The C^3 Theorem and a D^2 Algorithm for Large Scale Stochastic Integer Programming: Set Convexification

Suvrajeet Sen
Julia L. Higle

Dept. of Systems and Industrial Engineering
The University of Arizona
Tucson, AZ 85721

October 2000

Abstract

This paper considers the two stage stochastic integer programming problems, with an emphasis on problems in which integer variables appear in the second stage. Drawing heavily on the theory of disjunctive programming, we characterize convexifications of the second stage problem and develop a decomposition-based algorithm for the solution of such problems. In particular, we verify that problems with fixed recourse are characterized by scenario-dependent second stage convexifications that have a great deal in common. We refer to this characterization as the C^3 (Common Cut Coefficients) Theorem. Based on the C^3 Theorem, we develop an algorithmic methodology that we refer to as Disjunctive Decomposition (D^2). We show that when the second stage consists of 0-1 MILP problems, we can obtain accurate second stage objective function estimates after finitely many steps. We also set the stage for comparisons between problems in which the first stage includes only 0-1 variables and those that allow both continuous and integer variables in the first stage.

This research was funded by the National Science Foundation under grant 9978780.

The C^3 Theorem and a D^2 Algorithm for Large Scale Stochastic Integer Programming: Set Convexification

1. Introduction

One of the more formidable classes of optimization problems arises from the incorporation of uncertainty in integer linear programming models. Such models arise in a variety of applications ranging from location and network design models (see e.g. Wallace [1988]), to unit commitment problems (e.g. Nowak and Römisch [2000]) and modeling power forwards for electric utilities. Indeed, one can conceive of Stochastic Integer Programming (SIP) problems arising whenever a deterministic IP model leads to an inadequate model under uncertainty. Depending on when the integer decisions are made, relative to the observations of outcomes of the random variables, there are several classes of SIP problems that might arise. For example, in a facility location problem, location decisions are typically made in advance of demand realizations, while in production planning problems, scheduling decisions are typically made after demand has been realized. We refer to the case in which integer decisions only appear prior to the realization of the random variables, as SIP1. Similarly, we refer to problems for which some integer decisions are made after observing the outcome as SIP2.

Starting with the early paper by Wollmer [1980], there have been some attempts to combine ideas from integer programming with those from stochastic programming to solve SIP1. As an example, Norkin, Ermoliev and Ruszczynski [1995] combine sample-based function evaluations with a branch and bound algorithm for SIP1. Note that since bounding in such a method is based on statistical estimates, one must be careful not to delete portions of the tree where the estimates are not very accurate. While the need for, and the novelty of such statistically motivated branch and bound methods is clear, SIP1 remains a class of problems in which the value function of the second stage LP can be approximated via standard techniques, such as those used in deterministic Benders' decomposition (see Laporte and Louveaux [1993]). However, the situation is substantially different for SIP2, in which the integer restrictions appear after an outcome of the random variable has been observed. In general, the SIP2 model may be stated in the following manner:

$$\operatorname*{Min}_{x \in X} c^\top x + E[f(x, \tilde{\omega})], \tag{1}$$

where $X \subseteq \Re^{n_1}$ is a set of feasible first stage decisions x (possibly continuous and/or integer), $\tilde{\omega}$ is a random variable defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and

$$f(x,\omega) = \text{Min } \; g_u^\top u + g_z^\top z \tag{2.1}$$

$$\text{s.t.} \quad W_u u + W_z z \;\; \geq r(\omega) - T(\omega)x \tag{2.2}$$

$$u \in \Re_+^{n_u}, z \; \in \mathcal{Z}_+^{n_z} \tag{2.3}$$

Within the stochastic programming literature, a realization of $\tilde{\omega}$ is known as a "scenario," and we shall adopt that terminology here. As such, the second stage problem that appears in (2) is often referred to as a "scenario subproblem."

Note that the scenario subproblem, (2), is a mixed-integer linear program (MILP) with the variables denoted z capturing the integer restrictions in the second stage. Note that (2.2) varies with both the first stage vector, x, as well as the scenario. We also note that in general, uncertainty may impact any data element in (2). For the purposes of this paper, we restrict our attention to the case in which the second stage cost vector and technology matrix, g and W respectively, are not subject to uncertainty.

In this paper, we develop algorithmic concepts that may be used to solve SIP2 problems, as stated in (1),(2). Implicitly, evaluation of the objective function in (1) involves the solution of (2) for each scenario, $\omega \in \Omega$, which can be a formidable task. Thus, in order to pose a manageable problem, we assume

A1. Ω is a finite set
A2. X is a closed set
A3. $f(x,\omega) < \infty$ for all $(x,\omega) \in X \times \Omega$.

Note that A1 affords the opportunity for an alternate formulation of SIP2, in which all possible scenarios are considered simultaneously, as follows:

$$\begin{aligned} \min_{x \in X} \; & cx + \sum_{\omega \in \Omega} (g_u^\top u_\omega + g_z^\top z_\omega) p_\omega \\ \text{s.t.} \; & T(\omega)x + W_u u_\omega + W_z z_\omega \geq r(\omega) \qquad \forall\, \omega \in \Omega \\ & u_\omega \in \Re_+^{n_u}\,, \; z_\omega \in Z_+^{n_z} \qquad \forall\, \omega \in \Omega \end{aligned} \tag{3}$$

where $p_\omega = \mathcal{P}\{\tilde{\omega} = \omega\}$. We refer to the statement (3) as the "deterministic equivalent" statement of SIP2, and note that (1, 2.1-3) and (3) are equivalent problem representations. Assumption A3, which explicitly requires that (2) is feasible for all $(x,\omega) \in X \times \Omega$ is quite standard in the stochastic programming literature, and is known as relatively complete (integer) recourse (Wets [1974]).

Despite the fairly large array of applications that lead to SIP2 problems, it is fair to suggest that very few attempts have been made to understand and characterize the structure of these problems, and even fewer attempts have been made to utilize structure for their solution. Notable exceptions to this remark are the papers of Caroe and Tind [1997,1998], Klein Haneveld, Stougie, and van der Vlerk [1995, 1996], Schultz, Stougie and van der Vlerk [1998], and Schultz [1993]. Indeed, the papers by Klein Haneveld et al [1995, 1996] provide an elegant approach to a specially structured SIP2 problem, called the *simple integer recourse (SIR)* problem. Here, the second stage integer program is the integer analog of the continuous simple recourse problem (see Birge and Louveaux [1997]), and is applicable in situations where the recourse decision involves a penalty for straying from the forecast. Just as the continuous simple recourse models arise in the newsvendor problem, one can envision the SIR problem arising in planning for "large ticket" items such as aircrafts, ships etc., or in planning with severe penalties for over/under production.

In the case of the general SIP2 model, results and algorithms are somewhat sparse. A survey of results in the area is provided in Klein Haneveld and van der Vlerk [1999]. Perhaps the most comprehensive algorithmic treatise that has appeared to date is the doctoral dissertation of Caroe [1998], and this work includes papers that appeared as Caroe and Tind [1997, 1998]. These authors must be congratulated for taking the first steps towards bringing IP results to bear on SIP problems. Nevertheless, much more research is necessary for this approach to mature. Caroe and Tind [1997] present SIP2 in terms of the deterministic equivalent, (3). More recently, Sherali and Fraticelli [2000] develop cutting plane methods when all variables (first and second stage) are binary. These methods are motivated by the fact that the inclusion of cutting planes within the deterministic equivalent problem retains the block angular structure of the large scale MILP. Note that while cutting plane methods are not particularly effective on their own, the use of these inequalities in branch and cut methods is known to be quite effective (see Martin [1999]).

Caroe and Tind [1998] use IP duality to represent the second stage objective as a subadditive function in the master program of a decomposition method. However, the value function of an integer program is a far more complicated object than is the value function of an LP. As shown in Blair and Jeroslow [1982], and Blair [1995], IP value functions may be composed from a nested combination of two operations: matrix multiplications and rounding. Due to the nesting of several rounding operations, these functions are not only difficult to construct, but difficult to optimize as well.

An approach based on polynomial ideal theory may prove beneficial in the context of SIP2. This approach to integer programming, pioneered by Conti and Traverso [1991], can be interpreted in geometric terms by associating monomials with lattice points in the set of integers. In the geometric setting, it is convenient to work with the notion of *test sets* (Thomas [1995]). Geometrically, a test set provides a finite set of directions, such that for any integer feasible point, one need only scan these directions to either obtain a better (integer) solution or declare the feasible solution optimal. One of the key observations that makes this approach attractive for SIP2 is that the characterization of a test set depends only on the cost and technology matrix of the second stage. Consequently, identification of a test set immediately opens the door to solutions of an entire family of problems that differ only in the right hand side. Schultz et al [1998] noted this advantage for problems with the structure of SIP2. While the ideas underlying this approach are elegant, the computational scale for such methods is unclear at this time. Moreover, the literature for this approach has thus far been restricted to pure integer problems.

Cutting plane theory for deterministic IP has seen several important advances (e.g. Balas [1979], Sherali and Adams [1990], etc.) Our goal is to bring together advances in IP cutting plane theory with SP decomposition methods to address problems with the SIP2 structure. In essence, we will be interested in developing strong lower bounding approximations for the scenario subproblem, (2). As in deterministic IP, the benefits from using strong bounds for SIP2 cannot be overestimated. Thus, if the first stage of SIP2 also contains integer first stage variables, and we are interested in using the stochastic branch and bound algorithm of Norkin, Ermoliev and Ruszczynski [1995], then better bounds will lead to more reliable pruning of the tree. One might also wish to use the lower bounds to assess the quality of a solution that may have been obtained using a heuristic. For instance, Lokkentangen and Woodruff [1996] have proposed a heuristic in which they have combined the progressive hedging (or scenario aggregation) algorithm of Rockafellar and Wets [1991], with the Tabu search (see Glover and Laguna [1997]). While such methods are able to provide solutions without excessive computational effort, it is often difficult to ascertain the quality of solutions provided by the algorithm. By developing good lower bounding estimates, it is possible to provide solutions with some guarantees about their quality.

