A Cutting Surface Method for Uncertain Linear Programs with Polyhedral Stochastic Dominance Constraints

Tito Homem-de-Mello
Sanjay Mehrotra
Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60208
tito@northwestern.edu, mehrotra@iems.northwestern.edu

March 12, 2009

Abstract

In this paper we study linear optimization problems with a newly introduced concept of multi-dimensional polyhedral linear second-order stochastic dominance constraints. By using the polyhedral properties of this dominance condition we present a cutting-surface algorithm, and show its finite convergence. The cut generation problem is a difference of convex functions (DC) optimization problem. We exploit the polyhedral structure of this problem to present a novel branch-and-cut algorithm that incorporates concepts from concave minimization and binary integer programming. A linear programming problem is formulated for generating concavity cuts in our case, where the polyhedra is unbounded. We also present duality results for this problem relating the dual multipliers to utility functions, without the need to impose constraint qualifications, which again is possible because of the polyhedral nature of the problem. Numerical examples are presented showing the nature of solutions of our model.

Key Words: Linear Programming, Stochastic Ordering, Stochastic Dominance, Utility Functions, Convex Programming, Cutting Plane Algorithms

1 Introduction

The concept of stochastic dominance is fundamental when comparing two random variables. In particular, this concept allows us to define preference of one random variable over another. Several different notions of stochastic dominance exist and have been studied in the literature. For example, in the univariate case we say that a random variable ξ *stochastically dominates* ψ *in the first order*, denoted by $\xi \succeq_{(1)} \psi$, if

$$F(\xi; a) \ \leq \ F(\psi; a) \tag{1}$$

for all $a \in \mathbb{R}$, where $F(\xi; \cdot)$ and $F(\psi; \cdot)$ are the cumulative distribution functions of respectively ξ and ψ. Similarly, we say that ξ stochastically dominates ψ *in the second order*, denoted by $\xi \succeq_{(2)} \psi$, if

$$F_2(\xi; a) := \int_{-\infty}^{a} F(\xi; t)\, dt \ \leq \ \int_{-\infty}^{a} F(\psi; t)\, dt =: F_2(\psi; a) \tag{2}$$

for all $a \in \mathbb{R}$.

The concept of stochastic dominance is also related to *utility theory* (von Neumann and Morgenstern, 1947), which hypothesizes that for each rational decision maker there exists a utility function u such that the (random) outcome X is preferred to the (random) outcome Y if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$. Since we do not know decision maker's utility function, we impose $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all u. If we have more information on the decision maker (e.g., if our decision maker is risk averse) we can restrict the set from which u is taken (e.g., the set of increasing concave functions). Some notions of stochastic dominance correspond to particular classes of utility functions. For example, the first order dominance corresponds to the set of non-decreasing functions for which the expectations exist, whereas second order corresponds to non-decreasing concave functions.

Extensions of the concept of stochastic dominance to random vectors have been developed as well. For example, a random vector X is said to dominate Y in positive linear second order (written $X \succeq_{(2)}^{\mathrm{Plin}} Y$) if:[1]

$$v^T X \succeq_{(2)} v^T Y \qquad \text{for all } v \in \mathbb{R}^m_+. \tag{3}$$

Although the theory of stochastic dominance is well developed (see, e.g., Shaked and Shanthikumar 1994 and Müller and Stoyan 2002 for comprehensive discussions), the introduction of stochastic dominance as constraints for optimization problems is recent; see Dentcheva and Ruszczyński (2003, 2004). The results in these papers were obtained in the univariate context using the notion of second order stochastic dominance, more specifically for the problem

$$\begin{aligned} &\min f(x) && \text{(UniSDC)} \\ &\text{s. t. } g(x) \succeq_{(2)} y, && (4) \\ &\qquad x \in \mathcal{X} \subseteq \mathbb{R}^n. \end{aligned}$$

[1] Dentcheva and Ruszczyński (2003) define this notion as linear second order stochastic dominance; the concept is also related to the definition of positive linear convex order found in the literature, see for instance Müller and Stoyan (2002).

In the above problem, $g(\cdot)$ is a mapping from $\mathbb{R}^n$ to the space of random variables $\mathcal{L}_1^1$ ($\mathcal{L}_1^m$ is the space of integrable mappings from the underlying probability space to $\mathbb{R}^m$), and y is a *benchmark* random variable. Dentcheva and Ruszczyński (2003) considered a variant of (UniSDC) in which the decision variables are random variables, i.e., the problem takes the form $\min\{f(\xi) : \xi \trianglerighteq_{(2)} \psi, \ \xi \in \Xi\}$. They showed that when ξ and ψ are random variables with finite support, under mild conditions the feasible region in (UniSDC) is reformulated by using a finite number of variables and linear constraints, which are explicitly given. In a subsequent paper Dentcheva and Ruszczyński (2009) studied the multi-variate problem:

$$\begin{aligned} &\min f(x) && \text{(MultiSDC)} \\ &\text{s. t. } G(x) \trianglerighteq_{(2)} Y, && (5) \\ &\qquad x \in \mathcal{X} \subseteq \mathbb{R}^n, \end{aligned}$$

where $G(\cdot) : \mathbb{R}^n \mapsto \mathcal{L}_1^m$ and $Y \in \mathcal{L}_1^m$ (i.e., $G(x)$ is a random vector for each x, and Y is a benchmark random vector). By using the concept of positive linear second order dominance, Dentcheva and Ruszczyński (2009) developed duality results for (MultiSDC). While useful, those results do not yield an algorithm that can solve (MultiSDC).

In this paper we address the issue of developing an algorithm to solve a class of optimization problems with multi-variate stochastic dominance constraints. Such a class is more strict than (MultiSDC) in the sense that we consider linear problems, but on the other hand we consider a more general notion of dominance that includes positive linear dominance as a particular case. The precise notion of dominance we use is defined below:

Definition 1 *Given a (possibly unbounded) non-empty polyhedron $\mathcal{P}$, a random vector $X \in \mathcal{L}_1^m$ is said to dominate $Y \in \mathcal{L}_1^m$ in* polyhedral linear second order with respect to $\mathcal{P}$ *(written $X \trianglerighteq_{(\mathcal{P})} Y$ and called $\mathcal{P}$-dominance in short) if*

$$v^T X \trianglerighteq_{(2)} v^T Y \qquad \textit{for all } v \in \mathcal{P}. \tag{6}$$

The idea behind this definition is that one wants to impose one-dimensional stochastic dominance between certain combinations of the components of X and the same combinations of the components of Y. The set $\mathcal{P}$ represents a collection of weights used to combine the various criteria represented by the vectors X and Y. Some particular cases of polyhedral order are listed below:

1. By taking $\mathcal{P} = \mathbb{R}_+^m$, we have $X \trianglerighteq_{(\mathcal{P})} Y \equiv X \trianglerighteq_{(2)}^{\text{Plin}} Y$, i.e., the positive linear second order dominance is a special case of polyhedral second order dominance.

2. Suppose there are two criteria (i.e., $n = 2$) and one wishes to consider weights for each criteria ranging respectively from α to β and $1 - \alpha$ to $1 - \beta$. Then $\mathcal{P}$ is the line segment connecting $(\alpha, 1 - \alpha)$ to $(\beta, 1 - \beta)$.

3. By taking $\mathcal{P}$ to be the convex hull of the vectors $(1, 0, \ldots, 0), (1, 1, 0, \ldots, 0), \ldots, (1, 1, 1, \ldots, 1)$, we obtain a second-order version of the partial sum stochastic ordering described in Chang et al. (1991).

The notion of $\mathcal{P}-$dominance allows us greater flexibility than the requirements imposed from the positive linear second order dominance. In particular, by taking $\mathcal{P}$ to be a subset of the positive orthant, $\mathcal{P}$-dominance may provide a larger set of feasible solutions (i.e., it is less conservative). Moreover, although obvious, it is important to point out that one can specify $\mathcal{P}$ either by using a pre-defined set of vertices, or through a set of linear constraints. This has the potential to increase wider practical applicability and use of optimization with dominance constraints.

The following characterization shows that without loss of generality we can assume that the set $\mathcal{P}$ is compact. This property is useful in our analysis. Note also that this proposition remains valid for a general convex set.

Proposition 1 *Let $\mathcal{P}$ be a non-empty convex set. Then,* (6) *holds if and only if $v^T X \trianglerighteq_{(2)} v^T Y$ for all $v \in \widetilde{\mathcal{P}} := \mathrm{cl\ cone}(\mathcal{P}) \cap \Delta$, where* cl *denotes the closure of a set,* cone *denotes the conical hull of a set (defined as* $\mathrm{cone}(S) := \{\sum_{i=1}^k \lambda_i x_i \ : \ \lambda_i \geq 0,\ x_i \in S,\ i = 1, \dots, x\}$*), and* $\Delta := \{v \in \mathbb{R}^m \mid \|v\|_1 \leq 1\}$.

Proof: First notice that, given any two random variables ξ and ψ, we have $\xi \trianglerighteq_{(2)} \psi$ if and only if $\alpha\xi \trianglerighteq_{(2)} \alpha\psi$ for all $\alpha \geq 0$.

Suppose that (6) holds, and let $\bar{v}$ be an arbitrary point in $\mathrm{cone}(\mathcal{P})$. We first show that $\bar{v}^T X \trianglerighteq_{(2)} \bar{v}^T Y$. Assume that $\bar{v} \neq 0$ (the case $\bar{v} = 0$ is trivial). Since $\bar{v} \in \mathrm{cone}(\mathcal{P})$, there exist $v^1, \dots, v^k \in \mathcal{P}$ and coefficients $\lambda_1, \dots, \lambda_k$ with $\lambda_j > 0$ such that $\bar{v} = \sum_{j=1}^k \lambda_j v^j$. Now let $\alpha := 1/\sum_{j=1}^k \lambda_j$. Then, convexity of $\mathcal{P}$ implies that $\alpha\bar{v} \in \mathcal{P}$, so from (6) we have that $\alpha\bar{v}^T X \trianglerighteq_{(2)} \alpha\bar{v}^T Y$ and therefore $\bar{v}^T X \trianglerighteq_{(2)} \bar{v}^T Y$.

Next, suppose that $\bar{v} \in \widetilde{\mathcal{P}} = \mathrm{cl\ cone}(\mathcal{P})$ (note that $\bar{v}$ may not belong to $\mathrm{cone}(\mathcal{P})$). Then, there exists a sequence of points $\{\bar{v}^k\} \subset \mathrm{cone}(\mathcal{P})$ such that $\bar{v}^k \to \bar{v}$. As shown above, we have that $(\bar{v}^k)^T X \trianglerighteq_{(2)} (\bar{v}^k)^T Y$ for all k. We claim that this implies that $\bar{v}^T X \trianglerighteq_{(2)} \bar{v}^T Y$. Indeed, let $Z \in \mathcal{L}_1^m$ be an arbitrary random vector. Since $\bar{v}^k \to \bar{v}$, it follows that $(\bar{v}^k)^T Z \to \bar{v}^T Z$ with probability one and thus $(\bar{v}^k)^T Z$ converges in distribution to $\bar{v}^T Z$. Hence, $F((\bar{v}^k)^T Z; a) \to F(\bar{v}^T Z; a)$ for all continuity points of $F(\bar{v}^T Z; \cdot)$, where as before $F(\xi; \cdot)$ denotes the cumulative distribution function of a random variable ξ. It follows from the bounded convergence theorem that $F_2((\bar{v}^k)^T Z; a) \to F_2(\bar{v}^T Z; a)$ for all a and thus, by using X and Y in place of Z, we conclude that $\bar{v}^T X \trianglerighteq_{(2)} \bar{v}^T Y$.

