Journal of Global Optimization **22:** 319–342, 2002.

© 2002 *Kluwer Academic Publishers. Printed in the Netherlands.*

A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse

HANIF D. SHERALI* and BARBARA M.P. FRATICELLI
*Grado Department of Industrial and Systems Engineering (0118), Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA *Corresponding author (E-mail: hanifs@vt.edu)*

Abstract. In this paper, we modify Benders' decomposition method by using concepts from the Reformulation-Linearization Technique (RLT) and lift-and-project cuts in order to develop an approch for solving discrete optimization problems that yield integral subproblems, such as those that arise in the case of two-stage stochastic programs with integer recourse. We first demonstrate that if a particular convex hull representation of the problem's constrained region is available when binariness is enforced on only the second-stage (or recourse) variables, then the regular Benders' algorithm is applicable. The proposed procedure is based on sequentially generating a suitable partial description of this convex hull representation as needed in the process of deriving valid Benders' cuts. The key idea is to solve the subproblems using an RLT or lift-and-project cutting plane scheme, but to generate and store the cuts as functions of the first-stage variables. Hence, we are able to re-use these cutting planes from one subproblem solution to the next simply by updating the values of the first-stage decisions. The proposed Benders' cuts also recognize these RLT or lift-and-project cuts as functions of the first-stage variables, and are hence shown to be globally valid, thereby leading to an overall finitely convergent solution procedure. Some illustrative examples are provided to elucidate the proposed approach. The focus of this paper is on developing such a finitely convergent Benders' approach for problems having 0-1 mixed-integer subproblems as in the aforementioned context of two-stage stochastic programs with integer recourse. A second part of this paper will deal with related computational experiments.

Key words: Stochastic programming; Mixed-integer programming; Benders' decomposition; Reformulation-Linearization Technique (RLT); Lift-and-project cuts

1. Introduction

Benders' decomposition has proven to be a powerful technique for solving specially-structured large-scale linear and mixed-integer programs since its introduction in 1962 (see Nemhauser and Wolsey (1999), for example). The main idea behind this approach is to project the original problem onto the space of certain 'complicating variables' in order to derive an equivalent *master problem*, and to solve this latter problem via a relaxation or branch-and-cut strategy by generating Benders' feasibility and optimality cuts as prompted by the *subproblems* that are obtained by fixing the complicating variables. Stochastic mathematical programs constitute one class of problems for which solution procedures rely heavily on the premise of Benders' decomposition. The literature on stochastic programs focuses largely on two-stage stochastic programs with recourse. (In theory, multi-stage programs can be handled in a similar fashion via a nested approach, but in practice, this process is cumbersome to implement.) In these problems, the first-stage decisions must be made before the relevant random components of the environment are realized, with the provision that a set of second-stage (or recourse) variables can be subsequently used to compensate for the ensuing effect of the environment. The goal of the stochastic program is to optimize the first-stage costs plus the expected recourse costs. Some notable applications of stochastic programming include scheduling (Birge and Dempster, 1996), financial planning (Carino et al., 1994), power generation (Murphy et al., 1982), facility location (Laporte et al., 1994), and vehicle routing (Laporte et al., 1992). For more information on stochastic programming in general, we refer the reader to recent books on stochastic programming by Birge and Louveaux (1997) and Kall and Wallace (1994).

There are several popular methods for solving two-stage stochastic LPs with recourse, and most of these rely on the underlying principle of Benders' decomposition. The first-stage investment, resource acquisition, or location-type decisions represent the complicating variables, and the subproblems determine the best recourse actions for any given first-stage decisions. A common practice is to approximate continuous distributions with discrete ones, which allows the expected recourse function to be calculated as a simple weighted sum. In the case of stochastic programs with integer recourse, Schultz (1995) has shown that, under mild conditions, discrete distributions can effectively approximate continuous ones to any given accuracy. Consequently, assume that there are L possible environments or scenarios, ξ^l, $l = 1, \ldots, L$, each occurring with a respective probability of p_l. The set of constraints that couples the first- and second-stage decisions, $x \in R^n$ and $y \in R^m$, respectively, is generally expressed as

$$W^l y^l = h^l - T^l x ,$$

where the (technology) matrix T^l and the (resource) vector h^l are known for each possible scenario ξ^l, $l = 1, \ldots, L$. The matrix W^l (which is often assumed to be fixed in order to yield an exploitable subproblem structure, but in general, could be stochastic as well, i.e., dependent on ξ^l) is known as the recourse matrix, and it determines the set of recourse actions, y^l, that are governed by the net outcome $h^l - T^l x$. Given this notation, a typical Benders' decomposition for the two-stage stochastic program with recourse would view the given problem in the form

$$\begin{aligned} &\text{minimize} \quad && cx + \sum_{i=1}^{L} p_l Q(x, \xi^l) \\ &\text{subject to} \quad && x \in X , \end{aligned}$$

where $Q(x, \xi^l) = \min\{q^l y^l : W^l y^l = h^l - T^l x, y^l \geq 0\}$ for $l = 1, \ldots, L$, and where X is some nonempty polytope in R^n, with approximations for the optimal value functions

$Q(x, \xi^l)$, $l = 1, \ldots, L$ being generated via Benders' cuts. In some cases, it can be shown that every (feasible) first-stage solution results in a feasible second-stage problem. These problem instances are said to have (*relatively*) *complete recourse*, and in such cases, only optimality cuts are generated in this process.

Several algorithms have been designed to solve *stochastic linear programs* with recourse (see Ruszczynski (1999) for a thorough review of this subject). Most of these methods can be considered as extensions of the L-shaped algorithm that was proposed by Van Slyke and Wets (1969). For each solution of a suitable relaxed master problem, the L-Shaped Algorithm solves one subproblem for each of the L outcomes. If any of the subproblems are infeasible, a feasibility cut is added to the master problem. Otherwise, the optimal dual multipliers for the set of subproblems are used to create a single optimality cut for the master problem. Birge and Louveaux (1988) developed a multicut enhancement to the L-Shaped Algorithm, in which a separate optimality cut is constructed for each subproblem. Higle and Sen have used Stochastic Decomposition (1991) and Conditional Stochastic Decomposition (1994) to greatly reduce the total number of subproblems required to be solved. At each iteration of these methods, only one subproblem, associated with a randomly generated sample point, is solved. Optimality and feasibility cuts are generated as before, and the coefficients of these cuts are updated as more observations of the sample points are obtained.

Stochastic integer programs are stochastic programs in which some of the variables are restricted to be integer-valued. The integrality restriction can apply to the first- and/or second-stage variables. When the second-stage (recourse) variables are restricted to be integral, the resulting problem is referred to as a *stochastic program with integer recourse*. In this case, the problem complexity increases significantly, since the subproblem for any random outcome is an integer program whose parameters depend on the fixed first-stage decisions. Moreover, the optimal value recourse objective function now becomes nonconvex and discontinuous in general.