Our bridge between integer programming and stochastic programming is the theory of disjunctive programming (Balas [1979], Sherali and Shetty [1980]). In this paper, we introduce a new class of algorithms in which both master and subproblems result from convexifications of two coupled disjunctive programs. We refer to this class of methods as Disjunctive Decomposition, or D^2, algorithms. This class of methods provides a general framework in which decomposition methods for stochastic integer programming can be devised. The approach studied in this paper begins with sequential convexifications of the second stage problem, and continues with an exploration of the manner in which this convexification impacts the first stage objective function.

This paper is organized as follows. In §2, we summarize some key results from disjunctive programming, and in §3 we present the theoretical framework from which the D^2 algorithm is derived. An important observation here is that for problems with fixed recourse (i.e., where $W = [W_u, W_z]$ in (2.2) is fixed), the second stage convexifications associated with different scenarios have a great deal in common. We refer to this characterization as the C^3 (Common Cut Coefficients) Theorem. The class of algorithms proposed in this paper are designed to take advantage of the C^3 Theorem. A basic D^2 method is presented in §4, and its extensions are discussed in §5. We emphasize that at this point, the class of algorithms presented in this paper is best construed as conceptual. There are a variety of computational challenges that must be explored before the approach becomes reality.

2. Disjunctive Programming

Disjunctive programming provides a rather general setting for the study of the convex hull of feasible points in integer programming and related problems (Sherali and Shetty [1980]). This line of work originated with Balas [1975], and Blair and Jeroslow [1978]. Their focus was on characterizing the convex hull of disjunctive sets of the form

$$\mathcal{S} = \cup_{h \in H} S_h \tag{4}$$

where H is a finite index set, and the sets S_h are polyhedral sets represented as

$$S_h = \{y \mid G_h y \geq r_h, y \geq 0\}. \tag{5}$$

Within the context of SIP2, the vector $y = (u, z)$ defined in (2) and r_h includes $r(\omega) - T(\omega)x$, which varies with the first stage decision, x, and the scenario, ω. In this notation, we put $n_2 = n_u + n_z$, and so $y \in \Re^{n_2}$. A disjunction stated as in (4),(5) is said to be in *disjunctive normal form* (i.e., none of the terms S_h contain any disjunction). It is important to recognize that the set of feasible solutions of any mixed-integer (0-1) program can be written as the union of polyhedra as in (4),(5) above. However, the number of elements in H can be exponentially large, thus making an explicit representation computationally impractical. If one is satisfied with weaker relaxations, then more manageable disjunctions can be stated. For example, the lift-and-project inequalities of Balas, Ceria and Cornuéjols [1993] use conjunctions associated with a linear relaxation together with one disjunction of the form: $z_j \leq 0$ or $z_j \geq 1$. (Of course, z_j is assumed to be a binary variable.) For such a disjunctive set, the cardinality of H is two, and the disjunction is manageable. Indeed, there is a hierarchy of disjunctions that one may use in developing relaxations of the integer program. Assuming that we have chosen some convenient level within the hierarchy, the index set H is given, and we may proceed to obtain convex relaxations of the nonconvex set.

A convex relaxation of a nonconvex set may be stated via a collection of *valid inequalities* which may be defined as follows.

Definition: An inequality $\pi^\top y \geq \pi_0$ is said to be a *valid inequality* for the set $\mathcal{S}$ if it is satisfied by all $y \in \mathcal{S}$; that is, $\mathcal{S} \subseteq \{y \mid \pi^\top y \geq \pi_0.\}$ ◇◇

The following result is known as the disjunctive cut principle. The forward part of this theorem is due to Balas [1975], and the converse is due to Blair and Jeroslow [1978]. In the following, the column vector G_{hj} denotes the j^{th} column of the matrix G_h.

Theorem 1. *Let $\mathcal{S}$ and S_h be defined as in (4),(5) respectively. If $\lambda_h \geq 0$ for all $h \in H$, then*

$$\sum_j \{\underset{h\in H}{\text{Max}}\, \lambda_h^\top G_{hj}\} y_j \geq \underset{h\in H}{\text{Min}}\, \lambda_h^\top r_h \tag{6}$$

is a valid inequality for $\mathcal{S}$. Conversely, suppose that $\pi^\top y \geq \pi_0$ is a valid inequality, and $H^* = \{h \in H | S_h \neq \emptyset\}$. There exists nonnegative $\{\lambda_h\}_{h\in H^*}$ such that

$$\pi_j \geq \underset{h\in H^*}{\text{Max}}\, \lambda_h^\top G_{hj}, \qquad \text{and} \qquad \pi_0 \leq \underset{h\in H^*}{\text{Min}}\, \lambda_h^\top r_h \tag{7}$$

Armed with this characterization of valid inequalities for the disjunctive set $\mathcal{S}$, we can develop a variety of relaxations of a mixed-integer linear program. The quality of the relaxations will, of course, depend on the choice of disjunction used, and the subset of valid inequalities used in the approximation. For the purposes of this section, we assume that the disjunction $\mathcal{S} = \cup_{h \in H} S_h$ is fixed, and we are interested in understanding which valid inequalities are most likely to lead to a strong relaxation. Thus we are interested in valid inequalities that yield the closure of the convex hull of $\mathcal{S}$, denoted clconv($\mathcal{S}$). The following result of Balas [1979] provides an important characterization of the facets of clconv($\mathcal{S}$).

Theorem 2. *Let the reverse polar of $\mathcal{S}$, denoted $\mathcal{S}^{\#}$, be defined as*

$$\mathcal{S}^{\#} = \{(\pi, \pi_0) \mid \text{there are nonnegative vectors } \{\lambda_h\}_{h \in H} \text{ such that (7) is satisfied}\}.$$

Assuming that $S_h \neq \emptyset$ and is full dimensional for all $h \in H$, every extreme point of the reverse polar can be associated with a facet of clconv($\mathcal{S}$). Conversely, with every facet of clconv($\mathcal{S}$) we can associate at least one extreme point of the reverse polar. ■

Theorem 2 provides access to a sufficiently rich collection of valid inequalities to permit clconv($\mathcal{S}$) to be obtained algorithmically. For cases in which the sets S_h have some special structure, this characterization can be useful for constructing the convex hull efficiently (see Sen and Sherali [1986]). Within the context of stochastic programming, this result has been used to characterize the convex hull of feasible points in probabilistically constrained problems with discrete random variables (Sen [1992]). In general however, generating all facets of clconv($\mathcal{S}$) can become a computational nightmare. Consequently, one adopts a sequential process in which only those facets that are deemed necessary are generated.

In studying the behavior of sequential cutting plane methods, it is important to recognize that without appropriate safeguards, one may not, in fact, recover the convex hull of the set of feasible integer points (see Jeroslow [1980], Sen and Sherali [1985]). In such cases, the cutting plane method may not converge. We maintain however, that this is essentially a theoretical concern since practical schemes use cutting planes in conjunction with a branch and bound method, which are of course convergent.

Before closing this summary, we discuss a certain special class of disjunctions for which sequential convexification (one variable at a time) does yield the requisite closure of the convex hull of integer feasible points. This class of disjunctions gives rise to facial disjunctive sets, which are described next.

A disjunctive set in *conjunctive normal form* may be stated in the form

$$\mathcal{S} = Y \cap_{\ell \in L} D_\ell$$

where Y is a polyhedron, and each set D_ℓ is a set defined by the union of finitely many halfspaces. The set $\mathcal{S}$ is said to possess the facial property if for each ℓ, every hyperplane used in the definition of D_ℓ contains some face of Y. It is not difficult to see that a 0-1 MILP is a facial disjunctive program. For these problems Y is a polyhedral set that includes the "box" constraints $0 \leq z \leq 1$, and using indices ℓ associated with $z \in \Re^{n_z}$, the disjunctive sets D_ℓ are defined as follows.

$$D_\ell = \{y \mid z_\ell \leq 0\} \cup \{y \mid z_\ell \geq 1\}.$$

Balas [1979] has shown that for sets with the facial property, one can recover the set $\mathrm{clconv}(\mathcal{S})$ by generating a sequence of convex hulls recursively by initializing $Q_0 = Y$, and for $\ell = 1, \ldots, n$

$$Q_\ell = \mathrm{clconv}(Q_{\ell-1} \cap D_\ell). \tag{8}$$

3. Foundations of Decomposition and Convexification

In designing a decomposition method for SIP2 problems, the notion of convexification (relaxation) can be used in a variety of ways. The methods considered in this paper are based on a combination of results from the theory of valid inequalities and decomposition methods for stochastic programming. Bringing these ideas into the realm of SP decomposition methods permits the integration of effective computational approaches from both domains. As shown in this and subsequent sections, disjunctive programming provides many of the tools necessary to build this bridge.

We consider a class of algorithms in which the scenario subproblem value function, $f(x,\omega)$ defined in (2), is approximated by the value function of a linear program. This approximation will be constructed sequentially by including valid inequalities in the second stage. We begin by presenting the Common Cut Coefficients (C^3) Theorem, which will allow us to build convex approximations recursively.

