Mathematical Programming manuscript No.
(will be inserted by the editor)

Darinka Dentcheva · Andrzej Ruszczyński

Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints

Received: January 29, 2003 / Revised version: date

Abstract. We consider a new class of optimization problems involving stochastic dominance constraints of second order. We develop a new splitting approach to these models, optimality conditions and duality theory. These results are used to construct special decomposition methods.

Key words. Stochastic Programming – Stochastic Dominance – Optimality – Duality – Decomposition

1. Introduction

In stochastic decision models our decisions affect various random outcomes. There are different ways to formalize our preferences and our objective involving these outcomes. Many decision models involve expected values of the random outcomes, or probabilities of exceeding some threshold values.

One of established ways to formalize preferences among random outcomes is the relation of *stochastic dominance*. We refer to [8] and to [13] for a more general perspective. In recent publications [5, 6], we have introduced a new stochastic optimization model involving stochastic dominance relations as constraints. These constraints allow us to use random reference outcomes, instead of fixed thresholds. We have discovered the role of utility functions as Lagrange multipliers associated with dominance constraints.

Darinka Dentcheva: Stevens Institute of Technology, Department of Mathematics, Hoboken, NJ 07030 e-mail: ddentche@stevens-tech.edu

Andrzej Ruszczyński: Rutgers University, Department of Management Science and Information Systems, Piscataway, NJ 08854, USA, e-mail: rusz@rutcor.rutgers.edu

In this paper we shall consider the following general problem setting:

$$\max \mathbb{E}[H(z)] \tag{1}$$

$$\text{subject to } G_i(z) \succeq_{(2)} Y_i, \quad i = 1, \dots, m, \tag{2}$$

$$z \in Z. \tag{3}$$

Here Z is a convex subset of a separable locally convex Hausdorff vector space $\mathcal{Z}$, and G_i, $i = 1, \dots, m$, and H are *continuous* operators from $\mathcal{Z}$ to the space of real random variables $\mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R})$. The operators G_i and H are assumed to be concave in the following sense: for P-almost all $\omega \in \Omega$ the functions $[G_i(\cdot)](\omega)$, $i = 1, \dots, m$, and $[H(\cdot)](\omega)$ are *concave and continuous* on $\mathcal{Z}$.

Relation (2) is the second order stochastic dominance relation between the random variables $G_i(z)$ and Y_i. The random variables $Y_i \in \mathcal{L}_1$ play the role of *fixed reference outcomes*.

The relation of stochastic dominance is defined as follows. For a random variable $X \in \mathcal{L}_1$ we consider its distribution function, $F(X; \eta) = P[X \le \eta]$, and the function

$$F_2(X; \eta) = \int_{-\infty}^{\eta} F(X; \alpha)\, d\alpha \quad \text{for } \eta \in \mathbb{R}. \tag{4}$$

As an integral of a nondecreasing function, F_2 is a convex function of η. We say that a random variable $X \in \mathcal{L}_1$ *dominates* in the second order a random variable $Y \in \mathcal{L}_1$ if

$$F_2(X; \eta) \le F_2(Y; \eta) \quad \text{for all } \eta \in \mathbb{R}.$$

We focus on the second order dominance relation as the most relevant in applications. Our problem is formulated as a maximization problem with concave functions, because the stochastic dominance relation is associated with concave nondecreasing utility functions and it usually appears in the context of maximization.

In our earlier paper [6] we have considered a basic version of this problem with pure dominance constraints and we have developed necessary and sufficient optimality conditions. Our objective in this paper is to extend this analysis to more involved models in which our decisions affect in a nonlinear way many random outcomes subjected to dominance constraints. We develop a new optimality and duality theory which will allow us to create a decomposition approach to the problem.

In the next section we introduce a split-variable formulation of the problem. Section 3 is devoted to the development of necessary and sufficient optimality conditions. In Section 4 we present the duality and decomposition theory. Section 5 refines the results in the finite-dimensional case. In Section 6 we have a numerical illustration on a large real-world portfolio problem.

2. The split variable formulation

Let us introduce some notation used throughout the paper. An abstract probability space is denoted by $(\Omega, \mathcal{F}, P)$. The expected value operator is denoted by $\mathbb{E}$. The standard symbol $\mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ (shortly $\mathcal{L}_p^n$) denotes the space of all measurable mappings X from Ω to $\mathbb{R}^n$ for which $\mathbb{E}[|X|^p] < \infty$. For $p = 0$ we shall understand it as the space of all measurable mappings, and for $p = \infty$ as the space of all essentially bounded mappings. If the values are taken in $\mathbb{R}$ the superscript n will be omitted. The space of continuous functions on $[a, b] \subset \mathbb{R}$ is denoted $\mathcal{C}([a, b])$.

For a measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and an element $y \in \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, we shall understand $f(y)$ as a real random variable v with realizations $v(\omega) = f(y(\omega))$, $\omega \in \Omega$. If $\varphi : \mathcal{X} \to \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ then we write $\varphi(x)(\omega)$ to denote the realization of the random vector $\varphi(x)$ at an elementary event $\omega \in \Omega$.

The notation $\langle \theta, y \rangle$ is always used to denote the value of a linear functional $\theta \in \mathcal{X}^*$ at the point $y \in \mathcal{X}$, where $\mathcal{X}^*$ is the topological dual space to the topological vector space $\mathcal{X}$. The symbol $\|\cdot\|$ is used to denote the norm in the corresponding space; sometimes we use $\|\cdot\|_{\mathcal{X}}$ to stress the corresponding space $\mathcal{X}$.

The extended real line $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ is denoted by $\overline{\mathbb{R}}$. For a concave function $f : \mathcal{X} \to \overline{\mathbb{R}}$ we use the symbol $\partial f(x)$ to denote its subdifferential at x: the set of $\theta \in \mathcal{X}^*$ such that $f(y) \le f(x) + \langle \theta, y - x \rangle$ for all $y \in \mathcal{X}$.

We concentrate on the analysis of a relaxed version of problem(1)–(3):

$$\max \mathbb{E}[H(z)] \tag{5}$$

$$\text{subject to } F_2(G_i(z); \eta) \le F_2(Y_i; \eta) \quad \text{for all } \eta \in [a_i, b_i], \quad i = 1, \dots, m, \tag{6}$$

$$z \in Z, \tag{7}$$

where $[a_i, b_i]$, $i = 1, \dots, m$, are bounded intervals. If all $G_i(z)$, $z \in Z$, have uniformly bounded distributions, (6) is equivalent to (2) for appropriately chosen a_i and b_i. However, if the distributions are not uniformly bounded, (6) is a relaxation of (2).

The key constraint is the relation (6). Therefore we start from the characterization of the set $A(Y)$ of random variables $X \in \mathcal{L}_1^m$ satisfying:

$$F_2(X_i; \eta) \le F_2(Y_i; \eta) \quad \text{for all } \eta \in [a_i, b_i], \quad i = 1, \dots, m. \tag{8}$$

Changing the order of integration in (4) we get (see, e.g., [14])

$$F_2(X_i; \eta) = \mathbb{E}[(\eta - X_i)_+], \quad i = 1, \dots, m. \tag{9}$$

Therefore, an equivalent representation of (8) is:

$$\mathbb{E}[(\eta - X_i)_+] \le \mathbb{E}[(\eta - Y_i)_+] \quad \text{for all } \eta \in [a_i, b_i], \quad i = 1, \dots, m. \tag{10}$$

The following lemma is a slightly modified version of Proposition 2.3 of [6] and its proof is omitted here.

Lemma 1. *For every $Y \in \mathcal{L}_1^m$ the set $A(Y)$ is convex and closed. Furthermore, its recession cone has the form*

$$A^\infty(Y) = \{H \in \mathcal{L}_1^m : H \geq 0 \text{ a.s.}\}.$$

Let us consider a split-variable formulation of problem (5)–(7):

$$\max \mathbb{E}[H(z)] \tag{11}$$

$$\text{subject to } F_2(X_i;\eta) \leq F_2(Y_i;\eta) \quad \text{for all } \eta \in [a_i,b_i], \quad i=1,\dots,m, \tag{12}$$

$$G_i(z) \geq X_i \quad \text{a.s.}, \quad i=1,\dots,m, \tag{13}$$

$$z \in Z, \quad X_i \in \mathcal{L}_1, \quad i=1,\dots,m. \tag{14}$$

Introducing the variables $X_i \in \mathcal{L}_1$ we have separated the dominance constraints from the nonlinear functions G_i and we have put them in the pure form $X_i \succeq_{(2)} Y_i$ in $[a_i,b_i]$. This has two advantages. First, we can apply and develop the ideas from [6] to pure dominance constraints. Secondly, the splitting facilitates the decomposition approach of section 4. On the other hand, constraints (13) cannot be readily handled by the available optimization theory, because of the empty interior of the nonnegative cone in $\mathcal{L}_1$. In the next section we develop a dedicated approach to overcome this difficulty.

Let us observe that the assumptions that Z is a convex set and G_i and H are concave a.s., together with Lemma 1, imply that problem (11)–(14) is a convex optimization problem.