Conversely, suppose that $v^T X \trianglerighteq_{(2)} v^T Y$ for all $v \in \widetilde{\mathcal{P}}$. Let $\bar{v}$ be an arbitrary point in $\mathcal{P}$, and again assume without loss of generality that $\bar{v} \neq 0$. Let $\alpha := 1/\|\bar{v}\|_1$. Then, $\alpha\bar{v} \in \widetilde{\mathcal{P}}$, so $\alpha\bar{v}^T X \trianglerighteq_{(2)} \alpha\bar{v}^T Y$ and therefore $\bar{v}^T X \trianglerighteq_{(2)} \bar{v}^T Y$. $\square$

In this paper we study a linear version of (MultiSDC):

$$\min d^T x \tag{ULP}$$

$$\text{s. t.} \sum_{\ell=1}^{n} x_\ell a_\ell \trianglerighteq_{(\mathcal{P})} c, \tag{7}$$

where $a_\ell,\ \ell = 1, \dots, n$ and c are $m-$dimensional random vectors defined on a common probability space $(\Omega, \mathcal{F}, P)$. The sample space Ω is assumed to be finite. We show that (ULP) can also be reformulated as a linear program. However, this reformulation requires an exponential number of constraints. Consequently, we develop a cut-based algorithm to solve (ULP). The cut generation problem is a difference of convex functions (DC) optimization problem. We exploit the polyhedral structure of this problem to develop a branch-and-cut algorithm that combines concepts from concave minimization and binary integer programming. A linear programming problem is formulated for generating concavity cuts in our case, where the polyhedra is unbounded. Also, by exploiting the problem structure, we develop a dive-and-search method to find a local minimum of our polyhedral-DC. We also present duality results for (ULP). For simplicity we have omitted the deterministic constraint set $\mathcal{X}$ while defining (ULP). The algorithm presented in this paper remains valid in the presence of these constraints provided that the "master problems" are solved exactly. For example, these results are valid when $\mathcal{X}$ imposes integral requirements on decision variables.

The remainder of the paper is organized as follows. In Section 2 we analyze the polyhedral dominance problem, show a finite-constraint formulation for this problem, and illustrate the ideas with the help of a few numerical examples. In Section 3 we discuss duality results and the connection between dual multipliers and utility functions. In Section 4 we present our cut-based algorithm to solve (ULP), together with a discussion on how to solve the subproblems efficiently. Some conclusions and directions for future research are presented in Section 5.

2 Linear Optimization with Polyhedral Second Order Dominance

Let us consider (ULP) with $A = [a_1, \dots, a_n]$, and write (7) as $Ax \trianglerighteq_{(\mathcal{P})} c$ for a given polyhedron $\mathcal{P}$. We denote the realizations of the random vector c and random matrix A by $c^1, \dots, c^r$, and $A^1, \dots, A^t$, respectively. The corresponding probabilities are denoted by $q^1, \dots, q^r$ and $p^1, \dots, p^t$, respectively. Dentcheva and Ruszczyński (2003) showed the following result for dominance of univariate random variables, which is used in our subsequent developments.

Proposition 2 *Assume that a random variable ζ has a discrete distribution with realizations $\zeta^i, i = 1, \dots, r$, and corresponding probabilities $q_i, i = 1, \dots, r$. Let $\mathcal{U}$ be the set of all non-decreasing concave functions u such that $\lim_{t\to-\infty} u(t)/t < \infty$. Then $\xi \trianglerighteq_{(2)} \zeta$ if and only if*

$$\mathbb{E}[u(\xi)] \geq \mathbb{E}[u(\zeta)], \text{ for all } u \in \mathcal{U}. \tag{8}$$

Furthermore, (8) *is equivalent to*

$$\mathbb{E}[(\zeta^i - \xi)_+] \leq \mathbb{E}[(\zeta^i - \zeta)_+], i = 1, \dots, r, \tag{9}$$

where $(\cdot)_+$ indicates $\max\{\cdot, 0\}$.

By observing that in condition (3) $v^T X$ and $v^T Y$ are univariate random variables for a fixed v, Proposition 2 gives the following two formulations of (ULP):

$$\begin{array}{lll} \min & d^T x & \\ \text{s. t.} & E[u(v^T Ax)] \ \geq \ E[u(v^T c)] & \text{for all } u \in \mathcal{U} \text{ and all } v \in \widetilde{\mathcal{P}}, \end{array} \tag{10}$$

and

$$\min d^T x \qquad \text{(SILP)}$$

$$\text{s. t.} \sum_{j=1}^{t} p_j \, (v^T c^i - v^T A^j x)_+ \;\leq\; \sum_{l=1}^{r} q_l \, (v^T c^i - v^T c^l)_+, \quad i = 1, \dots, r, \text{ for all } v \in \widetilde{\mathcal{P}}, \tag{11}$$

where $\widetilde{\mathcal{P}}$ is the set defined in Proposition 1 as a function of $\mathcal{P}$, i.e., $\widetilde{\mathcal{P}} = \text{cl cone}(\mathcal{P}) \cap \Delta$. Note that $\widetilde{\mathcal{P}}$ is a polytope (cf. Theorem 19.7 in Rockafellar 1970).

The following theorem shows that in (SILP) it is sufficient to write constraints (11) for a finite number of vectors v.

Theorem 1 *Let*

$$\mathcal{P}_i := \{(v, y) \mid y_l \geq v^T (c^i - c^l),\ y_l \geq 0,\ v \in \widetilde{\mathcal{P}},\ l = 1, \dots, r\},\ i = 1, \dots, r. \tag{12}$$

Then, the semi-infinite constraints (11) *are equivalent to*

$$\sum_{j=1}^{t} p_j \, (v^{ik^T} c^i - v^{ik^T} A^j x)_+ \;\leq\; \sum_{l=1}^{r} q_l \, (v^{ik^T} c^i - v^{ik^T} c^l)_+, \quad i = 1, \dots, r,\ k = 1, \dots, \nu_i, \tag{13}$$

where v^{ik} are the v-components of the vertex solutions of $\mathcal{P}_i$.

Proof: Obviously all x satisfying (11) also satisfy (13). Now suppose we have an $\hat{x}$ which satisfies (13) but not (11), i.e., this $\hat{x}$ violates (11) for some v. Equivalently, there exists some $i \in \{1, \dots, r\}$ such that the problem

$$\min_{v \in \widetilde{\mathcal{P}}} \sum_{l=1}^{r} q_l \, (v^T c^i - v^T c^l)_+ - \sum_{j=1}^{t} p_j \, (v^T c^i - v^T A^j \hat{x})_+ \qquad (\text{DCP}_i)$$

has a negative objective value. The objective function in the problem (DCP_i) is a difference of two piecewise linear convex functions. We can reformulate (DCP_i) as a concave minimization problem as follows.

$$\min_{v,y} f(v, y) := \sum_{l=1}^{r} q_l \, y_l - \sum_{j=1}^{t} p_j \, (v^T c^i - v^T A^j \hat{x})_+ \qquad (\text{SepCP}_i)$$
$$\text{s. t. } (v, y) \in \mathcal{P}_i.$$

We first argue that (SepCP_i) has a minimizer. To see that, note that the rightmost term in the objective function of (SepCP_i) is bounded, since it is a continuous function involving only the v-components and v is restricted to the compact set $\widetilde{\mathcal{P}}$. The leftmost term is always non-negative. Thus, since this is a minimization problem, we can embed the y components into a compact set as well, which then implies that (SepCP_i) has a minimizer. It follows from Corollary 32.3.4 in Rockafellar (1970) that at least one of the vertex solutions of $\mathcal{P}_i$ is a minimizer of (SepCP_i). $\square$

Theorem 1 allows us to reformulate (ULP) as (FLP):

$$\min\ d^T x \tag{FLP}$$

$$\text{s. t.} \sum_{j=1}^{t} p_j\,({v^{ik}}^T c^i - {v^{ik}}^T A^j x)_+ \ \le\ \sum_{l=1}^{r} q_l\,({v^{ik}}^T c^i - {v^{ik}}^T c^l)_+,\ \ i=1,\dots,r,\ \ k=1,\dots,\nu_i. \tag{14}$$

However, (FLP) may have exponential number of constraints since the number of vertices in $\mathcal{K}_i$ may be exponential. Nevertheless, by introducing z^{ijk} to represent the shortfall value $({v^{ik}}^T c^i - {v^{ik}}^T A^j x)_+$ we can reformulate (FLP) as a linear program (FullLP):

$$\min\ d^T x \tag{FullLP}$$

$$\begin{aligned}
\text{s. t.} \sum_{j=1}^{t} p_j\, z^{ijk} &\le \sum_{l=1}^{r} q_l\,({v^{ik}}^T c^i - {v^{ik}}^T c^l)_+, \quad i=1,\dots,r,\ k=1,\dots,\nu_i \\
z^{ijk} &\ge ({v^{ik}}^T c^i - {v^{ik}}^T A^j x), \quad i=1,\dots,r,\ j=1,\dots,r,\ k=1,\dots,\nu_i \qquad (15)\\
z^{ijk} &\ge 0, \quad i=1,\dots,r,\ j=1,\dots,r,\ k=1,\dots,\nu_i.
\end{aligned}$$

Note that if x^* is an optimal solution of (FLP), then obviously $(x^*, z^{ijk*}) = (x^*, ({v^{ik}}^T c^i - {v^{ik}}^T A^j x^*)_+)$ is a feasible solution of (FullLP). Conversely, if (x^*, z^{ijk*}) is an optimal solution of (FullLP), then x^* is a feasible solution of (FLP). The latter follows because $p_j \ge 0$. Hence, we may use either (FLP) or (FullLP) to find x^*. In Section 4 we will discuss a cut-generation strategy that does not require enumerating all vertices v^{ik} of $\mathcal{P}_i$ in advance.

In the univariate case, we have the following corollary. The result is straightforward but it illustrates an application of Theorem 1.

Corollary 1 *(Dentcheva and Ruszczyński 2003) Let $\mathcal{P} \subseteq \mathbb{R}^1_+$, $\mathcal{P} \ne \{0\}$. Then we can solve* (ULP) *by solving:*

$$\begin{aligned}
&\min\ d^T x \\
&s.\ t. \sum_{j=1}^{t} p_j\, z^{ij} \ \le\ \sum_{l=1}^{r} q_l\,(c^i - c^l)_+, \\
&\qquad z^{ij} \ge (c^i - A^j x), \quad z^{ij} \ge 0, \quad i,j=1,\dots,r.
\end{aligned}$$

Proof: $\mathcal{P} \subseteq \mathbb{R}^1_+$, $\mathcal{P} \ne \{0\}$ implies that $\mathcal{P} = [a,b]$ for some $0 \le a \le b$ with $b > 0$. By Proposition 1, we have $vX \trianglerighteq_{(2)} vY$ for all $v \in \mathcal{P}$ if and only if $vX \trianglerighteq_{(2)} vY$ for all $v \in [0,1]$. Theorem 1 then ensures that the latter condition holds if and only if $X \trianglerighteq_{(2)} Y$ (it is easy to check that the v-components of the vertices of the polyhedron $\mathcal{P}_i$ in the theorem are either 0 or 1). It follows that, in (FullLP), we have $\nu_i = 1$ and $v^{ik} = 1$ for all i. $\square$

2.1 General Multivariate Dominance

Although in this paper we focus on the particular notion of $\mathcal{P}$-dominance introduced in (6), a different notion can be defined by extending the characterization via expected utility in (8) to the multivariate case. Following Müller and Stoyan (2002), we say that a random *vector* X stochastically dominates a random vector Y in second order (denoted $X \trianglerighteq_{(2)} Y$) if and only if

$$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)] \tag{16}$$

for all concave non-decreasing functions u. Note that the positive linear second order dominance is a special case of second order dominance. Unfortunately, such characterization requires verifying (16) for a large class of functions, and a simpler characterization such as (9) does not seem to be available at the present moment. This issue in part justifies the use of alternative orders such as the polyhedral order introduced in Definition 1.

The situation becomes much easier when the vectors X and Y have *independent* components. In that case, the multivariate dominance defined in (16) reduces to the univariate case. This result is shown in Huang et al. (1978), but we state it here for completeness. A proof is presented in the Appendix.

Theorem 2 *Let X and Y be random vectors in $\mathbb{R}^n$ with* independent *components. Then,*

$$X \trianglerighteq_{(2)} Y \iff X_\ell \trianglerighteq_{(2)} Y_\ell, \quad \ell = 1, \ldots, n.$$

A consequence of Theorem 2 in the context of $\mathcal{P}$-dominance defined in Definition 1 is given below.