Theorem 3 (The C^3 Theorem). *Given (x,ω), let $Y(x,\omega) = \{y = (u,z) \mid Wy \geq r(\omega) - T(\omega)x\ ,\ u \in \Re_+^{n_u}\ ,\ z \in \Re_+^{n_z}\}$, the set of mixed-integer feasible solutions for the second stage MILP. Suppose that $\{C_h, d_h\}_{h \in H}$, is a finite collection of appropriately dimensioned matrices and vectors such that for all $(x,\omega) \in X \times \Omega$,*

$$Y(x,\omega) \subseteq \cup_{h \in H} \{y \in \Re_+^{n_2} \mid C_h y \geq d_h\}.$$

Let

$$S_h(x,\omega) = \{y \in \Re_+^{n_2} \mid Wy \geq r(\omega) - T(\omega)x,\ C_h y \geq d_h\},$$

and let

$$\mathcal{S}(x,\omega) = \cup_{h\in H} S_h(x,\omega).$$

Let $(\bar{x},\bar{\omega})$ be given, and suppose that $S_h(\bar{x},\bar{\omega})$ is nonempty for all $h \in H$ and $\pi^\top y \geq \pi_0(\bar{x},\bar{\omega})$ is a valid inequality for $\mathcal{S}(\bar{x},\bar{\omega})$. Then there exists a function, $\pi_0 : X \times \Omega \to \Re$ such that for all $(x,\omega) \in X \times \Omega$, $\pi^\top y \geq \pi_0(x,\omega)$ is a valid inequality for $\mathcal{S}(x,\omega)$.

Proof. Let G_{hj} denote the vector obtained by concatenating W_j with C_{hj}, and let $r_h(x,\omega)$ denote the vector obtained by concatenating $r(\omega) - T(\omega)x$ with d_h. Since $\pi^\top y \geq \pi_0(\bar{x},\bar{\omega})$ is a valid inequality for $\cup_{h\in H} S_h(\bar{x},\bar{\omega})$ and $S_h(\bar{x},\bar{\omega})$ is nonempty for all $h \in H$, Theorem 1 ensures the existence of nonnegative vectors $\{\lambda_h\}_{h\in H}$ such that

$$\pi_j \geq \operatorname*{Max}_{h\in H} \lambda_h^\top G_{hj} \quad ; \quad \pi_0(\bar{x},\bar{\omega}) \leq \operatorname*{Min}_{h\in H} \lambda_h^\top r_h(\bar{x},\bar{\omega}).$$

Let $(x,\omega) \in X \times \Omega$ be given. Since $y \geq 0$, we have

$$\pi^\top y \geq \sum_j \operatorname*{Max}_{h\in H} \lambda_h^\top G_{hj} y_j$$

and thus

$$\begin{aligned} \pi^\top y &\geq \lambda_h^\top G_h y \\ &\geq \lambda_h^\top r_h(x,\omega) && \forall\ h \in H \\ \Rightarrow \pi^\top y &\geq \operatorname*{Min}_{h\in H} \lambda_h^\top r_h(x,\omega) && \forall\ h \in H \\ &\equiv \pi_0(x,\omega). \end{aligned}$$

It follows that $\pi^\top y \geq \pi_0(x,\omega)$ is a valid inequality for $\cup_{h\in H} S_h(x,\omega)$. ■

The C^3 Theorem ensures that with a simple translation, valid inequalities derived for one pair $(\bar{x},\bar{\omega})$ may be used to derive valid inequalities for any other pair (x,ω). As such, we may obtain a lower bound approximation for the scenario subproblem objective function as follows:

$$\begin{aligned} f(x,\omega) \geq f_0(x,\omega) \equiv \text{Min}\ & g^\top y \\ \text{s.t.}\ & Wy \geq r(\omega) - T(\omega)x \\ & \pi^\top y \geq \pi_0(x,\omega) \\ & y \geq 0. \end{aligned} \tag{9}$$

As a result of the C^3 Theorem, only the right hand side of the subproblem constraints is affected by a change in x or ω, a property that is typically exploited by methods such as Benders' decomposition. However, the right hand side element $\pi_0(x,\omega)$ depends on both of its arguments, and it is important to characterize its behavior as a function of x.

Corollary 4. *Let H be a finite index set, and let $\{S_h(x,\omega)\}_{h\in H}$ and $\mathcal{S}(x,\omega)$ be defined as in Theorem 3. Let $\pi^\top y \geq \pi_0(x,\omega)$ denote a valid inequality for $\mathcal{S}(x,\omega)$. Then, for $h \in H$ there exist vectors $(\bar{\nu}_h(\omega), \bar{\gamma}_h(\omega)) \in \Re^{n_1+1}$ such that*

$$\pi_0(x,\omega) = \underset{h\in H}{\text{Min}}\{\bar{\nu}_h(\omega) - \bar{\gamma}_h(\omega)^\top x\}.$$

Proof. Since $\pi^\top y \geq \pi_0(x,\omega)$ is a valid inequality for $\mathcal{S}(x,\omega)$, Theorem 2 ensures that there exists $\lambda_h \geq 0$, $h \in H$ such that

$$\pi_0(x,\omega) = \underset{h\in H}{\text{Min}}\, \lambda_h^\top r_h(x,\omega),$$

where $r_h(x,\omega) = [r(\omega) - T(\omega)x; d_h]$ as in the proof of Theorem 3. Hence

$$\lambda_h^\top r_h(x,\omega) = \bar{\nu}_h(\omega) - \bar{\gamma}_h(\omega)x$$
$$\text{where} \quad \bar{\nu}_h(\omega) = \lambda_h^\top [r(\omega); d_h] \text{ and } \bar{\gamma}_h(\omega) = [T(\omega); 0]^\top \lambda_h,$$

and the result follows. ■

We note that the function $\pi_0(x,\omega)$ is a piecewise linear concave function of the first argument. Consequently, the lower bound approximations suggested in (9) will, in general, be non-convex. This contrasts sharply with standard resource-directive methods, such as Benders' decomposition (Benders [1962]), Kelley's method (Kelley [1960), and the L-Shaped method (Van Slyke and Wets [1969]), which capitalize on the linearity of the right-hand-side vector and piecewise linear convexity of the LP value function to develop piecewise linear convex approximations of the scenario subproblem objective function. Fortunately, when X is a polyhedral set, Theorem 2 permits the development of a convexification of $\pi_0(x,\omega)$. This strategy is borrowed from reverse convex programming in which disjunctive programming is used to provide facets of the convex hull of reverse convex sets (Sen and Sherali [1987]).

To begin, suppose that X is a polyhedral set, so that

$$X = \{x \in \Re^{n_1}_+ \mid Ax \geq b\},$$

where $A \in \Re^{m_1 \times n_1}$ and $b \in \Re^{m_1}$. Let

$$\Pi_X(\omega) = \{(\alpha, x) \mid x \in X, \alpha \geq \pi_0(x, \omega)\}, \tag{10}$$

the epigraph of $\pi_0(\cdot, \omega)$ restricted to $x \in X$. Finally, let

$$E_h(\omega) = \{(\alpha, x) \mid \alpha \geq \bar{\nu}_h(\omega) - \bar{\gamma}_h(\omega)x, Ax \geq b, x \geq 0\}.$$

Then $\Pi_X(\omega)$ can be defined in disjunctive normal form as

$$\Pi_X(\omega) = \cup_{h \in H} E_h(\omega).$$

We define an epi-reverse polar of this set, which we denote as $\Pi_X^\dagger(\omega)$, as follows.

$$\begin{aligned}
\Pi_X^\dagger(\omega) = \{\sigma_0(\omega) \in \Re, \sigma(\omega) \in \Re^{n_1}, \delta(\omega) \in \Re \mid \forall\, h \in H,\ \exists\, \tau_h \in \Re^{m_1}\ ,\ \tau_{0h} \in \Re \text{ s.t.}& \\
\sigma_0(\omega)\ \geq\ \tau_{0h} \quad \forall h \in H& \\
\sum_h \tau_{0h} \quad = 1& \\
\sigma_j(\omega) \geq \tau_h^\top A_j\ +\ \tau_{0h}\bar{\gamma}_{hj}(\omega) \quad \forall h \in H, j = 1, \ldots, n_1& \\
\delta(\omega) \leq \tau_h^\top b\ +\ \tau_{0h}\bar{\nu}_h(\omega) \quad \forall h \in H& \\
\tau_h \geq 0,\ \ \tau_{0h} \geq 0 \quad \forall h \in H\}&
\end{aligned} \tag{11}$$

Note that $\sigma_0(\omega) \geq \text{Max}_h\, \tau_{0h} > 0$. Hence, the epi-reverse polar only allows those facets (of the $\Pi_X(\omega)$) that have positive coefficient for the variable α. With this observation, the following is a direct consequence of Theorem 2.

Corollary 5. *Let $\{\bar{\nu}_h, \bar{\gamma}_h\}_{h \in H}$ be given, and let $\Pi_X(\omega)$ and $\Pi_X^\dagger(\omega)$ be as defined in (10) and (11), respectively. Then,*

$$\Pi_X(\omega) = \{(\alpha, x) \mid x \in X, \alpha \geq (\frac{\delta(\omega)}{\sigma_0(\omega)}) - (\frac{\sigma^\top(\omega)}{\sigma_0(\omega)})\ x,\ \forall(\sigma_0(\omega), \sigma(\omega), \delta(\omega)) \in \Pi_X^\dagger(\omega)\}$$

Note that when the first stage decisions include integer restrictions, Corollary 5 can be used by replacing X with stricter relaxations, such as those obtained through the Reformulation-Linearization Technique (Sherali and Adams [1990]).