We denote $X = (X_1,\dots,X_m)$ and $G(z) = (G_1(z),\dots,G_m(z))$.

Proposition 1. *For every optimal solution $\hat{z}$ of problem* (5)–(7) *the point $(\hat{z}, G(\hat{z}))$ is an optimal solution of* (11)–(14)*. For every optimal solution $(\hat{z}, \hat{X})$ of* (11)–(14)*, the point $\hat{z}$ is an optimal solution of problem* (5)–(7)*.*

Proof. Let $\hat{z}$ be an optimal solution of problem (5)–(7). Then $(\hat{z}, G_1(\hat{z}),\dots,G_m(\hat{z}))$ is feasible for (11)–(14). On the other hand, for any optimal solution $(\hat{z}, \hat{X}_1,\dots,\hat{X}_m)$ of (11)–(14), we have

$$G_i(\hat{z}) \geq \hat{X}_i \quad \text{a.s.}, \quad i=1,\dots,m.$$

Therefore $G(\hat{z}) - \hat{X} \in A^\infty(Y)$, by virtue of Lemma 1. Since $\hat{X} \in A(Y)$ then $G(\hat{z}) \in A(Y)$. Consequently, $\hat{z}$ is feasible for (5)–(7). □

3. Optimality

We start our analysis of problem (11)–(14) from a version of necessary and sufficient conditions of optimality for a special convex stochastic optimization problem.

Theorem 1. *Assume that the function $\varphi : \mathcal{Z} \to \mathcal{L}_1^n$ is continuous and such that the function $\varphi(\cdot)(\omega)$ is concave and continuous on $\mathcal{Z}$ for P-almost all $\omega \in \Omega$. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is concave and monotone with respect to the usual partial order in $\mathbb{R}^n$, with uniformly bounded subdifferentials. Let Z be a convex set in $\mathcal{Z}$.*

A point $\hat{z}$ is a solution of the problem

$$\max_{z \in Z} \mathbb{E} f(\varphi(z)),$$

if and only if there exists $\theta \in \mathcal{L}_\infty^n$, $\theta(\omega) \in \partial f(\varphi(\hat{z})(\omega))$ for P-almost all $\omega \in \Omega$, such that $\hat{z}$ is a solution of the problem

$$\max_{z \in Z} \langle \theta, \varphi(z) \rangle.$$

Proof. Define $l : \mathcal{Z} \to \mathcal{L}_1$ as $l(z) = f(\varphi(z))$, and let $\bar{l}(z) = \mathbb{E}[l(z)]$. The functional $l(\cdot)(\omega)$ is concave for P-almost all $\omega \in \Omega$, owing to the monotonicity of f, and to the concavity of φ and f. Thus $\bar{l}$ is concave as well. We have

$$\begin{aligned} l(z_1)(\omega) - l(z_2)(\omega) &= f(\varphi(z_1)(\omega)) - f(\varphi(z_2)(\omega)) \\ &\le \langle \lambda(\omega), \varphi(z_1)(\omega) - \varphi(z_2)(\omega) \rangle, \end{aligned}$$

where $\lambda(\omega) \in \partial f(\varphi(z_2)(\omega))$ a.s.. Taking the expected values we obtain

$$\begin{aligned} \bar{l}(z_1) - \bar{l}(z_2) &= \mathbb{E}[f(\varphi(z_1)(\omega)) - f(\varphi(z_2)(\omega))] \\ &\le \mathbb{E} \langle \lambda(\omega), \varphi(z_1)(\omega) - \varphi(z_2)(\omega) \rangle. \end{aligned}$$

By assumption, there exists c such that $\|\lambda(\omega)\|_{\mathbb{R}^n} \le c$. Therefore

$$l(z_1)(\omega) - l(z_2)(\omega) \le c \|\varphi(z_1)(\omega) - \varphi(z_2)(\omega)\|_{\mathbb{R}^n},$$

$$\bar{l}(z_1) - \bar{l}(z_2) \le c \|\varphi(z_1) - \varphi(z_2)\|_{\mathcal{L}_1^n}.$$

Interchanging the role of z_1 and z_2 and using the continuity assumptions about φ we conclude that the functionals $l(\cdot)(\omega)$ and $\bar{l}(\cdot)$ are continuous. Therefore, they are subdifferentiable everywhere (see, e.g, [3, Th. I-27]).

It follows that at the solution $\hat{z}$ a subgradient $v \in \partial \bar{l}(\hat{z})$ exists, $v \in \mathcal{Z}^*$, such that

$$\langle v, z - \hat{z} \rangle \le 0 \quad \text{for all } z \in Z. \tag{15}$$

By the generalization of Strassen's theorem provided in [12, Thm. 1.1], we can interchange the subdifferentiation and integration operators in the calculation of v (see also [21,3]). It follows that there exists a weakly* measurable mapping $g : \Omega \to \mathcal{Z}^*$, such that

$$g(\omega) \in \partial l(\hat{z})(\omega) \quad \text{for } P\text{-almost all } \omega \in \Omega$$

and, for all $z \in \mathcal{Z}$,

$$\langle v, z - \hat{z} \rangle = \mathbb{E} \, \langle g(\omega), z - \hat{z} \rangle. \tag{16}$$

In the formula above the expected value is well-defined for all z, that is, v is a weak* expected value of g. Invoking [4, Thm. 2.3.9] for each ω, we can represent $g(\omega)$ as follows:

$$g(\omega) = \sum_{i=1}^{n} \theta_i(\omega) d_i(\omega) \quad \text{a.s.},$$

where $\theta(\omega) \in \partial f(\varphi(\hat{z})(\omega))$, and $d_i(\omega) \in \partial \varphi_i(\hat{z})(\omega)$, $i = 1, \dots, n$, for P-almost all $\omega \in \Omega$. Here $\varphi_i(z)(\omega)$ is the i-th component of $\varphi(z)(\omega)$. Consider the multifunction $M : \Omega \rightrightarrows \mathbb{R}^n \times (\mathcal{Z}^*)^n$ defined by

$$M(\omega) = \Big\{ (\kappa(\omega), s(\omega)) \in \partial f(\varphi(\hat{z})(\omega)) \times \big(\mathop{\times}_{i=1}^{n} \partial \varphi_i(\hat{z})(\omega) \big) : \sum_{i=1}^{n} \kappa_i(\omega) s_i(\omega) = g(\omega) \Big\}.$$

It is measurable with respect to the weak* topology on the target space, and it has nonempty, convex, and weakly* compact values. Since $\mathcal{Z}$ is separable, using [11] we obtain that M admits a weakly* measurable selection $(\theta(\omega), d(\omega))$. The selection $\theta(\omega)$ is measurable, because its values are in $\mathbb{R}^n$. By the monotonicity of $f(\cdot)$, the random variables θ_i are nonnegative. Since the subgradients of f are bounded, $\theta \in \mathcal{L}_\infty^n$. Substituting $\sum_{i=1}^n \theta_i(\omega) d_i(\omega)$ for $g(\omega)$ in (16) we obtain the equation:

$$\langle v, z - \hat{z} \rangle = \mathbb{E} \sum_{i=1}^{n} \theta_i(\omega) \langle d_i(\omega), z - \hat{z} \rangle \quad \text{for all} \quad z \in \mathcal{Z}. \tag{17}$$

If follows from the concavity assumptions about φ that the following inequalities hold for all z and P-almost all $\omega \in \Omega$:

$$\varphi_i(z)(\omega) \le \varphi_i(\hat{z})(\omega) + \langle d_i(\omega), z - \hat{z} \rangle, \quad i = 1, \dots, n.$$

Multiplying by $\theta_i(\omega)$, summing over i, and taking expected values we obtain

$$\langle \theta, \varphi(z) \rangle \le \langle \theta, \varphi(\hat{z}) \rangle + \mathbb{E} \sum_{i=1}^{n} \theta_i(\omega) \langle d_i(\omega), z - \hat{z} \rangle.$$

Using (17) we get

$$\langle \theta, \varphi(z) \rangle \le \langle \theta, \varphi(\hat{z}) \rangle + \langle v, z - \hat{z} \rangle \quad \text{for all } z.$$

By virtue of (15), the point $\hat{z}$ maximizes $\langle \theta, \varphi(z) \rangle$ in Z, as required.

To prove the converse implication, let us assume that $\hat{z}$ is a maximizer of $\langle \theta, \varphi(z) \rangle$ in Z, with $\theta(\omega) \in \partial f(\varphi(\hat{z})(\omega))$, $\theta \in \mathcal{L}_\infty^n$. By the concavity of f,

$$f(\varphi(z)(\omega)) \le f(\varphi(\hat{z})(\omega)) + \langle \theta(\omega), \varphi(z)(\omega) - \varphi(\hat{z})(\omega) \rangle$$

for all z. Taking the expected values we obtain

$$\mathbb{E}[f(\varphi(z))] \le \mathbb{E}[f(\varphi(\hat{z}))] + \langle \theta, \varphi(z) - \varphi(\hat{z}) \rangle.$$

By assumption, for all $z \in Z$ we have $\langle \theta, \varphi(z) \rangle \le \langle \theta, \varphi(\hat{z}) \rangle$. Thus $\hat{z}$ maximizes $\mathbb{E}[f(\varphi(z))]$ in Z. $\square$

We define the set $\mathcal{U}_1([a,b])$ of functions $u(\cdot)$ satisfying the following conditions:

$u(\cdot)$ is concave and nondecreasing;

$u(t) = 0$ for all $t \geq b$;

$u(t) = u(a) + c(t-a)$, with some $c > 0$, for all $t \leq a$.