Corollary 2 *Let X and Y be random vectors in $\mathbb{R}^m$ with* independent *components, and let $\mathcal{P} = \mathbb{R}^m_+$, which is the context of the positive linear second order dominance defined in* (3)*. Then, $X \trianglerighteq_{(\mathcal{P})} Y$ if and only if $X_\ell \trianglerighteq_{(2)} Y_\ell$, $\ell = 1, \ldots, m$.*

Proof: The "if" part follows from Theorem 2 by noticing that functions of the form $u(v^T x)$, where u is concave non-decreasing in $\mathbb{R}^1$ and $v \in \mathcal{P}$, are a subset of the set of concave non-decreasing functions in $\mathbb{R}^m$. The "only if" part is immediate since $e_\ell \in \mathcal{P}$ (where e_ℓ is the vector with the ℓth component equal to one and the remaining ones equal to zero). $\square$

Corollary 2 has an immediate application in the context of problem (ULP). Consider the matrix $A = [a_1, \ldots, a_n]$, where the a_ℓ are the column vectors in (7), and suppose that the rows of A are mutually independent. Suppose in addition that the vector c has independent components as well. In that case, if $\mathcal{P} = \mathbb{R}^m_+$ (i.e., positive linear second order dominance is used), then condition (7) becomes

$$(a^\ell)^T x \trianglerighteq_{(2)} c^\ell, \quad \ell = 1, \ldots, m,$$

where a^ℓ is the ℓth row of A. In other words, the only relevant vertices given by Theorem 1 are the vertices $e_1, \ldots, e_m$ of the simplex $\widetilde{\mathcal{P}} = \{v \in \mathbb{R}^m_+ \,|\, \sum v_i \leq 1, v \geq 0\}$. More vertices may exist, but they will necessarily generate redundant constraints. We shall see an illustration of this phenomenon in Section 2.2.

2.2 Examples

We illustrate now the result in Theorem 1 by describing an example in detail. Consider the linear program

$$\begin{aligned} &\max\ 3x_1 + 2x_2 \\ &\text{s. t.} \\ &-\begin{bmatrix} 4 & 2 \\ 2 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \ \geq\ -\begin{bmatrix} 200 \\ 160 \\ 40 \end{bmatrix}. \end{aligned} \tag{Ex1-LP}$$

Suppose there is uncertainty in some of the coefficients, which we want to model using stochastic dominance. More specifically, consider the ULP

$$\begin{aligned} &\max\ 3x_1 + 2x_2 \\ &\text{s. t.} \\ &-\begin{bmatrix} 4 \pm \alpha & 2 \\ 2 & 2 \pm \alpha \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \ \trianglerighteq^{\text{Plin}}_{(2)}\ -\begin{bmatrix} 200 \pm 10\beta \\ 160 \\ 40 \pm 5\beta \end{bmatrix}, \end{aligned} \tag{Ex1-ULP}$$

where $\trianglerighteq^{\text{Plin}}_{(2)}$ denotes the positive linear second order dominance defined in (3) (recall that such an order corresponds to the polyhedral order $\trianglerighteq_{(\mathcal{P})}$ with $\mathcal{P}$ defined as a normalization of the cone $\mathbb{R}^m_+$). In the above, we write $(a \pm b)$ to indicate that the actual value is random, with two outcomes $a + b$ and $a - b$. For the values in the matrix on the left-hand side, these outcomes have probability respectively equal to p and $1 - p$; for the values in the vector on the right-hand side, the outcomes have probability respectively equal to q and $1 - q$. The parameters α and β control the degree of uncertainty on respectively the left- and right-hand sides; the bigger those values, the bigger the degree of uncertainty (so $\alpha = \beta = 0$ corresponds to the original LP). It is assumed that all uncertain quantities are independent, so there are 16 scenarios in (Ex1-ULP). Let A^i and c^i denote the values of respectively the matrix on the left-hand side and the vector on the right hand-side for the ith scenario, and let us number the scenarios in such a way that $A_1 = A_2 = A_3 = A_4$, $c_1 = -(200 - 10\beta, 160, 40 - 5\beta)$, $c_2 = -(200 - 10\beta, 160, 40 + 5\beta)$, $c_3 = -(200 + 10\beta, 160, 40 - 5\beta)$, $c_4 = -(200 + 10\beta, 160, 40 + 5\beta)$, $A_5 = A_6 = A_7 = A_8$, $c_5 = c_1$, $c_6 = c_2$, $c_7 = c_3$, $c_8 = c_4$, and so on.

Let us write explicitly the polyhedra $\mathcal{P}_i$ defined in (12). From the numbering of the scenarios, it is clear that $\mathcal{P}_1 = \mathcal{P}_5 = \mathcal{P}_9 = \mathcal{P}_{13}$ and similarly for the other scenarios, so we only need to describe $\mathcal{P}_1, \ldots, \mathcal{P}_4$. Let $\mathcal{Q}$ be defined as

$$\mathcal{Q} \ \equiv\ \{(v, y) \,:\, y \geq 0,\ v \geq 0,\ v_1 + v_2 + v_3 \leq 1\}.$$

Then we have

$$
\begin{aligned}
\mathcal{P}_1 &= \{(v,y) : y_2, y_6, y_{10}, y_{14} \geq 10\beta v_3,\ y_3, y_7, y_{11}, y_{15} \geq 20\beta v_1,\ y_4, y_8, y_{12}, y_{16} \geq 20\beta v_1 + 10\beta v_3\} \cap \mathcal{Q} \\
\mathcal{P}_2 &= \{(v,y) : y_1, y_5, y_9, y_{13} \geq -10\beta v_3,\ y_3, y_7, y_{11}, y_{15} \geq 20\beta v_1 - 10\beta v_3,\ y_4, y_8, y_{12}, y_{16} \geq 20\beta v_1\} \cap \mathcal{Q} \\
\mathcal{P}_3 &= \{(v,y) : y_1, y_5, y_9, y_{13} \geq -20\beta v_1,\ y_2, y_6, y_{10}, y_{14} \geq -20\beta v_1 + 10\beta v_3,\ y_4, y_8, y_{12}, y_{16} \geq 10\beta v_3\} \cap \mathcal{Q} \\
\mathcal{P}_4 &= \{(v,y) : y_1, y_5, y_9, y_{13} \geq -20\beta v_1 - 10\beta v_3,\ y_2, y_6, y_{10}, y_{14} \geq -20\beta v_1,\ y_3, y_7, y_{11}, y_{15} \geq -10\beta v_3\} \cap \mathcal{Q}.
\end{aligned}
$$

Let $V(\mathcal{P})$ denote the set of points obtained by projecting the vertices of the polyhedron $\mathcal{P}$ onto the space of the v variables. It is possible to show (by enumeration) that

$$
\begin{aligned}
V(\mathcal{P}_1) &= \{(0,0,0),\ (1,0,0),\ (0,1,0),\ (0,0,1)\} \\
V(\mathcal{P}_2) &= V(\mathcal{P}_1) \cup \{(1/3, 0, 2/3)\} \\
V(\mathcal{P}_3) &= V(\mathcal{P}_2) \\
V(\mathcal{P}_4) &= V(\mathcal{P}_1)
\end{aligned}
$$

regardless of the value of β. It follows that (Ex1-ULP) can be written as the following linear program (the constraints that are obviously redundant have been eliminated):

$$
\begin{aligned}
& \max\ 3x_1 + 2x_2 \\
& \text{s. t.} && \text{(Ex1-SDLP)} \\
& \quad (4+\alpha)x_1 + 2x_2 - s_1 \leq 200 - 10\beta \\
& \quad (4-\alpha)x_1 + 2x_2 - s_2 \leq 200 - 10\beta \\
& \quad ps_1 + (1-p)s_2 \leq 20\beta q \\
& \quad 2x_1 + (2+\alpha)x_2 \leq 160 \\
& \quad x_1 \leq 40 - 5\beta(1-2q) \\
& \quad (6+\alpha)x_1 + 2x_2 - s_3 \leq 280 && (17) \\
& \quad (6-\alpha)x_1 + 2x_2 - s_4 \leq 280 && (18) \\
& \quad ps_3 + (1-p)s_4 \leq 20\beta q^2 && (19) \\
& \quad (4+\alpha)x_1 + 2x_2 \leq 200 + 10\beta \\
& \quad x, s \geq 0.
\end{aligned}
$$

The optimal solution of problem (Ex1-SDLP) (with $p = q = 1/2$, $\alpha = \beta = 1$) is $x^{\text{ULP}} = (28.18, 34.55)$ and the corresponding objective value is $\nu^{\text{ULP}} = 153.44$, whereas the optimal solution and optimal value of (Ex1-LP) are respectively $x^{\text{LP}} = (20, 60)$, $\nu^{\text{LP}} = 180$. We can see the effect of taking the uncertainty into account. In particular, for any fixed α the optimal value of (Ex1-SDLP) increases as β increases, i.e., adding uncertainty to the right-hand side improves the optimal value as it makes the right-hand side less attractive to a risk averse decision maker because of increased uncertainty. Increasing α for a fixed β leads to the opposite phenomenon. Figure 1 shows the optimal value of (Ex1-SDLP) as a function of α and β.

It is interesting to study the effect of dependence among the random variables in this example. Since all the random variables in the example are assumed to be independent, Corollary 2 ensures that the positive linear dominance constraint in (Ex1-ULP) is equivalent to univariate second order

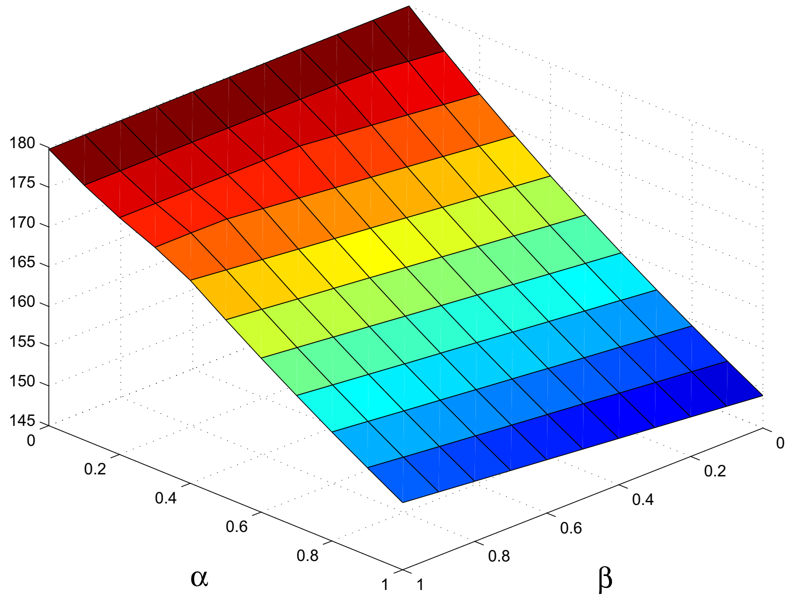

Figure 1: Optimal value of (Ex1-SDLP), as a function of α and β.

dominance involving each component. This leads to an LP which is identical to (Ex1-SDLP) except that it does not contain constraints (17)-(19) — which are generated by the non-simplex vertex $(1/3, 0, 2/3)$. In other words, Corollary 2 guarantees, via a probabilistic argument, that those constraints are redundant. Indeed, in the case $p = q = 1/2$, $\alpha = \beta = 1$, when we solve (Ex1-SDLP) without constraints (17)-(19) and with the objective function replaced with $7x_1 + 2x_2$, the objective value is 290. When the objective function is replaced with $5x_1 + 2x_2$, the objective value is 210. That is, in (Ex1-SDLP) we have $s_3 \leq 10$, $s_4 = 0$, so constraints (17)-(19) are always satisfied.