Although some solution strategies have been developed for specific applications of stochastic IPs, relatively few techniques have been developed to solve general stochastic IPs. We comment here on some recent algorithmic advances that employ decomposition techniques, and refer the reader to Klein Haneveld and van der Vlerk (1999) and Schultz et al. (1996) for further discussions. Laporte and Louveaux (1993) developed the integer L-shaped algorithm (a combination of the L-shaped method and branch-and-bound) to solve stochastic IPs having binary first-stage variables and complete (mixed-integer) recourse. This extension constructs optimality cuts based on independent evaluations of the recourse value function. For efficiency in an enumerative search process, certain lower bounding functionals on this recourse value function are also derived. Caroe and Tind (1998) have used duality theory to develop a more general extension of the L-shaped decomposition method to solve two-stage stochastic programs with integer recourse. Previously, Caroe and Tind (1997) had developed a Lagrangian dual approach based on applying variable splitting to the first-stage decisions, and then dualizing the resultant equal-value nonanticipatory constraints. This approach was shown to be equivalent to computing a hull relaxation in the context of disjunctive programming, and solving this via the lift-and-project cutting plane technique of Balas et al. (1993). Cuts derived for one subproblem were lifted to derive valid inequalities for other subproblems. However, in order to preserve facetial properties in this lifting process, a separate linear program needed to be solved. We note here that in our approach, which is geared toward solving the original problem itself (rather than its relaxation), we show how cuts derived for one subproblem can be directly used for other subproblems without any intermediate lifting step or auxiliary problem solution (other than a simple substitution). Moreover, facetial properties are preserved in a manner that induces finite convergence. For the specific case of simple integer recourse where $W^l = [I, -I]$, and with a fixed technology matrix and discretely distributed right-hand sides, Klein Haneveld et al. (1996) have used theoretical properties of the recourse objective value function to derive a convex hull representation for the problem. (See Klein Haneveld and van der Vlerk (1999) for a summary of several other techniques for simple integer recourse problems.) Caroe and Schultz (1999) have used scenario decomposition and Lagrangian relaxation within a branch-and-bound framework to solve two-stage stochastic IPs, and this approach can readily be extended to multistage stochastic programs. Ahmed et al. (2000) consider two-stage stochastic programs having pure integer second-stage variables, but mixed-integer first-stage variables. They employ a transformation that induces a special structure in the discontinuities of the second-stage optimal value function and based on a characterization of this structure, they design a finitely convergent branch-and-bound algorithm for the original problem. Promising computational results are provided on several classes of problems. A specialized approach for two-stage stochastic IPs with mixed-integer recourse that is similar to ours in concept, but uses an alternative sequential convexification process based on a different asymptotically exact cutting plane approach for solving the subproblems, has been proposed by Higle and Sen (2000). In a different vein, Schultz et al. (1998) have used Grobner basis techniques within a implicit enumeration strategy to address the class of problems having integer recourse. Although Grobner bases are typically expensive to compute, their use becomes relatively more effective when the same problem is re-solved for different right-hand side values, which is the case for recourse problems.

Our focus in this paper is to develop a Benders' decomposition strategy to solve discrete optimization problems where both the inner and outer state decisions might involve 0–1 variables, such as those encountered in a two-stage stochastic program with integer recourse. In Section 2, we begin by describing a relatively similar conceptual case for which a suitable convex hull representation can be constructed that permits a finite regular application of Benders' methodology. This lays the groundwork for the more usual case discussed in Section 3, where such a representation is only partially generated in a sequential fashion as needed within the context of a Benders' branch-and-cut approach. This viewpoint facilitates the generation of valid inequalities during the solution of any given subproblem in a form that renders them valid for all other subproblems by merely substituting the revised first-stage decisions in a derived linear functional term, and also enables the derivation of suitable Benders' cuts that induce finite convergence. Some numerical examples are presented to illustrate the proposed methodology. Section 4 addresses finite convergence issues related to the proposed cutting plane approach for solving the sub-problems, and Section 5 contains our conclusions and suggestions for future research.

2. Benders' partitioning based on a suitable convex hull representation

Since the proposed methodology is relevant to many types of discrete optimization problems in addition to stochastic IPs, we describe our approach in terms of a generic problem P that is given below in (1). Although this form does not specifically correspond to the notation used for stochastic IPs, it should be evident from the foregoing discussion that the structure of (1) subsumes this class of problems. (Note that in this context, it would be computationally facile, but not necessary, to have constant technology and recourse matrices, as variously assumed in the literature—for example, see Caroe and Tind (1997).)

$$\mathbf{P}:\ \text{Minimize}\quad cx + dy \tag{1a}$$

$$\text{subject to}\quad Ax + Dy \geq b \tag{1b}$$

$$x \in X,\ x \in \{0, 1\}^n,\ y \in Y \tag{1c}$$

where X represents a nonempty polytope in R^n that is defined in terms of the binary variables x, and Y is a compact subset of R^m and represents some linear restrictions on the y-variables, in addition to binary restrictions on a *subset* (say, $y_1, \ldots, y_p$) of the variables. By appropriately incorporating an artificial (interval-bounded) variable column within the y-variable set, we will assume that P is feasible for any fixed $x \in X$, x binary, and moreover, we will also assume that an optimum exists for P.

Now, let us define (denoting e as a compatible vector of ones)

$$Z = \text{conv}\{(x, y) : Ax + Dy \geq b,\ 0 \leq x \leq e,\ y \in Y\} \tag{2a}$$

$$\equiv \{(x, y) : Gx + Hy + Fw \geq f\},\ \text{say}\ , \tag{2b}$$

where for convenience, we have also absorbed any simple bound restrictions within the inequalities describing (2b). Note that the description (2b) is assumed to be derived in a higher dimensional space (including a set of new w-variables), as for example by using the Reformulation-Linearization Technique (RLT) process (see Sherali and Adams, 1990, 1994, 1999). Note also that aside from the bounding constraints $0 \leq x \leq e$ on the x-variables, the other constraining restrictions $x \in X$ on these variables are not included in the definition of Z. (This might be computationally advantageous in deriving the convex hull representation for Z; also, see Proposition 2 below for details on how this can be further exploited in the presence of dual-angular special structures.) Later, we will discuss a sequential scheme for partially generating this system as needed, but for now, assume that the entire description of Z is at hand.

Consider the problem

$$\mathbf{P}': \text{ Minimize } \quad cx + dy \tag{3a}$$

$$\text{subject to } \quad Gx + Hy + Fw \geq f \tag{3b}$$

$$x \in X, \, x \in \{0, 1\}^n \, . \tag{3c}$$

PROPOSITION 1. *P' has an optimal solution, and moreover, it is equivalent to P in the sense that if (x^*, y^*, w^*) solves P', where (y^*, w^*) is an extreme point optimum to P' for x fixed at x^*, then (x^*, y^*) solves P.*

Proof. By our assumptions on P, the set Z given by (2) is bounded and P$'$ is feasible. Hence P$'$ has an optimum (x^*, y^*, w^*) where (y^*, w^*) satisfies the condition stated in the proposition. Moreover, since P$'$ is a relaxation of P, and its constraints imply $Ax + Dy \geq b$, $x \in X$, and the linear constraints describing $y \in Y$, it is sufficient to show that y^* satisfies the required binary restrictions on its subcomponents. From (2), any extreme point $(\bar{x}, \bar{y})$ of Z satisfies $\bar{y} \in Y$ (including the binary restrictions). Furthermore, if we define $Z(x^*) = Z \cap \{(x, y) : x = x^*\}$, then since $Z(x^*)$ is a face of Z, any extreme point $(x^*, \bar{y})$ of $Z(x^*)$ has $\bar{y} \in Y$ as well. Noting that $Z(x^*)$ defines the feasible region of P$'$ when x is fixed at x^*, and that (x^*, y^*) is a vertex of $Z(x^*)$, we have (x^*, y^*) is feasible, and therefore optimal, to P. This completes the proof. $\square$

2.1. SPECIALIZED MODIFICATIONS FOR PROBLEMS HAVING A DUAL ANGULAR STRUCTURE

Before proceeding further, it is instructive to comment on a modified derivation of the equivalent representation P$'$ when the original problem P exhibits a dual-angular structure (as in the special case of two-stage stochastic IPs). This analysis also lends further insights into the flexibility of constructing only partial convex hull representations in deriving an equivalent restatement of the problem to which Benders' decomposition method is applicable. Toward this end, suppose that P possesses a dual-angular structure as revealed by the coefficient matrices given in the form

$$A \equiv \begin{bmatrix} A^1 \\ \vdots \\ A^S \end{bmatrix}, \qquad D \equiv \begin{bmatrix} D^1 & & \\ & \ddots & \\ & & D^S \end{bmatrix}, \qquad b \equiv \begin{bmatrix} b^1 \\ \vdots \\ b^S \end{bmatrix}, \quad \text{and } d \equiv \begin{bmatrix} d^1 \\ \vdots \\ d^S \end{bmatrix}, \tag{4a}$$

where the vector y is also accordingly partitioned into components y^s, for $s = 1, \ldots, S$, with $y \in Y$ being replaced by

$$y^s \in Y_s \quad \forall s = 1, \ldots, S . \tag{4b}$$

Here, for $s = 1, \ldots, S$, each Y_s is assumed to impose certain polyhedral restrictions on the (recourse) variables y^s (pertaining to scenario s), including binary restrictions on a subset of variables.