It is evident that $\mathcal{U}_1([a,b])$ is a convex cone. Moreover, the subgradients of each function $u \in \mathcal{U}_1([a,b])$ are bounded for all $t \in \mathbb{R}$. We denote by $\mathcal{U}_1^m$ the product $\mathcal{U}_1([a_1,b_1]) \times \cdots \times \mathcal{U}_1([a_m,b_m])$.

Let us introduce the Lagrangian, $L : \mathcal{Z} \times \mathcal{L}_1^m \times \mathcal{U}_1^m \times \mathcal{L}_\infty^m \to \mathbb{R}$, associated with problem (12)-(14):

$$L(z,X,u,\theta) := \mathbb{E}\Big[H(z) + \sum_{i=1}^{m}\Big(u_i(X_i) - u_i(Y_i) + \theta_i(G_i(z) - X_i)\Big)\Big].$$

Definition 1. Problem (11)–(14) satisfies the *uniform dominance condition* if there exists a point $\tilde{z} \in Z$ such that

$$\inf_{\eta\in[a_i,b_i]}\Big\{F_2(Y_i;\eta) - F_2(G_i(\tilde{z});\eta)\Big\} > 0, \quad i = 1,\ldots,m.$$

Theorem 2. *Assume that the uniform dominance condition is satisfied. If* $(\hat{z},\hat{X})$ *is an optimal solution of* (11)–(14) *then there exist* $\hat{u} \in \mathcal{U}_1^m$ *and* $\hat{\theta} \in \mathcal{L}_\infty^m$ *such that*

$$L(\hat{z},\hat{X},\hat{u},\hat{\theta}) = \max_{(z,X)\in Z\times\mathcal{L}_1^m} L(z,X,\hat{u},\hat{\theta}), \tag{18}$$

$$\mathbb{E}[\hat{u}_i(\hat{X}_i)] = \mathbb{E}[\hat{u}_i(Y_i)], \quad i = 1,\ldots,m, \tag{19}$$

$$\hat{\theta}_i(\hat{X}_i - G_i(\hat{z})) = 0, \quad i = 1,\ldots,m, \quad \hat{\theta} \geq 0 \quad \textit{a.s.}. \tag{20}$$

Conversely, if for some function $\hat{u} \in \mathcal{U}_1^m$ *and for* $\hat{\theta} \in \mathcal{L}_\infty^m$, $\hat{\theta} \geq 0$ *a.s., an optimal solution* $(\hat{z},\hat{X})$ *of* (18) *satisfies* (12)–(13) *and* (19)–(20), *then* $(\hat{z},\hat{X})$ *is an optimal solution of* (11)–(14).

Proof. Let us define the operators $\Gamma_i : \mathcal{L}_1 \to \mathcal{C}([a_i,b_i])$ as

$$\Gamma_i(X_i)(\eta) := F_2(Y_i;\eta) - F_2(X_i;\eta), \quad \eta \in [a_i,b_i], \quad i = 1,\ldots,m.$$

Let K be the cone of nonnegative functions in $\mathcal{C}([a_i,b_i])$. Each operator Γ_i is concave with respect to the cone K, that is, for any X_i^1, X_i^2 in $\mathcal{L}_1$ and for all $\lambda \in [0,1]$,

$$\Gamma_i(\lambda X_i^1 + (1-\lambda)X_i^2) - [\lambda\Gamma_i(X_i^1) + (1-\lambda)\Gamma_i(X_i^2)] \in K.$$

Furthermore, we define the convex set

$$C = \{(z, X) \in \mathcal{Z} \times \mathcal{L}_1^m : z \in Z,\ X_i \le G_i(z) \text{ a.s.},\ i = 1, \dots, m\}.$$

We can rewrite (11)–(14) in the general form:

$$\begin{aligned} &\max \mathbb{E}[H(z)] \\ &\text{subject to } \varGamma_i(X_i) \in K, \quad i = 1, \dots, m, \\ &(z, X) \in C. \end{aligned} \tag{21}$$

By the Riesz representation theorem, the space dual to $\mathcal{C}([a_i, b_i])$ is the space $\mathbf{rca}([a_i, b_i])$ of regular countably additive measures on $[a_i, b_i]$ having finite variation (see, e.g., [7]). Let us define the space $\mathcal{M} = \mathbf{rca}([a_1, b_1]) \times \cdots \times \mathbf{rca}([a_m, b_m])$. We introduce the Lagrangian $\varLambda : \mathcal{Z} \times \mathcal{L}_1^m \times \mathcal{M} \to \mathbb{R}$,

$$\varLambda(z, X, \mu) := \mathbb{E}[H(z)] + \sum_{i=1}^m \int_{a_i}^{b_i} \varGamma_i(X_i)(\eta)\, d\mu_i(\eta). \tag{22}$$

Let us observe that the uniform dominance condition implies that for $\tilde{X}_i = G_i(\tilde{z})$ the following generalized Slater condition is satisfied:

$$\varGamma_i(\tilde{X}_i) \in \operatorname{int} K, \quad i = 1, \dots, m.$$

Moreover, $(\tilde{z}, \tilde{X}) \in C$. By [2, Prop. 2.106], this is equivalent to the regularity condition:

$$0 \in \operatorname{int} \bigcup_{(z,X) \in C} [\varGamma_i(X_i) - K], \quad i = 1, \dots, m.$$

Therefore we can use the necessary conditions of optimality for problem (21) (see, e.g., [2, Thm. 3.4]). We conclude that there exists a vector of nonnegative measures $\hat{\mu} \in \mathcal{M}$ such that

$$\varLambda(\hat{z}, \hat{X}, \hat{\mu}) = \max_{(z,X) \in C} \varLambda(z, X, \hat{\mu}) \tag{23}$$

and

$$\int_{a_i}^{b_i} [F_2(Y_i; \eta) - F_2(\hat{X}_i; \eta)]\, d\hat{\mu}_i(\eta) = 0, \quad i = 1, \dots, m. \tag{24}$$

We shall derive from these conditions the required relations (18)–(20).

Every measure $\mu \in \mathbf{rca}([a, b])$ can be extended to the whole real line by assigning measure 0 to Borel sets not intersecting $[a, b]$. A function $u : \mathbb{R} \to \mathbb{R}$ can be associated with every nonnegative measure μ as follows:

$$u(t) = \begin{cases} -\displaystyle\int_t^b \mu([\tau, b])\, d\tau & t < b, \\ 0 & t \ge b. \end{cases}$$

Since $\mu \geq 0$, the function $\mu([\cdot, b])$ is nonnegative and nonincreasing, which implies that $u(\cdot)$ is nondecreasing and concave. Consequently, $u \in \mathcal{U}_1([a,b])$. We have shown in [5, 6] that for any $X \in \mathcal{L}_1$ the function u defined above satisfies the equation

$$\int_a^b F_2(X;\eta)\,d\mu(\eta) = -\mathbb{E}[u(X)]. \tag{25}$$

Thus, the measures $\hat{\mu}_i$ correspond to functions $\hat{u}_i \in \mathcal{U}_1([a_i, b_i])$, $i = 1, \ldots, m$. Relations (25) for $\hat{X}_i$, $\hat{\mu}_i$ and $\hat{u}_i$ and equations (24) imply the complementarity condition (19).

In a similar manner, our Lagrangian (22) can be expressed as

$$\Lambda(z, X, \mu) = \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(u_i(X_i) - u_i(Y_i)\Big)\Big].$$

It follows that there exists $\hat{u} \in \mathcal{U}_1^m$ such that the optimal pair $(\hat{z}, \hat{X})$ is the solution of the problem:

$$\begin{aligned} &\max \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(\hat{u}_i(X_i) - \hat{u}_i(Y_i)\Big)\Big] \\ &\text{subject to } X_i \leq G_i(z) \quad \text{a.s.}, \quad i = 1, \ldots, m, \\ &z \in Z, \quad X_i \in \mathcal{L}_1, \quad i = 1, \ldots, m. \end{aligned} \tag{26}$$

By the monotonicity of $\hat{u}_i(\cdot)$, the point $\hat{z}$ is also the solution of

$$\max_{z \in Z} \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(\hat{u}_i(G_i(z)) - \hat{u}_i(Y_i)\Big)\Big].$$

We can now invoke Theorem 1 with

$$\varphi(z) = (H(z), G_1(z), \ldots, G_m(z)),$$

$$f(y_0, y_1, \ldots, y_m) = y_0 + \sum_{i=1}^{m} \hat{u}_i(y_i).$$

Since $df/dy_0 = 1$, we conclude there there exists $\hat{\theta} \in \mathcal{L}_\infty^m$, $\hat{\theta} \geq 0$ a.s., such that the point $\hat{z}$ is a solution of the problem

$$\max_{z \in Z} \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \hat{\theta}_i G_i(z)\Big]. \tag{27}$$

Moreover, $\hat{\theta}_i \in \partial \hat{u}_i(G_i(\hat{z}))$ a.s., $i = 1, \ldots, m$.

