to the inclusion of the proposed optimality cuts. This was to be expected given that feasibility cuts do not bound the θ variables in the master problem. Moreover, the generation of the proposed feasibility cuts increases the time requirements. For these reasons, a deterioration of the obtained root node gaps is observed. One can thus anticipate that the proposed feasibility cuts will be most beneficial for problems where finding feasible solutions is a challenge. However, given the marginal impact of these cuts in the present case, we henceforth only use the classical feasibility cuts to solve the FMCND instances.

Table 2 Performance of the Proposed Feasibility Cuts versus the Classical Ones

	Classical Feasibility Cuts				Proposed Feasibility Cuts			
	BDD_1		BDD_3		BDD_1		BDD_3	
	Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)	Gap(%)	Time(s.)
r04	19.52	49.46	1.64	900.98	19.36	43.84	1.50	892.88
r05	13.17	239.90	1.14	2160.62	13.87	228.36	1.00	2281.52
r06	14.63	1026.42	9.72	2977.36	14.45	1004.79	10.18	3055.83
r07	16.50	70.10	3.58	1048.47	16.21	74.50	4.18	1034.03
r08	17.35	376.39	7.12	2673.75	17.45	303.84	7.55	2667.76
r09	16.99	1683.00	12.81	3681.55	16.74	1567.31	12.90	3842.33
r10	-	-	-	-	-	-	-	-
Mean	**16.36**	**574.21**	**6.00**	**2240.46**	**16.35**	**537.11**	**6.22**	**2295.72**

6.1.2. Quality of the upper bound We now assess the quality of the upper bounds obtained from solving MIP subproblems at the root node when the different BDD variants or the LDD method are used. The percentage gaps of the obtained upper bounds with respect to the optimal solutions are summarized in Table 3. It should be noted that the computation times required to obtain these upper bounds were reported in Table 1. Also, we again recall that, in the case of the $r10$ instances, no results are reported for the different BDD variants given that the methods were unable to solve the LP relaxations in the maximum allotted time of 1 hour.

From Table 3, we observe that the obtained upper bounds by the BDD_3 variant are very close to the optimal solutions. When compared to BDD_4, BDD_3 finds better upper bounds, which can be explained by the fact that, in this variant, the integer requirements are imposed on all the copied variables in the subproblems. It thus generates more integer solutions. Similarly, given their respective definition, BDD_3 and BDD_4 yield a much larger pool of integer solutions when compared to the BDD_1 and BDD_2 variants. Consequently, these variants greatly improve the quality of the upper bounds generated. Moreover, the upper bounds obtained by BDD_3 and BDD_4 (with the exception of the FMCND instances) are much better than those generated by the LDD method, which clearly indicates that our method provides an improved search mechanism. Last but not least, the quality of the primal bounds reported in Table 3 shows that the proposed enhancements significantly alleviate the primal inefficiencies of the BD method.

Table 3 Average Percentage Gap of the Upper Bound Obtained by Different Method from the Optimal Value

Prob.	Inst.	Dual decomposition	The proposed BDD method			
			BDD_1	BDD_2	BDD_3	BDD_4
FMCND	r04	2.03	24.86	27.70	0.77	3.09
	r05	1.74	18.86	22.41	0.36	5.08
	r06	2.00	15.13	15.80	3.61	13.58
	r07	3.06	29.16	29.99	2.06	5.65
	r08	3.29	35.48	35.29	2.59	2.93
	r09	5.32	26.01	23.89	4.84	20.79
	r10	(22.44)	-	-	-	-
	Mean	**2.91**	**24.92**	**25.85**	**2.37**	**8.52**
CFL-S	101-104	1.34	6.69	6.69	0.07	0.23
	111-114	5.37	15.66	15.66	0.06	0.18
	121-124	8.61	14.27	14.27	0.08	0.15
	131-134	19.00	15.36	15.36	1.41	0.28
	Mean	**8.58**	**13.00**	**13.00**	**0.40**	**0.21**
SNI	3-30	15.82	24.47	22.71	2.94	3.72
	3-40	20.54	28.78	34.13	6.44	6.44
	3-50	27.61	40.92	39.50	8.90	7.93
	3-60	36.22	46.61	47.13	11.36	11.36
	3-70	43.92	44.63	44.41	16.89	16.89
	3-80	54.33	46.88	46.43	12.59	12.78
	3-90	57.66	47.25	42.23	15.22	19.93
	4-30	18.71	37.05	37.05	3.58	5.92
	4-40	26.89	51.27	51.27	6.46	6.88
	4-50	34.27	61.00	56.89	11.65	11.99
	4-60	44.03	78.67	78.67	13.55	13.55
	4-70	57.01	78.01	78.01	13.75	14.12
	4-80	75.08	82.48	82.48	16.87	17.16
	4-90	88.26	76.89	76.89	15.61	16.19
	Mean	**42.88**	**53.21**	**52.70**	**11.13**	**11.78**

6.2. Computational results with Branch-and-cut

As previously observed, our method increases the time that is spent at the root node of the branch-and-bound tree. Thus, to resolve the issue of whether or not it is computationally beneficial to apply the proposed method, we now compare the variants BDD_3 and BDD_4 to both the BD and LDD methods by running the algorithms for a time limit of 10 hours (where at most 5 hours are dedicated to solve the LP at the root node). These numerical results are summarized in Table 4 where, in addition to the average total running times and gaps obtained upon completion, the number of solved instances within an optimality gap of 1% (i.e., column #Sol.) are also reported.

From Table 4, we observe that our algorithms reach much better optimality gaps and solve more instances in noticeably shorter CPU times. Superiority of the proposed method compared to the classical BD algorithm is without exception. These results are explained by the fact that our method generates considerably smaller branch-and-bound trees even though a larger amount of time is spent at the root node. Moreover, the LDD method (due to its slow progression) fails to reach to the same solution quality than both BDD_3 and BDD_4 obtain at the root node (see Tables 1 and 3).

When comparing the two proposed variants, one observes that BDD_3 performs better than BDD_4 when solving the ND and FL instances. These results can be explained by two main reasons. First, when the BDD_3 variant is applied, stronger cuts are generated and, due to the integrality requirements being imposed on all the copied variables in the subproblems, a more extensive pool of integer solutions is obtained. Therefore, tighter lower and upper bounds are generated at the root node (see Tables 1 and 3). The second reason explaining these results is that, due to the proposed variable fixing strategy, the time requirements to solve a subproblem in both BDD_3 and BDD_4 are quite comparable as a consequence of fixing the indicator variables. Finally, we observe that BDD_4 outperforms BDD_3 on the SNI instances. However, the explanation for this is that the variable fixing strategy, in this case, cannot be applied due to the presence of a budget constraint. Therefore, when solving the SNI instances, there is an added time requirement that is needed to solve the subproblems when applying the BDD_3 variant, when compared to BDD_4.