Now, let us define

$$Z_s = \text{conv}\{(x, y^s) : A^s x + D^s y^s \geq b^s,\ 0 \leq x \leq e,\ y^s \in Y_s\} \tag{5a}$$

and let

$$Z' = \{(x, y) : (x, y^s) \in Z_s \text{ for each } s = 1, \ldots, S\} . \tag{5b}$$

Note that in general, $Z \subseteq Z'$, and that is is relatively easier to characterize Z' than it is to construct Z. Moreover, Z' retains the separability of the (recourse) variables y^s, $s = 1, \ldots, S$. The following results asserts that the equivalence of P′ and P as stated in Proposition 1 remains valid when Z is replaced by Z' under (4). In this context, similar to (2b), the construction (5) would yield P′ in the form given by (3) where the coefficient matrices in (3b) would possess the structure

$$\begin{gathered} G \equiv \begin{bmatrix} G^1 \\ \vdots \\ G^S \end{bmatrix}, \qquad H \equiv \begin{bmatrix} H^1 & & \\ & \ddots & \\ & & H^S \end{bmatrix}, \\ F \equiv \begin{bmatrix} F^1 & & \\ & \ddots & \\ & & F^S \end{bmatrix}, \quad \text{and } f \equiv \begin{bmatrix} f^1 \\ \vdots \\ f^S \end{bmatrix}, \end{gathered} \tag{6}$$

and where the higher-dimensional vector w is also decomposed into the corresponding components w^s, for $s = 1, \ldots, S$.

PROPOSITION 2. *Suppose that P has a dual angular structure as given by (4), and let P′ be defined by replacing Z with the set Z′ given by (5). Then P′ is equivalent to P in the sense asserted by Proposition 1.*

Proof. Let (x^*, y^*, w^*) solve P′, where (y^*, w^*) is as stated in the proposition. Note from (6) that when we fix $x = x^*$, the problem P′ separates into S problems (by scenarios) given as follows:

$$\text{minimize}\{d^s y^s : (x^*, y^s) \in Z_s\} . \tag{7}$$

Again, because (5) includes the hypercube restrictions $0 \leq x \leq e$, we have that $Z_s(x^*) \equiv Z_s \cap \{(x, y^s) : x = x^*\}$ is a face of Z_s, and therefore, its extreme points satisfy the required binary restrictions on y^s. Noting that $Z_s(x^*)$ is the feasible region of (7), this completes the proof. □

In what follows, for the sake of simplicity in notations and generality, we will assume that the set Z conforms with Z' whenever we have the dual angular structure exhibited by (4), with the system (2b) possessing the structure exhibited by (6). Hence, whenever we employ (2b), or develop lower-level RLT relaxations for the system $\{\cdot\}$ in (2a), we assume via Proposition 2 that in the presence of a dual-angular structure, we respectively have the structure (6), or that we correspondingly apply the lower-level RLT relaxation to the system $\{\cdot\}$ in (5a) for each $s = 1, \ldots, S$. We will periodically make some related comments in the sequel to re-emphasize this feature.

2.2. DERIVATION OF A BENDERS' APPROACH FOR PROBLEM P′

Assuming tentatively that we have explicitly constructed the equivalent formulation P′, we can apply Benders' partitioning to solve this problem as follows:

$$\underset{x \in X \cap \{0,1\}^n}{\text{Minimize}} \{cx + \text{minimum}\{dy : Hy + Fw \geq f - Gx\}\} \tag{8a}$$

i.e.,

$$\underset{x \in X \cap \{0,1\}^n}{\text{Minimize}} \{cx + \text{maximum}\{\pi(f - Gx) : \pi H = d,\ \pi F = 0,\ \pi \geq 0\}\}\,. \tag{8b}$$

Since we have assumed that the inner problem in (8) is feasible and bounded for any fixed $x \in X \cap \{0, 1\}^n$ letting

$$\{\pi^q,\ q = 1, \ldots, Q\} \equiv \text{vert}(\Lambda)\,, \tag{9a}$$

where

$$\Lambda \equiv \{\pi : \pi H = d,\ \pi F = 0,\ \pi \geq 0\}\,, \tag{9b}$$

we obtain the following projected form of P′.

$$\begin{aligned} &\text{Minimize} \quad z && (10a)\\ &\text{subject to} \quad z \geq cx + \pi^q(f - Gx) \text{ for } q = 1, \ldots, Q && (10b)\\ &\qquad\qquad\quad x \in X \cap \{0, 1\}^n\,. && (10c) \end{aligned}$$

Recall that (10) is the Benders' (overall) *master program*, and the inner minimization problem in (8a), or its dual in (8b), for any fixed x is referred to as the Benders' *subproblem*. This subproblem generates the *Benders' cuts* (10b) (along with upper bounds on the problem).

REMARK 1. In case we do not incorporate suitable artificial variable column(s) as needed to ensure that the inner problem in (8a) is feasible for any fixed $x \in X \cap \{0, 1\}^n$, we would also need to generate feasibility or extreme direction cuts in (10) of the following type, where δ^r, $r = 1, \ldots, R$, are extreme directions of the polyhedron Λ that is defined in (9).

$$\delta^r(f - Gx) \leq 0\,, \quad \text{for } r = 1, \ldots, R\,. \tag{11}$$

REMARK 2. Note that in a practical implementation, we need not solve the relaxed

Benders' master programs to optimality at each iteration. Rather, a branch-and-cut approach could be adopted, with the enumeration process set up only once, and with the current relaxed master program (RMP, say) being used to determine lower bounds, the subproblem (SP, say) providing upper bounds, and the (globally valid) Benders' cuts (10b) being generated as needed, i.e., whenever an incumbent solution to the current relaxed master program is found that has an objective value less than the present upper bound on the overall problem. Geoffrion and McBride (1978) and Adams and Sherali (1993) provide details for such an approach. Any actual application of Benders' method discussed here can be adapted to follow such a scheme.

EXAMPLE 1. As an illustration, consider the following example.

$$\textbf{P}: \text{Minimize} \quad -x_1 - 2y_1 \tag{12a}$$

$$\text{subject to} \quad -4x_1 - 3y_1 \geq -6 \tag{12b}$$

$$(x_1, y_1) \text{ binary}. \tag{12c}$$

Figure 1 depicts the solution of this problem and identifies the set Z, along with the key facet that describes this set. By (2), this set Z is given by

$$Z = \text{conv}\{(x_1, y_1) : -4x_1 - 3y_1 \geq -6,\ 0 \leq x_1 \leq 1,\ y_1 \text{ binary}\}. \tag{13}$$

Since there is only one y-variable for this problem, we can develop the complete RLT representation of Z by multiplying each of the constraints in (13) by the two bound-factors associated with y_1. This yields the following equivalent Problem P′ as defined by (3):

$$\text{Minimize} \quad -x_1 - 2y_1 \tag{14a}$$

$$\text{subject to} \quad 3y_1 - 4w \geq 0 \tag{14b}$$

$$-4x_1 - 6y_1 + 4w \geq -6 \tag{14c}$$

$$y_1 - w \geq 0 \tag{14d}$$

$$x_1 - w \geq 0 \tag{14e}$$

$$-x_1 - y_1 + w \geq -1 \tag{14f}$$

$$w \geq 0 \tag{14g}$$

$$x_1 \text{ binary}. \tag{14h}$$

Note that (14b) and (14c) are obtained by the RLT product of $-4x_1 - 3y_1 \geq -6$ with y_1 and $(1 - y_1)$, respectively, and (14d)–(14g) are bound-factor RLT product constraints obtained via the products of the bounding inequalities $0 \leq x_1 \leq 1$ with y_1 and with $(1 - y_1)$. Observe that the surrogate of (14b) and (14f) according to

$$(3y_1 - 4w) + 4(-x_1 - y_1 + w + 1) \geq 0 \tag{15a}$$

produces the required key facet of Z identified in Figure 1 as

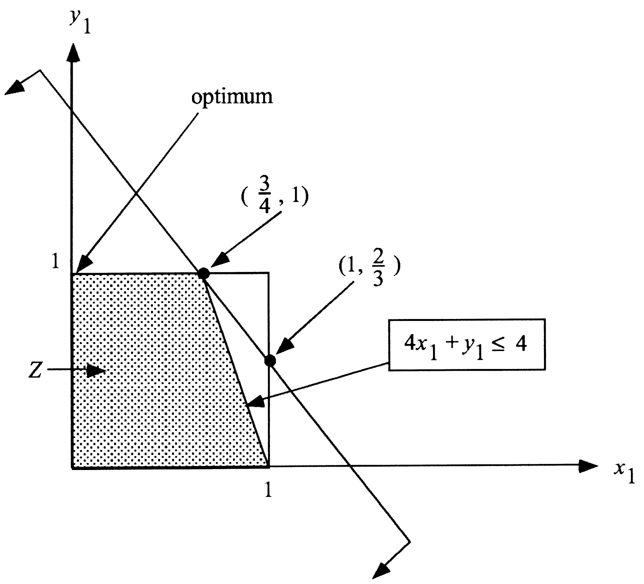

Figure 1. Illustration for Example 1.