The situation changes if the independence assumption is dropped. For example, consider the same example as above but suppose there are only two scenarios: in scenario 1 (the probability of which is p) we have $A_1 = \begin{bmatrix} 4+\alpha & 2 \\ 2 & 2-\alpha \\ 1 & 0 \end{bmatrix}$, $c_1 = (200 - 10\beta, 160, 40 + 5\beta)^T$, and in scenario 2 (the probability of which is $1-p$) we have $A_2 = \begin{bmatrix} 4-\alpha & 2 \\ 2 & 2+\alpha \\ 1 & 0 \end{bmatrix}$, $c_2 = (200 + 10\beta, 160, 40 - 5\beta)^T$. Then, by doing similar calculations as before, we obtain the following LP:

$$
\begin{array}{lrcl}
\max 3x_1 + 2x_2 & & & \\
\text{s. t.} & & & \qquad\qquad \text{(Ex1dep-SDLP)} \\
& (4+\alpha)x_1 + 2x_2 - s_1 & \leq & 200 - 10\beta \\
& (4-\alpha)x_1 + 2x_2 - s_2 & \leq & 200 - 10\beta \\
& ps_1 + (1-p)s_2 & \leq & 20\beta(1-p) \\
& 2x_1 + (2+\alpha)x_2 & \leq & 160 \\
& x_1 & \leq & 40 - 5\beta(1-2p) \\
& (6+\alpha)x_1 + 2x_2 & \leq & 280 \qquad\qquad (20) \\
& (4+\alpha)x_1 + 2x_2 & \leq & 200 + 10\beta \\
& x, s & \geq & 0.
\end{array}
$$

To compare this problem and (Ex1-SDLP), consider again the case $p = q = 1/2$, $\alpha = \beta = 1$. Then, without constraint (20), the feasible region of problem (Ex1dep-SDLP) coincides with that of (Ex1-SDLP) except that it does not contain constraints (17)-(19). However, constraint (20) is *not* redundant — indeed, when we solve (Ex1dep-SDLP) without constraint (20) and with the objective function replaced with $7x_1 + 2x_2$, the objective value is 290, which is the same value as that of (Ex1-SDLP) also with objective function $7x_1 + 2x_2$. With constraint (20), of course, the optimal value of (Ex1dep-SDLP) cannot be bigger than 280. Therefore, the feasible regions of (Ex1-SDLP) and (Ex1dep-SDLP) are different.

3 Duality Results for the Uncertain Linear Program

Dentcheva and Ruszczyński (2009) give very general duality results for the optimization problems with second order linear stochastic dominance constraints in the vector case. These duality results show the existence of certain dual functions under Slater-type constraint qualification conditions. In the following we give analogous results in our case. As we shall see below, because of the polyhedral nature of the problem, constraint qualification conditions are not required.

Theorem 3 *A solution x^* is an optimal solution of* (FLP) *if and only if there exist multipliers $\pi^{ik} \geq 0$, $i = 1, \dots, r$, $k = 1, \dots, \nu_i$, such that*

$$d + \sum_{i=1}^{r} \sum_{k=1}^{\nu_i} \pi^{ik} g^{ik} = 0, \tag{21}$$

and

$$\pi^{ik} = 0 \quad \text{if} \quad \sum_{j=1}^{t} p_j \, ({v^{ik}}^T c^i - {v^{ik}}^T A^j x^*)_+ < \sum_{l=1}^{r} q_l \, ({v^{ik}}^T c^i - {v^{ik}}^T c^l)_+. \tag{22}$$

In the above, $g^{ik} = \sum_{j=1}^{t} p_j s^{ijk}$*, and*

$$s^{ijk} \in \begin{cases} \{0\} \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) < 0 \\ \{{A^j}^T v^{ik}\} \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) > 0 \\ \text{conv}(\{0, {A^j}^T v^{ik}\}) \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) = 0. \end{cases} \tag{23}$$

Proof: ($\Rightarrow$) Let x^* be an optimal solution of (FLP). Then, there exists z^* such that (x^*, z^*) solves (FullLP), so there exist non-negative multipliers $(\lambda^{ik}, \mu^{ijk}, \theta^{ijk})$, $i = 1, \dots, r$, $j = 1, \dots, t, k = 1, \dots, \nu_i$ for constraints (15) respectively, satisfying:

$$\begin{pmatrix} d \\ 0 \end{pmatrix} = \begin{pmatrix} -\sum_{i=1}^{r}\sum_{k=1}^{\nu_i}\sum_{j=1}^{t} \mu^{ijk} {A^j}^T v^{ik} \\ \sum_{i=1}^{r}\sum_{k=1}^{\nu_i}\sum_{j=1}^{t} (\mu^{ijk} + \theta^{ijk}) e_{ijk} \end{pmatrix} - \begin{pmatrix} 0 \\ \sum_{i=1}^{r}\sum_{k=1}^{\nu_i} \left(\lambda^{ik} \sum_{j=1}^{t} p_j e_{ijk}\right) \end{pmatrix}, \tag{24}$$

i.e.,

$$d = -\sum_{i=1}^{r}\sum_{k=1}^{\nu_i}\sum_{j=1}^{t} \mu^{ijk} {A^j}^T v^{ik} \tag{25}$$

$$p_j \lambda^{ik} = \mu^{ijk} + \theta^{ijk}, \tag{26}$$

and also

$$\mu^{ijk} = 0 \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) < 0 \tag{27a}$$

$$\theta^{ijk} = 0 \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) > 0 \tag{27b}$$

$$\lambda^{ik} = 0 \ \text{ if } \sum_{j=1}^{t} p_j \, z^{ijk*} < \sum_{l=1}^{r} q_l \, ({v^{ik}}^T c^i - {v^{ik}}^T c^l)_+. \tag{27c}$$

Now, define $\alpha^{ijk} := \frac{\mu^{ijk}}{\mu^{ijk} + \theta^{ijk}}$, where we adopt the convention that $0/0 = 0$. Note that from (27) we have

$$\alpha^{ijk} \ \in \begin{cases} \{0\} \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) < 0 \\ \{1\} \ \text{ if } {v^{ik}}^T(c^i - A^j x^*) > 0 \\ [0, 1] \ \text{ otherwise,} \end{cases}$$

which implies that $s^{ijk} := \alpha^{ijk} {A^j}^T v^{ik}$ satisfies (23).

Next, let $\pi^{ik} := \lambda^{ik}$. Then from (26) we have $\pi^{ik} p_j \alpha^{ijk} = \mu^{ijk}$ and thus condition (25) can be written as

$$d \ = \ -\sum_{i=1}^{r}\sum_{k=1}^{\nu_i}\sum_{j=1}^{t} \pi^{ik} p_j \alpha^{ijk} {A^j}^T v^{ik} \ = \ -\sum_{i=1}^{r}\sum_{k=1}^{\nu_i} \pi^{ik} \sum_{j=1}^{t} p_j s^{ijk} \ = \ -\sum_{i=1}^{r}\sum_{k=1}^{\nu_i} \pi^{ik} g^{ik},$$

which shows (21). Finally, it is easy to see that, when the condition in (22) holds, we have $\lambda^{ik} = 0$ from (27c) and therefore $\pi^{ik} = 0$.

($\Leftarrow$) Notice that s^{ijk} defined in (23) is a subgradient of the function $({v^{ik}}^T c^i - {v^{ik}}^T A^j x)_+$ at x^*, and consequently $g^{ik} := \sum_{j=1}^{t} p_j s^{ijk}$ is a subgradient of (14) at x^*. The result follows from the fact that (21) and (22) correspond to the KKT conditions for problem (FLP) at x^*, which in turn implies the optimality of x^* (see, e.g., Theorem VII.2.2.4 in Hiriart-Urruty and Lemarechal 1993). $\square$

It is interesting to put the result stated in Theorem 3 into the context of the general duality results of Dentcheva and Ruszczyński (2009), which we briefly review here for ease of reference. Consider again problem (MultiSDC), and notice that when positive linear dominance is used, the stochastic dominance constraint can be written as

$$\mathbb{E}[u(v^T G(x))] \leq \mathbb{E}[u(v^T Y)] \quad \text{for all } v \in \Delta \text{ and all } u \in \mathcal{U}, \tag{28}$$

with Δ and $\mathcal{U}$ as defined in Propositions 1 and 2. Consider the class of functions $\mathbb{R}^m \mapsto \mathbb{R}$ of the form

$$\phi_{Q,\mu}(w) \;=\; \int_\Delta [Q(v)](v^T w)\, \mu(dv), \tag{29}$$

where Q maps a vector $v \in \Delta$ into the space $\mathcal{U}$ — in such a way that the mapping $q(v,w) := [Q(v)](v^T w)$ is Lebesgue measurable — and μ is a finite non-negative measure in Δ. Define the following functional for problem (MultiSDC):

$$\mathcal{L}(x,\phi) \;=\; \mathbb{E}\left[f(x) + \phi(G(x)) - \phi(Y)\right], \tag{30}$$

and assume that f and G are concave. Dentcheva and Ruszczyński (2009) show that, under a certain Slater-type constraint qualification, if x^* is an optimal solution of (MultiSDC) then there exists a function $\phi^*_{Q,\mu}$ of the form (29) such that $L(x^*, \phi^*_{Q,\mu}) = \max_{x\in\mathcal{X}} \mathcal{L}(x, \phi^*_{Q,\mu})$ and $\mathbb{E}[\phi^*_{Q,\mu}(G(x^*))] = \mathbb{E}[\phi^*_{Q,\mu}(Y)]$. That is, the functional $\mathcal{L}$ plays the role of a Lagrangian function, and the optimal multiplier $\phi^*_{Q,\mu}$ corresponds to a "weighted average" of the utility functions in (28) that yields equality.

To view Theorem 3 in light of the above results, consider the Lagrangian function of problem (FLP):

$$L(x,\pi) \;=\; d^T x + \sum_{i=1}^{r}\sum_{k=1}^{\nu_i} \pi^{ik} \left[\sum_{j=1}^{t} p_j \,({v^{ik}}^T c^i - {v^{ik}}^T A^j x)_+ - \sum_{l=1}^{r} q_l \,({v^{ik}}^T c^i - {v^{ik}}^T c^l)_+ \right]. \tag{31}$$

Suppose we represent v^{ik} as an element of $\mathbb{R}^{m+1}$ by adding the index i as the $m+1$-component. Moreover, let us view each c^i as an element of $\mathbb{R}^{m+1}$ by adding a zero as the $m+1$-component, and each A^i as an element of $\mathbb{R}^{m+1} \times \mathbb{R}^n$ by adding a row of zeros as the $m+1$-row. Clearly, such reformulations do not change the value of $L(x,\pi)$. However, this allows us to uniquely label all the v-components of the vertices of $\mathcal{P}_i$, $i = 1, \ldots, r$. Define now the mapping $\hat{Q} : \Delta \times \{1, \ldots, r\} \mapsto \mathcal{U}$ as

$$[\hat{Q}(v)](z) \;:=\; \sum_{j=1}^{r}\sum_{\ell=1}^{\nu_j} \mathrm{I}_{\{v=v^{j\ell}\}} (v^T c^j - z)_+. \tag{32}$$

It is clear that the mapping $q(v,w) = [\hat{Q}(v)](v^T w)$ is Lebesgue measurable. Define the measure $\hat{\mu}$

on $\Delta \times \{1, \ldots, r\}$ as the atomic measure with mass on $\{v^{ik} : i = 1, \ldots, r,\ k = 1, \ldots, \nu_i\}$ such that $\hat{\mu}(v^{ik}) = \hat{\pi}^{ik}$, where $\hat{\pi}$ is the vector of optimal multipliers given by Theorem 3. Then, the function $\phi_{\hat{Q},\hat{\mu}}$ given in (29) can be written as

$$\begin{aligned} \phi_{\hat{Q},\hat{\mu}}(w) &= \int_{\Delta \times \{1,\ldots,r\}} [\hat{Q}(v)](v^T w)\, \hat{\mu}(dv) \;=\; \sum_{i=1}^{r} \sum_{k=1}^{\nu_i} \hat{\pi}^{ik} [\hat{Q}(v^{ik})](v^{ik^T} w) \\ &= \sum_{i=1}^{r} \sum_{k=1}^{\nu_i} \hat{\pi}^{ik} (v^{ik^T} c^i - v^{ik^T} w)_+. \end{aligned}$$

Note that in the above calculation it was important to ensure that, when computing the function $\hat{Q}(v^{ik})$, exactly one term inside the sum in (32) is nonzero; this is the reason why we extended the vectors v^{ik} to include the index i, otherwise it could happen that $v^{ik} = v^{j\ell}$ for $i \neq j$ for some k, ℓ.