Let $\{(\sigma_0^i(\omega), \sigma^i(\omega), \delta^i(\omega))\}_{i \in I}$ denote the set of extreme points of $\Pi_X^\dagger(\omega)$. Let $\nu_i(\omega) = \delta^i(\omega)/\sigma_0^i(\omega)$ and $\gamma_i(\omega) = \sigma^i(\omega)/\sigma_0^i(\omega)$, and define $\pi_c : X \times \Omega \to \Re$, where

$$\pi_c(x, \omega) = \underset{i \in I}{\text{Max}}\{\nu_i(\omega) - \gamma_i^\top(\omega)x\}.$$

That is, $\{(\alpha, x) \mid x \in X, \alpha \geq \pi_c(x, \omega)\}$, the epigraph of $\pi_c(x, \omega)$ restricted to $x \in X$, agrees with $\text{clconv}(\Pi_X(\omega))$. For this reason, we refer to $\pi_c(\cdot, \omega)$ as the convex hull approximation of $\pi_0(\cdot, \omega)$. For future reference, it is worth noting that $\pi_0(x, \omega) = \pi_c(x, \omega)$ whenever x is an extreme point of X.

In the forthcoming section, we discuss algorithmic approaches to decomposition and convexification. In doing so, convexification will take place in an iterative fashion. To facilitate the algorithmic presentation, it is convenient to collect some of the quantities we have defined into a matrix notation. The C^3 Theorem ensures that we may represent valid inequalities in the form $\pi^\top y \geq \pi_c(x, \omega)$. That is, only the right hand side varies with x and ω. If we have iteratively identified k such inequalities, $\{\pi^t, \pi_c^t(x, \omega)\}_{t=1}^k$, then the matrix of coefficients in the second stage, which is W augmented by $\{(\pi^t)^\top\}_{t=1}^k$, will be denoted as W^k. Similarly, the right hand side vector, which consists of $r(\omega) - T(\omega)x$ augmented by $\{\pi_c^t(x, \omega)\}_{t=1}^k$, will be denoted as $\rho_c^k(x, \omega)$. Thus, we define

$$f_c^k(x, \omega) = \text{Min } \; g^\top y \tag{12.1}$$

$$\text{s.t.} \quad W^k y \; \geq \rho_c^k(x, \omega) \tag{12.2}$$

$$y \; \in \Re_+^{n_2} \tag{12.3}$$

and note that $f_c^k(x, \omega) \leq f(x, \omega)$ for all $(x, \omega) \in X \times \Omega$. Of course, if x is an extreme point, and the solution to (12) satisfies the integrality constraints, then $f_c^k(x, \omega) = f(x, \omega)$.

Our results thus far have focussed exclusively on the development of convexifications of the scenario subproblems. Before continuing on to the next section, a few comments on the manner in which these convexifications impact a decomposition procedure are in order. As we have suggested throughout, the problem SIP2 naturally lends itself to a temporal decomposition based on its stages (1) and (2). In such a decomposition, the challenge lies in the development of approximations of the second stage objective function to be used in the first stage or "master" problem. In SIP2, this challenge is further compounded by the need to iteratively improve the quality of the subproblem approximations by the addition of valid inequalities of the form $(\pi^k)^\top y \geq \pi_c^k(x, \omega)$. Note that the second stage objective function is a weighted sum of all of the scenario subproblem objective functions, $E[f(x, \tilde{\omega})] = \sum_{\omega \in \Omega} f(x, \omega) p_\omega$. In (12), these have been approximated by the convex lower bounds, $\{f_c^k(x, \omega)\}_{\omega \in \Omega}$. Note that unlike $f(x, \omega)$, $f_c^k(x, \omega)$ involves only continuous variables, and hence is easier to compute. Furthermore, the convexity of $\rho_c^k(\cdot, \omega)$ ensures that if we pass a Benders'-type optimality cut (i.e a subgradient of f_c^k) to the master program, the resulting approximation (in the master) remains a lower bound on the function $f(x, \omega)$.

When forming the objective function to be used in the master program, one must specify the manner in which scenarios are aggregated. At one end of the spectrum (of aggregations) is an approximation based on the expectation operator, in which all scenarios are combined to obtain a single functional approximation. At the other end of the spectrum is the multicut method where the master program maintains separate approximations for each scenario. Between these two extremes are a variety of aggregations, and specific algorithms are realized based on the level of aggregation chosen in the master program. Without specifying this level of aggregation, let $F^k(x)$ denote the approximation of the subproblem objective function used in the k^{th} master program, and suppose that $F^k(x) \le E[f(x,\tilde{\omega})]$ for all $x \in X$, for all k. The decomposition methods we consider generate a sequence of first stage solutions $\{x^k\}$ such that

$$x^k \in \operatorname{argmin}\{c^\top x + F^{k-1}(x) \mid x \in X\}. \tag{13}$$

For continuous stochastic programming problems, well-trodden paths to convergent decomposition algorithms rely on either epi-convergence (Rockafellar and Wets [1997]) or epigraphical nesting (Higle and Sen [1992], [1995]). Both of these approaches call for some version of asymptotic accuracy of the approximations. The following result provides a sufficient condition for convergence of solutions generated through (13).

Theorem 6. *Let $\{F^k\}$ denote the sequence of second stage objective function approximations used in the master program, and let*

$$x^k \in \operatorname{argmin}\{c^\top x + F^{k-1}(x) \mid x \in X\}.$$

If X is a closed set, $F^k(x) \le E[f(x,\tilde{\omega}))]$ for all $x \in X$, for all k, and

$$\{x^k\}_{\mathcal{K}} \to \bar{x} \;\Rightarrow\; \lim_{k\in\mathcal{K}} F^{k-1}(x^k) = E[f(\bar{x},\tilde{\omega})] \tag{14}$$

then every accumulation point of $\{x^k\}$ is an optimal solution to (1).

Proof. Let $v^* = \operatorname{Min}\{c^\top x + E[f(x,\tilde{\omega})] \mid x \in X\}$, and suppose that $\{x^k\}_{\mathcal{K}} \to \bar{x}$. Since X is a closed set, $\bar{x} \in X$. It follows that

$$c^\top x^k + F^{k-1}(x^k) \le v^* \le c^\top \bar{x} + E[f(\bar{x},\tilde{\omega})].$$

Since $\{F^{k-1}(x^k)\}_{\mathcal{K}} \to E[f(\bar{x},\tilde{\omega})]$, the result follows. ■

Note that (14) requires that the lower bounding approximations attain local accuracy asymptotically as iterates converge. In decomposition methods for stochastic linear programming (SLP) problems, one often has $F^k(x^k) = E[f(x^k, \tilde{\omega})]$, so that the main effort in proving convergence amounts to studying the difference between $F^{k-1}(x^k)$ and $F^k(x^k)$ or $E[f(x^k, \tilde{\omega})]$. In deterministic decomposition methods, this difference vanishes, whereas in stochastic decomposition methods, this difference vanishes with probability one. For SIP problems, the complexity associated with solving integer programs at each iteration may preclude the condition $F^k(x^k) = E[f(x^k, \tilde{\omega})]$ for all k. Instead, we will require the approximations to agree with the actual values after finitely many iterations. How this may be accomplished is addressed in the following section.

4. Algorithmic Aspects of Decomposition and Convexification

In developing algorithmic approaches to decomposition and convexification, the C^3 Theorem prompts us to consider methods in which outcomes are able to share cut data structures among alternative scenarios. In the following, we discuss issues related to approximations of the second stage feasible region as a function of (x, ω). As before, we will assume that the complete integer recourse assumption, A3, is in effect, so that the second stage MILP is feasible for each scenario $\omega \in \Omega$.

In drawing connections between this section and §3, it is useful to recognize the iterative nature of the decomposition-convexification. In the k^{th} iteration,

- A master problem forwards a first stage solution, which we denote as x^k.
- Given x^k, and the convex approximations developed in the first $k-1$ iterations, the k^{th} refinement of the approximation of the scenario subproblems is developed, and an updated representation of the second stage objective function is obtained.

Because the scenario subproblem approximations are dynamic throughout the course of the algorithm, many of the entities that appear to be static in §3 will carry iteration indices in this section. Thus, for example,

- $\pi_0(x, \omega)$ and $\pi_c(x, \omega)$ will be replaced by $\pi_0^k(x, \omega)$ and $\pi_c^k(x, \omega)$, respectively
- $(\bar{\nu}(\omega), \bar{\gamma}(\omega))$ will be replaced by $(\bar{\nu}^k(\omega), \bar{\gamma}^k(\omega))$, etc.

4.1 A Basic D^2 Algorithm

We begin by discussing the manner in which the scenario subproblems are approximated. For a given x^k in iteration k, these are of the form

$$\begin{aligned} f_c^k(x^k,\omega) = \text{Min } & g^\top y \\ \text{s.t. } & W^k y \geq \rho_c^k(x^k,\omega) \\ & y \in \Re_+^{n_2}, \end{aligned}$$

where $\rho_c^k(x^k,\omega) = r^k(\omega) - T^k(\omega)x^k$. Referring to (2), this subproblem is initialized with $W^1 = W$, $r^1(\omega) = r(\omega)$, and $T^1(\omega) = T(\omega)$. These elements are updated as iterations progress. Let $y^k(\omega) \in \text{argmin } \{g^\top y \mid W^k y \geq \rho_c^k(x^k,\omega),\ y \in \Re_+^{n_2}\}$. If $z^k(\omega)$, the value assigned to integer variables in $y^k(\omega)$ is integer for all ω, then no update is necessary, and $W^{k+1} = W^k$, $r^{k+1}(\omega) = r^k(\omega)$ and $T^{k+1}(\omega) = T^k(\omega)$.