$$-4x_1 - y_1 \geq -4\,. \tag{15b}$$

In essence, by projecting the region of (14) onto the (x_1, y_1) space (only for illustrative purposes; this combinatorial step would not be performed in actual implementations), we get that (14) can equivalently be written as follows.

$$\begin{aligned} &\text{Minimize} && -x_1 - 2y_1 && (16a)\\ &\text{subject to} && -4x_1 - y_1 \geq -4 && (16b)\\ & && x_1 \text{ binary}, \ 0 \leq y_1 \leq 1\,. && (16c) \end{aligned}$$

We could now apply Benders' partitioning to solve (14), which in essence, would be tantamount to applying this method to (16). For the sake of convenience, we apply it directly to (16) and obtain the decomposition

$$\min_{x_1 \in \{0,1\}} \{-x_1 + \max\{\pi_1(4x_1 - 4) - \pi_2 : -\pi_1 - \pi_2 \leq -2, (\pi_1, \pi_2) \geq 0\}\}\,. \tag{17}$$

Noting that the extreme points of the inner maximization problem in (17) are $(\pi_1, \pi_2) = (2, 0)$ and $(0, 2)$, and that (12) is feasible for any binary x_1, the complete Benders' master program is derived as follows.

$$\begin{aligned} &\text{Minimize} && z && (18a)\\ &\text{subject to} && z \geq 7x_1 - 8 && (18b)\\ & && z \geq -x_1 - 2 && (18c)\\ & && x_1 \text{ binary}\,. && (18d) \end{aligned}$$

The optimum to (18) (which would ultimately be generated via the usual process of applying Benders' methodology) is given by $x_1^* = 0$ and $z^* = -2$. Solving (16) (or (14)) with x_1 fixed at $x_1^* = 0$ yields $y_1^* = 1$ (and $w^* = 0$), with $v(x_1^*) = z^* = -2$. Since the relaxed master problem and subproblem have the same objective values, we have obtained an optimal solution to (12).

3. Benders' partitioning using a sequential convex hull constructive process

The approach (8)–(10) is based on an *a priori* generation of the convex hull representation Z defined in (2) (or Z' defined by (5) under the structure (4)). If the size of the problem permits this construction (in particular, if we have few y-variables, or each partitioned constraint set in (5a) has a relatively simple structure), then this is a viable option, and leads to a usual application of Benders' decomposition as per Remark 2. Otherwise, we can generate a partial representation for Z as needed in a sequential convexification process, as discussed below. The following remark first highlights a key concept that is used in developing our proposed solution process.

REMARK 3. Let $\bar{Y}$ denote the continuous relaxation of Y, and let $J^* = \{j : y_j$ is restricted to be binary in $Y\}$. For any $J \subseteq J^*$, define

$$Z^J = \text{conv}\{(x, y) : Ax + Dy \geq b,\ 0 \leq x \leq e,\ y \in \bar{Y},\ y_j \text{ binary } \forall j \in J\}\,. \tag{19}$$

Note that $Z^\emptyset$ along with $x \in X$ represents the continuous relaxation of (1), and $Z \equiv Z^{J^*}$. Since $Z \subseteq Z^J$ for each $J \subseteq J^*$, valid Benders' cuts can be derived from any such set Z^J. In fact, using the RLT process, we can construct a higher dimensional representation of Z^J for any $J \subseteq J^*$ that could be characterized as a surrogate of the representation (2b) for Z using suitable nonnegative multipliers (see Sherali and Adams, 1990, 1994). Hence, Benders' cuts derived via the relaxation Z^J substituted in place of Z would correspond to cuts obtained via some feasible, though not necessarily extreme point, solution to Λ. Likewise, Benders' cuts derived via lower-level RLT applications to $Z^\emptyset$ (levels less than $|J|$ for the case of Z^J) based on considering binariness on the variables y_j for $j \in J$, but not necessarily having constructed the entire convex hull representation Z^J, would be valid as well. Moreover, since the description of such a lower level representation can be obtained by surrogating the constraints of Z^J, and hence those of Z, the resulting cuts can also be viewed as implicitly obtained from feasible, nonextremal solutions to Λ.

Based upon these insights, we now develop a finitely convergent method for solving Problem P, or Problem P′ via (8)–(10), by sequentially constructing a partial convex hull representation as needed. In this approach, for any fixed $\bar{x}$, the corresponding Benders' subproblem in (8b) that is reproduced below as

$$\mathbf{SP}\text{: maximize}\,\{\pi(f - G\bar{x}) : \pi H = d,\ \pi F = 0,\ \pi \geq 0\}\,, \tag{20}$$

is solved *implicitly* via an RLT-based or lift-and-project cutting plane approach (see Balas et al., 1993; Sherali et al., 2000). In the proposed method, we explicitly generate appropriate surrogated versions of Z as needed to derive valid RLT or lift-and-project cutting planes as needed for solving the subproblems. The key idea is that these generated cuts are characterized as functions of x, and can therefore be updated and re-used for subsequent subproblems based on the corresponding fixed value of x. Likewise, the Benders' cuts derived via the solution of the subproblems using such a cutting plane approach recognize these cuts as function of x, and are hence shown to be globally valid. This leads to an overall finitely convergent solution process.

REMARK 4. To set ideas, let us first consider a preliminary rudimentary approach for solving Problem P′ via Benders' decomposition. This simple approach solves various restricted versions of the subproblems (20) (or relaxed versions of its dual) as follows. For the first instance of Problem SP, we let $k = 0$ and take $J_k = \emptyset$. Using $Z^{J_k} = Z^{\emptyset}$ as the current RLT representation within the inner minimization in (8a), we solve SP and generate the associated Benders' constraint for the relaxed master problem. At each subsequent visit to SP, if the current subproblem yields a binary y-solution, we use this solution to update the incumbent solution and to generate a Benders' cut. Otherwise, we increment k and take $J_k = J_{k-1} \cup \{j\}$ where y_j is restricted to be binary, but currently has a fractional value. We then construct Z^{J_k} as the *updated* RLT representation using the scheme described in Sherali and Adams (1990, 1994), solve SP, and generate the associated Benders' constraint for the relaxed master problem.

Note that this process creates a nested sequence of sets $J_0 \subseteq J_1 \subseteq J_2 \subseteq \cdots$ leading up to J^* in the worst case. Within a finite number of visits to SP, this procedure generates cuts based on Z via either a partial or full representation of this set, thereby deriving valid upper bounds from each such SP, and resulting in an overall finitely convergent algorithm based on the finiteness of the set $X \cap \{0, 1\}^n$. Alternatively, we could derive valid upper bounds from each subproblem by continuing to expand the set J_k at each iteration k to include fractionating y-variable indices until an integer feasible y-solution is obtained. This alternative is more in the conceptual spirit of the proposed approach as explained below.

Clearly, the approach described in Remark 4 of sequentially generating approximations leading up to Z is computationally intensive because of the potentially exponential size of these (partial) convex hull representations. The procedure we propose below instead relies on generating cuts as needed to solve each subproblem SP based upon its fractionating variables, rather than generating full (partial) convex hull representations. More importantly, it characterizes these cuts in a fashion that permits them to be re-used in a suitably modified form for other subsequent subproblems. Furthermore, the cuts are generated in the original dimensional space, and previously generated cuts can be retained or deleted as desired.