It follows that the Lagrangian function in (31) can be written as

$$L(x, \hat{\pi}) \;=\; d^T x + \mathbb{E}\left[\phi_{\hat{Q},\hat{\mu}}(Ax)\right] - \mathbb{E}\left[\phi_{\hat{Q},\hat{\mu}}(c)\right] \;=\; \mathcal{L}(x, \phi_{\hat{Q},\hat{\mu}}),$$

hence we see that the standard Lagrangian coincides with the general functional of Dentcheva and Ruszczyński (2009). Moreover, (21) ensures that, if $\hat{x}$ solves (FLP), then $L(\hat{x}, \hat{\pi}) = \max_{x \in \mathbb{R}^n} L(x, \hat{\pi})$. Finally, from (22) we have that

$$\sum_{i=1}^{r} \sum_{k=1}^{\nu_i} \pi^{ik} \left[\sum_{j=1}^{t} p_j \,(v^{ik^T} c^i - v^{ik^T} A^j x^*)_+ - \sum_{l=1}^{r} q_l \,(v^{ik^T} c^i - v^{ik^T} c^l)_+ \right] \;=\; 0,$$

i.e., $\mathbb{E}\left[\phi_{\hat{Q},\hat{\mu}}(Ax)\right] = \mathbb{E}\left[\phi_{\hat{Q},\hat{\mu}}(c)\right]$. Thus, the results in Dentcheva and Ruszczyński (2009) apply to our case without the need to impose constraint qualifications.

4 A Cut-Generation Algorithm for Linear Optimization Problems with Polyhedral Second Order Dominance Constraints

We discuss now an algorithm to solve problem (FLP). The fact that the constraints in (14) are generated using the vertices of $\mathcal{P}_i$ suggests the use of a cut-generation approach for solving that problem, instead of adding all the constraints up front. In the cut-generation approach we solve a sequence of relaxations of (FLP), over a subset of constraints in (14). The relaxed problems are solved using their linear programming reformulation as given in (FullLP). At a solution $\hat{x}$ of a relaxed problem we consider the subproblems (SepCP$_i$) defined in the proof of Theorem 1. If all (SepCP$_i$) have a non-negative objective value, we have a solution of (FLP). Otherwise, we have a vertex solution $\hat{v}$ of (SepCP$_i$) with a negative objective value. Corresponding to this vertex, the constraint $\sum_{j=1}^{t} p_j \,(\hat{v}^T c^i - \hat{v}^T A^j x)_+ \;\leq\; \sum_{l=1}^{r} q_l \,(\hat{v}^T c^i - \hat{v}^T c^l)_+$, is a valid cut for $\hat{x}$.

Algorithm 1 below outlines the basic steps. In a similar fashion to what we did when discussing the duality results in Section 3, we store the generated vertices as elements of $\mathbb{R}^m \times \mathbb{R}^r \times \{1, \ldots, r\}$, where the last component is the index i of the corresponding polyhedron $\mathcal{P}_i$ defined in Theorem 1.

Theorem 4 *Algorithm 1 terminates after a finite number of steps with either an optimal solution to* (ULP)*, or a proof of infeasibility (or unboundedness) of* (ULP)*.*

Algorithm 1 A Cutting Surface Algorithm for Linear Optimization with Polyhedral Second Order Dominance

0. Set $s := 0$, $v^0 :=$ an arbitrary vertex of $\mathcal{K}_i$, where $i \in \{1, \ldots, r\}$ is also chosen arbitrarily. Set $\mathcal{V}^0 := (v^0, 0, i)$.

1. Solve a linear programming reformulation of the problem

$$\begin{array}{ll} \min & d^T x \\ \text{s.t.} & \sum_{j=1}^{t} p_j (v^T c^i - v^T A^j x)_+ \leq \sum_{l=1}^{r} q_l (v^T c^i - v^T c^l)_+, \ (v, y, i) \in \mathcal{V}^s. \end{array} \tag{33}$$

   If the problem is infeasible, **stop**; if it is unbounded, then let $\hat{x}$ and $\hat{h}$ be respectively a solution and a direction that generate a ray and go to Step 2. Otherwise, let $\hat{x}$ be an optimal solution to (33) and go to Step 3.

2. For each $j = 1, \ldots, t$, solve the linear program

$$\begin{array}{ll} \min & v^T A^j \hat{h} \\ \text{s.t.} & v \in \widetilde{\mathcal{P}}. \end{array} \tag{34}$$

   If any of the problems (34) has negative objective value, let $\tilde{v}$ be a vertex optimal solution to that problem and choose $i \in \{1, \ldots, r\}$ arbitrarily; let $\mathcal{V}^{s+1} := \mathcal{V}^s \cup \{(\tilde{v}, 0, i)\}$ and go to Step 5.

   Otherwise (i.e., if the problems (34) have non-negative objective values for all j), go to Step 3.

3. Solve problems (SepCP$_i$) to find one or more vertex solution(s) $(v, y) \in \mathcal{K}_i$, for some $i \in \{1, \ldots, r\}$, such that

$$\sum_{l=1}^{r} q_l y_l - \sum_{j=1}^{t} p_j (v^T c^i - v^T A^j \hat{x})_+ < 0. \tag{35}$$

   Let (v^{ik}, y^{ik}), $k = 1, \ldots, k_i$ be these identified vertices.

4. If no vertex solution is found in Step 3, **stop**; otherwise, let

$$\mathcal{V}^{s+1} := \mathcal{V}^s \cup \{(v^{ik}, y^{ik}, i), k = 1, \ldots, k_i\}.$$

5. Set $s := s + 1$ and go to Step 1.

Proof: First notice that, if (33) is infeasible, then the original problem is infeasible as well. Suppose now that (33) is unbounded, so $\hat{x}$ and $\hat{h}$ generate a ray, and suppose there exists j_0 such that (34) has negative objective value for $j = j_0$. Since $\tilde{v}^T A^{j_0} \hat{h} < 0$, given any $i \in \{1, \ldots, r\}$ there exists $\alpha_i > 0$ such that

$$\sum_{j=1}^{t} p_j(\tilde{v}^T c^i - \tilde{v}^T A^j(\hat{x} + \alpha_i \hat{h}))_+ \; > \; \sum_{l=1}^{r} q_l(\tilde{v}^T c^i - \tilde{v}^T c^l)_+.$$

It follows that $\hat{x}$ and $\hat{h}$ will not generate a ray for (33) in the next iteration, so $\tilde{v}$ yields a new valid cut.

Suppose next that (33) is unbounded, so $\hat{x}$ and $\hat{h}$ generate a ray, but (34) has non-negative objective value for all j. That is, $v^T A^j \hat{h} \geq 0$ for all $v \in \widetilde{\mathcal{P}}$ and all j. It follows that, for any $i \in \{1, \ldots, r\}$, the term $(v^T c^i - v^T A^j(\hat{x} + \alpha \hat{h}))_+$ is a non-increasing function of α, so to check whether a cut can be generated it suffices to check it for $\alpha = 0$, which is what is done in Step 3.

Finally, if Step 3 identifies one or more vertices satisfying (35), then $\hat{x}$ will not be feasible for (33) in the next iteration, so at least one new valid cut will be generated. If no vertices satisfying (35) can be found, then $\hat{x}$ is feasible for (FLP), hence we have obtained either an optimal solution $\hat{x}$ to (ULP), or a solution $\hat{x}$ and a direction $\hat{h}$ that generate a ray, showing that (ULP) is unbounded. Since cuts are never repeated, the algorithm terminates after a finite number of steps. □

Step 3 of Algorithm 1 requires solving (SepCP$_i$). As discussed earlier, (SepCP$_i$) is a reformulation of (DCP$_i$), which minimizes a difference of two convex polyhedral functions over a polyhedral set. Such problems are called polyhedral DC programming problems, and have been a subject of theoretical and algorithmic study (see e.g., An and Tao 2005, and references therein). The algorithm for DC programming proposed in An and Tao (2005) converges to a local minimum. However, in order to solve (FLP) we may be required to solve (DCP$_i$) to optimality.

(SepCP$_i$) is also a concave minimization problem. The problem of minimizing a concave function over a polyhedral set has also received considerable attention and several approaches have been developed to solve such problems. The methods for solving concave minimization fall into three categories: enumeration methods, successive partitioning methods, and successive approximation (cutting-plane) methods (Horst and Pardalos, 1994; Horst et al., 1995; Al-Khayyal and Sherali, 2000; Locatelli and Thoai, 2000; Porembski, 2002, 2004). Although these methods are developed for general concave objective functions, one may adapt them to exploit the polyhedral structure of the objective function in (SepCP$_i$). Unfortunately, the methods for concave minimization problem are designed to achieve an $\epsilon-$optimal solution, and an exact minimum is possible only if certain conditions are satisfied. For example, Porembski (2002) considers the problem:

$$\begin{array}{l} \min \; f(x) \\ x \in \mathcal{K}, \end{array} \tag{36}$$

where $\mathcal{K}$ (in our case $\mathcal{P}_i$) is assumed to be a bounded full-dimensional polyhedral set and $f(x)$ is a concave function. The finite convergence of the cone adaptation cutting plane method of Porembski

(2002) for (36) is proved under a finite-convergence (FC) condition

$$\begin{array}{c}\text{For any } x_1 \text{ and } x_2 \text{ lying on the edges of } \mathcal{K} \text{ with } \hat{f} \leq \min\{f(x_1), f(x_2)\} \\ \text{we have } \operatorname{conv}(\{x_1, x_2\}) \cap \operatorname{bd}(\mathcal{L}(\hat{f})) \subseteq \{x_1, x_2\},\end{array} \tag{37}$$

where $\hat{f}$ is the objective value of incumbent solutions in the cutting plane algorithm, and $\operatorname{bd}(\mathcal{L}(\hat{f}))$ represents the boundary of $\mathcal{L}(\hat{f}) := \{x \mid f(x) \geq \hat{f}\}$. This condition implies that the face of $\mathcal{P}_i$ that is completely contained in $\operatorname{bd}(\mathcal{L}(\hat{f}))$ has to be a vertex solution of $\mathcal{P}_i$. Such a condition is satisfied when $f(\cdot)$ is a strictly concave function. Unfortunately, the function in (SepCP$_i$) is not strictly concave, and it is possible to construct an example violating the FC condition (37). Hence, using the algorithm in Porembski (2002) we can only expect to generate an $\epsilon-$optimal solution. One way to overcome this problem is to process such an $\epsilon-$optimal solution to an exact optimum — indeed, later in this section we will develop a procedure that generates a vertex solution with an improved objective value starting from an arbitrary feasible solution of (SepCP$_i$). Starting from an $\epsilon-$optimal solution we can use this procedure to generate a vertex solution (SepCP$_i$) for a sufficiently small choice of ϵ.

In what follows we discuss an approach that simultaneously exploits the concavity of the objective function in (SepCP$_i$) and its polyhedral structure. We first formulate (SepCP$_i$) as an integer program, and then present a branch-and-cut method that solves this problem to optimality by exploiting the structural properties of the objective function in (SepCP$_i$). The method is novel in that it optionally generates two types of cuts. It generates "concavity cuts" in the space of $\mathcal{P}_i$, and standard integer programming based cuts in the binary reformulation of the problem. The algorithm is enhanced by four subroutines: a dive-and-search method that finds a local minimizer, a procedure to convert the local minimizer into a *star vertex* solution[2], a routine that yield concavity cuts, and a routine that yields cuts leading to the convex hull of the set given by mixed-integer inequalities. Each of these subroutines will be discussed in detail.

4.1 A Branch and Cut Method for a Class of Polyhedral DCP

We now present a branch-and-cut method to solve (DCP$_i$), while using its formulation as in (SepCP$_i$). Problem (DCP$_i$) belongs to the class of polyhedral DCP problems whose objective function is given by $\max\{0, l(x)\}$, where $l(x)$ is a linear function. The branching in the algorithm exploits the polyhedral structure of the objective functions in these problems, by considering a mixed integer binary linear programming formulation. The cut generation exploits the concavity of the objective function while using the cut generation methodology described in Benson (1999) and Porembski (2002, 2004). In addition it combines the use of "convexity" cuts known from the theory of mixed integer programming.

[2] A vertex of a polyhedron is a *star solution* if its objective value is no higher than the objective values of all of its neighboring vertices.