Let us consider problem (26) for a fixed $z = \hat{z}$. It splits into independent problems:

$$\max_{X_i \leq G_i(\hat{z}) \text{ a.s.}} \mathbb{E}[\hat{u}_i(X_i)], \quad i = 1, \ldots, m. \tag{28}$$

The points $\hat{X}_i$ are their solutions. The monotonicity of $\hat{u}_i(\cdot)$ implies that the points $G_i(\hat{z})$ are optimal as well. Therefore

$$\hat{u}_i(\hat{X}_i) = \hat{u}_i(G_i(\hat{z})) \quad \text{a.s.,} \quad i = 1, \dots, m.$$

For any other X_i, by the concavity of $\hat{u}_i$ and by the definition of $\hat{\theta}_i$,

$$\hat{u}_i(X_i) \le \hat{u}_i(G_i(\hat{z})) + \hat{\theta}_i(X_i - G_i(\hat{z})) \quad \text{a.s..}$$

Thus, for each $i = 1, \dots, m$,

$$u_i(X_i) - \hat{\theta}_i X_i \le \hat{u}_i(G_i(\hat{z})) - \hat{\theta}_i G_i(\hat{z}) \le \hat{u}_i(\hat{X}_i) - \hat{\theta}_i \hat{X}_i \quad \text{a.s..}$$

Therefore the point $\hat{X}_i$ maximizes the expression at the left hand side, for P-almost all $\omega \in \Omega$. At this point the last displayed inequalities are satisfied as equations and therefore the second group of complementarity conditions hold true:

$$\hat{\theta}_i[G_i(\hat{z}) - \hat{X}_i] = 0 \quad \text{a.s.,} \quad i = 1, \dots, m.$$

If follows that each point $\hat{X}_i$, $i = 1, \dots, m$, is a maximizer of the corresponding problem:

$$\max_{X_i \in \mathcal{L}_1} \mathbb{E}\left[u_i(X_i) - \hat{\theta}_i X_i - \hat{u}_i(Y_i)\right]. \tag{29}$$

Putting together (27) and (29) we conclude that the pair $(\hat{z}, \hat{X})$ maximizes

$$\mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(\hat{u}_i(X_i) + \hat{\theta}_i(G_i(z) - X_i) - \hat{u}_i(Y_i)\Big)\Big],$$

Therefore the pair $(\hat{z}, \hat{X})$ is the solution of (18).

Let us now prove the converse. If $u_i \in \mathcal{U}_1([a_i, b_i])$ then the left derivative of u_i,

$$(u_i)'_-(t) = \lim_{\tau \uparrow t} [u_i(t) - u_i(\tau)]/(t - \tau),$$

is well-defined, nonincreasing and continuous from the left. By the classical result (see, e.g., [7, Thm 3.1.3]), after an obvious adaptation, there exists a unique regular nonnegative measure μ_i satisfying

$$\mu_i([t, b]) = (u_i)'_-(t).$$

Thus the correspondence between nonnegative measures in $\mathbf{rca}([a, b])$ and functions in $\mathcal{U}_1([a, b])$ is a bijection and formula (25) is always valid. For every X_i satisfying (12) we obtain

$$\mathbb{E}[u_i(X_i)] - \mathbb{E}[u_i(Y_i)] = -\int_{a_i}^{b_i} F_2(X_i; \eta)\, d\mu_i(\eta) + \int_{a_i}^{b_i} F_2(Y_i; \eta)\, d\mu_i(\eta) \ge 0.$$

Thus, at every (z, X), which is feasible for (11)–(14), and for every $u \in \mathcal{U}_1^m$ and $\theta \in \mathcal{L}_\infty^m$, $\theta \ge 0$ a.s., we have

$$L(z, X, u, \theta) = \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(u_i(X_i) - u_i(Y_i) + \theta_i(G_i(z) - X_i)\Big)\Big] \ge \mathbb{E}[H(z)].$$

If the maximizer $(\hat{z}, \hat{X})$ of the Lagrangian is feasible, and complementarity conditions (19)–(20) are satisfied, we obtain

$$\mathbb{E}[H(\hat{z})] = L(\hat{z}, \hat{X}, u, \theta) \geq L(z, X, u, \theta) \geq \mathbb{E}[H(z)],$$

for any feasible (z, X). Consequently, the point $(\hat{z}, \hat{X})$ is optimal for the original problem (11)–(14). □

4. Duality and Decomposition

Let us define the dual functional $D : \mathcal{U}_1^m \times \mathcal{L}_\infty^m \to \overline{\mathbb{R}}$ associated with problem (11)-(14) as follows:

$$\begin{aligned} D(u,\theta) &:= \sup_{z \in Z, X \in \mathcal{L}_1^m} L(z, X, u, \theta) \\ &= \sup_{z \in Z, X \in \mathcal{L}_1^m} \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(u_i(X_i) - u_i(Y_i) + \theta_i(G_i(z) - X_i)\Big)\Big]. \end{aligned} \tag{30}$$

We also define the dual problem:

$$\min \{D(u,\theta) : u \in \mathcal{U}_1^m,\ \theta \in \mathcal{L}_\infty^m,\ \theta \geq 0 \text{ a.s.}\}. \tag{31}$$

As a direct consequence of Theorem 2 we obtain the duality theorem.

Theorem 3. *Assume that the uniform dominance condition is satisfied. If problem* (11)–(14) *has an optimal solution, then the dual problem* (31) *has an optimal solution and the optimal values of both problems coincide. Furthermore, for every solution $(\hat{u}, \hat{\theta})$ of the dual problem, any optimal solution $(\hat{z}, \hat{X})$ of* (18) *satisfying* (12)–(13) *and* (19)–(20), *is an optimal solution of the primal problem* (11)–(14).

Proof. At every (z, X), which is feasible for problem (11)–(14), and for every (u, θ) feasible for problem (31), we have

$$L(z, X, u, \theta) = \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \Big(u_i(X_i) - u_i(Y_i) + \theta_i(G_i(z) - X_i)\Big)\Big] \geq \mathbb{E}[H(z)].$$

Therefore the weak duality relation holds:

$$D(u, \theta) \geq \mathbb{E}[H(z)].$$

Let $(\hat{z}, \hat{X})$ be an optimal solution of the primal problem. It follows from Theorem 2 that there exist $(\hat{u}, \hat{\theta})$, which are feasible for (31), such that

$$D(\hat{u}, \hat{\theta}) = \mathbb{E}[H(\hat{z})].$$

This proves the equality of the optimal values of both problems.

Let $(\hat{u}, \hat{\theta})$ be a solution of the dual problem and $(\hat{z}, \hat{X})$ be the corresponding maximizer of the Lagrangian. If the complementarity conditions (19)–(20) are satisfied, we obtain

$$\mathbb{E}[H(\hat{z})] = L(\hat{z}, \hat{X}, u, \theta) \geq L(z, X, u, \theta) \geq \mathbb{E}[H(z)],$$

for any feasible (z, X). Consequently, if the point $(\hat{z}, \hat{X})$ is feasible, it is optimal for the primal problem (11)–(14). □

It follows from (30) that the dual functional can be decomposed into the sum

$$D(u, \theta) = D_0(\theta) + \sum_{i=1}^{m} D_i(u_i, \theta_i), \tag{32}$$

where the functions $D_0 : \mathcal{L}_\infty^m \to \overline{\mathbb{R}}$ and $D_i : \mathcal{U}_1([a_i, b_i]) \times \mathcal{L}_\infty \to \overline{\mathbb{R}}$ are defined as

$$D_0(\theta) := \sup_{z \in Z} \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \theta_i G_i(z)\Big], \tag{33}$$

and

$$D_i(w, \zeta) := \sup_{X \in \mathcal{L}_1} \mathbb{E}\Big[w(X) - w(Y_i) - \zeta X\Big], \quad i = 1, \ldots, m. \tag{34}$$

The function D_0 has the structure of the dual function associated with the standard Lagrangian,

$$L_0(z, \theta) = \mathbb{E}\Big[H(z) + \sum_{i=1}^{m} \theta_i G_i(z)\Big],$$

for a stochastic optimization problem with almost sure constraints. Under the assumptions of Theorem 3, $D_0(\cdot)$ is a proper convex function. Moreover, if for a given θ a solution $\hat{z}(\theta)$ of the problem at the right hand side of (33) exists, then the random vector

$$g = (G_1(\hat{z}), \ldots, G_m(\hat{z}))$$

is a subgradient of D_0 at θ. By the definition of the operators G_i, we have $g \in \mathcal{L}_1^m$.

Let us concentrate on the properties of the functions D_i, $i = 1, \ldots, m$. For a concave function $v : \mathbb{R} \to \overline{\mathbb{R}}$ we define its Fenchel conjugate in a symmetrical fashion to the conjugate of a convex function:

$$v^*(\xi) = \inf_t [\xi t - v(t)].$$

Alternatively, we could work with the usual definition of a conjugate of a convex function $-v$. The results would be the same but with less convenient notation.