As alluded above, the proposed method implicitly generates an appropriate surrogated representation of Z as needed for each individual SP via an RLT cutting plane approach as follows. Suppose that we are solving SP for a given $\bar{x}$. In essence, we wish to solve

$$v(\bar{x}) = c\bar{x} + \text{minimum}\{dy : Dy \geq b - A\bar{x},\ y \in Y\} \tag{21}$$

but we conceive the implicit solution of this via the problem

$$v(\bar{x}) = c\bar{x} + \text{minimum}\{dy : Hy + Fw \geq f - G\bar{x}\} \tag{22}$$

from (8a), so that we can derive a valid Benders' cut. (Note that in the presence of a dual angular structure, (22) would yield a separable system as per (6).) Now suppose that we adopt a sequential convexification lift-and-project type of cutting plane scheme to solve (21), using RLT cuts based on enforcing binariness on one variable as in Balas et al. (1993), or on multiple variables as in Sherali et al. (2000). (See Section 4 for details on the finite convergence of such a cutting plane algorithm.) Suppose that we obtain the final cut-enhanced problem that solves (21) as given by (23) below, where (23c) represents the continuous relaxation $\bar{Y}$, and where (23d) represents the set of RLT or lift-and-project cuts generated.

$$\begin{aligned} v(\bar{x}) = c\bar{x} + \text{minimum} \quad & dy && (23a)\\ \text{subject to} \quad & Dy \geq b - A\bar{x} && (23b)\\ & \Gamma y \geq \gamma && (23c)\\ & \alpha_t y \geq \beta_t - \phi_t \bar{x} \text{ for } t = 1, \ldots, T\,. && (23d) \end{aligned}$$

Each of the cuts $t = 1, \ldots, T$ in (23d) is derived via the following steps.

Step 1. Based on some current fractional solution $\bar{y}$, generate an appropriate RLT enhancement of $Z^{\emptyset}$ given as follows (by enforcing binariness on one or more variables—see Section 4, and in particular, Remark 6 given later for some additional details):

$$G_t x + H_t y + F_t w \geq f_t\,. \tag{24}$$

(In the presence of dual-angularity, this system would have a structure similar to that in (6).)

Step 2. Fix $x = \bar{x}$, and determine dual multipliers $\pi_t \geq 0$ for (24) that solves the following separation problem, where e is a conformable vector of ones, and where (25c) is a normalization constraint (that can be imposed separably in the context of dual-angular structures).

$$\begin{aligned} \text{Minimize} \quad & \pi_t(H_t\bar{y}) - \pi_t(f_t - G_t\bar{x}) && \text{(25a)} \\ \text{subject to} \quad & \pi_t F_t = 0 && \text{(25b)} \\ & e \cdot \pi_t = 1 && \text{(25c)} \\ & \pi_t \geq 0 \,. && \text{(25d)} \end{aligned}$$

Note that by virtue of the RLT process, an appropriate representation (24) can be generated that yields a negative value in (25). Let $\bar{\pi}_t$ be the solution of (25). Then we have that

$$\tilde{\pi}_t H_t y \geq \tilde{\pi}_t(f_t - G_t\bar{x}) \tag{26}$$

deletes the current fractional solution $\bar{y}$. Furthermore, with the substitution

$$\alpha_t \equiv \tilde{\pi}_t H_t,\ \beta_t \equiv \tilde{\pi}_t f_t, \text{ and } \phi_t \equiv \tilde{\pi}_t G_t \,, \tag{27}$$

we have that (26) is of the form (23d).

The final representation (23) can be used to derive a valid Benders' cut, as shown in Proposition 3. This leads to a finitely convergent algorithm, as demonstrated subsequently in Proposition 4. Following this, we will comment on the re-use of previously generated cuts for new subproblems (21)–(23) solved for revised values for $\bar{x}$.

PROPOSITION 3. *Consider Problem (23), and let ψ_1, ψ_2, and (ψ_{3t}, $t = 1, \ldots, T$) be the optimal nonnegative dual multipliers obtained for the constraints (23b), (23c), and (23d), respectively. Then, noting (27), the inequality*

$$z \geq cx + \psi_1(b - Ax) + \psi_2\gamma + \sum_{t=1}^{T} \psi_{3t}(\beta_t - \phi_t x) \tag{28}$$

is a valid Benders cut.

Proof. Consider the system (3b) that is derived from (2). Since the original constraints in (2a) are implied by (2b) via a suitable surrogation process, and noting the definition of (23c), there exist nonnegative surrogate multiplier matrices τ_1 and τ_2 such that

$$\tau_1[G, H, F] = [A, D, 0], \text{ with } \tau_1 f \geq b \,, \tag{29}$$

and

$$\tau_2[G, H, F] = [0, \Gamma, 0], \text{ with } \tau_2 f \geq \gamma \,. \tag{30}$$

Similarly, since any lower-level or partial RLT application such as (24) is implied by (3b) via a surrogation process, there exist nonnegative surrogate multiplier matrices τ_{3t}, $t = 1, \ldots, T$, such that

$$\tau_{3t}[G, H, F] = [G_t, H_t, F_t], \text{ with } \tau_{3t} f \geq f_t \,, \quad \forall t = 1, \ldots, T \,. \tag{31}$$

Now, let us define

$$\bar{\pi} = \psi_1 \tau_1 + \psi_2 \tau_2 + \sum_{t=1}^{T} \psi_{3t} \tilde{\pi}_t \tau_{3t} \tag{32}$$

where $\tilde{\pi}_t$ is obtained as an optimum to (25) and satisfies (26). Note that $\bar{\pi} \geq 0$, and from (26), (29)–(32), we get

$$\bar{\pi} H = \psi_1 D + \psi_2 \Gamma + \sum_{t=1}^{T} \psi_{3t} \tilde{\pi}_t H_t$$

i.e.

$$\bar{\pi} H = \psi_1 D + \psi_2 \Gamma + \sum_{t=1}^{T} \psi_{3t} \alpha_t = d \tag{33}$$

via duality in (23). Moreover, we have from (25b), (29)–(32) that

$$\bar{\pi} F = \psi_1(0) + \psi_2(0) + \sum_{t=1}^{T} \psi_{3t} \tilde{\pi}_t F_t = 0\,. \tag{34}$$

Hence, $\bar{\pi} \in \Lambda$ as defined in (9), and so the constraint

$$z \geq cx + \bar{\pi}(f - Gx) \tag{35a}$$

is a valid Benders' inequality. But from (26), (29)–(32), we have,

$$\bar{\pi}(f - Gx) \geq \psi_1(b - Ax) + \psi_2 \gamma + \sum_{t=1}^{T} \psi_{3t} \tilde{\pi}_t (f_t - G_t x)$$

i.e.

$$\bar{\pi}(f - Gx) \geq \psi_1(b - Ax) + \psi_2 \gamma + \sum_{t=1}^{T} \psi_{3t} \tilde{\pi}_t (\beta_t - \phi_t x)\,. \tag{35b}$$

Noting (35a) and (35b), we have that (28) is a valid Benders' cut, and this completes the proof. □

REMARK 5. Note that the key insignt above is that although the right-hand sides in (23) are real numbers in the process of solving the underlying subproblem, the Benders' inequality generated from its optimal dual solution via (28) needs to recognize the right-hand sides of both (23b) and (23d) as *functions* of x, much as in the usual Benders approach. In particular, we need to store the constant β_t and the vector ϕ_t for each cut $t = 1, \ldots, T$ in (23d). Note that the parent matrices or RLT representations that generated these cuts need not be stored. Furthermore, because of the global validity of the inequality

$$\alpha_t y \geq \beta_t - \phi_t x \tag{36}$$

for *any* x by virtue of (24) and the surrogation of the type in (26), we can impose the previously generated cuts of type (23d) in any subsequent subproblem solution, *simply by modifying its right-hand side according to the current* $\bar{x}$. (This would occur for each separable subproblem component in the presence of dual angularity, with the facility of sharing cuts between subproblems being also available in this context.) This re-use opportunity can greatly benefit the solution procedure. Section 4 further addresses the finite convergence issues related to such a cutting plane process applied to any given subproblem.

Despite the fact that we might not be generating extreme points of Λ in the cuts (28), the following result establishes finite convergence of the overall algorithm, assuming that each subproblem is solved finitely (as discussed in Section 4 below).