Let us reformulate (SepCP$_i$) as a binary integer program as follows:

$$\begin{aligned}
\min_{v,y,g,h,b} \; & \sum_{l=1}^{r} q_l \, y_l - \sum_{j=1}^{t} p_j \, g_j && \text{(SepIP}_i\text{)} \\
\text{s. t. } \; & (v,y) \in \mathcal{P}_i, \\
& g_j - h_j \;=\; v^T c^i - v^T A^j \hat{x}, \quad j = 1, \ldots, t && (38) \\
& \alpha_j b_j \;\geq\; g_j, \quad \beta_j (1 - b_j) \;\geq\; h_j, \quad j = 1, \ldots, t \\
& g_j \geq 0, \; h_j \geq 0, \; b_j \in \{0,1\}, \quad j = 1, \ldots, t,
\end{aligned}$$

where $\alpha_j := \max\{\max_{v \in \widetilde{\mathcal{P}}} v^T c^i - v^T A^j \hat{x}, 0\}$, and $\beta_j := -\min\{\min_{v \in \widetilde{\mathcal{P}}} v^T c^i - v^T A^j \hat{x}, 0\}$. The coefficients α_j, β_j together with the binary variable b_j are introduced in the (SepIP$_i$) formulation to ensure that only one of the variables g_j or h_j is positive at a feasible solution of (SepIP$_i$). Note that both α_j and β_j are easily computed by solving a linear program.

Now consider a node in the branch-and-cut tree $\mathcal{T}$, where a subset of the binary variables are fixed to either zero or one. Each node $\mathcal{N}$ of the branch-and-cut tree corresponds to a partition of the binary variables. Let $\mathcal{L}$, $\mathcal{G}$, be the set of variables fixed at zero and one respectively, and let $\mathcal{R}$ be the set of variables to be relaxed. Let $\mathcal{B} := (\mathcal{G}, \mathcal{L}, \mathcal{R})$, and represent that node as $\mathcal{N}^{\mathcal{B}}$. The node corresponds to the following binary linear program (where we use (38) to substitute for variables $g_j, j \in \mathcal{R}$):

$$\begin{aligned}
\min_{v,y,h,b} \; & \sum_{l=1}^{r} q_l y_l - \sum_{j \in \mathcal{G}} p_j (v^T c^i - v^T A^j \hat{x}) - \sum_{j \in \mathcal{R}} p_j (v^T c^i - v^T A^j \hat{x} + h_j) && \text{(SepIP}^{\mathcal{N}}\text{)} \\
\text{s. t. } \; & (v,y) \in \hat{\mathcal{P}}_i \\
& v^T c^i - v^T A^j \hat{x} \;\geq\; 0, \quad j \in \mathcal{G} \\
& v^T c^i - v^T A^j \hat{x} \;\leq\; 0, \quad j \in \mathcal{L} \\
& \hat{\alpha}_j b_j \;\geq\; v^T c^i - v^T A^j \hat{x} + h_j, \quad \hat{\beta}_j (1 - b_j) \;\geq\; h_j, \quad j \in \mathcal{R} \\
& v^T c^i - v^T A^j \hat{x} + h_j \geq 0, \; h_j \geq 0, \; b_j \in \{0,1\}, \quad j \in \mathcal{R}.
\end{aligned}$$

In the above, $\hat{\alpha}_j := \max\{\max_{(v,y) \in \hat{\mathcal{P}}_i} v^T c^i - v^T A^j \hat{x}, 0\}$, $\hat{\beta}_j := -\min\{\min_{(v,y) \in \hat{\mathcal{P}}_i} v^T c^i - v^T A^j \hat{x}, 0\}$, and

$$\hat{\mathcal{P}}_i \;=\; \mathcal{P}_i \cap \{(v,y) \mid D^T v \leq d\}, \tag{39}$$

where $\{v \mid D^T v \leq d\}$ represents the cuts that may have been added to $\mathcal{P}_i$. Note that we may use a bound on $\hat{\alpha}_j$ and $\hat{\beta}_j$ in this formulation. In particular, $\hat{\alpha}_j = \alpha_j$, and $\hat{\beta}_j = \beta_j$ can be used since $\hat{\mathcal{P}}_j \subseteq \mathcal{P}_j$. However, this may result in a significantly lower value of the lower bound on the objective value of (SepCP$_i$) generated from the linear programming relaxation of (SepIP$^{\mathcal{N}}$), since $\hat{\alpha}_j$ and $\hat{\beta}_j$ can be much smaller than α_j and β_j respectively, resulting in a smaller linear relaxation region.

The branch-and-cut algorithm is presented in Algorithm 2. In the spirit of branch-and-cut algorithms in mixed integer programming the convergence of this algorithm is ensured due to branching on binary variables. However, the size of the branch-and-cut tree is managed by adding cuts as necessary. As a consequence of this approach, we avoid $\epsilon-$convergence arguments in the analysis of the algorithms based on concave programming or DC-programming (An and Tao, 2005; Porembski, 2004).

Algorithm 2 A Branch-and-Cut Algorithm for DCP_i

Input $(c, A, \hat{x}, i)$

0. Initialization. $\mathcal{R} := \{1, \ldots, t\}$, $\mathcal{L} := \emptyset$, $\mathcal{G} := \emptyset$, $\mathcal{B} := (\mathcal{G}, \mathcal{L}, \mathcal{R}), \mathcal{T} := \{\mathcal{N}^{\mathcal{B}}\},\ z^U := \infty$.
1. Consider a node $\mathcal{N}$ of $\mathcal{T}$ with smallest objective value (call that value z^L) and let $(v^{\mathcal{N}}, y^{\mathcal{N}}, b^{\mathcal{N}}, h^{\mathcal{N}})$ be an optimal solution of the linear relaxation of $(\text{SepIP}^{\mathcal{N}})$, call it $(\text{LP}^{\mathcal{N}})$. If $z^U \leq z^L$, stop and return (v^U, y^U), else go to Step 2.
2. Let $\mathcal{R}_+ := \{j \in \mathcal{R} \,|\, {v^{\mathcal{N}}}^T c^i - {v^{\mathcal{N}}}^T A^j \hat{x} \geq 0\}$, and $\mathcal{R}_- := \{j \in \mathcal{R} \,|\, {v^{\mathcal{N}}}^T c^i - {v^{\mathcal{N}}}^T A^j \hat{x} \leq 0\}$. Call the routine $\text{DIVE}(\mathcal{N}^{\mathcal{B}}, \mathcal{R}_+, \mathcal{R}_-, \mathcal{L}, \mathcal{G})$ described in Algorithm 3. If $z^{\text{DIVE}} \leq z^U$, then $(v^U, y^U) := (v^{\text{DIVE}}, y^{\text{DIVE}})$ and $z^U := z^{\text{DIVE}}$.
3. If $z^U \leq z^L$, stop and return (v^U, y^U), else go to Step 4.
4. Convert the local minimum $(v^{\text{DIVE}}, y^{\text{DIVE}})$ from $\text{DIVE}(\mathcal{N}^{\mathcal{B}}, \mathcal{R}_+, \mathcal{R}_-, \mathcal{L}, \mathcal{G})$ to a *star vertex solution* $(v^{\text{VERT}}, y^{\text{VERT}})$ of $\hat{\mathcal{P}}_i$ by running $\text{VERT}(v^{\text{DIVE}}, y^{\text{DIVE}})$ described in Algorithm 4.
5. Call $\text{ConcaveCUTS}(v^{\text{VERT}}, y^{\text{VERT}})$ to either generate concavity cut(s) (as described in Section 4.4) that cut away $(v^{\text{VERT}}, y^{\text{VERT}})$, or claim global optimality of $(v^{\text{VERT}}, y^{\text{VERT}})$.
6. Call ConvexCUTS() (as described in Section 4.5) to generate cuts that cut away $(v^{\mathcal{N}}, y^{\mathcal{N}}, b^{\mathcal{N}}, h^{\mathcal{N}})$ if it is not already done so by ConcaveCUTS().
7. Repeat Steps 1 to 6 until no more progress is achieved.
8. Choose a variable $j \in \mathcal{R}$ of node $\mathcal{B}$ to branch. Let $\mathcal{B}^1 := (\mathcal{G} \cup \{j\}, \mathcal{L}, \mathcal{R} \setminus \{j\})$, $\mathcal{B}^2 := (\mathcal{G}, \mathcal{L} \cup \{j\}, \mathcal{R} \setminus \{j\})$, and let $\mathcal{N}^1 := \mathcal{N}^{\mathcal{B}^1}$, $\mathcal{N}^2 := \mathcal{N}^{\mathcal{B}^2}$. Compute $z^{\mathcal{N}^1}$ and $z^{\mathcal{N}^2}$ by solving $\text{LP}^{\mathcal{N}^1}$ and $\text{LP}^{\mathcal{N}^2}$. Delete $\mathcal{N}^1$ if $z^{\mathcal{N}^1} \geq z^U$, or if the problem is infeasible, else set $\mathcal{T} := \mathcal{T} \cup \mathcal{N}^1$, and record the corresponding objective value $z^{\mathcal{N}^1}$. Similarly for $\mathcal{N}^2$. Go to Step 1.

4.2 Dive and Search Method for a Class of Polyhedral DCP

In this section we present a dive and search method for finding a local minimum of (SepCP_i) at any node of the branch-and-cut tree. This method is different from the method in (An and Tao, 2005) for finding a local minimum. In particular, it exploits the polyhedral and max structure of the objective function. In this method we start with a fixed partition $\mathcal{R}$, $\mathcal{G}$ and $\mathcal{L}$ of the binary variables and consider the problem in the space of $\mathcal{P}_i$, i.e., (v, y) space. The procedure iteratively approximates the non-convex portion of the objective function in (SepCP_i) with a linear function (Step 1). As it proceeds, depending on the value of $v^T c^i - v^T A^j \hat{x}$ at the solution in the previous iteration, it updates the objective function, and adds constraints when $v^T c^i - v^T A^j \hat{x} \leq 0$ (Step 2). It allows the possibility of dropping previously added constraints in Step 4. We note that the update of objective and addition of constraints in Step 2 can be performed one constraint at a time, or for all the constraints at once.

Algorithm 3 A Dive and Search Method DIVE($\cdot$)

Input. $\mathcal{N}^{\mathcal{B}}$, $\mathcal{R}_+$, $\mathcal{R}_-$, $\mathcal{L}$, and $\mathcal{G}$

Output. A local minimum solution $(v^{\text{DIVE}}, y^{\text{DIVE}})$, and corresponding objective z^{DIVE}.

0. Initialization. $\mathcal{R}_+^0 := \mathcal{R}_+$, $\mathcal{R}_-^0 := \mathcal{R}_-$, $k := 0$.

1. Let (v^k, y^k) be an optimal solution of the following problem:

$$\min L^k(v,y) := \sum_{l=1}^{r} q_l y_l - \sum_{j \in \mathcal{G} \cup \mathcal{R}_+^k} p_j (v^T c^i - v^T A^j \hat{x}) \tag{LP$_i^k$}$$

$$\begin{aligned} &(v,y) \in \hat{\mathcal{P}}_i \\ &v^T c^i - v^T A^j \hat{x} \ \geq \ 0, \quad j \in \mathcal{G} \\ &v^T c^i - v^T A^j \hat{x} \ \leq \ 0, \quad j \in \mathcal{R}_-^k \cup \mathcal{L}. \end{aligned} \tag{40}$$

   Set $f^k := \sum_{l=1}^{r} q_l y_l^k - \sum_{j \in \mathcal{G} \cup \mathcal{R}_+^k} p_j ({v^k}^T c^i - {v^k}^T A^j \hat{x})_+$, and let λ_k, $j \in \mathcal{R}_-^k$, be the dual solution (Lagrange multipliers) associated with the binding (if any) constraints in (40) at (v^k, y^k).

2. $\mathcal{R}_+^{k+1} := \mathcal{R}_+^k$, and $\mathcal{R}_-^{k+1} := \mathcal{R}_-^k$. For all $j \in \mathcal{R}_+^k$, if $({v^k}^T c^i - {v^k}^T A^j \hat{x}) \leq 0$,
   then $\mathcal{R}_+^{k+1} := \mathcal{R}_+^{k+1} \setminus \{j\}$, and $\mathcal{R}_-^{k+1} := \mathcal{R}_-^{k+1} \cup \{j\}$.