On the otherhand, suppose that the subproblems do not yield integer optimal solutions. Let $j(k)$ denote an index, j, for which $z_j^k(\omega)$ is non-integer for some $\omega \in \Omega$. Let $\bar{z}_{j(k)}$ denote one of the non-integer values $\{z_j^k(\omega)\}_{\omega\in\Omega}$. For example, $\bar{z}_{j(k)}$ could be either the "Min" or the "Max" among these non-integer values. To eliminate this non-integer solution, a disjunction of the form

$$\mathcal{S}_k(x^k,\omega) = S_{0,j(k)}(x^k,\omega) \cup S_{1,j(k)}(x^k,\omega), \tag{15}$$

where

$$\begin{aligned} S_{0,j(k)}(x^k,\omega) =\{ & y \in \Re_+^{n_2} \text{ such that} \\ & W^k y \geq \rho_c^k(x^k,\omega) && (16.1) \\ & - z_{j(k)} \geq -\lfloor \bar{z}_{j(k)} \rfloor \} && (16.2) \end{aligned}$$

$$\begin{aligned} S_{1,j(k)}(x^k,\omega) =\{ & y \in \Re_+^{n_2} \text{ such that} \\ & W^k y \geq \rho_c^k(x^k,\omega) && (17.1) \\ & z_{j(k)} \geq \lceil \bar{z}_{j(k)} \rceil \} && (17.2) \end{aligned}$$

may be used.

We will refer to $j(k)$ as the "disjunction variable" for iteration k. Since the disjunction will be based on an either-or condition, we will use $H = \{0, 1\}$. Note that when the integer restrictions are binary, the right hand side of (16.2) is zero, and the right hand side of (17.2) is one, and as indicated earlier, this is precisely the disjunction used in lift-and-project cuts of Balas, Ceria and Cornuéjols [1993].

In forming a valid inequality for the disjunction (15), the multipliers associated with (16.1) will be denoted $\lambda_{0,1}$ and the scalar multiplier associated with (16.2) will be denoted $\lambda_{0,2}$. Similarly, we associate $\lambda_{1,1}$ and $\lambda_{1,2}$ with (17.1) and (17.2) respectively. Following the standard approach of generating valid inequalities in disjunctive programming (see, e.g. Sherali and Shetty [1980]), we can develop a linear program that optimizes some measure of distance of the current solutions $y^k(\omega)$ from the cut. It is interesting to note that within the context of SIP2, we may actually pose this linear program as a particular type of *stochastic* linear program, whose structure has been well-studied.

We begin by assuming that the sets defined in (16) and (17) are non-empty for all $\omega \in \Omega$. Let

$$I_j^k = \begin{cases} 0 & \text{if } j \neq j(k) \\ 1 & \text{otherwise.} \end{cases}$$

The following LP/SLP may be used to generate cut coefficients in iteration k.

$$\text{Max} \quad E[\pi_0(\tilde{\omega})] - E[y^k(\tilde{\omega})]^\top \pi \tag{18.1}$$

$$\text{s.t.} \quad \pi_j \geq \lambda_{0,1}^\top W_j^k - I_j^k \lambda_{0,2} \quad \forall j \tag{18.2}$$

$$\pi_j \geq \lambda_{1,1}^\top W_j^k + I_j^k \lambda_{1,2} \quad \forall j \tag{18.3}$$

$$\pi_0(\omega) \leq \lambda_{0,1}^\top \rho_c^k(x^k, \omega) - \lambda_{0,2} \lfloor \bar{z}_{j(k)} \rfloor \quad \forall\, \omega \in \Omega \tag{18.4}$$

$$\pi_0(\omega) \leq \lambda_{1,1}^\top \rho_c^k(x^k, \omega) + \lambda_{1,2} \lceil \bar{z}_{j(k)} \rceil \quad \forall\, \omega \in \Omega \tag{18.5}$$

$$-1 \leq \pi_j \leq 1, \forall j, \quad -1 \leq \pi_0(\omega) \leq 1, \forall \omega \in \Omega \tag{18.6}$$

$$\lambda_{0,1}, \lambda_{0,2}, \lambda_{1,1}, \lambda_{1,2} \geq 0 \tag{18.7}$$

The bounds included in (18.6) "scale" the cut coefficients, and do not eliminate any valid inequalities.

Note that this problem is essentially a stochastic version of the linear program used to generate the lift-and-project cuts. Within the context of stochastic programming problems, (18) has a well-studied structure. The variables π and λ are constant with respect to ω, a characteristic known as nonanticipativity. The remaining variables, $\{\pi_0(\omega)\}_{\omega\in\Omega}$ are not. Note that for a given collection of λ values, $\{\pi_0(\omega)\}_{\omega\in\Omega}$ are trivially determined. Indeed, one sees that (18) is a simple recourse problem (Wets [1974]). Such problems are generally regarded as being among the easiest SLP's to solve.

As with linear programming problems, it is not difficult to show that there exists an extreme point $(\pi^k, \pi_0^k(\omega))$ of the reverse polar $S^{\#}(x^k, \omega)$ such that $(\pi^k, \{\pi_0^k(\omega)\}_{\omega\in\Omega})$ is an optimal solution to (18). Note that if the optimal objective value is positive, a valid inequality that eliminates the current non-integer solution is given by $(\pi^k)^\top y \geq \pi_0^k(\omega)$. The reader should note that in order to view these terms relative to our results in §3, the scalar values $\{\pi_0^k(\omega)\}_{\omega\in\Omega}$ may be interpreted as $\{\pi_0^k(x^k, \omega)\}_{\omega\in\Omega}$. Here, the superscript k reflects the iterative nature of the convexification process. However, in order to preserve convex approximations in the first stage, the cut we propose will have the form $(\pi^k)^\top y \geq \pi_c^k(x^k, \omega)$, as previously discussed.

Remark: In order to solve (18), we note that it is more convenient to work with its dual, rather than (18) itself. The set of dual feasible solutions may be interpreted as one that chooses points in (16) and (17) that create the convex hull associated with the disjunction. Hence, identifying feasibility of (16) and (17) becomes part of a Phase 1 procedure in solving the dual to (18). Note that under the complete recourse assumption at least one of the sets defined in (16) or (17) must be non-empty. If both are non-empty, then we use them in the LP/SLP as specified in (18). On the other hand, if there is an outcome ω for which one of the sets is infeasible, then that outcome is not relevant with respect to the disjunction, and consequently does not effect the cut. It is deleted in (18). Let Ω' denote the subset of outcomes for which (16) and (17) are both feasible. Then (18) will be modified to include only those outcomes that belong to Ω'. We will refer to this LP/SLP as $(18')$. In the interest of brevity, we do not state $(18')$ explicitly. $\diamondsuit\diamondsuit$

We now turn to calculations associated with $\pi_c^k(x, \omega)$, which requires the specification of the coefficients $\nu^k(\omega)$ and $\gamma^k(\omega)$. These coefficients will be used to define the updated function $r^{k+1}(\omega) = [r^k(\omega), \nu^k(\omega)]$, as well as the updated matrix $T^{k+1}(\omega)$, which is obtained by appending the row $\gamma^k(\omega)^\top$ to the matrix $T^k(\omega)$. In deriving these quantities, Theorem 1 ensures that valid inequalities may be derived from any nonnegative multiplier vector. Thus, although $(18')$ includes only $\omega \in \Omega'$, the vector $(\lambda_{h,1}, \lambda_{h,2})_{h\in H}$ obtained from the solution of $(18')$ may be used to derive valid inequalities for $\mathcal{S}^k(x^k, \omega)$.

The parameters $\{\nu^k(\omega), \gamma^k(\omega)\}$ can be obtained via Corollary 4 and Corollary 5. Since the disjunction used for cut formation has $H = \{0, 1\}$, the epigraph of $\pi_0(x, \omega)$ is a union of two polyhedral sets. Following the proof of Corollary 4, for all $\omega \in \Omega$

$$\bar{\nu}_0^k(\omega) = \lambda_{0,1}^\top r^k(\omega) - \lambda_{0,2}\lfloor \bar{z}_{j(k)} \rfloor, \quad \bar{\nu}_1^k(\omega) = \lambda_{1,1}^\top r^k(\omega) + \lambda_{1,2}\lceil \bar{z}_{j(k)} \rceil,$$

and

$$[\bar{\gamma}_h^k(\omega)]^\top = \lambda_{h,1}^\top T^k(\omega), \quad h \in H.$$

Using these parameters as data for the polyhedron $\Pi_X^\dagger(\omega)$ as defined in (11) the following LP can be solved to define $\pi_c^k(x,\omega)$ for each $\omega \in \Omega$.

$$\begin{aligned} &\text{Max } \delta(\omega) - \sigma_0(\omega) - (x^k)^\top \sigma(\omega) \\ &(\delta(\omega), \sigma_0(\omega), \sigma(\omega)) \in (\Pi_X^\dagger(\omega))^k, \\ &\quad \|(\delta(\omega), \sigma_0(\omega), \sigma(\omega))\|_\infty \leq 1 \end{aligned} \tag{19}$$

Here, $(\Pi_X^\dagger(\omega))^k$ denotes the epi-reverse polar (i.e set of facets of the convex hull of the epigraph) of $\pi_0^k(x,\omega)$. Let $(\sigma_0^k(\omega), \sigma^k(\omega), \delta^k(\omega))$ denote an optimal solution to (19). Then for $\omega \in \Omega$, we use Corollary 5 to obtain

$$\nu^k(\omega) = \frac{\delta^k(\omega)}{\sigma_0^k(\omega)} \tag{20.1}$$

and

$$\gamma^k(\omega) = \frac{\sigma^k(\omega)}{\sigma_0^k(\omega)}. \tag{20.2}$$

Finally, for each $\omega \in \Omega$, $\pi_c^k(x^k,\omega) = \nu^k(\omega) - \gamma^k(\omega)^\top x^k$ approximates $\pi_c(x^k,\omega)$ in the right hand side of the new row of the updated second stage LP for outcome ω. Note that if there exists an $\omega' \in \Omega'$ such that $\pi_c^k(\omega') - y^k(\omega')^\top \pi^k > 0$ then a cut of the form $(\pi^k)^\top y \geq \pi_c^k(\omega')$ in the scenario subproblem approximation eliminates the point $(x^k, y^k(\omega'))$ from the LP relaxation of the deterministic equivalent. That is, the cuts used in the D^2 algorithm obey the requirements imposed by Caroe and Tind [1997]. However, since our development promotes a partitioning approach (in cut generation), both the size and the number of cut-generation LPs is reduced dramatically. Note also that unlike the cut generation problem for the second stage convexification, the size of the LP in (19) remains fixed at roughly $2n_1$ structural variables, and the same number of constraints. On the other hand, because the matrix W is augmented sequentially, the cut generation LP/SLP in (18) must grow in size. We note that this is true of all sequential convexification methods, including the one proposed by Caroe and Tind [1997]. However, by using the C^3 Theorem, we are able to restrict the growth of the cut generation LP/SLP. We now summarize a D^2 (Disjunctive Decomposition) algorithm based on set convexification of the second stage (MILP).