PROPOSITION 4. *Suppose that we implement Benders' algorithm in the traditional fashion as follows. At each iteration, we solve the relaxed master program (10), where the Benders' cuts (10b) are replaced by the current set of cuts of type (28). Let* $(\bar{z}, \bar{x})$ *be an optimal solution to this relaxed master program, where* $\bar{x} \in X \cap \{0, 1\}^n$. *Next, we solve the subproblem (23) to determine the value* $v(\bar{x})$ *of Problem P when x is fixed at* $\bar{x}$, *and accordingly, either terminate if* $\bar{z} \geq v(\bar{x})$ *(equivalently,* $\bar{z} = v(\bar{x})$*), or else, generate a Benders' cut (28) if* $\bar{z} \leq v(\bar{x})$. *Then, this process will converge finitely with an optimum for Problem P.*

Proof. Note that by the validity of (28) in Proposition 3, the result holds true if we show that we will finitely obtain the termination criterion $\bar{z} \geq v(\bar{x})$. Observe that by duality in (23), the right-hand side of (28) evaluated at $x = \bar{x}$ yields $v(\bar{x})$. Hence, whenever a previous $\bar{x}$ is regenerated by the master program, the termination criterion would hold true. Since there are only a finite number of solutions in $X \cap \{0, 1\}^n$, this must occur finitely, and the proof is complete. □

As mentioned previously, an actual implementation would follow Remark 2. Figure 2 provides a flow-chart for such a process.

EXAMPLE 2. Consider the problem of Example 1. To illustrate the concept of the proposed approach, suppose that we have a relaxed master program RMP that currently has the Benders' inequality (18c), but not (18b). This problem yields the solution $\bar{x}_1 = 1$ and $\bar{z} = -3$. We now solve for $v(\bar{x}_1)$ via the following problem, using a cutting plane process in the spirit of (23).

$$v(\bar{x}_1) = -\bar{x}_1 + \text{minimum}\{-2y_1 : -3y_1 \geq 4\bar{x}_1 - 6,\ y_1 \text{ binary}\}\,. \tag{37}$$

The continuous optimum for (37) is $\bar{y}_1 = 2/3$. At Step 1 of the cut generation process, let the RLT constraints (24) be given by (14b)–(14g) as in the lift-and-project scheme of Balas et al. (1993). The corresponding separation problem (25) at Step 2 is given as follows, where (for 't' = 1), $\pi_{11}, \ldots, \pi_{16}$ denote the surrogate multipliers with respect to the constraints (14b)–(14g), respectively.

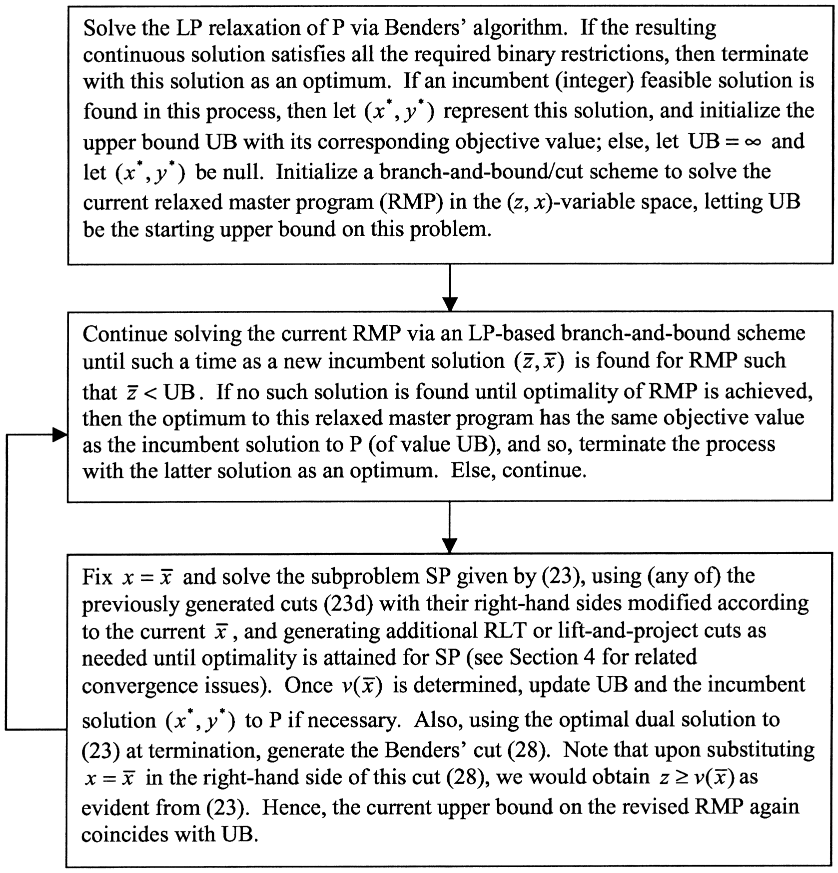

Figure 2. Flow-chart of an implementation for the proposed Benders' Algorithm.

$$\begin{aligned}
\text{Minimize} \quad & 2\pi_{11} - 2\pi_{12} + \frac{2}{3}\pi_{13} + \pi_{14} - \frac{2}{3}\pi_{15} \\
\text{subject to} \quad & -4\pi_{11} + 4\pi_{12} - \pi_{13} - \pi_{14} + \pi_{15} + \pi_{16} = 0 \\
& \pi_{11} + \pi_{12} + \pi_{13} + \pi_{14} + \pi_{15} + \pi_{16} = 1 \\
& (\pi_{11}, \ldots, \pi_{16}) \geq 0\,.
\end{aligned}$$

This problem yields the solution $\tilde{\pi}_{11} = 1/5$, $\tilde{\pi}_{15} = 4/5$, $\tilde{\pi}_{12} = \tilde{\pi}_{13} = \tilde{\pi}_{14} = \tilde{\pi}_{16} = 0$, with an objective value of $-2/15$, thereby indicating that a cut is generated. From (27), this cut yields

$$\alpha_1 = -\frac{1}{5},\ \beta_1 = -\frac{4}{5},\ \text{and}\ \phi_1 = -\frac{4}{5}. \tag{38}$$

The globally valid cut of type (32) is then given via (26) as

$$-\frac{1}{5}y_1 \geq -\frac{4}{5} + \frac{4}{5}x_1 \tag{39}$$

which corresponds to the facet of Z depicted in Figure 1. The particular cut (23d) that is incorporated within (37) is obtained by fixing $x_1 = \bar{x}_1 \equiv 1$ in (39). This yields the inequality $-y_1 \geq 0$, thereby producing (23) as

$$\bar{v}(\bar{x}_1) = -\bar{x}_1 + \text{minimum} \quad -2y_1 \tag{40a}$$

$$\text{subject to} \quad -3y_1 \geq 4\bar{x}_1 - 6 = -2 \tag{40b}$$

$$-y_1 \geq -4 + 4\bar{x}_1 = 0 \tag{40c}$$

$$0 \leq y_1 \leq 1\,. \tag{40d}$$

The optimal solution is given by $\bar{y}_1 = 0$, with the dual multipliers with respect to (40b) and (40d) being zeroes and with respect to (40c) being 2, yielding $v(\bar{x}_1) = -1 > \bar{z} = -3$. Hence, we generate the Benders' cut (28) as

$$z \geq -x_1 + 2(-4 + 4x_1)$$

i.e.

$$z \geq 7x_1 - 8\,.$$

This produces the revised relaxed Benders' master program given by (18) as in Example 1, which results in an optimal solution being detected as before.

4. Finite convergence of a cutting plane procedure for solving subproblems

In the foregoing section, we have developed a Benders' partitioning approach for Problem P of the type (1) based on the use of a suitable cutting plane approach for solving each subproblem (21) via (23). The cuts derived via (24)–(27) were generated to be directly valid for Z itself, but were then imposed on the current subproblem by fixing $x = \bar{x}$, where $\bar{x}$ corresponds to the given first-stage decision for the present subproblem. This not only permitted their re-use for other subproblems, but also enabled the derivation of the required Benders' cuts that induced an overall finitely convergent process. In this section, we now address the issue of designing a finitely convergent cutting plane procedure of this type for computing $v(\bar{x})$ defined in (21) via (23). (As alluded variously in the foregoing section, in the context of dual-angular structures, the separability of (21) and the partial convex hull requirement stipulated by Proposition 2 can be exploited below with obvious modifications.)