Table 4 Comparing the Proposed Decomposition Method to the Classical Primal and Dual Decomposition Methods

Prob.	Inst.	Benders decomposition			Dual decomposition			BDD_3			BDD_4		
		Time(s.)	Gap(%)	#Sol.	Time(s.)	Gap(%)	#Sol.	Time(s.)	Gap(%)	#Sol.	Time(s.)	Gap(%)	#Sol.
FMCND	r04	6045.47	1.08	5/5	10946.13	2.84	4/5	1233.76	0.88	5/5	952.08	1.06	5/5
	r05	21836.23	8.40	2/5	22376.22	1.90	2/5	1994.71	0.81	5/5	2145.48	0.58	5/5
	r06	29114.44	11.27	1/5	24045.43	8.30	2/5	17156.79	4.04	3/5	16498.10	3.03	3/5
	r07	25743.46	6.03	2/5	22853.89	5.54	3/5	4956.88	0.96	5/5	5350.06	0.98	5/5
	r08	36315.18	12.20	1/5	23755.06	7.59	2/5	9188.43	1.57	4/5	8840.45	1.33	4/5
	r09	30972.97	14.54	2/5	30686.53	13.81	1/5	23780.16	6.69	3/5	21000.60	5.59	3/5
	r10	31445.58	15.73	1/5	33104.18	16.96	1/5	33175.08	11.23	2/5	34025.04	11.10	1/5
	Mean	**25924.76**	**9.89**	**14/35**	**23966.78**	**8.14**	**15/35**	**13069.40**	**3.74**	**27/35**	**12687.40**	**3.38**	**26/35**
CFL-S	101-104	36796.25	14.62	0/4	5292.66	1.05	4/4	1100.41	0.75	4/4	645.68	0.64	4/4
	111-114	36254.33	8.20	0/4	28951.78	2.27	1/4	2489.96	0.94	4/4	2739.68	0.99	4/4
	121-124	36227.73	17.66	0/4	18430.40	1.08	4/4	4278.08	1.04	4/4	4799.71	1.04	4/4
	131-134	36294.90	19.87	0/4	31224.45	1.19	2/4	18803.08	1.10	3/4	14031.66	1.26	3/4
	Mean	**36393.30**	**15.09**	**0/16**	**20974.82**	**1.40**	**11/16**	**6667.88**	**0.96**	**15/16**	**5554.18**	**0.98**	**15/16**
SNI	3-30	2762.35	1.10	4/5	36471.92	14.67	0/5	2121.51	0.97	5/5	1936.05	1.07	5/5
	3-40	30982.14	1.80	0/5	36534.46	18.95	0/5	8639.26	1.10	4/5	9424.96	1.10	4/5
	3-50	27298.87	2.47	2/5	36381.64	23.25	0/5	6535.48	1.10	4/5	7963.64	1.06	4/5
	3-60	29300.58	2.66	0/5	36466.10	24.93	0/5	4693.26	1.10	4/5	4795.40	1.03	5/5
	3-70	32136.92	4.54	0/5	36460.78	29.92	0/5	3896.33	1.06	4/5	4493.22	1.04	5/5
	3-80	33614.04	5.32	1/5	36620.18	34.55	0/5	8982.22	1.10	2/5	19887.93	1.14	2/5
	3-90	36669.58	8.06	0/5	36496.58	39.25	0/5	16600.52	1.52	3/5	18786.01	1.54	3/5
	4-30	2253.55	1.10	3/5	36389.18	15.82	0/5	3060.10	1.10	4/5	2219.62	1.06	5/5
	4-40	12231.08	1.10	1/5	36441.30	18.89	0/5	6582.58	1.10	2/5	6281.24	1.10	4/5
	4-50	20818.73	2.16	2/5	36291.34	24.05	0/5	7417.27	0.69	5/5	7212.67	1.10	4/5
	4-60	21378.10	2.40	1/5	36459.86	28.04	0/5	10257.96	1.10	2/5	9263.12	0.90	4/5
	4-70	22587.84	1.72	3/5	36493.82	32.99	0/5	10466.20	0.89	3/5	6048.30	1.10	5/5
	4-80	25946.64	1.28	2/5	36648.72	38.75	0/5	4081.00	1.06	5/5	4607.59	0.77	4/5
	4-90	21177.38	1.74	1/5	36321.86	41.89	0/5	7831.69	1.10	5/5	14077.00	2.06	3/5
	Mean	**22796.98**	**2.67**	**20/70**	**36462.70**	**27.57**	**0/70**	**7226.10**	**1.07**	**52/70**	**8356.91**	**1.15**	**57/70**

6.3. Comparison with a state-of-the-art optimization solver

In this last section, we will assess how our proposed method (i.e., specifically the BDD_3 and BDD_4 variants) compares with a state-of-the-art optimization solver (i.e., CPLEX 12.7). To do so, we add a family of classical inequalities to the subproblems of the FMCND and CFL-S instances to provide comparable results to CPLEX which extensively exploits the special structures of these classical problems. These inequalities, which are redundant in the original models but may help to strengthen the relaxations, take the form of $x_a^k \leq \min\{d^k, u_a\} y_a$, for the arcs a in the network. They state that the amount of flow on each connecting arc a for each customer k has to be lower than the minimum between the customer's demand d_k and the arc's capacity u_a. It should be noticed that the use of these inequalities does not yield any complication in our implementations given the presence of both the integer and continuous variables in the subproblems. Moreover, they do not require a separation and lifting procedure. However, these inequalities are not added to the extensive formulation, there is an exponential number of them and their inclusion noticeably slows down the CPLEX solver. The results for this numerical comparison are provided in Table 5. We again report the average running times, optimality gaps and the number of solved instances by each method.

Before analyzing the results reported in Table 5, a general comparison between these results and those provided in Table 4 shows that a simple and straightforward use of classical valid inequalities noticeably improves the performance of our method. The overall insight that can be gained from this comparison is that the bulk of the literature on the acceleration techniques that have been developed for the BD method (Rahmaniani et al. 2017a) can be applied here also and one can expect that this would noticeably enhance the numerical performance of the proposed method.