A Basic D^2 Algorithm

0. Initialize. Let $\epsilon > 0$ and $x^1 \in X$ be given. Let $k \leftarrow 1$ and initialize an upper bound $V_0 = \infty$, and a lower bound $v_0 = -\infty$. Put $W^1 = W, T^1(\omega) = T(\omega), r^1(\omega) = r(\omega)$.

1. Solve one LP Subproblem for each $\omega \in \Omega$
Put $V_k \leftarrow V_{k-1}$. Use the matrix W^k as well as the right hand side vector $\rho_c^k(x^k, \omega) = r^k(\omega) - T^k(\omega)x^k$ to solve (12.1)-(12.3) for each $\omega \in \Omega$. If $y^k(\omega)$ satisfy the integer restrictions for all $\omega \in \Omega$, $V_k \leftarrow \text{Min}\{c^\top x^k + E[f(x^k, \tilde{\omega})], V_k\}$, and go to step 4.

2. Solve Multiplier/Cut Generation LP/SLP (18′) and Perform Updates
   (i) Choose a disjunction variable $j(k)$ and formulate (18′). This process identifies Ω', as defined in the Remark. For $\omega \notin \Omega'$ record which "branch" (either (16.2) or (17.2)) to include in the subproblem. Solve (18′) to obtain π^k, and define W^{k+1} by appending the new row $[\pi^k]^\top$ to W^k.
   (ii) Using the multipliers λ_0^k, λ_1^k and the value $\bar{z}_{j(k)}$ obtained in (i) solve (19) for each outcome ω. The solution defines $\nu^k(\omega)$ and $\gamma^k(\omega)$ which are then used to update $r^{k+1}(\omega) = [r^k(\omega), \nu^k(\omega)]$ and $T^{k+1}(\omega)$. The latter is obtained by appending $[\gamma^k(\omega)]^\top$ to the matrix $T^k(\omega)$.

3. Update and Solve one LP Subproblem for each $\omega \in \Omega$
For each $\omega \in \Omega'$ solve the updated LP using W^{k+1} and $\rho_c^{k+1}(x^k, \omega)$ in (9). For $\omega \notin \Omega'$ solve the subproblem associated with the "branch" identified in step 2(i). If $y^k(\omega)$ satisfy the integer restrictions for all $\omega \in \Omega$, $V_k \leftarrow \text{Min}\{c^\top x^k + E[f(x^k, \tilde{\omega})], V_k\}$.

4. Update and Solve the Master Problem
Using the dual multipliers from the most recently solved subproblem (either step 1 or step 3) update the approximation F^k by adopting a standard decomposition method (e.g. Benders [1962]). Let $x^{k+1} \in \text{argmin}\ \{c^\top x + F^k(x) \mid x \in X\}$, and let v^k denote the optimal value of the master problem. If $V_k - v_k \leq \epsilon$, stop. Otherwise, $k \leftarrow k+1$ and repeat from 1.

When some of the first stage variables x are also restricted to be integer, the set X should be replaced by some convexification. Furthermore in this case, it is not advisable to solve the master program in Step 4 to optimality (see McDaniel and Devine [1977]).

Instead, it is recommended that LP relaxations of the MILP master problem be solved until the LP solutions begin to stabilize, at which point, MILP master programs should be solved. Note that if such a strategy is adopted, we should suspend the updates of V_k in those iterations at which the first stage decision is non-integer.

4.2 On Guaranteeing an Optimal First Stage Solution

We note that approximations used within a D^2 algorithm satisfy $F^k(x) \leq E[f(x, \tilde{\omega})]$ for all $x \in X$ for all k. Theorem 6 ensures that the points identified in Step 4 accumulate at optimal solutions, provided that $\{F^{k-1}(x^k)\}_{\mathcal{K}} \to E[f(\bar{x}, \tilde{\omega})]$ whenever $\{x^k\}_{\mathcal{K}} \to \bar{x}$. Unfortunately, this may not happen in general. That is, without appropriate safeguards, the cutting plane method may fail to yield IP solutions. Our safeguards are in the form of an algorithmic modification, and require additional assumptions on the problem. In addition to assumptions A.1-A.3 stated at the outset, we assume that the second stage integer variables are binary, so that the scenario subproblems are 0-1 MILPs.

Because a 0-1 MILP is a facial disjunctive program, one is tempted to use (8), where all valid inequalities associated with a disjunction from one variable are generated before inequalities from the next disjunction (e.g. Balas [1979]). However, as noted in Balas [1997], such a scheme is not very practical since all basic solutions of a (scaled) reverse polar (associated with one disjunction variable) must be generated at once, and in a subsequent iteration, a similar list of basic solutions (associated with some other disjunction variable) is created. For such an approach, the concerns are not only limited to the complexity of enumerating exponentially many basic solutions and the complexity of solving rapidly growing LPs; there is also the concern that such a scheme, being driven by worst-case analysis, will attain the worst-case (exponential) bound for even those instances which are solvable without all basic solutions.

In the remainder of this subsection we discuss a somewhat more implementable approach that also guarantees that the convex hull can be generated, if necessary. This discussion is inspired by the convergence results of Blair [1980], Jeroslow [1980], and Sen and Sherali [1985]. However, unlike the methods of those papers, the disjunctions that we allow are restricted to use at most two atoms (i.e., $|H| = 2$), namely ($z_j \leq 0$ or $z_j \geq 1$). As discussed earlier, this restriction curtails the proliferation of cuts. Rather than construct the entire set of valid inequalities for one disjunction variable at a time, as in (8), the modification permits the disjunction variables to be considered in an arbitrary sequence. However, to achieve the technical advantages afforded by the one-variable-at-a-time approach, we will control the structure of the cut-identification problem with which each disjunctive variable is associated. In doing so, we will achieve finite convergence of the subproblems.

We begin by defining a vector, $m^k \in \Re^{n_z}$. If the integer variable j is used as the disjunction variable in the k^{th} iteration, the j^{th} component of this vector, denoted as $m^k(j)$, is used to specify the matrix to be used in Step 2 to identify a new cut. In particular, $m^k(j)$ corresponds to an iteration index, so that $W^{m^k(j)}$ is the matrix used in the cut formation problem.

Modification of Steps 2 and 3 for Convergence

2. Solve a Multiplier/Cut Generation LP/SLP, and Perform Updates

(i_a) If $z^k(\omega)$ is binary for all ω, go to 3(ii).
Else, calculate as follows. When $k = 1$, we define $m^k(j) = 1, \forall j$. For all other k, m^k is assumed to have been calculated previously. $j \leftarrow 1$, $J_1^k(\omega) \leftarrow \emptyset$, $J_0^k(\omega) \leftarrow \emptyset$.

(i_b) If there exists $\omega \in \Omega$ such that $z_j^k(\omega)$ is non-integer, then formulate $(18')$ using $W^{m^k(j)}$ (instead of W^k), and identify Ω'_j.
For all $\omega \notin \Omega'_j$, include j in $J_0^k(\omega)$ (or $J_1^k(\omega)$) if (16) (or (17)) is feasible. For these scenarios, revise $y^k(\omega)$ by solving (12) using appropriately fixed variables identified in $J_0^k(\omega)$ and $J_1^k(\omega)$.
If $\Omega'_j \neq \emptyset$ and there exists $\omega' \in \Omega'_j$, such that $\pi_c^k(x^k, \omega') - [y^k(\omega')]^\top \pi^k > 0$, then go to $(2i_c)$. (Here $(\pi^k, \pi_c^k(x^k, \omega))$ are obtained by solving $(18')$ and (19).)
If $j = n, W^{k+1} = W^k, T^{k+1} = T^k$ and $r^{k+1} = r^k$. Go to 3(ii).
Otherwise, increment j and repeat $(2i_b)$.

(i_c) $j(k) = j$ (obtained in $(2i_b)$). For $j > j(k)$, put $m^{k+1}(j) \leftarrow k + 1$.

(ii) Update W^{k+1}, T^{k+1} and r^{k+1} using parameters identified in $(2i_b)$.

3. Update and Solve one LP Subproblem for each $\omega \in \Omega$

(i) For each $\omega \in \Omega'_{j(k)}$ setup the updated LP, fix the variables in $J_1^k(\omega)$ to be one, and those in $J_0^k(\omega)$ to be zero, and solve the resulting LP. For $\omega \notin \Omega'_{j(k)}$ simply fix the variables in $J_1^k(\omega)$ to be one, and those in $J_0^k(\omega)$ to be zero, and solve the resulting LP.

(ii) If $z^k(\omega)$ is binary for all $\omega \in \Omega$, $V_k \leftarrow \text{Min}\{c^\top x^k + E[f(x^k, \tilde{\omega})], V_k\}$.