Note that in practice, one could use a variety of lift-and-project or RLT cuts as presented in Balas et al. (1993) and Sherali et al. (2000) to implement (23). However, in order to ensure that such a process finitely solves the underlying 0–1 mixed-integer program, some care needs to be exerised while sequentially constructing the (partial) convex hull representation that is necessary to solve this problem.

As in Balas et al.'s (1993) lift-and-project cutting plane algorithm, we rely on Jeroslow's (1980) cutting plane game concept for facial disjunctive programs. (Note that (21), and likewise Problem P given by (1), is a *facial disjunctive program* in that it involves the conjunction of the disjunctions that $y_j \leq 0$ or $y_j \geq 1$ (in concert with $0 \leq y_j \leq 1$) for each $j = 1, \ldots, p$, along with the facial property that the intersection of either of these disjunctive restrictions with the continuous feasible region of (21) defines a face of this region.) However, there is one important variation in the standard process that we need to account for, in that we are generating cuts that are valid for Z of Eq. (2) in our context, and then imposing these cuts in (23) by fixing $x = \bar{x}$. As Proposition 5 below establishes, the key element that validates this variation is that for any binary feasible solution $\bar{x}$, if we denote the convex hull of the feasible region of the subproblem (21) as $Z(\bar{x})$ and view this region in the form

$$Z(\bar{x}) = \text{conv}\{(x, y) : Dy \geq b - Ax,\ y \in Y,\ \text{and}\ x = \bar{x}\}\,, \tag{42a}$$

then we effectively have that

$$Z(\bar{x}) = Z \cap \{(x, y) : x = \bar{x}\} \tag{42b}$$

since the right-hand side in (42b) defines a face of Z because Z includes the restrictions $0 \leq x \leq e$ in its definition. Consequently, we can derive the required description of the facial structure of $Z(\bar{x})$ given by (42a) that is necessary for solving the subproblem (21) by generating appropriate valid inequalities for Z, and then restricting $x = \bar{x}$. Figure 3 provides a flow-chart for such a cutting plane process in the context of lift-and-project cuts of Balas et al. (1993), and Remark 6 below provides comments on using more general RLT cuts along with some implementation suggestions. The following result establishes finite convergence of the procedure presented in Figure 3.

PROPOSITION 5. *The cutting plane procedure of Figure 3 finitely solves the subproblem (21) via (23), yielding a family of valid inequalities (23d) that can be re-used for any other subproblem by revising the corresponding first-stage decision* $\bar{x}$.

Proof. First of all, note that the cut generation process of Section 3 is based on deriving valid inequalities for relaxations of Z of the type (24), obtained by applying RLT while enforcing binariness on a single variable y_q to some system of type (43) (see Figure 3). Hence, inductively, each inequality generated of the form $\phi_t x + \alpha_t y \geq \beta_t$ is valid for Z, and therefore, can be imposed for any subproblem by fixing the x-variables to the corresponding first-stage decision values.

Next, let us view the subproblem (21) that is to be solved in the following form (augmented with an initial set of valid cuts), where x is declared to be a variable, but the parameter M is assumed to be sufficiently large so that we necessarily have $x = \bar{x}$ at optimality in this problem (44), as well as in its LP relaxation.

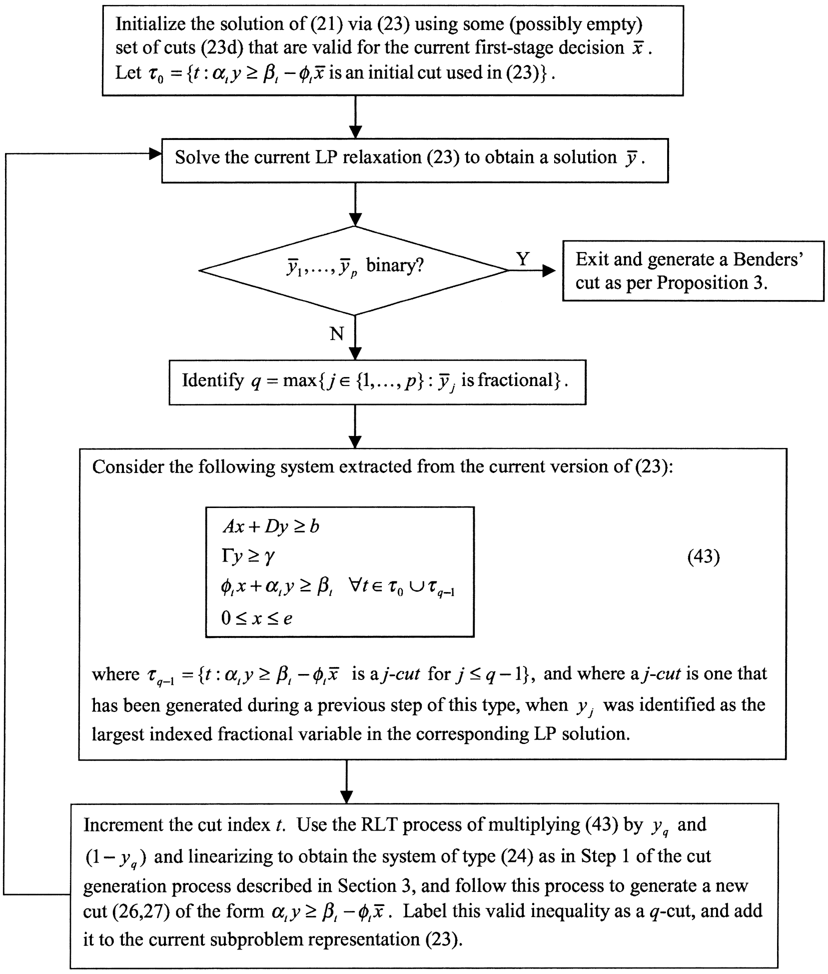

Figure 3. Cutting plane procedure for solving any subproblem.

$$
\begin{aligned}
v(\bar{x}) = c\bar{x} + \text{minimum} \quad & dy + M\left[\sum_{j:\bar{x}_j=0} x_j + \sum_{j:\bar{x}_j=1} (1 - x_j)\right] \\
\text{subject to} \quad & Ax + Dy \geq b \\
& \Gamma y \geq \gamma \\
& \phi_t x + \alpha_t y \geq \beta_t \ \ \forall t \in \tau_0 \\
& 0 \leq x \leq e, \ y_i \in \{0, 1\} \ \forall i = 1, \ldots, p\,.
\end{aligned}
\tag{44}
$$

Now, suppose that we apply the lift-and-project cutting plane procedure described in Balas et al. (1993) to Problem (44). By making M sufficiently large, we can assume that each LP relaxation solved in the (finite) iterative process will continue to yield

$x = \bar{x}$, so that each of these LP relaxations can effectively be solved via (23) by fixing $x = \bar{x}$ as in the flow-chart of Figure 3. Note that if $\bar{y}$ is a resulting extreme point solution, then $(\bar{x}, \bar{y})$ is a vertex of the continuous relaxation to (44) augmented with any additional cuts, since $x = \bar{x}$ describes a face of this latter region. Consequently, the procedure of Figure 3 is precisely the lift-and-project cutting plane scheme that is proven in Theorem 3.1 of Balas et al. (1993) to converge finitely as applied to Problem (44), and this completes the proof. □

REMARK 6. Note that the lift-and-project cutting plane procedure of Balas et al. (1993) is predicated on generating cuts based on enforcing binariness on 0–1 variables one at a time. A more general RLT process of Sherali and Adams (1990, 1994) could be used to devise a cut generation scheme that likewise enforces binariness on more than one variable at a time. In such a process, the 0–1 variables can be grouped into batches containing one or more variables per batch, perhaps based on the initial LP solution. A similar scheme as in Figure 3 could then be followed, in which the relaxation (24) of Z is generated by applying RLT while enforcing binariness on the highest indexed batch of variables that contains some fractionating variable(s), to a system (43) that contains cuts generated previously for lower-indexed batches. The convergence of such a problem would follow from Jeroslow's (1980) cutting plane game as in Proposition 5. Of course, the advantage of considering batches of cardinality one is that the associated separation problems are relatively easier to solve. However, Sherali et al. (2000) have recently demonstrated how stronger RLT cuts accruing from the simultaneous consideration of multiple variables can be efficiently generated by using suitably restricted projections of the associated dual cone. Furthermore, in practical implementations, one could employ all the retained cuts in (43) of the procedure of Figure 3 or consider the deletion of cuts based on certain filtering criteria as well. In addition, as alluded in Remarks 3 and 4, and as evident from the foregoing discussion, we could prematurely abort the solution of any particular subproblem for a given $x = \bar{x}$ via the described cutting plane scheme, and generate a corresponding valid Benders' cut. This might entail regenerating a previous $\bar{x}$, while not yet having solved Problem P. However, so long as complete subproblem solutions are enforced after a finite number of iterations or even finitely often, we would obtain an overall finitely convergent process. Investigations of this type require extensive computational experimentations that we hope to pursue in future research.