As for the specific results provided in Table 5, one first observes that the only cases where CPLEX performs better are when it is used to solve the $r04$ and $r07$ instances (i.e., CPLEX solves these instances in less than 10 minutes). For the rest of the instances, CPLEX is not competitive with either the BDD_3 and BDD_4 variants, both of which optimally solve a larger subset of the instances and in much shorter computation times. Moreover, CPLEX fails to solve any of the SNI instances after 10 hours of computational effort, while the two variants of our proposed method solve more that 74% of these instances in approximately 2 hours. Finally, when directly comparing the two developed variants, one finds that BDD_4 generally outperforms BDD_3 (in terms of the average computation times, gaps obtained and number of instances solved). Overall, this is explained by the fact that the BDD_4 variant is able to solve the root node faster, albeit while

Table 5 Comparing the Proposed Method to the State-of-the-Art Optimization Solver

Prob.	Inst.	CPLEX			BDD_3			BDD_4		
		Time(s.)	Gap(%)	#Sol.	Time(s.)	Gap(%)	#Sol.	Time(s.)	Gap(%)	#Sol.
FMCND	r04	258.10	0.57	5/5	410.57	0.44	5/5	385.64	0.71	5/5
	r05	1718.28	0.72	5/5	1485.92	0.56	5/5	1168.67	0.56	5/5
	r06	16076.07	2.92	3/5	12803.33	0.90	5/5	4897.01	0.61	5/5
	r07	699.17	0.89	5/5	1384.93	0.69	5/5	1231.60	0.76	5/5
	r08	8529.26	1.72	4/5	7430.11	0.65	5/5	7392.03	0.66	5/5
	r09	21741.47	8.28	3/5	17370.16	1.67	4/5	11961.41	1.57	4/5
	r10	28973.48	11.52	1/5	28950.08	8.65	1/5	28908.78	6.72	1/5
	Mean	**11142.26**	**3.80**	**26/35**	**9976.44**	**1.94**	**30/35**	**7992.16**	**1.66**	**30/35**
CFL-S	101-104	50.09	0.00	4/4	25.14	0.00	4/4	24.15	0.00	4/4
	111-114	1252.44	0.05	4/4	647.24	0.01	4/4	421.70	0.01	4/4
	121-124	2345.20	0.15	4/4	640.15	0.00	4/4	267.45	0.00	4/4
	131-134	1399.11	0.08	4/4	196.53	0.00	4/4	130.22	0.00	4/4
	Mean	**1261.71**	**0.07**	**16/16**	**377.27**	**0.00**	**16/16**	**210.88**	**0.00**	**16/16**
SNI	30	36225.72	7.81	0/5	2121.51	0.97	5/5	1936.05	1.07	5/5
	40	36168.22	15.60	0/5	8639.26	1.10	4/5	9424.96	1.10	4/5
	50	36163.76	19.33	0/5	6535.48	1.10	4/5	7963.64	1.06	4/5
	60	36166.40	19.22	0/5	4693.26	1.10	4/5	4795.40	1.03	5/5
	70	36155.20	17.61	0/5	3896.33	1.06	4/5	4493.22	1.04	5/5
	80	36155.84	20.14	0/5	8982.22	1.10	2/5	19887.93	1.14	2/5
	90	36150.18	24.25	0/5	16600.52	1.52	3/5	18786.01	1.54	3/5
	30	36183.46	12.00	0/5	3060.10	1.10	4/5	2219.62	1.06	5/5
	40	36147.56	17.94	0/5	6582.58	1.10	2/5	6281.24	1.10	4/5
	50	36141.76	22.39	0/5	7417.27	0.69	5/5	7212.67	1.10	4/5
	60	36136.64	24.92	0/5	10257.96	1.10	2/5	9263.12	0.90	4/5
	70	36138.64	42.29	0/5	10466.20	0.89	3/5	6048.30	1.10	5/5
	80	36129.26	29.25	0/5	4081.00	1.06	5/5	4607.59	0.77	4/5
	90	36133.82	60.29	0/5	7831.69	1.10	5/5	14077.00	2.06	3/5
	Mean	**36156.89**	**23.79**	**0/70**	**7226.10**	**1.07**	**52/70**	**8356.91**	**1.15**	**57/70**

generating weaker cuts when compared to BDD_3. However, the inclusion of the valid inequalities mitigates the latter disadvantage of BDD_4 by reducing the overall differences observed between the lower bounds provided at the root node by the two variants. In turn, when the branch-and-bound process begins, the time reductions at the root node obtained via BDD_4 become the determining factor in providing a more efficient solution process for our proposed method.

7. Conclusions

In this paper, we have proposed a decomposition method that combines the complementary advantages of the classical Benders and Lagrangian dual decomposition approaches. It generates cuts that dominate the classical optimality and feasibility cuts at fractional points of the master problem. The proposed method also tightens the LP relaxation of the MP problem, which we have shown to be at least as tight as the best bound obtained by the Lagrangian decomposition method. Another important feature of our method lies in its enhanced capabilities to find high quality incumbent solutions at early iterations of the algorithm.

We have applied the proposed method to solve a wide range of hard combinatorial optimization instances. It was observed that, for a reasonable time limit, the proposed method was able to reach much tighter root node bounds when compared to the Lagrangian dual decomposition approach. Our algorithm was also capable of finding incumbent solutions very close to the optimal values at the early iterations of the search process. Furthermore, it was numerically shown that the developed method increases the time spent on the root nodes when compared to the classical Benders decomposition method. However, this added effort, which produces significantly tighter bounds at the root nodes, is largely compensated by the fact that the improved bounds then enable the algorithm to generate smaller branch-and-bound trees to solve the instances. Finally, we observed that our method also outperforms a state-of-the-art optimization solvers.

Going forward, there are many avenues of research to further improve the proposed algorithm. When applied on stochastic models, one could first take advantage of parallel computing to solve the subproblems. Furthermore, as originally presented, when our algorithm is used to solve a stochastic model, each subproblem is defined using a single scenario. However, it has been previously shown that having subproblems defined on clusters of scenarios yields both tighter lower bounds and better incumbent solutions. Thus, exploring how different clustering strategies applied on scenarios would impact the overall performance of our proposed method is definitely an interesting line of inquiry. In addition, there have been numerous studies dedicated to the improvement of both the Benders and Lagrangian decomposition methods. Investigating which of these strategies could be incorporated in our method to further enhance its numerical performance would be of interest. For example, we made use of a simple cutting plane method to update the Lagrangian multipliers. In this regard, more effective strategies could be employed to significantly improve the proposed algorithm. Last but not least, the proposed method guarantees convergence even when the Benders subproblems are nonlinear mixed-integer problems. Thus, numerically assessing how the proposed method would fare against classical algorithms, such as the integer L-shaped and outer approximation methods, is certainly worthwhile.

Acknowledgments

Partial funding for this project has been provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery Grant program and by the Fonds de recherche du Québec through its Team Grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their strategic infrastructure grants. S. Ahmed was supported in part by the National Science Foundation Grant #1633196 and the Office of Naval Research Grants N00014-15-1-2078 and N00014-18-1-2075.

References

Adulyasak Y, Cordeau JF, Jans R (2015) Benders decomposition for production routing under demand uncertainty. *Operations Research* 63(4):851–867.

Ahmed S (2013) A scenario decomposition algorithm for 0–1 stochastic programs. *Operations Research Letters* 41(6):565 – 569.

Beale EML (1965) Survey of integer programming. *Journal of the Operational Research Society* 16(2):219–228.

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik* 4(1):238–252.

Birge JR, Louveaux F (2011) *Introduction to stochastic programming* (Springer Science & Business Media).

Bodur M, Dash S, Günlük O, Luedtke J (2017) Strengthened Benders cuts for stochastic integer programs with continuous recourse. *INFORMS Journal on Computing* 29(1):77–91.

Boland N, Fischetti M, Monaci M, Savelsbergh M (2016) Proximity Benders: a decomposition heuristic for stochastic programs. *Journal of Heuristics* 22(2):181–198.