As before, for a real random variable ζ we shall understand $v^*(\zeta)$ as a random variable in $\overline{\mathbb{R}}$ with realizations $v^*(\zeta(\omega))$, $\omega \in \Omega$.

Theorem 4. *For every $v \in \mathcal{U}_1([a, b])$ and every $\zeta \in \mathcal{L}_\infty$ the following formula holds true:*

$$D_i(v, \zeta) = -\mathbb{E}\,[v^*(\zeta) + v(Y_i)].$$

Proof. For a function $v \in \mathcal{U}_1([a,b])$ and $\zeta \in \mathcal{L}_\infty$ let us consider the problem

$$\sup_{X \in \mathcal{L}_1} \mathbb{E}\,[v(X) - \zeta X].$$

Suppose that $P[\zeta < 0] = \varepsilon > 0$. Choosing $X_M = M 1\!\!1_{\{\zeta<0\}}$ and noting that $v(t) \to 0$ as $t \to \infty$ by definition, we see that

$$\mathbb{E}[v(X_M) - \zeta X_M] = (1-\varepsilon)v(0) + \varepsilon v(M) - M\varepsilon\,\mathbb{E}[\zeta | \zeta < 0] \xrightarrow{M\to\infty} \infty.$$

Suppose now that $P[\zeta > v'_-(a)] > 0$. Choosing $X_M = -M 1\!\!1_{\{\zeta > v'_-(a)\}}$ for $M \to \infty$ we obtain an infinite limit again.

It remains to consider the case when $0 \le \zeta \le v'_-(a)$ a.s.. In this case the function $v(t) - \zeta t$ has a maximizer in $[a,b]$, a.s.. By [18, Thm. 14.60] we have

$$\begin{aligned}\sup_{X \in \mathcal{L}_1} \mathbb{E}\,[v(X) - \zeta X] &\le \sup_{X \in \mathcal{L}_0} \mathbb{E}\,[v(X) - \zeta X] = \mathbb{E}\,\sup_{t\in\mathbb{R}}[v(t) - \zeta(\omega)t] \\ &= \mathbb{E}\,\sup_{t\in[a,b]}[v(t) - \zeta(\omega)t] = -\mathbb{E}\,[v^*(\zeta)].\end{aligned}$$

Since the maximizer X is bounded a.s., it is an element of $\mathcal{L}_1$. Thus all relations above are equalities and our assertion is true in this case. Moreover, we have shown that

$$\operatorname{dom} D_i = \{(v,\zeta) : 0 \le \zeta \le v'_-(a) \text{ a.s.}\}$$

Outside of the domain the asserted formula remains valid as well, because both sides are equal to $+\infty$. □

The proof of the theorem shows that the dual functionals $D_i(u_i, \theta_i)$ are finite whenever $0 \le \theta_i \le (u_i)'_-(a_i)$ a.s.. We shall show that they are subdifferentiable and we shall find a representation of some of their subgradients.

The key element of the analysis is the functional

$$f(v,\zeta) = -\mathbb{E}\,v^*(\zeta). \tag{35}$$

The specificity here is that v is considered as an argument and therefore we need an appropriate functional space for this argument of f. It is convenient to consider the space $\mathrm{Lip}(\mathbb{R})$ of Lipschitz continuous functions on $\mathbb{R}$, equipped with the norm

$$\|v\|_{\mathrm{Lip}} = |v(0)| + \sup_{t\neq s} \frac{|v(t) - v(s)|}{|t-s|}.$$

We shall treat the functional f as defined on $\mathrm{Lip}(\mathbb{R}) \times \mathcal{L}_1$. It is obvious that $\mathcal{U}_1([a,b])$ is a subset of $\mathrm{Lip}(\mathbb{R})$.

Recall that if $v \in \mathcal{U}_1([a,b])$ and ζ satisfies $0 \le \zeta \le v'_-(a)$ a.s. then there exists a measurable selection X such that

$$X(\omega) \in \operatorname*{argmax}_t\,[v(t) - \zeta(\omega)t] \quad \text{for } P\text{-almost all } \omega$$

(see e.g. Theorem 14.37, [18]). Moreover, $X \in [a,b]$ a.s.. We use the symbol P_X to denote the probability measure on $\mathbb{R}$ induced by the random variable X.

Theorem 5. *For every $\bar{v} \in \mathcal{U}_1([a,b])$ and every $\bar{\zeta} \in \mathcal{L}_1$ such that $0 \le \bar{\zeta} \le \bar{v}'_-(a)$ a.s., the functional $f(v,\zeta) = -\mathbb{E}\, v^*(\zeta)$ is subdifferentiable at $(\bar{v},\bar{\zeta})$. Moreover, for every measurable selection $X(\omega) \in \operatorname{argmax}_t[\bar{v}(t) - \bar{\zeta}(\omega)t]$, the pair $(P_X, -X)$ is a subgradient of f at $(\bar{v},\bar{\zeta})$, that is, the inequality*

$$f(v,\zeta) \ge f((\bar{v},\bar{\zeta}) + \int (v(t) - \bar{v}(t))\, dP_X(t) - \mathbb{E}\,[X(\zeta - \bar{\zeta})].$$

holds true for all $(v,\zeta) \in \mathrm{Lip}(\mathbb{R}) \times \mathcal{L}_1$.

Proof. Given $\bar{v}$, and $\bar{\zeta}$ and a selection X satisfying the assumptions, we have for every $(v,\zeta) \in \mathrm{Lip}(\mathbb{R}) \times \mathcal{L}_1$ and for every $\omega \in \Omega$:

$$\begin{aligned}
-v^*(\zeta(\omega)) &= \sup_t [v(t) - \zeta(\omega)t] \ge v(X(\omega)) - \zeta(\omega)X(\omega) \\
&= \bar{v}(X(\omega)) - \bar{\zeta}(\omega)X(\omega) + v(X(\omega)) - \bar{v}(X(\omega)) - X(\omega)(\zeta(\omega) - \bar{\zeta}(\omega)) \\
&= -\bar{v}^*(\bar{\zeta}(\omega)) + v(X(\omega)) - \bar{v}(X(\omega)) - X(\omega)(\zeta(\omega) - \bar{\zeta}(\omega)).
\end{aligned}$$

Taking the expected value of both sides we obtain:

$$\begin{aligned}
f(v,\zeta) &\ge f(\bar{v},\bar{\zeta}) + \mathbb{E}[v(X) - \bar{v}(X)] - \mathbb{E}\,[X(\zeta - \bar{\zeta})] \\
&= f(\bar{v},\bar{\zeta}) + \int [v(t) - \bar{v}(t)]\, dP_X(t) - \mathbb{E}\,[X(\zeta - \bar{\zeta})],
\end{aligned}$$

which is the required inequality. The selection X is included in $[a,b]$ a.s.. Thus $X \in \mathcal{L}_\infty$ and X is a continuous linear functional on $\mathcal{L}_1$.

It remains to prove that the measure P_X is a continuous linear functional on the space $\mathrm{Lip}(\mathbb{R})$. For any $v \in \mathrm{Lip}(\mathbb{R})$, denoting by c_v its Lipschitz constant, we obtain

$$\begin{aligned}
\int |v(t)|\, dP_X(t) &\le \int (|v(0)| + c_v|t|)\, dP_X(t) \\
&= |v(0)| + c_v \mathbb{E}[|X|] \le \|v\|(1 + \mathbb{E}[|X|]).
\end{aligned}$$

Since $\mathbb{E}[|X|]$ is finite, the functional P_X is continuous. This proves that the function f is subdifferentiable and $(P_X, -X)$ is a subgradient. □

Our analysis shows that the calculation of the dual functional $D(u,\theta)$ and of its subgradient splits into separate maximization problem with respect to z and with respect to $X_i(\omega)$, $\omega \in \Omega$, $i = 1,\dots,m$. This is crucial for the development of decomposition methods for solving dominance-constrained stochastic optimization problems.

5. Discrete Distributions

Let us now consider the case when the underlying probability space is finite, $\Omega = \{\omega_1,\dots,\omega_n\}$, with probabilities $p_j = P(\{\omega_j\})$, $j = 1,\dots,n$. Let $J = \{1,\dots,n\}$, $I = \{1,\dots,m\}$. For the split-variable problem (11)–(14) we introduce the following notation for $j \in J$ and $i \in I$:

$$h_j(z) = H(z)(\omega_j),\ g_{ij}(z) = G_i(z)(\omega_j),\ y_{ij} = Y_i(\omega_j),\ x_{ij} = X_i(\omega_j).$$

For each i the function (4) has the form

$$F_2(X_i;\eta) = \mathbb{E}[(\eta - X_i)_+] = \sum_{j=1}^{n} p_j(\eta - x_{ij})_+,$$

and the dominance constraints (12) can be expressed as

$$\sum_{j=1}^{n} p_j(\eta - x_{ij})_+ \le \sum_{j=1}^{n} p_j(\eta - y_{ij})_+, \quad \text{for all } \eta \in [a_i, b_i], \quad i \in I. \tag{36}$$

Lemma 2. *Assume that $a_i \le y_{ij} \le b_i$ for all $i \in I$ and $j \in J$. Then inequalities* (36) *are equivalent to*

$$\sum_{j=1}^{n} p_j(y_{ik} - x_{ij})_+ \le \sum_{j=1}^{n} p_j(y_{ik} - y_{ij})_+, \quad i \in I, \quad k \in J. \tag{37}$$

Proof. It is sufficient to consider a fixed i. Let $y_{i,[j]}$, $j \in J$, be ordered realizations y_{ij}, that is, $y_{i,[1]} \le y_{i,[2]} \le \dots \le y_{i,[n]}$. It is sufficient to prove that (37) imply that

$$F_2(X_i;\eta) \le F_2(Y_i;\eta) \quad \text{for all } \eta \in [a_i, b_i].$$

The function $F_2(Y_i;\cdot)$ is piecewise linear and has break points at $y_{i,[j]}$, $j \in J$. Let us consider three cases, depending on the value of η.