3. If $\mathcal{R}_+^{k+1} \neq \mathcal{R}_+^k$, set $k := k+1$, and go to Step 1; otherwise, go to Step 4.

4. If there exists $\iota \in \mathcal{R}_-^k$ such that $\lambda_\iota > 0$, then set $\mathcal{R}_-^{k+1} := \mathcal{R}_-^k \setminus \{j\}$, $\mathcal{R}_+^{k+1} := \mathcal{R}_+^k \cup \{j\}$, $k := k+1$, and go to Step 1. Otherwise, return (v^k, y^k) and f^k.

Proposition 3 *The function values f^k at the solution in Step 1 of Algorithm 3 are non-increasing. If the primal and dual optimal solutions (v^k, y^k, λ^k) of* (LP_i^k) *satisfy strict complementarity conditions, then (v^k, y^k) is a local minimum solution for the node $\mathcal{N}^{\mathcal{B}}$.*

Proof: Note that for all k

$$\begin{aligned} L^k(v^k, y^k) &= \textstyle\sum_{l=1}^r q_l y_l^k - \sum_{j\in\mathcal{G}\cup\mathcal{R}_+^k} p_j({v^k}^T c^i - {v^k}^T A^j \hat{x}) \\ &\geq \textstyle\sum_{l=1}^r q_j y_l^k - \sum_{j\in\mathcal{G}\cup\mathcal{R}_+^k} p_j({v^k}^T c^i - {v^k}^T A^j \hat{x})_+ = f^k. \end{aligned}$$

Furthermore, when returning to Step 1 from Step 3, $f^k = L^{k+1}(v^k, y^k) \geq L^{k+1}(v^{k+1}, y^{k+1}) \geq f^{k+1}$, since (v^k, y^k) is a feasible solution for (LP_i^{k+1}). Otherwise, we return to Step 1 from Step 4. In this case, since $\lambda_\iota > 0$, and the primal-dual solutions are strictly complementary, from duality theory we know that for some small $\epsilon > 0$, the problem:

$$\begin{aligned} \min L_\epsilon(v, z) := \quad & \sum_{l=1}^r q_l y_l - \sum_{j\in\mathcal{G}\cup\mathcal{R}_+^k} p_j(v^T c^i - v^T A^j \hat{x}) \\ \text{s. t.} \quad & (v, y) \in \hat{\mathcal{P}}_i \\ & v^T c^i - v^T A^j \hat{x} \geq 0, \quad j \in \mathcal{G} \\ & v^T c^i - v^T A^j \hat{x} \leq 0, \quad j \in \mathcal{R}_-^k \cup \mathcal{L}, j \neq \iota \\ & v^T c^i - v^T A^\iota \hat{x} \leq \epsilon, \end{aligned} \tag{41}$$

will have constraint (41) active at an optimal solution $(v_\epsilon^*, y_\epsilon^*)$. Moreover, the optimal objective value L_ϵ^* satisfies:

$$L_\epsilon^* \;\leq\; L(v^k, y^k) \;=\; f^k,$$

where the last equality follows because we enter into Step 4 only when $\mathcal{R}_+^{k+1} = \mathcal{R}_+^k$, i.e., when the LP solution coincides with the corresponding DCP solution for a DCP problem defined using $\mathcal{G} \cup \mathcal{R}_+^k$. It follows that

$$\begin{aligned} f^{k+1} \;\leq\; L^{k+1}(v^{k+1}, y^{k+1}) \;&\leq\; \sum_{l=1}^r q_l {y_\epsilon^*}_l - \sum_{j\in\mathcal{G}\cup\mathcal{R}_+^k} p_j({v_\epsilon^*}^T c^i - {v_\epsilon^*}^T A^j \hat{x}) - p_\iota({v_\epsilon^*}^T c^i - {v_\epsilon^*}^T A^\iota \hat{x}) \\ &= \; L_\epsilon^* - p_\iota \epsilon \;\leq\; f^k - p_\iota \epsilon. \end{aligned}$$

In the above, the second inequality follows from the fact that $(v_\epsilon^*, y_\epsilon^*)$ is a feasible solution for (LP_i^{k+1}), whereas the equality follows from the fact that (41) is active at $(v_\epsilon^*, y_\epsilon^*)$. The claim that (v^k, y^k) is a local minimum for $\mathcal{N}^{\mathcal{B}}$ follows because upon exit from Step 4 (v^k, y^k) is a global minimum for the problem

$$\begin{aligned} \min \quad & \sum_{l=1}^r q_l y_l - \sum_{j\in\mathcal{G}\cup\mathcal{R}_+^k} p_j(v^T c^i - v^T A^j \hat{x})_+ \\ & (v, y) \in \hat{\mathcal{P}}_i \\ \text{s. t.} \quad & v^T c^i - v^T A^j \hat{x} \geq 0, \quad j \in \mathcal{G} \\ & v^T c^i - v^T A^j \hat{x} \leq 0, \quad j \in \mathcal{R}_-^k \cup \mathcal{L}, \end{aligned}$$

and, moreover, ${v^k}^T c^i - {v^k}^T A^j \hat{x} < 0$ for all $j \in \mathcal{R}_-^k$ since $\lambda_j = 0$ for all $j \in \mathcal{R}_-^k$ and strict complementarity holds. $\square$

It is worthwhile mentioning that the strict complementarity assumption made in Proposition 3 is not a strong requirement — this will be satisfied if (LP_i^k) is solved using interior point methods (Mehrotra and Ye, 1993).

4.3 Generating a Vertex Solution

We discuss now a procedure to convert a local minimum solution of problem ($\text{SepIP}^{\mathcal{N}}$) into a star vertex solution of $\hat{\mathcal{P}}_i$ defined in (39). We can use the same procedure ot convert an $\epsilon-$optimal non-vertex solution to a better vertex solution in (SepCP_i). Let us follow the notation in Theorem 1. First observe that the set of extreme directions of $\hat{\mathcal{P}}_i$ is $0 \times R_+^r$, since $\widetilde{\mathcal{P}}$ is bounded. Furthermore, the objective value of (SepCP_i) is non-decreasing along any direction in $0 \times R_+^r$. Hence, without loss of generality we may assume that any solution $(\hat{v}, \hat{y})$ satisfies $\hat{y}_j = \max\{0, \hat{v}^T(c^i - c^j)\}$. Now assume that $(\hat{v}, \hat{y})$ is a non-vertex solution of $\hat{\mathcal{P}}_i$, satisfying a subset of constraint in $\hat{\mathcal{P}}_i$ as equality constraints. Let us represent these constraints by $A(v, y) = a$. Let $p = (p_v, p_y) \neq 0$ be a direction satisfying $Ap = 0$ (or any direction if A is empty), and consider points $(\bar{v}, \bar{y}) = (\hat{v}, \hat{y}) + \bar{\alpha}(p_v, p_y)$ and $(\tilde{v}, \tilde{y}) = (\hat{v}, \hat{y}) - \tilde{\alpha}(p_v, p_y)$, where $\bar{\alpha}$ and $\tilde{\alpha}$ are maximum step lengths that we can take along directions p and $-p$ without violating feasibility. Note that $p_v \neq 0$, otherwise, $p_y = 0$. Hence, such finite $\bar{\alpha}$ and $\tilde{\alpha}$ exist. Now, since $(\hat{v}, \hat{y}) = \frac{\tilde{\alpha}}{\tilde{\alpha}+\bar{\alpha}}(\bar{v}, \bar{y}) + \frac{\bar{\alpha}}{\tilde{\alpha}+\bar{\alpha}}(\tilde{v}, \tilde{y})$, due to concavity of $f(v, y)$ we have $f(\hat{v}, \hat{y}) \geq \frac{\tilde{\alpha}}{\tilde{\alpha}+\bar{\alpha}} f(\bar{v}, \bar{y}) + \frac{\bar{\alpha}}{\tilde{\alpha}+\bar{\alpha}} f(\tilde{v}, \tilde{y})$. We now set $(v^\circ, y^\circ) := \text{argmin}\,\{f(\bar{v}, \bar{y}), f(\tilde{v}, \tilde{y})\}$, and $A := \begin{bmatrix} A \\ B^1 \end{bmatrix}$, $a = \begin{pmatrix} a \\ a^1 \end{pmatrix}$ where $B^1(v^\circ, y^\circ) = a^1$ are additional binding constraints at (v°, y°). Because of the choice of $\bar{\alpha}$ and $\tilde{\alpha}$, at least one row of B^1 must be linearly independent of rows of A. Thus, the above procedure will terminate with a vertex solution (v°, y°) of $\hat{\mathcal{P}}_i$ after a finite number of repetitions. After such a vertex is obtained, simplex-type pivoting operations can be done until a vertex that is better than all its neighbors is found.

4.4 Concavity Cuts

In this section we give a basic approach to generating concavity cuts used in the concave minimization problem (SepCP_i). The discussion here is based on Benson (1999). However, there is an important difference. The methodology developed in Benson (1999) assumes that the polyhedral constraint set for the concave minimization is bounded. This assumption is not true in the case of (SepCP_i). However, we exploit the structure of the problem to give an extension of the method in Benson (1999) in our case. Porembski (2002) has given approaches for strengthening (deepening) concavity cuts. We refer the reader to Porembski (2002) for a further discussion on this topic.

Consider problem ($\text{SepIP}^{\mathcal{N}}$) defined in Section 4.1. The concavity cuts are defined on the space of (v, y), based on a given star vertex solution (v^*, y^*) of $\hat{\mathcal{P}}_i$. Let $\mathcal{E}$ be the set of edge directions of $\hat{\mathcal{P}}_i$ available at (v^*, y^*) (note that $\mathcal{E}$ is finite due to the polyhedral nature of the problem). The following theorem gives a system of inequalities for generating a concavity cut at (v^*, y^*).

Theorem 5 *Assume that (v^*, y^*) is a star vertex solution of $\hat{\mathcal{P}}_i$, and $z^* = f(v^*, y^*)$. Let $\{d^1, \ldots, d^s\} \subset \mathcal{E}$ be the set of edge directions of $\hat{\mathcal{P}}_i$ at (v^*, y^*) such that $z^* \leq f((v^*, y^*) + \theta d^j)$, for all $\theta > 0$.*

Algorithm 4 An algorithm for converting an interior solution into a star vertex solution VERT$(\cdot)$

Input $(\hat{v}, \hat{y})$

While $(\hat{v}, \hat{y})$ is not a vertex solution of $\hat{\mathcal{P}}_i$
{

1. Set $\hat{y}_j := \max\{0, \hat{v}^T(c^i - c^j)\}$, $j = 1, \ldots, r$.
2. Identify binding constraints $A(\hat{v}, \hat{y}) = a$ at $(\hat{v}, \hat{y})$
3. Find $p = (p_v, p_y) \neq 0$ satisfying $Ap = 0$ (take an arbitrary $p \neq 0$ if A is empty).
4. Compute $\bar{\alpha}, \tilde{\alpha}$ using minimum ratio tests along p and $-p$ respectively.
5. Set $(\hat{v}, \hat{y}) := \text{argmin}\{f(\bar{v}, \bar{y}), f(\tilde{v}, \tilde{y})\}$, where $(\bar{v}, \bar{y}) = (\hat{v}, \hat{y}) + \bar{\alpha}(p_v, p_y)$ and $(\tilde{v}, \tilde{y}) = (\hat{v}, \hat{y}) - \tilde{\alpha}(p_v, p_y)$.

}
Starting from $(\hat{v}, \hat{y})$, do simplex-type pivoting operations until a star vertex solution is found.