Before proceeding to the analysis, we observe that the quantities $m^k(j)$ calculated during the execution of step $(2i_b)$ in iteration k satisfy $m^k(j) \le k$. Hence, $W^{m^k(j(k))}$ is a well defined matrix and so is the associated LP/SLP $(18')$.

Lemma 7. *Suppose that* $X = \{x \in \Re_+^{n_1} \mid Ax \ge b\}$, *assumptions A1-A3 hold, and all second stage integer variables are binary. Suppose Step* $2i_b$ *of the modified version of the algorithm identifies extreme point solutions of (18′). Then there exists* $\bar{K} < \infty$ *such that for all* $k > \bar{K}$, $f_c^k(x^k, \omega) = f(x^k, \omega)$ *for all* $\omega \in \Omega$ *whenever* x^k *is an extreme point of* X.

Proof. By definition of π_c, if x^k is an extreme point of X, then $\pi_c(x^k, \omega) = \pi_0(x^k, \omega)$. Furthermore, all cuts generated via $(18')$ are facets of the closure of the convex hull of the disjunctive set defined in (15). Recall that (15) was defined via a choice of a disjunction variable $j(k)$ such that $z_{j(k)}(\omega)$ is noninteger for some $\omega \in \Omega$. Indeed, in the modified algorithm, we identify $j(k)$ as well as $\omega' \in \Omega'_{j(k)}$ in step $2(i_b)$ such that if $z_{j(k)}(\omega')$ is noninteger, then $y^k(\omega')$ is deleted. Suppose that we encounter an iteration in which no new cut is found in step $2(i_b)$ of the modified algorithm. Then, regardless of the choice of the index j, the scenario solutions $\{y^k(\omega)\}_{\omega \in \Omega'_j}$, satisfy $\pi^\top y^k(\omega) \ge \pi_0(x^k, \omega)$ for all $[\pi, \pi_0(x^k, \omega)]$ that are extreme points of the collection of reverse polars of $\{\mathcal{S}_k(x^k, \omega)\}_{\omega \in \Omega'_j}$. Furthermore, for $\omega \notin \Omega'_j$, z_j^k is fixed to be binary. Since these statements are true for all j, the facial property and (8) imply that if no new cut is generated in step $2(i_b)$, we must have $f_c^k(x^k, \omega) = f_0^k(x^k, \omega) = f(x^k, \omega)$ for all $\omega \in \Omega$.

It remains to be shown that there exists $\bar{K} < \infty$ such that no new cuts are generated in Step 2 for any iteration $k > \bar{K}$. Note that by construction, $m^k(1) = 1$ for all k. Hence, whenever the disjunction variable is $j = 1$, the (scaled) reverse polar uses the same matrix $W^1 = W$ in formulating $(18')$. For $j = 1$, and $W^k = W$, consider the following system of inequalities

$$\pi_j \ge \lambda_{0,1}^\top W_j^k - I_j^k \lambda_{0,2} \quad \forall j \tag{21.1}$$

$$\pi_j \ge \lambda_{1,1}^\top W_j^k + I_j^k \lambda_{1,2} \quad \forall j \tag{21.2}$$

$$-1 \le \pi_j \le 1, \forall j \tag{21.3}$$

$$\lambda_{0,1}, \lambda_{0,2}, \lambda_{1,1}, \lambda_{1,2} \ge 0 \tag{21.4}$$

Note that this system is essentially the same as $(18')$, except that the variables $\pi_0(\omega)$, and the associated rows do not appear in (21). Since every extreme point of $(18')$ can be associated with an extreme point of (21), it follows that after finitely many iterations, all necessary extreme points will have been generated. Moreover, the set in (19) also has only finitely many bases corresponding to facets of $\pi_c(x, \omega)$. Hence if $K_1 = \text{Max}\{k \, : \, j(k) = 1\}$, then $K_1 < \infty$. Note further that the last cut identified in Step $2i_b$ for which z_1 is the disjunction variable is identified in some iteration, $K_1^- \leq K_1$. Thus, for all $k > K_1$, $j(k) \geq 2$ and $m^k(2) = K_1^- + 1$. Recursively, we see that for all j, there exists $K_j^- \leq K_j < \infty$ such that $m^k(j) = K_j^-$ for all $k \geq K_j$. Since there are only finitely many disjunction variables, it follows that all necessary cuts are generated after finitely many iterations.

It follows that for all $k > \bar{K}$, $f_c^k(x^k, \omega) = f(x^k, \omega)$. ∎

Lemma 7 ensures, for example, that whenever first stage solutions to SIP2 can be assured to be extreme points of X and the integer variables in the scenario subproblems are all binary, then the modified D^2 algorithm will provide optimal first stage solutions. Thus, if the the optimization problem stated in (1) requires us to seek an optimal decision from among a subset of the vertices of X, the modified version of D^2 will identify optimal solutions. An example of such an instance is the problem in which the first stage decision is required to be binary. Since the constraints $0 \leq x \leq 1$ are assumed to be included in X, binary feasible solutions are extreme points of X. We state this observation in Theorem 8 below. An extension of these ideas which permits continuous first stage variables is discussed in §5.

Theorem 8. *Suppose that assumptions A1-A3 hold, and that D is a subset of the extreme points of* $X = \{x \in \Re_+^{n_1} \mid Ax \geq b\}$ *over which an optimal first stage decision is sought. Let the master program in iteration* k *solve* $\text{Min}\{c^\top x + F^k(x) \mid x \in D\}$. *Moreover, suppose that all second stage integer variables are binary. Then the modified version of the* D^2 *algorithm ensures that there exists* $\bar{K} < \infty$ *such that for all* $k > \bar{K}$, $f_c^k(x^k, \omega) = f(x^k, \omega)$ *for all* $\omega \in \Omega$.

5. Extensions and Conclusions

Theorem 8 addresses the identification of an optimal SIP solution under the condition that the master program solution is restricted to a subset of extreme points $D \subset X$. In this section, we study the more general case in which first stage solutions need not be restricted to be extreme points of X. We maintain the assumption that all integer variables in the second stage are binary. Hence, all second stage problems satisfy the facial disjunctive property, so that sequential convexification is possible for any scenario subproblem.

Lemma 7 points us in the direction of the necessary extension. From the proof of that result we note that because there are finitely many disjunction variables (z_j), there are only finitely many polyhedra of the form (21), and each of these polyhedra can have only finitely many extreme points. Let $\{(\pi^e, \lambda_0^e, \lambda_1^e)\}_{e=1}^M$ denote the collection of all such extreme points (from all possible polyhedra of the form (21)). With any pair $(\lambda_0^e, \lambda_1^e)$, we can associate the pair $(\bar{\nu}_0^e(\omega), \bar{\gamma}_0^e(\omega))$ and $(\bar{\nu}_1^e(\omega), \bar{\gamma}_1^e(\omega))$ as suggested by Corollary 4. Thus, for $e \in \{1, \ldots, M\}$

$$\pi_0^e(x, \omega) = \text{Min}\{\bar{\nu}_0^e(\omega) - \bar{\gamma}_0^e(\omega)^\top x, \bar{\nu}_1^e(\omega) - \bar{\gamma}_1^e(\omega)^\top x\}.$$

Therefore, using the sequential convexification property for facial disjunctive programs, it follows that one can recover the closure of the convex hull of second stage (0-1) mixed integer points by appending the constraints $(\pi^e)^\top y \geq \pi_0^e(x, \omega)$ for all e. Consequently,

$$\begin{aligned} f(x, \omega) = \text{Min}\{g^\top y \mid y \in \Re_+^{n_2}, Wy &\geq r(\omega) - T(\omega)x, \\ (\pi^e)^\top y &\geq \pi_0^e(x, \omega), \quad e = 1, \ldots, M\}. \end{aligned} \tag{22}$$

From (22) it is obvious that the set of dual extreme points associated with (22) is finite, and we can index them by $\{(\theta_\ell, \{\mu_\ell^e\}_{e=1}^M)\}_{\ell=1}^L$. Note that the list of vertices of the dual polyhedron is independent of the scenario ω. However, the optimal choice of the dual multiplier does depend on ω. It follows that the value function can be represented in the form

$$f(x, \omega) = \underset{\ell=1,\ldots,L}{\text{Max}} \; \theta_\ell(\omega)^\top [r(\omega) - T(\omega)x] + \sum_{e=1}^M \mu_\ell^e(\omega) \pi_0^e(x, \omega). \tag{23}$$

Thus, if we wish to devise convergent algorithms that yield an optimal first stage solution over $x \in X$, it is appropriate to develop approximations of the convex hull of (23).

In order to suggest an iterative scheme, one may develop approximations based on information revealed sequentially. Let $\{(\theta_\ell, \{\mu_\ell^e\}_{e=1}^M)\}_{\ell=1}^{L_k}$ denote the subset of dual vertices of (22) revealed through iteration k, and note that the restriction of (23) to these vertices yields a lower bound on $f(x,\omega)$. Moreover if $\{(\lambda_0^e,\lambda_1^e)\}_{e=1}^{M^k}$ denotes the subset of vertices of (21) identified through iteration k, then the LP approximation of (22) also includes fewer cuts than M. Hence, using L_k and M_k in (23) provides a lower bounding approximation of $f(x,\omega)$. Therefore, convexifying this approximation must also provide a lower bounding function. The details associated with such a convexification are provided below.