Cerisola S, Álvaro Baíllo, Fernández-López JM, Ramos A, Gollmer R (2009) Stochastic power generation unit commitment in electricity markets: A novel formulation and a comparison of solution methods. *Operations Research* 57(1):32–46.

Codato G, Fischetti M (2006) Combinatorial Benders' cuts for mixed-integer linear programming. *Operations Research* 54(4):756–766.

Contreras I, Cordeau JF, Laporte G (2011) Benders decomposition for large-scale uncapacitated hub location. *Operations Research* 59(6):1477–1490.

Cordeau JF, Pasin F, Solomon MM (2006) An integrated model for logistics network design. *Annals of Operations Research* 144(1):59–82.

Costa AM (2005) A survey on Benders decomposition applied to fixed-charge network design problems. *Computers & Operations Research* 32(6):1429 – 1450.

Costa AM, Cordeau JF, Gendron B, Laporte G (2012) Accelerating Benders decomposition with heuristic master problem solutions. *Pesquisa Operacional* 32(1):03–20.

Crainic TG, Frangioni A, Gendron B (2001) Bundle-based relaxation methods for multicommodity capacitated fixed charge network design. *Discrete Applied Mathematics* 112(1):73 – 99.

Crainic TG, Fu X, Gendreau M, Rei W, Wallace SW (2011) Progressive hedging-based metaheuristics for stochastic network design. *Networks* 58(2):114–124.

Crainic TG, Hewitt M, Rei W (2016) Partial Benders decomposition strategies for two-stage stochastic integer programs. Publication CIRRELT-2016-37, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal.

Fischetti M, Ljubić I, Sinnl M (2016) Benders decomposition without separability: A computational study for capacitated facility location problems. *European Journal of Operational Research* 253(3):557 – 569.

Fischetti M, Salvagnin D, Zanette A (2010) A note on the selection of Benders' cuts. *Mathematical Programming* 124(1):175–182.

Flippo OE, Rinnooy Kan AHG (1993) Decomposition in general mathematical programming. *Mathematical Programming* 60(1):361–382.

Fragkogios A, Saharidis GK (2018) Latest advances on Benders decomposition. *Encyclopedia of Information Science and Technology, Fourth Edition*, 5411–5421 (IGI Global).

Gendron B, Scutellà MG, Garroppo RG, Nencioni G, Tavanti L (2016) A Branch-and-Benders-cut method for nonlinear power design in green wireless local area networks. *European Journal of Operational Research* 255(1):151–162.

Geoffrion AM (1972) Generalized benders decomposition. *Journal of Optimization Theory and Applications* 10(4):237–260.

Geoffrion AM, Graves GW (1974) Multicommodity distribution system design by benders decomposition. *Management Science* 20(5):822–844.

Grothey A, Leyffer S, McKinnon K (1999) A note on feasibility in Benders decomposition. *Numerical Analysis Report NA/188, Dundee University* .

Holmberg K (1990) On the convergence of cross decomposition. *Mathematical Programming* 47(1):269–296.

Jünger M, Liebling TM, Naddef D, Nemhauser GL, Pulleyblank WR, Reinelt G, Rinaldi G, Wolsey LA (2009) *50 years of integer programming 1958-2008: From the early years to the state-of-the-art* (Springer Science & Business Media).

Klibi W, Martel A (2012) Scenario-based supply chain network risk modeling. *European Journal of Operational Research* 223(3):644 – 658.

Lin S, Lim GJ, Bard JF (2016) Benders decomposition and an IP-based heuristic for selecting IMRT treatment beam angles. *European Journal of Operational Research* 251(3):715 – 726.

Louveaux FV (1986) Discrete stochastic location models. *Annals of Operations Research* 6(2):21–34.

Magnanti TL, Wong RT (1981) Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research* 29(3):464–484.

McDaniel D, Devine M (1977) A modified Benders' partitioning algorithm for mixed integer programming. *Management Science* 24(3):312–319.

Mitra S, Garcia-Herreros P, Grossmann IE (2016) A cross-decomposition scheme with integrated primal–dual multi-cuts for two-stage stochastic programming investment planning problems. *Mathematical Programming* 157(1):95–119.

Naoum-Sawaya J, Elhedhli S (2013) An interior-point benders based branch-and-cut algorithm for mixed integer programs. *Annals of Operations Research* 210(1):33–55.

Nemhauser GL, Wolsey LA (1988) *Integer and Combinatorial Optimization* (New York, NY, USA: Wiley-Interscience), ISBN 0-471-82819-X.

Newman AM, Weiss M (2013) A survey of linear and mixed-integer optimization tutorials. *INFORMS Transactions on Education* 14(1):26–38.

Pacqueau R, Francois S, Le Nguyen H (2012) A fast and accurate algorithm for stochastic integer programming, appllied to stochastic shift scheduling. Publication G-2012-29, Groupe d'études et de recherche en analyse des décisions (GERAD), Université de Montréal, Montréal, QC, Canada.

Pan F, Morton DP (2008) Minimizing a stochastic maximum-reliability path. *Networks* 52(3):111–119.

Papadakos N (2008) Practical enhancements to the magnanti–wong method. *Operations Research Letters* 36(4):444 – 449.

Poojari CA, Beasley JE (2009) Improving Benders decomposition using a genetic algorithm. *European Journal of Operational Research* 199(1):89–97.

Rahmaniani R, Crainic TG, Gendreau M, Rei W (2017a) The Benders decomposition algorithm: A literature review. *European Journal of Operational Research* 259(3):801 – 817.

Rahmaniani R, Crainic TG, Gendreau M, Rei W (2017b) A Benders decomposition method for two-stage stochastic network design problems. Publication CIRRELT-2017-22, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal.

Rei W, Cordeau JF, Gendreau M, Soriano P (2009) Accelerating Benders decomposition by local branching. *INFORMS Journal on Computing* 21(2):333–345.

Rush AM, Collins M (2012) A tutorial on dual decomposition and lagrangian relaxation for inference in natural language processing. *Journal of Artificial Intelligence Research* 45(1):305–362.

Ruszczyński A (1997) Decomposition methods in stochastic programming. *Mathematical Programming* 79(1):333–353.

Sahinidis N, Grossmann I (1991) Convergence properties of generalized Benders decomposition. *Computers & Chemical Engineering* 15(7):481 – 491.

Santoso T, Ahmed S, Goetschalckx M, Shapiro A (2005) A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research* 167(1):96 – 115.

Sherali HD, Lunday BJ (2013) On generating maximal nondominated Benders cuts. *Annals of Operations Research* 210(1):57–72.

Van Roy TJ (1983) Cross decomposition for mixed integer programming. *Mathematical Programming* 25(1):46–63.

Wentges P (1996) Accelerating Benders' decomposition for the capacitated facility location problem. *Mathematical Methods of Operations Research* 44(2):267–290.