Case 1: If $\eta \le y_{i,[1]}$ we have

$$0 \le F_2(X_i;\eta) \le F_2(X_i;y_{i,[1]}) \le F_2(Y_i;y_{i,[1]}) = 0.$$

Therefore the required relation holds as an equality.

Case 2: Let $\eta \in [y_{i,[k]}, y_{i,[k+1]}]$ for some k. Since, for any X, the function $F_2(X;\cdot)$ is convex, inequalities (37) for k and $k+1$ imply that for all $\eta \in [y_{i,[k]}, y_{i,[k+1]}]$ one has

$$\begin{aligned} F_2(X;\eta) &\le \lambda F_2(X;y_{i,[k]}) + (1-\lambda)F_2(X;y_{i,[k+1]}) \\ &\le \lambda F_2(Y_i;y_{i,[k]}) + (1-\lambda)F_2(Y_i;y_{i,[k+1]}) = F_2(Y_i;\eta), \end{aligned}$$

where $\lambda = (y_{i,[k+1]} - \eta)/(y_{i,[k+1]} - y_{i,[k]})$.

Case 3: For $\eta > y_{i,[n]}$ we have

$$\begin{aligned} F_2(Y_i;\eta) &= F_2(Y_i;y_{i,[n]}) + \eta - y_{i,[n]} \\ &\geq F_2(X;y_{i,[n]}) + \int_{y_{i,[n]}}^{\eta} F(X;\alpha)\,d\alpha = F_2(X;\eta), \end{aligned}$$

as required. □

In fact, we have proved that inequalities (37) are equivalent to (36) for arbitrary $[a_i, b_i]$ covering the realizations of Y_i. Thus, they are equivalent to the dominance relation enforced on the entire real line.

It follows that in the case of finite distributions, problem (11)–(14) with sufficiently large intervals $[a_i, b_i]$ is equivalent to the following nonlinear programming problem

$$\max \sum_{j=1}^{n} p_j h_j(z) \tag{38}$$

$$\text{subject to } \sum_{j=1}^{n} p_j (y_{ik} - x_{ij})_+ \leq \sum_{j=1}^{n} p_j (y_{ik} - y_{ij})_+, \quad i \in I, \quad k \in J, \tag{39}$$

$$x_{ik} \leq g_{ik}(z), \quad i \in I, \quad k \in J, \tag{40}$$

$$z \in Z. \tag{41}$$

In addition, suppose for simplicity that $Z \subseteq \mathbb{R}^N$. Let us observe that for the smallest realization $y_{i,k^*(i)} = y_{i,[1]}$ of Y_i the corresponding dominance constraint becomes

$$\sum_{j=1}^{n} p_j (y_{i,k^*(i)} - x_{ij})_+ \leq \sum_{j=1}^{n} p_j (y_{i,k^*(i)} - y_{ij})_+ = 0.$$

The uniform dominance condition (Definition 1) cannot be satisfied, unless $a_i > y_{i,k^*(i)}$. Fortunately, the left hand sides of the dominance constraints (39) are convex polyhedral functions of x. The existence of Lagrange multipliers is guaranteed under the standard Slater condition: there exist $\tilde{z} \in \operatorname{relint} Z$ and $\tilde{X}_i$, $i \in I$, such that

$$\tilde{x}_{ik} < g_{ik}(\tilde{z}), \quad i \in I, \quad k \in J,$$

and the dominance constraints (39) are satisfied (see [16, Thm. 28.2]).

The set $\mathcal{V}_i \subset \mathcal{U}_1([a_i, b_i])$ of utility functions corresponding to the ith group of dominance constraints in (39) contains all concave nondecreasing functions $u(\cdot)$ which are piecewise–linear with break points at y_{ik}, $k \in J$, and which satisfy $u(y_{i,[n]}) = 0$.

The Lagrange multipliers θ_i corresponding to the splitting constraints (40) are nonnegative vectors in $\mathbb{R}^n$.

The Lagrangian takes on the form

$$
\begin{aligned}
L(z,X,u,\theta) = &\sum_{j=1}^{n} p_j \Big[h_j(z) + \sum_{i=1}^{m} \theta_{ij} g_{ij}(z) \Big] \\
&+ \sum_{i=1}^{m} \sum_{j=1}^{n} p_j \Big[u_i(x_{ij}) - u_i(y_{ij}) - \theta_{ij} x_{ij} \Big].
\end{aligned} \tag{42}
$$

The optimality conditions can be formulated as follows.

Theorem 6. *Assume that problem* (38)–(41) *satisfies the Slater constraint qualification condition. If* $(\hat{z}, \hat{X})$ *is an optimal solution of* (38)–(41)*, then there exist* $\hat{u}_i \in \mathcal{V}_i$ *and nonnegative vectors* $\hat{\theta}_i \in \mathbb{R}^n$*,* $i = 1, \ldots, m$*, such that*

$$
L(\hat{z}, \hat{X}, \hat{u}, \hat{\theta}) = \max_{(z,X) \in Z \times \mathbb{R}^{mn}} L(z, X, \hat{u}, \hat{\theta}), \tag{43}
$$

$$
\sum_{j=1}^{n} p_j [\hat{u}_i(\hat{x}_{ij}) - \hat{u}_i(y_{ij})] = 0, \quad i \in I, \tag{44}
$$

$$
\hat{\theta}_{ij}(\hat{x}_{ij} - g_{ij}(\hat{z})) = 0, \quad i \in I, \quad j \in J. \tag{45}
$$

Conversely, if for some functions $\hat{u}_i \in \mathcal{V}_i$ *and nonnegative vectors* $\hat{\theta}_i \in \mathbb{R}^n$*,* $i \in I$*, an optimal solution* $(\hat{z}, \hat{X})$ *of* (43) *satisfies* (39)–(40) *and* (44)–(45)*, then* $(\hat{z}, \hat{X})$ *is an optimal solution of* (38)–(41)*.*

Proof. Let us introduce Lagrange multipliers μ_{ik}, $i \in I$, $k \in J$, associated with constraints (39). The standard Lagrangian takes on the form:

$$
\begin{aligned}
\Lambda(z,X,\mu,\theta) = &\sum_{j=1}^{n} p_j \Big[h_j(z) + \sum_{i=1}^{m} \theta_{ij} (g_{ij}(z) - x_{ij}) \Big] \\
&+ \sum_{i=1}^{m} \sum_{k=1}^{n} \mu_{ik} \Big[\sum_{j=1}^{n} p_j (y_{ik} - y_{ij})_+ - \sum_{j=1}^{n} p_j (y_{ik} - x_{ij})_+ \Big].
\end{aligned}
$$

Rearranging the last sum we notice that

$$
\sum_{k=1}^{n} \mu_{ik} \sum_{j=1}^{n} p_j (y_{ik} - x_{ij})_+ = \sum_{j=1}^{n} p_j \sum_{k=1}^{n} \mu_{ik} (y_{ik} - x_{ij})_+ = -\sum_{j=1}^{n} p_j u_i(x_{ij}),
$$

where

$$
u_i(t) = -\sum_{k=1}^{n} \mu_{ik} (y_{ik} - t)_+. \tag{46}
$$

Substituting this into the Lagrangian $\Lambda(z,X,\mu,\theta)$ yields (42). Applying (46) to the standard complementarity conditions for the problem (38)–(41) we obtain the conditions (44)–(45). Consequently, our conditions follow from standard necessary optimality conditions for problem (38)–(41) (see, e.g., [16, Cor. 28.3.1]).

In order to show that the standard sufficient optimality conditions follow from conditions (43)–(45), we shall establish a correspondence between Lagrange multipliers μ_{ik}, $k \in J$, and concave nondecreasing utility functions in $\mathcal{V}_i$. We have shown that the Lagrange multipliers generate a utility function. Conversely, let us consider a utility function $v \in \mathcal{V}_i$, and let $t_1 < t_2 < \cdots < t_K$ be its break points. We can define

$$\nu_k = v'_-(t_k) - v'_+(t_k), \quad k = 1, \ldots, K.$$

For every $k = 1, \ldots, K$ we define $J(k) = \{j \in J : y_{ij} = t_k\}$. By the definition of $\mathcal{V}_i$, the sets $J(k)$ are nonempty and constitute a partition of the set J. Therefore, for every $j \in J$, there is unique $k = 1, \ldots, K$ with $j \in J(k)$, and we can define:

$$\mu_{ij} = \nu_k / |J(k)|,$$

where $|J(k)|$ denotes the cardinality of $J(k)$. It is a routine check to see that the μ_{ij} satisfy the equation (46). Thus, substituting $u_i(t)$ in (42) yields the standard Lagrangian $\Lambda(z, X, \mu, \theta)$. Similarly, we can transform the complementarity conditions.