Let $q \in \mathcal{B}^{M_k}$ denote a binary vector with elements $q_e \in \{0,1\}, e=1,\ldots,M_k$. Let H^k denote the index set of all combinations of such binary vectors, and let $h \in H^k$ represent the index associated with one such combination. Then for each $h \in H^k$, we can define the following polyhedral set,

$$\begin{aligned} S_h^k(\cdot,\omega) = \{(\eta,x) \mid\ & Ax \geq b, x \geq 0, \\ & \eta \ \geq\ \theta_\ell(\omega)^\top[r(\omega)-T(\omega)x] \\ & \quad + \sum_{e=1}^{M_k}\mu_\ell^e(\omega)[\bar{\nu}_{q_e}^e(\omega)-\bar{\gamma}_{q_e}^e(\omega)^\top x], \quad \ell=1,\ldots,L_k\}. \end{aligned} \tag{24}$$

and furthermore, the disjunctive set may be stated as follows.

$$\mathcal{S}(\cdot,\omega)=\cup_{h\in H^k}S_h^k(\cdot,\omega).$$

Thus, we can derive valid inequalities (as well as facets) by using the disjunctive cut principle. By noting that for $k > \bar{K}$ (as defined in Lemma 7 and Theorem 8), one ultimately obtains a sufficient representation of the convex hull of (23). This class of algorithms thus provides a comprehensive framework for SIP2 problems.

While it is satisfying to observe that the methodology can be extended to the case in which we are allowed to seek first stage optimality over $x \in X$, this extension comes at a price: the generation of approximations in the general case requires us to develop cuts from disjunctions with many atoms ($h \in H^k$), and each of the polyhedra (S_h) include all dual extreme points (θ^ℓ, $\{\mu_e^\ell\}$) generated through iteration k. The development of algorithmic procedures that do not become overburdened by these calculations will be the subject of future papers.

To summarize, we have presented a new decomposition method based on generating convexifications of both the master and subproblems in stochastic integer programming. The new method, referred to as the D^2 algorithm, provides the motivation for a variety of future studies. First and foremost, we mention the need to incorporate branch and bound methods within a decomposition setting. Doing so will allow us to design decomposition-based branch and cut algorithms. These algorithms should also be implemented, and computational experiments conducted. Given the degree of difficulty associated with the class of problems under consideration, it is important to consider high performance computing platforms that allow a network of processors to address various pieces of the decomposed problem. Finally, we can also foresee a stochastic version of the proposed method. As one can surmise, this line of research presents numerous promising avenues which will be explored in the future.

Bibliographic Notes. The C^3 theorem was first announced at the INFORMS conference in Seattle (Fall 1998). This paper has also formed the basis for a stochastic IP course offered in Spring 2000 at Arizona, as well lectures at the INFORMS conference in Salt Lake City (Spring 2000), the ISMP 2000 meeting (August 2000) and the West Coast Optimization Meeting in Seattle (October 2000).

References

Balas, E. [1975], "Disjunctive programming: cutting planes from logical conditions," in *Nonlinear Programming 2*, (O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds.), Academic Press, N.Y.

Balas, E. [1979], "Disjunctive programming," *Annals of Discrete Mathematics*, vol. 5, pp. 3-31.

Balas, E. [1997], "A modified lift-and-project procedure," *Mathematical Programming, Series B*, **79**, pp. 19-31.

Balas, E., S. Ceria and G. Cornuéjols [1993], "A lift-and-project cutting plane algorithm for mixed 0-1 programs," *Mathematical Programming*, vol. 58, pp. 295-324.

Benders, J.F. [1962], "Partitioning procedures for solving mixed-variable programming problems," *Numerische Mathematic*, vol. 4, pp. 238-252.

Blair [1980], "Facial disjunctive programs and sequence of cutting planes," *Discrete Applied Mathematics*, **2**, pp. 173-179.

Blair, C. [1995], "A closed-form representation of mixed-integer program value functions," *Mathematical Programming*, vol. 71, pp. 127-136.

Blair, C and R. Jeroslow [1978], "A converse for disjunctive constraints," *Journal of Optimization Theory and Applications*, **25**, pp. 195-206.

Blair, C and R. Jeroslow [1982], "The value function of an integer program," *Mathematical Programming*, **23**, pp. 237-273.

Birge, J.R. and F. Louveaux [1997], *Introduction to Stochastic Programming*, Springer.

Caroe, C. C. [1998], *Decomposition in Stochastic Integer Programming*, Ph.D. thesis, Institute of Mathematical Sciences, Dept. of Operations Research, University of Copenhagen, Denmark.

Caroe, C.C. and J. Tind [1997], "A cutting plane approach to mixed 0-1 stochastic integer programs," *European Journal of Operations Research*, **101**, pp. 306-316.

Caroe, C.C. and J. Tind [1998], "L-shaped decomposition of two-stage stochastic programs with integer recourse," *Mathematical Programming*, vol. 83, no. 3, pp. 139-152.

Conti, P. and C. Traverso [1991], "Buchberger algorithm and integer programming," *Proceedings of AAECC-9 New Orleans*, Springer Verlag LNCS 539, pp. 130-139.

Glover, F. and M. Laguna [1997], *Tabu Search*, Kluwer Academic Publishers.

Higle, J.L. and S. Sen [1992], "On the convergence of algorithms with implications for stochastic and nondifferentiable optimization," *Math. of Operations Research*, **17**, pp. 112-131.

Higle, J.L. and S. Sen [1995], "Epigraphical Nesting: a unifying theory for the convergence of algorithms," *Journal of Optimization Theory and Applications*, **84**, pp. 339-360.

Jeroslow, R. [1980], "A cutting plane game for facial disjunctive programs," *SIAM Journal on Control and Optimization*, **18**, pp. 264-281.

Kelley, J.E. [1960], The cutting plane method for convex programs, *Journal of SIAM*, **8**, pp.

703-712.

Klein Haneveld, W.K., L. Stougie and M.H. van der Vlerk [1995], "On the convex hull of the simple integer recourse objective function," *Annals of Operations Research*, **56**, pp. 209-224.

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**, pp. 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**, pp. 39-57.

Laporte, G. and F.V. Louveaux [1993], "The integer L-shaped methods for stochastic integer programs with complete recourse," *Operations Research Letters*, vol. 13, pp. 133-142.

Lokkentangen, A. and D. Woodruff [1996], "Progressive hedging and tabu search applied to mixed integer (0,1) multi-stage stochastic programming," *Journal of Heuristics* **2**, pp. 111-128.

Martin, R.K. [1999] *Large Scale Linear and Integer Optimization*, Kluwer Academic Publishers.

McDaniel, D. and M. Devine [1977], "A modified Benders' partitioning algorithm for mixed integer programming," *Management Science*, **24**, pp. 312-319.

Norkin, V.I., Y.M. Ermoliev, and A. Ruszczynski [1998], "On optimal allocation of indivisibles under uncertainty," *Operations Research* vol. 46, no. 3, pp. 381-395.

Nowak, M. and W. Römisch [2000], "Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty," to appear in *Annals of Operations Research.*

Rockafellar, R.T. and R. J-B Wets [1991], "Scenario and policy aggregation in optimization under uncertainty," *Mathematics of Operations Research*, **16**, pp. 119-147.

Rockafellar, R.T. and R. J-B Wets [1997], *Variational Analysis*, Springer.

Schultz, R. [1993], "Continuity properties of expectation functions in stochastic integer programming", *Mathematics of Operations Research*, vol. 18, pp. 578-589.

Schultz, R., L. Stougie, and M.H. van der Vlerk [1998], "Solving stochastic programs with integer recourse by enumeration: a framework using Grobner basis reduction," *Mathematical Programming*, vol. 83, no. 2, pp. 71-94.

Sen, S. [1992], "Relaxations for probabilistically constrained programs with discrete random variables," *Operations Research Letters*, **11**, pp. 81-86.

Sen, S. and H.D. Sherali [1985], "On the convergence of cutting plane algorithms for a class of nonconvex mathematical programs," *Mathematical Programming*, vol. 31, pp. 42-56.

Sen, S. and H.D. Sherali [1986], "Facet inequalities from simple disjunctions in cutting plane theory," *Mathematical Programming*, vol. 34, pp. 72-83.

Sen, S. and H.D. Sherali [1987], "Nondifferentiable reverse convex programs and facetial cuts via a disjunctive characterization," *Mathematical Programming*, **37**, pp. 169-183.

Sherali, H.D. and Adams [1990], "A hierarchy of relaxations between the continuous and convex hull representations for 0-1 programming problems," *SIAM J. on Discrete Mathematics*, **3**, pp. 411-430.

Sherali, H.D. and B.M.P Fraticelli [2000], "A modified Benders' partitioning approach for problems having discrete subproblems with application to stochastic programs with integer recourse," Working Paper, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Sherali, H.D. and C.M. Shetty [1980] *Optimization with Disjunctive Constraints*, Lecture Notes in Economics and Math. Systems, Vol. 181, Springer-Verlag, Berlin.

Thomas, R. [1995], "A geometric Buchberger algorithm for integer programming," *Mathematics of Operations Research*, **20**, pp.864-884.

van der Vlerk, M.H. [1995], *Stochastic Programming with Integer Recourse*, Thesis Rijksuniversiteit Groningen, Labyrinth Publication, The Netherlands.

van Slyke, R. and R.J-B. Wets [1969], L-Shaped linear programs with applications to optimal control and stochastic programming, *SIAM J. on Appl. Math.*, **17**, pp. 638-663.

Wallace, S.W. [1988], "A two stage stochastic facility location problem with time-dependent supply," in *Numerical Techniques for Stochastic Optimization*, (Yu. Ermoliev and R. J-B. Wets (eds.)) pp. 489-513.

Wets, R.J-B., [1974], Stochastic Programs with fixed recourse: the equivalent deterministic problem, *SIAM Review*, —bf 16, pp. 309-339.

Wollmer, R.M. [1980], "Two stage linear programming under uncertainty with 0-1 first stage variables," *Mathematical Programming*, vol. 19, pp. 279-288.