Zaourar S, Malick J (2014) Quadratic stabilization of Benders decomposition, URL `https://hal.archives-ouvertes.fr/hal-01181273`.

Zou J, Ahmed S, Sun XA (2017) Stochastic dual dynamic integer programming URL `http://www.optimization-online.org/DB\_HTML/2016/05/5436.html`.

Appendix A: Proof of the Theorems and Propositions

A.1. Proof of Proposition 1

To show the validity of the cut we need to demonstrate that the above inequality holds true for any $y \in F \bigcap \mathbb{Z}_+^n$, i.e., it does not cut off any feasible integer solution. We first assume that y^* is an integer feasible solution, i.e., $y^* \in F \bigcap \mathbb{Z}_+^n$. Accordingly, $\theta \geq \min_{(x,z)\in X}\{c^\top x : z = y^*\} = \min_{(x,z)\in X}\{c^\top x : z = y^*,\ z \in \mathbb{Z}_+^n\} \geq \max_{\lambda\in\mathbb{R}^n}\left\{\lambda^\top y^* + \min_{(x,z)\in X}\{c^\top x - \lambda^\top z : z \in \mathbb{Z}_+^n\}\right\}$ for any $y^* \in F \bigcap \mathbb{Z}_+^n \subseteq F$. We thus deduce that $\theta \geq \lambda^\top y + Q(\lambda)$ for any $\lambda \in \mathbb{R}^n$ is a valid optimality cut at any feasible integer solution, where $Q(\lambda) = \min_{(x,z)\in X}\{c^\top x - \lambda^\top z : z \in \mathbb{Z}_+^n\}$. Thus, if y^* is fractional and $y^* \in F$, the validity of (2.7) follows from setting λ to an arbitrary maximizer of $\max_{\lambda\in\mathbb{R}^n}\left\{\lambda^\top y^* + Q(\lambda)\right\}$.

A.2. Proof of Theorem 1

For a given $y^* \in F$ and $\hat{\lambda}$, we have $\theta \geq \max\limits_{\lambda\in\mathbb{R}^n}\left\{\lambda^\top y^* + \min\limits_{(x,z)\in X: z\in\mathbb{Z}_+^n}\{c^\top x - \lambda^\top z\}\right\} \geq \hat{\lambda}^\top y^* + \min\limits_{(x,z)\in X: z\in\mathbb{Z}_+^n}\{c^\top x - \hat{\lambda}^\top z\} \geq \hat{\lambda}^\top y^* + \min\limits_{(x,z)\in X}\{c^\top x - \hat{\lambda}^\top z\}$. This gives following two optimality cuts: (i) $\theta \geq \hat{\lambda}^\top y + \min\limits_{(x,z)\in X: z\in\mathbb{Z}_+^n}\{c^\top x - \hat{\lambda}^\top z\}$ and (ii) $\theta \geq \hat{\lambda}^\top y + \min\limits_{(x,z)\in X}\{c^\top x - \hat{\lambda}^\top z\}$. We observe that the optimality cut (i) is equal to fixing λ to $\hat{\lambda}$ in (2.6) to generate cut (2.7). While the optimality cut (ii) is equal to the cut that we extract from (2.5) by setting $\lambda = \hat{\lambda}$. The cuts (i) and (ii) are in parallel since they have the same slope λ. However, the constant part in the former is larger by an amount equal to the LP gap of the minimization problem which is equal to σ in (2.8).

A.3. Proof of Proposition 2

We first recall that following two formulations are equal

$$\min_{(x,z,v)\in\mathbb{R}_+^m\times\mathbb{R}_+^n\times\mathbb{R}_+^l}\{\mathbb{1}^\top v:\ Bz \geq b,\ Wx + Tz + v \geq h,\ z = y^*\} = \tag{4.5}$$

$$\min_{(x,z,v,u)\in\mathbb{R}_+^m\times\mathbb{R}_+^n\times\mathbb{R}_+^l\times\mathbb{R}_+^k}\{\mathbb{1}^\top v + \mathbb{1}^\top u:\ Bz + u \geq b,\ Wx + Tz + v \geq h,\ z = y^*\} \tag{4.6}$$

given that $u=0$ since $Bz \geq b$ is satisfied through $z=y^*$ and y^* is the master solution. Thus, following the same steps as in Proposition (1), we directly conclude that (2.10) provides a valid cut for any $\beta \in \mathbb{R}^n$. To show that the cut is violated, notice that $\mathbb{1}^\top \bar{v} > 0$ in (4.5) given that y^* is an infeasible solution for which there is no $x \in \mathbb{R}^n_+$ such that $Wx \geq h - Ty^*$. Thus,

$$0 < \min_{(x,z,v,u)\in\mathbb{R}^m_+\times\mathbb{R}^n_+\times\mathbb{R}^l_+\times\mathbb{R}^k_+} \{\mathbb{1}^\top v + \mathbb{1}^\top u:\ Bz+u \geq b,\ Wx+Tz+v \geq h,\ z=y^*\} =$$

$$\max_{\beta\in\mathbb{R}^n} \min_{(x,z,v,u)\in\mathbb{R}^m_+\times\mathbb{R}^n_+\times\mathbb{R}^l_+\times\mathbb{R}^k_+} \{\mathbb{1}^\top v + \mathbb{1}^\top u - \beta^\top(z-y^*):\ Bz+u \geq b,\ Wx+Tz+v \geq h,\ z=y^*\} \leq$$

$$\max_{\beta\in\mathbb{R}^n} \min_{(x,z,v,u)\in\mathbb{R}^m_+\times\mathbb{Z}^n_+\times\mathbb{R}^l_+\times\mathbb{R}^k_+} \{\mathbb{1}^\top v + \mathbb{1}^\top u - \beta^\top(z-y^*):\ Bz+u \geq b,\ Wx+Tz+v \geq h,\ z=y^*\}$$

A.4. Proof of Theorem 2

First, recall that due to the Theorem 6.2 in Nemhauser and Wolsey (1988)

$$\max_{\lambda\in\mathbb{R}^n} \left\{\lambda^\top y^* + \min_{(x,z)\in X}\{c^\top x - \lambda^\top z:\ z\in\mathbb{Z}^n_+\}\right\} \equiv \max_{\lambda\in\mathbb{R}^n} \left\{\lambda^\top y^* + \min_{(x,z)\in conv(\tilde{X})}\{c^\top x - \lambda^\top z\}\right\}$$

and

$$\max_{\beta\in\mathbb{R}^n} \left\{\beta^\top y^* + \min_{(x,z',v,u)\in H}\{\mathbb{1}^\top v + \mathbb{1}^\top u - \beta^\top z' : z'\in\mathbb{Z}^n_+\ : z'\in\mathbb{Z}^n_+\}\right\} \equiv \max_{\beta\in\mathbb{R}^n} \left\{\beta^\top y^* + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v + \mathbb{1}^\top u - \beta^\top z'\}\right\}$$