Consequently, our conditions are equivalent to the standard necessary and sufficient optimality conditions for problem (38)–(41). ☐

The dual functional $D : \mathop{\times}\limits_{i=1}^{m} \mathcal{V}_i \times \mathbb{R}^{mn} \to \overline{\mathbb{R}}$ associated with problem (38)-(41) has the form:

$$\begin{aligned} D(u,\theta) &= \sup_{z \in Z, X \in \mathbb{R}^{mn}} L(z, X, u, \theta) \qquad (47) \\ &= \sup_{z \in Z, X \in \mathbb{R}^{mn}} \sum_{j=1}^{n} p_j \Big[h_j(z) + \sum_{i=1}^{m} \Big(u_i(x_{ij}) - u_i(y_{ij}) + \theta_{ij}(g_{ij}(z) - x_{ij}) \Big) \Big]. \end{aligned}$$

The dual problem reads:

$$\min \{ D(u,\theta) : u \in \mathop{\times}\limits_{i=1}^{m} \mathcal{V}_i,\ \theta \in \mathbb{R}^{mn},\ \theta \geq 0 \}. \qquad (48)$$

As a direct consequence of Theorem 6 we obtain the duality theorem.

Theorem 7. *Assume that the Slater condition is satisfied. If problem* (38)–(41) *has an optimal solution then the dual problem* (48) *has an optimal solution and the optimal values of both problems coincide. Furthermore, for every solution* $(\hat{u}, \hat{\theta})$ *of the dual problem, any optimal solution* $(\hat{z}, \hat{X})$ *of* (43) *satisfying* (39)–(40) *and* (44)–(45)*, is an optimal solution of the primal problem* (38)–(41).

The dual functional (47) can be decomposed into the sum

$$D(u,\theta) = D_0(\theta) + \sum_{i=1}^{m} D_i(u_i, \theta_i), \qquad (49)$$

where the functions $D_0 : \mathbb{R}^{mn} \to \overline{\mathbb{R}}$ and $D_i : \mathcal{V}_i \times \mathbb{R}^n \to \overline{\mathbb{R}}$ are defined as follows:

$$D_0(\theta) = \sup_{z \in Z} \sum_{j=1}^{n} p_j \Big[h_j(z) + \sum_{i=1}^{m} \theta_{ij} g_{ij}(z) \Big], \qquad (50)$$

and

$$D_i(w,\zeta) = \sup_{x\in\mathbb{R}^n} \sum_{j=1}^{n} p_j \Big[w(x_j) - w(y_{ij}) - \zeta_j x_j \Big], \quad i = 1,\dots,m. \tag{51}$$

The function D_0 has the structure of the dual function associated with the standard Lagrangian,

$$L_0(z,\theta) = \sum_{j=1}^{n} p_j \Big[h_j(z) + \sum_{i=1}^{m} \theta_{ij} g_{ij}(z) \Big].$$

Under the assumptions of Theorem 7, $D_0(\cdot)$ is a proper convex function. Moreover, if for a given θ a solution $\hat{z}(\theta)$ of the problem at the right hand side of (50) exists, then the matrix $\Gamma(\hat{z})$ with entries

$$\gamma_{ij} = p_j g_{ij}(\hat{z}), \quad i \in I, \quad j \in J,$$

is a subgradient of D_0 at θ. If, additionally, the set Z is compact, then we have (see, e.g.,[4, Thm. 2.8.2])

$$\partial D_0(\theta) = \overline{\text{co}}\, \{\Gamma(\hat{z}) : L_0(\hat{z},\theta) = D_0(\theta)\}.$$

Here $\overline{\text{co}}\, A$ denotes the closed convex hull of the set A.

Now we shall describe the subdifferential of the functions D_i. For this purpose let us introduce the spaces PL_i of piecewise linear functions from $\mathbb{R}$ to $\mathbb{R}$ having break points at y_{ik}, $k \in J$. They are, clearly, finite dimensional.

We can represent the functions D_i as follows:

$$D_i(w,\zeta) = \sum_{j=1}^{n} p_j \sup_{x_j\in\mathbb{R}} \Big[w(x_j) - w(y_{ij}) - \zeta_j x_j \Big] = - \sum_{j=1}^{n} p_j d_{ij}(w,\zeta_j),$$

where

$$d_{ij}(w,\zeta_j) = w^*(\zeta_j) + w(y_{ij}).$$

Let us observe that

$$\text{dom}\, d_{ij} = \{(w,\zeta) : 0 \le \zeta_j \le w'_-(y_{i,[1]}),\ j \in J\}.$$

By the definition of the set $\mathcal{V}_i$, the following equation holds

$$\sup_{x_j\in\mathbb{R}} \Big[w(x_j) - w(y_{ij}) - \zeta_j x_j \Big] = \max_{k\in J} \Big[w(y_{ik}) - w(y_{ij}) - \zeta_j y_{ik} \Big] \tag{52}$$

whenever $d_{ij}(w,\zeta_j)$ is finite. The subdifferential of d_{ij} can be characterized as follows.

Lemma 3. *The function d_{ij} is a convex polyhedral function on $\text{Lip}(\mathbb{R}) \times \mathbb{R}$. Assume that $(\bar{w},\bar{\zeta}) \in \text{dom}\, d_{ij}$ and let $J^* = \{j : d_{ij}(\bar{w},\bar{\zeta}_j) = \bar{w}(y_{ik}) - \bar{w}(y_{ij}) - \bar{\zeta}_j y_{ik}\}$. The function d_{ij} is subdifferentiable at $(\bar{w},\bar{\zeta})$ and*

$$\text{co}\Big\{ \bigcup_{k\in J^*} (\delta_{y_{ik}} - \delta_{y_{ij}}, -y_{ik}) \Big\} \subseteq \partial d_{ij}(\bar{w},\bar{\zeta}),$$

where δ_t is the Dirac measure at t. Moreover, if $0 < \bar{\zeta}_j < \bar{w}'_-(y_{i,[1]})$ then the above formula is satisfied as equality.

Proof. Since for any fixed t the left derivative $w'_-(t)$ is a bounded linear functional on $\mathrm{Lip}(\mathbb{R})$, the domain of d_{ij} is determined by finitely many linear inequalities. As shown in (52), the function $d_{ij}(w,\zeta_j)$ is the maximum of finitely many linear functions of (w,ζ_j) in its domain. Therefore $d_{ij}(w,\zeta_j)$ is a convex polyhedral function. Its subdifferential contains the convex hull of the gradients of the functions

$$d_{ij}^k(w,\zeta_j) = w(y_{ik}) - w(y_{ij}) - \zeta_j y_{ik}, \quad k \in J^*.$$

Since $\nabla d_{ij}^k(\bar{w},\bar{\zeta}_j) = (\delta_{y_{ik}} - \delta_{y_{ij}}, -y_{ik})$ we obtain the required result.

At the boundary points of the domain of d_{ij}, when $\bar{\zeta}_j = 0$ or $\bar{\zeta}_j = \bar{w}'_-(y_{i,[1]})$, the subdifferential contains also all elements of the normal cone to the domain. □

6. Numerical Example

It follows from our analysis that the dual functional can be expressed as a weighted sum of $mn+1$ convex nonsmooth functions:

$$D(u,\theta) = D_0(\theta) + \sum_{i=1}^{m}\sum_{j=1}^{n} p_j d_{ij}(u_i,\theta_{ij}), \tag{53}$$

whose domains are known, and whose subgradients can be readily calculated. Furthermore, the functions d_{ij} are polyhedral. All these facts can be used for efficient numerical solution of the problem. The regularized decomposition method, which was developed in [19] for a similar purpose, can be adapted to this problem as well. It is a specialized bundle method [9, 10] which takes advantage of representation (53) to increase the efficiency and the numerical stability of the algorithm.

In order to apply any numerical method we need to decide about a finite dimensional representation of the utility functions u_i, $i = 1,\dots,m$. We shall represent each function u_i by its slopes. Let us denote the values of u_i at its break points by

$$u_{ij} = u_i(y_{ij}), \quad j = 1,\dots,n.$$

According to Lemma 3, a subgradient of d_{ij} within the domain is given by the formula

$$(\delta_{k^*} - \delta_j, -y_{ik^*})$$

where k^* is the maximizer of

$$u_{ik} - u_{ij} - \theta_{ij} y_{ik}, \quad k = 1,\dots,n,$$

and δ_j denotes the jth unit vector in $\mathbb{R}^n$.