5. Summary and conclusions

In this paper, we have modified Benders' decomposition method using RLT and lift-and-project cuts to develop a new method for solving discrete optimization problems that yield 0–1 mixed-integer subproblems, such as those encountered in stochastic programs with integer recourse. Viewing the problem *implicitly* in the light of a suitably defined convex hull representation, with appropriate modifications when the original problem exhibits a dual-angular structure, we have demonstrated how cutting planes could be generated to derive a partial description of this convex hull representation as needed in order to devise a finitely convergent solution procedure. Importantly, the classes of cuts used in the subproblems were derived in terms of *functions* of the first-stage x-variables, enabling them to be re-used in subsequent subproblems simply by revising them according to the corresponding x-solutions. Additionally, globally valid Benders' cuts were obtained by recognizing these cuts as functions of the first-stage variables. The ability to reuse cutting planes from one subproblem to the next in this fashion is useful from the viewpoint of potentially reducing the computational effort required to solve the discrete subproblems, while providing globally valid Benders' cuts that enhance the lower-bounding mechanism via the relaxed master program. The focus of this paper has been on developing the theory for such a modified Benders' approach. A follow-up paper will conduct a variety of computational tests, particularly in the context of stochastic programs with integer recourse.

Acknowledgements

This material is based upon research supported by the National Science Foundation under Grant Numbers DMI-9812047 and DMI-0094462.

References

Adams, W.P. and H.D. Sherali (1993), Mixed-Integer Bilinear Programming Problems. *Mathematical Programming* **59**(3), 279–305.

Ahmed, S., M. Tawarmalani, and N.V. Sahindis (2000), A Finite Branch and Bound Algorithm for Two-Stage Stochastic Integer Programs. Working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Balas, E., S. Ceria and G. Cornuejols (1993), A Lift-and-Project Cutting Plane Algorithms for Mixed 0-1 Programs. *Mathematical Programming* **58**, 295–324.

Birge, J.R. and M.A.H. Dempster (1966), Stochastic Programming Approaches to Stochastic Scheduling, *Journal of Global Optimization* **9**(3-4), 417–451.

Birge, J.R. and F.V. Louveaux (1988), A Multicut Algorithm for Two-Stage Stochastic Linear Programs. *European Journal of Operational Research* **34**(3), 384–392.

Birge, J.R. and F.V. Louveaux (1997), *Introduction to Stochastic Programming*. New York, NY: Springer.

Carino, D.R., T. Kent, D.H. Meyers, C. Stacy, M. Sylvanus, A.L. Turner, K. Watanabe, and W.T. Ziemba (1994), The Russell-Yasuda Kasai Model: An Asset/Liability Model for a Japanese Insurance Company Using Multistage Stochastic Programming. *Interfaces* **24**(1), 29–49.

Caroe, C.C. and R. Schultz (1999), Dual Decomposition in Stochastic Integer Programming. *Operations Research Letters* **24**(1), 37–45.

Caroe, C.C. and J. Tind (1997), A Cutting-Plane Approach to Mixed 0-1 Stochastic Integer Programs. *European Journal of Operational Research* **101**(2), 306–316.

Caroe, C.C. and J. Tind (1998), L-Shaped Decomposition of two-Stage Stochastic Programs with Integer Recourse. *Mathematical Programming* **83**(3), 451–464.

Geoffrion, A. and R. McBride (1978), Lagrangean Relaxation Applied to Capacitated Facility Location Problems. *AIIE Transactions* **10**(1), 40–47.

Higle, J.L. and S. Sen (1991), Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse. *Mathematics of Operations Research* **16**(3), 650–669.

Higle, J.L. and S. Sen (1994), Conditional Stochastic Decomposition: An Algorithmic Interface for Optimization and Simulation. *Operations Research* **42**(2), 311–322.

Higle, J.L. and S. Sen (2000), The C^3 Theorem and a D^2 Algorithm for Large Scale Stochastic Integer Programming: Set Convexification. Working paper, Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85721. Also presented at the 17th International Symposium on Mathematical Programming, Atlanta, GA, August 7–11, 2000.

Jeroslow, R. (1980), A Cutting Plane Game for Facial Disjunctive Programs. *SIAM Journal on Control and Optimization* **18**(3), 264–280.

Kall, P. and S.W. Wallace (1994), *Stochastic Programming*. Chichester, England: John Wiley and Sons.

Klein Haneveld, W.K., L. Stougie, and M.H. van der Vlerk (1996), An Algorithm for the Construction of Convex Hulls in Simple Integer Recourse Programming. *Annals of Operations Research* **64**, 67–81.

Klein Haneveld, W.K. and M.H. van der Vlerk (1999), Stochastic Integer Programming: General Models and Algorithms. *Annals of Operations Research* **85**, 39–57.

Laporte, G. and F.V. Louveaux (1993), The Integer L-Shaped Method for Stochastic Integer Programs with Complete Recourse. *Operations Research Letters* **13**(3), 133–142.

Laporte, G., F.V. Louveaux, and H. Mercure (1992), The Vehicle Routing Problem with Stochastic Travel Times. *Transportation Science* **26**(3), 161–170.

Laporte, G., F.V. Louveaux, and L. van Hamme (1994), Exact Solution of a Stochastic Location Problem b an Integer L-Shaped Algorithm. *Transportation Science* **28**(2), 95–103.

Murphy, F.H., S. Sen, and A.L. Soyster (1982), Electric Utility Capacity Expansion Planning with Uncertain Load Forecasts. *AIIE Transactions* **14**, 52–59.

Nemhauser, G.L. and L.A. Wolsey (1999), *Integer and Combinatorial Optimization*. New York, NY: Wiley-Interscience.

Ruszczynski, A. (1999), Some Advances in Decomposition Methods for Stochastic Linear Programming. *Annals of Operations Research* **85**, 153–172.

Schultz, R. (1995), On Structure and Stability in Stochastic Programs with Random Technology Matrix and Complete Integer Recourse. *Mathematical Programming* **70**(1), 73–90.

Schultz, R., L. Stougie, and M.H. van der Vlerk (1996), Two-Stage Stochastic Integer Programming: A Survey. *Statistica Neerlandica* **50**(3), 404–416.

Schultz, R., L. Stougie, and M.H. van der Vlerk (1998), Solving Stochastic Programs with Integer Recourse by Enumeration: A Framework using Grobner Basis Reductions. *Mathematical Programming* **83**(2), 229–252.

Sherali, H.D. and W.P. Adams (1990), A Hierarchy of Relaxations Between the Continuous and Convex Hull Representations for Zero-One Programming Problems. *SIAM Journal on Discrete Mathematics* **3**(3), 411–430.

Sherali, H.D. and W.P. Adams (1994), A Hierarchy of Relaxations and Convex Hull Characterizations for Mixed-Integer Zero-One Programming Problems. *Discrete Applied Mathematics* **52**(1), 83–106.

Sherali, H.D. and W.P. Adams (1999), *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Boston, MA: Kluwer Academic Publishing.

Sherali, H.D., Y. Lee, and Y. Kim (2000), Partial Convexification Cuts. Manuscript, Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. Also presented at the 17th International Symposium on Mathematical Programming, Atlanta, GA, August 7–11, 2000.

Van Slyke, R.M. and R. Wets (1969), L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming. *SIAM Journal of Applied Mathematics* **17**(4), 638–663.