where $\tilde{X} = X \cap \{z\in\mathbb{Z}^n_+\}$, $\tilde{H} = H\cap\{z'\in\mathbb{Z}^n_+\}$ and $conv(\Theta)$ represents the convex hull of a given set Θ. Notice that we use z' to represent the copied variables in the feasibility problem in order to distinguish them from the copied variables that are included in the optimality problem. We thus have the following reformulation of the LP relaxation of the MP,

$$\min_{\theta,y\in Y}\left\{f^\top y+\theta:\ \theta \geq \max_{\lambda\in\mathbb{R}^n}\left\{\lambda^\top y + \min_{(x,z)\in conv(\tilde{X})}\{c^\top x-\lambda^\top z\}\right\},\ 0 \geq \max_{\beta\in\mathbb{R}^n}\left\{\beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\right\}\right\} =$$

$$\min_{\theta,y\in Y}\left\{f^\top y+\theta:\ \theta \geq \lambda^\top y + \min_{(x,z)\in conv(\tilde{X})}\{c^\top x-\lambda^\top z\}\ \forall\lambda\in E_o,\ 0 \geq \beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\right\}$$

where E_o and E_r are the set of extreme points of the Lagrangian dual programs associated to the optimality and feasibility cuts. Accordingly, we have the following master formulation

$$\min_{\theta,y\in Y}\left\{f^\top y+\theta:\ \theta \geq \lambda^\top y + \min_{(x,z)\in conv(\tilde{X})}\{c^\top x-\lambda^\top z\}\ \forall\lambda\in E_o,\ 0 \geq \beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\right\} =$$

$$\min_{y\in Y}\left\{f^\top y+\max_{\lambda\in E_o}\left\{\lambda^\top y + \min_{(x,z)\in conv(\tilde{X})}\{c^\top x-\lambda^\top z\}\right\}:\ 0 \geq \beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\right\} =$$

$$\min_{y\in Y}\left\{\max_{\lambda\in E_o}\min_{(x,z)\in conv(\tilde{X})}\{f^\top y + c^\top x+\lambda^\top(y-z)\}:\ 0 \geq \beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\right\}$$

It should be noted that for any function $g: U\times V \to \mathbb{R}$, the following inequality between max-min and min-max operators is always satisfied $\sup_{u\in U}\inf_{v\in V} g(u;v) \leq \inf_{v\in V}\sup_{u\in U} g(z;v)$. Thus,

$$\min_{y\in Y}\left\{\max_{\lambda\in E_o}\min_{(x,z)\in conv(\tilde{X})}\{f^\top y + c^\top x+\lambda^\top(y-z)\}:\ 0 \geq \beta^\top y + \min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\right\} \geq$$

$$\max_{\lambda\in E_o}\min_{y\in Y}\Big\{\min_{(x,z)\in conv(\tilde{X})}\{f^\top y+c^\top x+\lambda^\top(y-z)\}:\ 0\geq\beta^\top y+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\Big\}=$$

$$\max_{\lambda\in E_o}\min_{y\in Y,(x,z)\in conv(\tilde{X})}\Big\{f^\top y+c^\top x+\lambda^\top(y-z):\ 0\geq\beta^\top y+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\Big\}\geq$$

$$\max_{\lambda\in E_o}\min_{y\in\mathbb{R}^n_+,(x,z)\in conv(\tilde{X})}\Big\{f^\top y+c^\top x+\lambda^\top(y-z):\ 0\geq\beta^\top y+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\Big\}$$

The last inequality is true given that we have removed the constraint set $By\geq b$. One should notice that the last max-min problem is the Lagrangian dual associated with

$$\min_{y\in\mathbb{R}^n_+,(x,z)\in conv(\tilde{X})}\{f^\top y+c^\top x:\ y=z,\ 0\geq\beta^\top y+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\},$$

where $y=z$ is priced into the objective function. This problem is equivalent to

$$\min_{(x,z)\in conv(\tilde{X})}\{f^\top z+c^\top x:\ 0\geq\beta^\top z+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\}.$$

Recalling that $conv(\tilde{X})=X\cap\{z\in\mathbb{Z}^n_+\}$, by expanding $conv(\tilde{X})$, we rewrite the previous minimization problem as

$$\min_{(x,z)\in\mathbb{R}^m_+\times\mathbb{Z}^n_+}\{f^\top z+c^\top x:\ Bz\geq b,\ Wx+Tz\geq h,\ 0\geq\beta^\top z+\min_{(x,z',v,u)\in conv(\tilde{H})}\{\mathbb{1}^\top v+\mathbb{1}^\top u-\beta^\top z'\}\ \forall\beta\in E_r\}.$$

In the above formulation the feasibility cuts become redundant since any z is guaranteed to satisfy $Wx+Tz\geq h$. We thus obtain

$$\min_{(x,z)\in\mathbb{R}^m_+\times\mathbb{Z}^n_+}\{f^\top z+c^\top x:\ Bz\geq b,\ Wx+Tz\geq h\}$$

which is equivalent to the original problem (1.1).

A.5. Proof of Proposition 3

We have $\theta\geq\max_{\lambda\in\mathbb{R}^n}\min_{(x,z)\in X}\{c^\top x-\lambda^\top(z-y^*)\ :\ z\in\mathbb{Z}^n_+\}\geq\max_{\lambda\in\mathbb{R}^n}\min_{(x,z)\in X}\{c^\top x-\lambda^\top(z-y^*)\ :\ z_I\in\mathbb{Z}^{|I|}_+\}\geq\max_{\lambda\in\mathbb{R}^n}\big\{\min_{(x,z)\in X}\{c^\top x-\lambda^\top(z-y^*)\ :\ z_I\in\mathbb{Z}^{|I|}_+\}\ :\ \lambda_{I'}=\hat{\lambda}_{I'}\big\}=\max_{\lambda\in\mathbb{R}^n}\big\{\min_{(x,z)\in X:z_I\in\mathbb{Z}^{|I|}_+}\{c^\top x-\lambda^\top(z-y^*)\}\ :\ \lambda_{I'}=\hat{\lambda}_{I'}\big\}\geq\max_{\lambda\in\mathbb{R}^n}\big\{\min_{(x,z)\in X:z_I\in\mathbb{Z}^{|I|}_+}\{c^\top x-\lambda^\top(z-y^*):z_{I'}=ub_{I'}\}:\lambda_{I'}=\hat{\lambda}_{I'}\big\}$. Note that the last inequality is true due to our assumption and the condition imposed on $\hat{\lambda}_{I'}$.