Let us now consider the ordered realizations $y_{i,[1]} \le y_{i,[2]} \le \cdots \le y_{i,[n]}$. We introduce the variables

$$s_{ik} = (u_i)'_-(y_{i,[k]}), \quad k = 1,\dots,n.$$

The vector $s_i \in \mathbb{R}^n$ is nonnegative and $s_{ik} \geq s_{i,k+1}$, $k = 1, \dots, n-1$. With this re-ordering of coordinates we can calculate the ordered values of u_i as follows

$$u_{i,[k]} = u_i(y_{i,[k]}) = -\sum_{\ell > k} s_{i\ell}(y_{i,[\ell]} - y_{i,[\ell-1]}).$$

A subgradient of the function d_{ij} with respect to (s_i, θ_{ij}) can be calculated accordingly:

$$\Big(-\sum_{\ell > k^*} \delta_\ell(y_{i,[\ell]} - y_{i,[\ell-1]}) + \sum_{y_{i,[\ell]} > y_{ij}} \delta_\ell(y_{i,[\ell]} - y_{i,[\ell-1]}), -y_{i,[k^*]} \Big),$$

where k^* is the index at which the maximum of

$$u_{i,[k]} - u_{ij} - \theta_{ij} y_{i,[k]}, \quad k = 1, \dots, n,$$

is attained. The domain of the dual problem is defined by:

$$0 \leq \theta_{ij} \leq s_{i1}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

and it can be directly taken into account within the method.

We tested our dual approach on the following financial optimization example. We have N assets with random returns $Q_1, \dots, Q_N$. If $z_1, \dots, z_N$ are the fractions of the initial capital invested in assets $1, \dots, N$, then the portfolio return has the form:

$$G(z) = Q_1 z_1 + \cdots + Q_N z_N.$$

The set of feasible allocations is defined as the simplex

$$Z = \{z \in \mathbb{R}^N : z_1 + \cdots + z_N = 1,\ z_k \geq 0,\ k = 1, \dots, N\}.$$

Let $\tilde{z} \in Z$ represent a reference portfolio and let $Y = G(\tilde{z})$. We consider the following problem

$$\max \mathbb{E}[G(z)]$$

$$\text{subject to } G(z) \succeq_{(2)} Y,$$

$$z \in Z.$$

In our experiment we have assumed that the returns of the assets have a discrete distribution with n realizations. Let us observe that the set Z is a convex polyhedron. Furthermore, the function G is linear, and the dominance constraint becomes a convex constraint involving a polyhedral function, as discussed in section 5 for problem (38)–(41). Therefore we do not need to verify the Slater condition here.

Our calculations were carried out for a basket of 719 real-world assets, and for 616 possible realizations of their joint returns [20]. Historical data on weekly returns in the 12 years from Spring 1990 to Spring 2002 were used as equally likely realizations. More specifically, if q_{jk} denotes the historical return of asset k in week j, the vector

$(q_{j1}, \dots, q_{jN})$, where $N = 719$, is considered as the jth realization of the vector of returns. Therefore

$$g_j(z) = \sum_{k=1}^{N} q_{jk} z_k$$

is the jth realization of the portfolio return, attained with probability $p_j = 1/n$, where $n = 616$.

Function (50) has the form

$$\begin{aligned} D_0(\theta) &= \sup_{z \in Z} \sum_{j=1}^{n} p_j (1+\theta_j) g_j(z) = \sup_{z \in Z} \sum_{j=1}^{n} \sum_{k=1}^{N} p_j (1+\theta_j) q_{jk} z_k \\ &= \max_{1 \le k \le N} \sum_{j=1}^{n} p_j (1+\theta_j) q_{jk}. \end{aligned}$$

In the last expression we have used the fact that a linear form attains its maximum over a simplex at one of the vertices. The value of D_0 can be easily calculated by enumeration, and a subgradient with respect to θ is given by the vector Γ with coordinates $\gamma_j = p_j q_{jk^*}$, $j = 1, \dots, n$, with k^* representing the best vertex.

The dual problem of minimizing (53) has 1335 decision variables: the utility function u, represented by the vector of slopes $s \in \mathbb{R}^N$, and the multiplier $\theta \in \mathbb{R}^n$. The number of functions in (53) equals 617. It is a rather hard nonsmooth optimization problem, for present standards. As indicated earlier, we have used for its solution a new version of the regularized decomposition method of [19]. After the solution $(\hat{u}, \hat{\theta})$ of the dual problem is found, the optimal solution of the primal problem can be recovered from the subgradients of the dual function satisfying the optimality conditions.

We have selected as the reference portfolio the equally weighted portfolio of the 200 fastest growing companies in this 12-year period. The expected weekly return of this portfolio equals 0.0071. Of course, it has been selected *ex post*, but our objective here is just to illustrate the effect of the dominance constraint.

The method solved the problem in 163 iterations in *ca.* 38 min CPU time on a personal computer having a 1.6 GHz clock. The optimality conditions were satisfied with the accuracy of 10^{-8}.

The optimal portfolio contains 22 assets with weights ranging from 0.00095 to 0.0922. Its expected return equals 0.0116, as compared to 0.0164 of the fastest growing asset. It is interesting to note that the fastest growing asset participates in the optimal portfolio with the weight of 7% only.

The optimal utility function associated with the dominance constraint is illustrated in Figure 1. The data points in the figure are the points at which the slope of the utility function changes.

Acknowledgements. The authors are indebted to the two Referees for their very insightful comments.

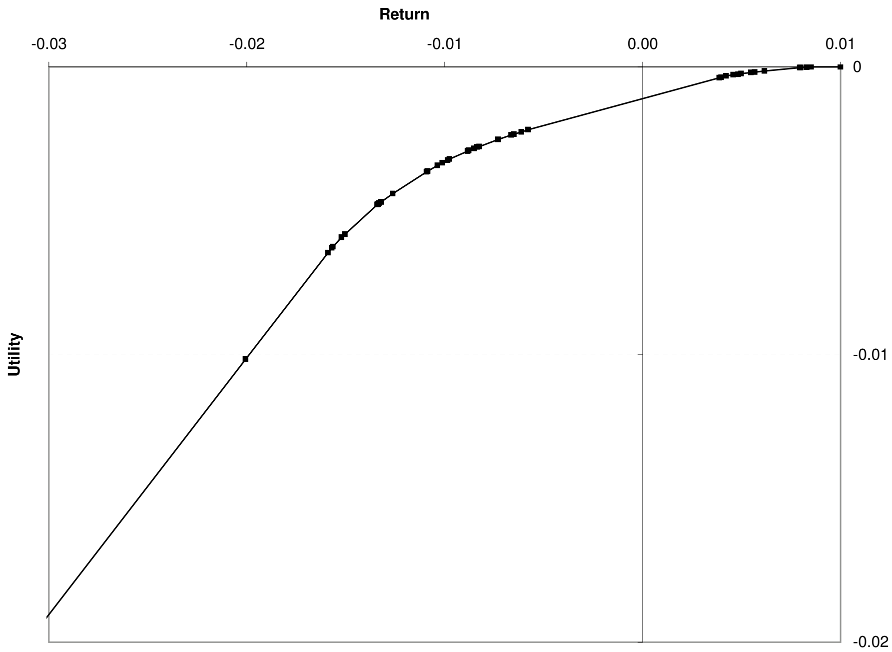

Fig. 1. The optimal utility function.

References

1. J.R Birge and F.V. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 1997.
2. J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.
3. C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Springer-Verlag, Berlin, 1977.
4. F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
5. D. Dentcheva and A. Ruszczyński, Optimization under linear stochastic dominance, *Comptes Rendus de l'Academie Bulgare des Sciences* 56 (2003), No. 6, pp. 6–11.
6. D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, 2002, manuscript submitted for publication.
7. R.M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.
8. J. Hadar and W. Russell, Rules for ordering uncertain prospects, *The American Economic Review* 59 (1969) 25–34.
9. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.

10. K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133, Springer-Verlag, Berlin, 1985.
11. K. Kuratowski and C. Ryll-Nardzewski, A general theorem on selectors, *Bull. Acad. Pol. Sc.* 13 (1965) 397–403.
12. V.L. Levin, *Convex Analysis in Spaces of Measurable Functions and Its Applications in Economics*, Nauka, Moscow, 1985 (in Russian).
13. K. Mosler and M. Scarsini (Eds.), *Stochastic Orders and Decision Under Risk*, Institute of Mathematical Statistics, Hayward, California, 1991.
14. W. Ogryczak and A. Ruszczyński, On consistency of stochastic dominance and mean–semideviation models, *Mathematical Programming* 89 (2001) 217–232.
15. A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, Boston, 1995.
16. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
17. R.T. Rockafellar and R.J-B. Wets, Stochastic convex programming: relatively complete recourse and induced feasibility, *SIAM J. Control Optimization* 14 (1976), 574–589.
18. R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
19. A. Ruszczyński, A regularized decomposition method for minimizing a sum of polyhedral functions, *Mathematical Programming* 35 (1986) 309-333.
20. A. Ruszczyński and R.J. Vanderbei, Frontiers of stochastically nondominated portfolios, *Econometrica*, 2003 (accepted for publication).
21. V. Strassen, The existence of probability measures with given marginals, *Annals of Mathematical Statistics* 38 (1965) 423–439.