---

1. *If $\mathcal{E}\backslash\{d^1, \ldots, d^s\} = \emptyset$, then (v^*, y^*) is a global minimum of* (SepIP$^{\mathcal{N}}$)*;*
2. *Otherwise, let $\{\eta^1, \ldots, \eta^u\} := \mathcal{E} \setminus \{d^1, \ldots, d^s\}$, $\theta_j := \sup\{\theta > 0 : f((v^*, y^*) + \theta\eta^j) \geq f(v^*, y^*)\}, j = 1, \ldots, u$ (note that $\theta_j < \infty$), and let π be a solution of the system of equations*

$$\pi^T\eta^j \geq 1/\theta_j, \; j = 1, \ldots, u, \quad \pi^T d^j \geq 0, \; j = 1, \ldots, s. \tag{42}$$

Then $\pi^T(v, y) \geq \pi^T(v^*, y^*) + 1$ is a valid cut for $\hat{\mathcal{P}}_i$, i.e., $f(v, y) \geq z^*$ for all $(v, y) \in \mathcal{C} := \{\hat{\mathcal{P}}_i \cap \{(v, y) \mid \pi^T(v, y) \leq \pi^T(v^*, y^*) + 1\}\}$. Furthermore, (42) *is non-empty.*

Proof: The first part of the result follows from concavity of $f(\cdot)$. To show the second part, let us take a $(v, y) \in \mathcal{C}$. Note that any $(v, y) \in \hat{\mathcal{P}}_i$ can be written as

$$(v, y) \;=\; (v^*, y^*) + \sum_{j=1}^{u} \rho_j\eta^j + \sum_{j=1}^{s} \gamma_j d^j \tag{43}$$

for some non-negative coefficients $\{\rho_j\}$ and $\{\gamma_j\}$. By multiplying the above equation by an arbitrary vector π satisfying (42) we obtain

$$\begin{aligned} \pi^T(v, y) \;&=\; \pi^T(v^*, y^*) + \sum_{j=1}^{u} \rho_j\pi^T\eta^j + \sum_{j=1}^{s} \gamma_j\pi^T d^j \\ &\geq\; \pi^T(v^*, y^*) + \sum_{j=1}^{u} \frac{\rho_j}{\theta_j} \end{aligned}$$

and thus, since $(v,y) \in \mathcal{C}$, we conclude that $\delta := \sum_{j=1}^{u} \frac{\rho_j}{\theta_j} \leq 1$. Now, rewrite (43) as

$$(v,y) \;=\; (1-\delta)\left((v^*,y^*) + \frac{1}{1-\delta}\sum_{j=1}^{s}\gamma_j d^j\right) + \sum_{j=1}^{u}\left[\frac{\rho_j}{\theta_j}\left((v^*,y^*)+\theta_j\eta^j\right)\right].$$

Concavity of f implies that

$$f(v,y) \;\geq\; (1-\delta)\, f\left((v^*,y^*) + \frac{1}{1-\delta}\sum_{j=1}^{s}\gamma_j d^j\right) + \sum_{j=1}^{u}\frac{\rho_j}{\theta_j}\, f\left((v^*,y^*)+\theta_j\eta^j\right)$$

and since $f((v^*,y^*) + \frac{\gamma_j}{1-\delta}d^j) \geq f(v^*,y^*)$ and $f((v^*,y^*)+\theta_j\eta^j) \geq f(v^*,y^*)$ for all j, we have

$$f(v,y) \;\geq\; (1-\delta)f(v^*,y^*) + \sum_{j=1}^{u}\frac{\rho_j}{\theta_j} f(v^*,y^*) \;=\; f(v^*,y^*) = z^*.$$

Now from (Benson, 1999, Theorem 2.1) we have that

$$\pi^T\eta^j \geq 1/\theta_j,\; j=1,\ldots,u, \quad \pi^T d^j \geq \epsilon,\; j=1,\ldots,s$$

is feasible for some $\epsilon > 0$. Hence, the feasibility of (42) follows immediately. $\square$

Remark: Note that in our context in Theorem 5 we can get a deeper valid cut by replacing z^* with $\min\{z^U, 0\}$, where z^U is the best known objective value of (SepIP$_i$) used in Algorithm 2. Also, the assumption that (v^*,y^*) a star vertex solution is not essential — the arguments in the proof carry over to the case where (v^*,y^*) is just a locally optimal vertex solution.

4.5 Convexity Cuts

Two related approaches based on the theory of mixed binary linear programming are possible to generate tighter relaxations of (SepIP$^{\mathcal{N}}$). These are: (i) adding "lift-and-project" cuts generated using the methods of Balas et al. (1993) and (ii) the relaxation linearization technique (RLT) of Sherali and Adams (1994). Unfortunately, the number of constrains grow exponentially when defining the RLT hierarchy. Instead of generating linear programs with an increasing hierarchy of constraints, Balas et al. (1993) find inequalities that consider their projection in the space of original variables. Efficient methods for generating cuts in this fashion are studied in Balas and Perregaard (2002).

5 Conclusions

We have studied an uncertain linear programming problem, where the constraints are defined using a concept of $\mathcal{P}$-dominance and the data is given over a finite support. We have shown that this uncertain linear program can be reformulated as a finite linear program. We have presented a cutting-surface algorithm for solving this problem where the cuts are generating using a difference of convex function minimization problem. We have given a novel algorithm for the difference of convex function minimization problem by exploiting the polyhedral properties of this problem. This algorithm is enhanced by four subroutines: a dive-and-search method that finds a local minimizer, a method to convert a local minimizer into a star vertex solution, a subroutine that yields concavity cuts, and a subroutine that yields standard convexity (e.g., disjunctive) cuts. The concavity cuts remove a local minimum (or star) solution of the difference of convex function minimization problem. A local minimizer is useful in terminating our search for an optimal solution of (SepCP$_i$) early. By cutting away previously generated local minima, the concavity cuts have the potential value of allowing us to generate a desirable solution (with negative objective value) of this problem quickly, without having to solve the difference of convex function minimization problem to optimality. The convexity cuts are used to obtain tighter linear programming relaxations of the mixed integer linear programming formulation of the difference of convex function minimization problem. The convexity cuts are beneficial in reducing the size of the branch-and-bound tree, and in proving optimality of a feasible solution.

The algorithm presented in this paper remains valid if decision variables have additional restrictions, such as integrality requirements. A generalization of our results to the case where the problem data is defined using continuous distributions, and an extension of the proposed approach to more general problems are topics of a forthcoming paper Hu et al. (2009). An efficient implementation and numerical testing of the proposed algorithms is a topic of future research.

Acknowledgement

The authors thank Jian Hu for valuable discussions that led us to re-think the presentation of our results. We also thank two anonymous referees for their comments and suggestions. This research has been supported by NSF Grant CMMI-0727532.

References

F. A. Al-Khayyal and H. D. Sherali. On finitely terminating branch-and-bound algorithms for some global optimization problems. *SIAM J. on Optimization*, 10:1049–1057, 2000.

L. T. H. An and P. D. Tao. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, 133:23–46, 2005.

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

E. Balas and M. Perregaard. Lift-and-project for mixed 0-1 programming: recent progress. *Discrete Appl. Math.*, 123(1-3):129–154, 2002. ISSN 0166-218X.

H. P. Benson. Generalized γ-valid cut procedure for concave minimization. *Journal of Optimization Theory and Applications*, 102(2):289–298, 1999.

C.-S. Chang, X. Chao, M. Pinedo, and J. G. Shanthikumar. Stochastic convexity for multidimensional processes and its applications. *IEEE Transactions on Automatic Control*, 36(12): 1347–1355, 1991.

D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM J. Optim.*, 14(2):548–566, 2003.

D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Math. Programming*, 99:329–350, 2004.

D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Math. Programming*, 117:111–127, 2009.

J. B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*, volume I. Springer-Verlag, Berlin, Germany, 1993.

R. Horst and P. M. Pardalos. *Handbook of Global Optimization*. Boston, 1994.

R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization*. Boston, 1995.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Sample average approximation and methods for stochastic dominance constrained programs, 2009.

C. C. Huang, D. Kira, and I. Vertinsky. Stochastic dominance rules for multi-attribute utility functions. *The Review of Economic Studies*, 45(3):611–615, 1978.

M. Locatelli and N. V. Thoai. Finite exact branch-and-bound algorithms for concave minimization over polytopes. *Journal of Global Optimization*, 18:107–128, 2000.

S. Mehrotra and Y. Ye. Finding an interior point in the optimal face of linear programs. *Mathematical Programming (A)*, 63(3):497 – 515, 1993.

A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, 2002.

M. Porembski. Cone adaption strategies for a finite and exact cutting plane algorithm for concave minimization. *Journal of Global Optimization*, 24:89–107, 2002.

M. Porembski. A new successive partition algorithm for concave minimization based on cone decomposition and decomposition cuts. *Journal of Global Optimization*, 29:191–224, 2004.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

M. Shaked and J. G. Shanthikumar. *Stochastic Orders and their Applications*. Academic Press, Boston, 1994.

H. D. Sherali and W. P. Adams. A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems. *Discrete Applied Mathematics*, 52:83–106, 1994.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2nd. edition, 1947.

Appendix

Proof of Theorem 2: The necessity part is immediate, noticing that the function $u(x_1, \ldots, x_n) := u_\ell(x_\ell)$ (where u_ℓ is concave and increasing in $\mathbb{R}^1$) is concave and increasing in $\mathbb{R}^n$.

To show sufficiency, we use induction in n. For $n = 1$ the result is trivial. Suppose it holds for an arbitrary n. Consider a concave increasing function u in $\mathbb{R}^{n+1}$. Given $x_{n+1} \in \mathbb{R}$, define the function $u_{x_{n+1}}$ in $\mathbb{R}^n$ as $u_{x_{n+1}}(x_1, \ldots, x_n) := u(x_1, \ldots, x_{n+1})$. Clearly, $u_{x_{n+1}}$ is concave increasing in $\mathbb{R}^n$. Thus, by the induction hypothesis, given two random vectors X and Y in $\mathbb{R}^{n+1}$ we have

$$\begin{aligned} \mathbb{E}[u(X_1, \ldots, X_n, x_{n+1})] &= \mathbb{E}[u_{x_{n+1}}(X_1, \ldots, X_n)] \geq \mathbb{E}[u_{x_{n+1}}(Y_1, \ldots, Y_n)] \\ &= \mathbb{E}[u(Y_1, \ldots, Y_n, x_{n+1})]. \end{aligned} \tag{44}$$

Let $F_{X_{n+1}}(\cdot)$ be the cumulative distribution function of X_{n+1}. By integrating both sides of the above inequality we obtain

$$\begin{aligned} \mathbb{E}[u(X_1, \ldots, X_n, X_{n+1})] &= \int_{\mathbb{R}} \mathbb{E}[u(X_1, \ldots, X_n, x_{n+1})] \, dF_{X_{n+1}}(x_{n+1}) \\ &\geq \int_{\mathbb{R}} \mathbb{E}[u(Y_1, \ldots, Y_n, x_{n+1})] \, dF_{X_{n+1}}(x_{n+1}). \end{aligned} \tag{45}$$

Next, define the function $\tilde{u}$ in $\mathbb{R}$ as $\tilde{u}(\cdot) := \mathbb{E}[u(Y_1, \ldots, Y_n, \cdot)]$. Notice that the right-most term in (45) can be written as

$$\int_{\mathbb{R}} \mathbb{E}[u(Y_1, \ldots, Y_n, x_{n+1})] \, dF_{X_{n+1}}(x_{n+1}) = \int_{\mathbb{R}} \tilde{u}(x_{n+1}) \, dF_{X_{n+1}}(x_{n+1}) = \mathbb{E}[\tilde{u}(X_{n+1})]. \tag{46}$$

Moreover, $\tilde{u}$ is concave increasing since so is u. Thus, since $X_{n+1} \trianglerighteq_{(2)} Y_{n+1}$ by assumption, we have

$$\begin{aligned} \mathbb{E}[\tilde{u}(X_{n+1})] &\geq \mathbb{E}[\tilde{u}(Y_{n+1})] \\ &= \int_{\mathbb{R}} \tilde{u}(y_{n+1}) \, dF_{Y_{n+1}}(y_{n+1}) \\ &= \int_{\mathbb{R}} \mathbb{E}[u(Y_1, \ldots, Y_n, y_{n+1})] \, dF_{Y_{n+1}}(y_{n+1}) \\ &= \mathbb{E}[u(Y_1, \ldots, Y_n, Y_{n+1})]. \end{aligned} \tag{47}$$

The result then follows by combining inequalities (45), (46) and (47). $\square$