A.6. Proof of Proposition 4

Let us assume that $(\bar{x},\bar{z})$ is the actual optimal solution of the subproblem and let $Q(\bar{x},\bar{z})$ and $Q(\hat{x}(\varepsilon),\hat{z}(\varepsilon))$ be the objective value of the minimization problem for the optimal and ε-optimal solutions, respectively. Since the solution $(\bar{x}(\varepsilon),\bar{z}(\varepsilon))$ is ε-optimal, we have $Q(\hat{x}(\varepsilon),\hat{z}(\varepsilon))-Q(\bar{x},\bar{z})\leq\varepsilon$ or equivalently $Q(\bar{x},\bar{z})\geq Q(\hat{x}(\varepsilon),\hat{z}(\varepsilon))-\varepsilon$. We know that $\theta\geq Q(\bar{x},\bar{z})+y^\top\lambda^*$ for all $y\in Y$ and thus $\theta\geq Q(\hat{x}(\varepsilon),\hat{z}(\varepsilon))-\varepsilon+y^\top\lambda^*=c^\top\hat{x}(\varepsilon)+(y-\hat{z}(\varepsilon))^\top\lambda^*-\varepsilon$ for all $y\in Y$.

Appendix B: Lagrangian dual decomposition method

Without lose of generality, we rewrite problem (1.1) as follows

$$\min_{y,x,z}\{f^\top y + c^\top x :\ By \geq b,\ Wx + Tz \geq h,\ z = y,\ \ y \in \mathbb{Z}^n_+,\ z \in \mathbb{Z}^n_+,\ x \in \mathbb{R}^m_+\}$$

Pricing out the equality constraint $z = y$ into the objective function using dual multiplier $\lambda \in \mathbb{R}^n$ results into following Lagrangian dual problem

$$\max_{\lambda\in\mathbb{R}^n}\ \min_{y,x,z}\{f^\top y + c^\top x + \lambda^\top(y - z) :\ By \geq b,\ Wx + Tz \geq h,\ \ y \in \mathbb{Z}^n_+,\ z \in \mathbb{Z}^n_+,\ x \in \mathbb{R}^m_+\} =$$

$$\max_{\lambda\in\mathbb{R}^n}\ \min_{y,x,z}\{(f+\lambda)^\top y + c^\top x - \lambda^\top z :\ By \geq b,\ Wx + Tz \geq h,\ \ y \in \mathbb{Z}^n_+,\ z \in \mathbb{Z}^n_+,\ x \in \mathbb{R}^m_+\}$$

for which the minimization problem can be optimized separately over y and (z,x) variables.

Appendix C: Example

In this part, we provide the details of solving the toy example (3.1) when the strengthened Benders cuts and the Lagrangian cuts are used.

C.1. Strengthened Benders cuts

Considering the binary variable as the complicating one, we derive the following relaxed LP master problem,

$$\min_{y,\theta}\{\theta : 1 \geq y \geq 0\}, \tag{A.1}$$

and the following subproblem

$$\min_{z,x}\{x : x + 15z \geq 8,\ 3x + 10z \geq 13,\ x + 10z \geq 7,\ 2x - 10z \geq -1,\ 2x - 70z \geq -49,\ z = \bar{y},\ z \in [0,1]\} \tag{A.2}$$

where z is a copy of the y variable and $\bar{y}$ is the current master solution. Solving the master problem (A.1) yields $\bar{y} = 0$. This is an integer point and thus we only generate the (generalized) Benders cut by solving subproblem (A.2) for $\bar{y} = 0$. This gives $x = 8$ and $\lambda = -15$, where λ is the multiplier associated to the constraint $z - \bar{y} = 0$. We can thus generate an optimality cut and updates the master problem (A.1) as follows:

$$\min_{y,\theta}\{\theta :\ \theta \geq 8 - 15y,\ 1 \geq y \geq 0\} \tag{A.3}$$

In next iteration, solving the master problem (A.3) gives $y = 1$ and solving the subproblem (A.2) for $\bar{y} = 1$ yields $x = 10.5$ and $\lambda = 35$. Thus, we can generate a new optimality cut and update the master problem as follows:

$$\min_{y,\theta}\{\theta : \theta \geq 8 - 15y,\ \theta \geq -\frac{49}{2} + 35y,\ 1 \geq y \geq 0\} \tag{A.4}$$

Solving the above master problem we get a lower bound of -1.75 and $y = 0.65$. We get $\lambda = 5$ by solving subproblem (A.2) for $\bar{y} = 0.65$, resulting the Benders cut $\theta \geq -0.5 + 5y$. Since the master solution is fractional, we can generate the strengthened Benders cut by solving the following Lagrangian dual subproblem (obtained from relaxing $z = \bar{y}$ into objective function and imposing integrality requirement on the z variable):

$$\min_{z,x}\{x + 5(\bar{y} - z) : x + 15z \geq 8,\ 3x + 10z \geq 13,\ x + 10z \geq 7,\ 2x - 10z \geq -1,\ 2x - 70z \geq -49,\ z \in \{0,1\}\} \tag{A.5}$$

the optimal solution of the above problem is $z = 1$ with objective value of 5.5. This gives the strengthened Benders cuts $\theta \geq 5.5 + 5y$ which is 6 units tighter than the classical Benders at any master solution. Adding this cut to the master problem (A.4), we get

$$\min_{y,\theta}\{\theta : \theta \geq 8 - 15y,\ \theta \geq -\frac{49}{2} + 35y,\ \theta \geq 5.5 + 5y,\ 1 \geq y \geq 0\} \tag{A.6}$$

with $y = 0.125$ and the lower bound of 6.125. Executing the next iteration, we observe that this is the best solution that we can get for the LP relaxation of the master problem.

C.2. Lagrangian cuts

Note from section C.1 that the master problem generates integer solutions for the first two iterations. As we mentioned earlier, the classical Benders cuts are the tightest at the integer points. We thus start directly from following master problem to avoid reduplication of the results,

$$\min_{y,\theta}\{\theta : \theta \geq 8 - 15y,\ \theta \geq -\frac{49}{2} + 35y,\ 1 \geq y \geq 0\} \tag{A.7}$$

which gives the lower bound of -1.75 and $y = 0.65$. To generate the Lagrangian cut, we need to solve the following Lagrangian dual problem:

$$\max_{\lambda}\left\{\lambda\bar{y} + \min_{z,x}\left\{x - \lambda z : x + 15z \geq 8, 3x + 10z \geq 13, x + 10z \geq 7, 2x - 10z \geq -1, 2x - 70z \geq -49, z \in \{0,1\}\right\}\right\} \tag{A.8}$$

To solve (A.8), we use the subgradient method. To initiate the λ value, we solve problem (A.2) with $\bar{y} = 0.65$ which gives $\lambda = 5$. For this λ value, we solve the inner minimization problem of (A.8), i.e.,

$$\min_{z,x}\{x - 5z : x + 15z \geq 8,\ 3x + 10z \geq 13,\ x + 10z \geq 7,\ 2x - 10z \geq -1,\ 2x - 70z \geq -49,\ z \in \{0,1\}\}, \tag{A.9}$$

which gives $z = 1$ and $x = 5.5$. Notice that $z = 1$ is a feasible solution to the original problem and its associated cost is 10.5 units. We can thus update the upper bound at this step. Given the lower bound of -1.75, upper bound of 10.5, a step size of $\frac{10}{49}$, $\bar{y} = 0.65$, and $z = 1$, we update the dual multiplier as follows $\lambda = 5 - \frac{10}{49}\frac{(10.5+1.75)}{||1-0.65||_2^2}(1 - 0.65) = 2.5$. We solve again the Lagrangian subproblem (A.9) with the new λ multiplier, i.e.,

$$\min_{z,x}\left\{x - 2.5z : x + 15z \geq 8, 3x + 10z \geq 13, x + 10z \geq 7, 2x - 10z \geq -1, 2x - 70z \geq -49, z \in \{0,1\}\right\}. \tag{A.10}$$

This gives $x = 8$, $z = 0$, and an optimality cut of $\theta \geq 8 + 2.5y$. It is easy to observe that this cut gives the convex hull representation of the master problem. We thus terminate the subgradient method. Moreover, $z = 0$ is a new integer solution which is feasible to the original problem, we can thus update the upper bound value from 10.5 to 8.

Appendix D: The test problems

In this section, we give the mathematical formulation of our test problems.

D.1. Multicommodity Capacitated Network Design

This problem is defined on a directed graph with a set of potential arcs A and the node set N. The goal is to chose the most appropriate subset of the arcs such that each commodity k, from set K, can flow from its unique origin node $O(k) \in N$ to its unique destination node $D(k) \in N$, while the total cost is minimum. Three parameters are associated to each arc $a \in A$, i.e., fixed cost f_a, capacity u_a and flow cost c_a^k per unit of flow for commodity $k \in K$. Each commodity $k \in K$ is associated with a stochastic demand characterized by d_s^k for each observation of the uncertainty $s \in S$. To model this problem, we define binary variables y_a (equal to 1 if arc $a \in A$ is chosen and 0 otherwise) and continuous variables $x_s^{a,k}$ (to measure the amount of flow on arc $a \in A$ under observation $s \in S$ for commodity $k \in K$). Thus, the extensive formulation of this problem is

$$ND := \min_{y \in \{0,1\}^{|A|},\ x \in \mathbb{R}_+^{|A|\times|K|\times|S|}} \sum_{a \in A} f_a y_a + \sum_{s \in S} \sum_{a \in A} \sum_{k \in K} p_s c_a^k x_s^{a,k} \tag{B.1}$$

$$\text{s.t.} \sum_{a \in A^+(i)} x_s^{a,k} - \sum_{a \in A^-(i)} x_s^{a,k} = d_s^i \qquad \forall i \in N, k \in K, s \in S \tag{B.2}$$

$$\sum_{k \in K} d_s^k x_s^{a,k} \leq u_a y_a \qquad \forall a \in A, s \in S \tag{B.3}$$

where $A^+(i)/A^-(i)$ indicates the set of outward/inward arcs to node $i \in N$, p_s is probability of scenario s and d_s^i is equal to d_s^k $\left(-d_s^k\right)$ if $i = O(k)$ $\big(D(k)\big)$, and 0 otherwise.

D.2. Stochastic Capacitated Facility Location Problem

To define this problem, let N be the set of potential locations for the facilities and M the set of customers. The goal is to open sufficient facilities to satisfy all the demand at minimum cost. Each customer may be served by one or several facilities. At each potential location $i \in N$, at most one facility with service capacity of u_i can be opened which entails a fixed cost of f_i units. The routing cost from customer $j \in M$ to facility $i \in N$ per unit of flow is c_{ij}. Each customer $j \in M$ has a stochastic demand d_j^s, where $s \in S$ is a specific realization of the uncertainty with probability of p_s such that $\sum_{s \in S} p_s = 1$.

Let binary variable y_i take 1 if a facility is opened at location $i \in N$ and 0 otherwise, and let $x_{ij}^s \geq 0$ indicate the amount of flow from customer $j \in J$ to facility $i \in I$ under realization $s \in S$. We thus use following formulation of this problem

$$SFL := \min_{y \in \{0,1\}^{|N|},\, x \in \mathbb{R}_+^{|N|\times|M|\times|S|}} \sum_{i \in N} f_i y_i + \sum_{s \in S} \sum_{i \in N} \sum_{j \in M} p_s c_{ij} x_{ij}^s \tag{B.4}$$

$$\text{s.t.} \sum_{i \in N} x_{ij}^s \geq d_j^s \qquad \forall j \in M, s \in S \tag{B.5}$$

$$\sum_{j \in M} x_{ij}^s \leq u_i y_i \qquad \forall i \in N, s \in S \tag{B.6}$$

$$\sum_{i \in N} u_i y_i \geq \max_{s \in S} \sum_{j \in M} d_j^s \tag{B.7}$$

The objective function minimizes the total fixed costs of opening facility plus the expected flow cots. First constraint set imposes the demand satisfaction for each customer in every scenario. Second constraint set imposes the capacity restriction on each facility and the last constraint is to add the relatively complete recourse property to the problem.

D.3. Stochastic Network Interdiction

This problem is defined over a directed graph consisting a set of nodes N, arcs A, and a subset of candidate arcs $L \subseteq A$ on which sensors can be installed. The goal is to maximize the probability of catching an intruder that traverses some path in the network. The first-stage decisions determine whether or not to install a sensor on arc $a \in L$ knowing the probability of the intruder avoiding detection with and without a sensor on this arc (denoted by r_a and q_a). A scenario $s \in S$, with probability of p_s, corresponds to the origin O_s and destination D_s of the intruder. Cost of installing of a sensor on arc $a \in D$ is c_a units and there is a total budget of b units to install sensors. In the second-stage the intruder choses a maximum-reliability path from its origin to the targeted node that maximizes the probability of avoiding detection. The maximum-reliability path from node $j \in N$ to destination D_s is represented by ψ_s^j.

The first-stage binary variables y_a takes value of 1 if a sensor is installed on arc $a \in L$ and 0 otherwise. The second-stage variables π_s^i, $\forall i \in N, s \in S$ model the probability of reaching destination D_s from node i without being detected. The formulation of this problem is as follows:

$$SNI := \min_{y \in \{0,1\}^{|L|}, \pi \in \Re_+^{|N||S|}} \sum_{s \in S} p_s \pi_s^{O_s} \tag{B.8}$$

$$\text{s.t.} \sum_{a \in L} c_a y_a \leq b \tag{B.9}$$

$$\pi_s^{D_s} = 1 \qquad \forall s \in S \tag{B.10}$$

$$\pi_s^{a^+} - q_a\pi_s^{a^-} \geq 0 \qquad \forall a \in L, s \in S \tag{B.11}$$

$$\pi_s^{a^+} - r_a\pi_s^{a^-} \geq 0 \qquad \forall a \in A \backslash L, s \in S \tag{B.12}$$

$$\pi_s^{a^+} - r_a\pi_s^{a^-} \geq -(r_a - q_a)\psi_s^{a^-} y_a \quad \forall a \in L, s \in S \tag{B.13}$$

where a^+ and a^- indicate the tail and head of arc a. Constraint (B.9) imposes the busted restriction. The remaining constraints calculate the least probability that the intruder can reach to the destination undetected.