MATHEMATICS OF OPERATIONS RESEARCH
Vol. 31, No. 3, August 2006, pp. 433–452
ISSN 0364-765X | EISSN 1526-5471 | 06 | 3103 | 0433
DOI 10.1287/moor.1050.0186

Optimization of Convex Risk Functions

Andrzej Ruszczyński
Department of Management Science and Information Systems, Rutgers University, Piscataway, New Jersey 08854,
rusz@business.rutgers.edu, http://www.rusz.rutgers.edu

Alexander Shapiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332,
ashapiro@isye.gatech.edu, http://www2.isye.gatech.edu/people/faculty/Alex_Shapiro/

We consider optimization problems involving convex risk functions. By employing techniques of convex analysis and optimization theory in vector spaces of measurable functions, we develop new representation theorems for risk models, and optimality and duality theory for problems with convex risk functions.

Key words: convex analysis; stochastic optimization; risk measures; duality
MSC2000 subject classification: Primary: 90C15, 90C48; secondary: 91B30, 90C46
OR/MS subject classification: Primary: stochastic programming; secondary: risk
History: Received January 31, 2004; revised January 24, 2005, and November 20, 2005.

1. Introduction. Comparison of uncertain outcomes is central for decision theory. If the outcomes have a probabilistic description, a wealth of concepts and techniques from the theory of probability can be employed. We can mention here the expected utility theory, stochastic ordering, and various mean-risk models. Our main objective is to contribute to this direction of research, by exploiting relations between risk models and optimization theory.

We assume that Ω is a certain space of elementary events and that an uncertain outcome is represented by a function $X\colon \Omega \to \mathbb{R}$. To focus attention, from now on, we assume that the smaller the values of X, the better; for example, X may represent an uncertain cost. It will be obvious how to translate our results to other situations.

By a *risk function*, we understand a function ρ, which assigns to an uncertain outcome X a real value $\rho(X)$. To make this concept precise and to obtain some meaningful results, one has to define the space $\mathcal{X}$ of allowable uncertain outcomes and to restrict the class of considered functions $\rho(\cdot)$. We assume that Ω is a measurable space equipped with σ-algebra $\mathcal{F}$ of subsets of Ω, and that $\mathcal{X}$ is a linear space of $\mathcal{F}$-measurable functions $X\colon \Omega \to \mathbb{R}$. Also, we consider risk functions, which can take values in the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. For $X, Y \in \mathcal{X}$, the notation $Y \succeq X$ means that $Y(\omega) \geq X(\omega)$ for all $\omega \in \Omega$.

For an extended real valued function $\rho\colon \mathcal{X} \to \overline{\mathbb{R}}$ consider the following conditions:

(A1) **Convexity:** $\rho(\alpha X + (1-\alpha)Y) \leq \alpha\rho(X) + (1-\alpha)\rho(Y)$ for all $X, Y \in \mathcal{X}$, and $\alpha \in [0, 1]$.
(A2) **Monotonicity:** If $X, Y \in \mathcal{X}$ and $Y \succeq X$, then $\rho(Y) \geq \rho(X)$.
(A3) **Translation equivariance:** If $a \in \mathbb{R}$ and $X \in \mathcal{X}$, then $\rho(X + a) = \rho(X) + a$.
(A4) **Positive homogeneity:** If $t > 0$ and $X \in \mathcal{X}$, then $\rho(tX) = t\rho(X)$.

These conditions were introduced, and real valued functions $\rho\colon \mathcal{X} \to \mathbb{R}$ satisfying (A1)–(A4) were called *coherent measures of risk* in the pioneering paper by Artzner et al. [1]. We reserve the term "measure" for its classical meaning of a countably additive set function, and refer to functions $\rho\colon \mathcal{X} \to \overline{\mathbb{R}}$ as *risk functions*.

A related research direction investigates mean-risk or mean-deviation models. In these models, the objective is a combination of a certain mean outcome (calculated with respect to some fixed probability measure $\bar{\mu}$), and some dispersion or deviation statistics, representing the uncertainty of the outcome. Most notable are here the works on the mean-variance model by Markowitz [13, 14], but many efforts have been made to use other deviation measures, like semideviation and deviations from quantiles (Ogryczak and Ruszczyński [16, 17, 18]). Recently, Rockafellar et al. [26] also developed an axiomatic approach to deviation measures. The emphasis in that paper is more on a connection between risk and deviation measures and is less on the monotonicity property (A2). Whenever appropriate, we compare their approach with the one presented in that paper.

The main contribution of our paper is the analysis of the relations of the theory of risk functions and optimization theory.

Section 2 has a synthetic character. We exploit general results of convex analysis in topological vector spaces of measurable functions to derive properties of convex risk functions. In this way, we generalize dual representation theorems given in Artzner et al. [1], Cheridito et al. [6], Delbaen [7], Föllmer and Schied [8], and Rockafellar et al. [26].

In applications, uncertain outcomes usually result from actions, or decisions, undertaken in some uncertain systems. Formally $X = F(z)$, where z is an element of some vector space $\mathscr{Z}$, and $F\colon \mathscr{Z} \to \mathscr{X}$. This creates the need to consider *composite risk functions*, of the form $\rho(F(z))$, and associated optimization problems

$$\operatorname*{Min}_{z\in S} \rho(F(z)), \tag{1.1}$$

where S is a convex subset of $\mathscr{Z}$. Section 3 is devoted to the analysis of differentiability properties of risk functions. In particular, we obtain representation of subgradients and directional derivatives of composite risk functions. In §4, we analyze risk functions resulting from several classical mean-risk models. In §5, we introduce the notion of risk aversion for risk functions and we characterize it with the use of the theoretical results of the first two sections.

The optimization problem (1.1) is discussed in §6. We analyze the implications of properties of the risk function ρ and of F on properties of problem (1.1) and its solutions. We also derive necessary and sufficient conditions of optimality. In §7, we introduce the concept of *risk value of perfect information*, for problem (1.1), and discuss its properties. Finally, in §8, we develop a duality relation for optimization problems involving risk functions and nonanticipativity constraints.

2. Conjugate duality of risk functions. Let $(\Omega, \mathscr{F})$ be a measurable space and $\mathbb{Y}$ be the (linear) space of all signed finite measures on $(\Omega, \mathscr{F})$. For $\mu \in \mathbb{Y}$, we denote by $|\mu|$ the corresponding total variation measure, i.e., $|\mu| = \mu^+ + \mu^-$, where $\mu = \mu^+ - \mu^-$ is the Jordan decomposition of μ.

Consider a linear space $\mathscr{X}$ of $\mathscr{F}$-measurable functions $X\colon \Omega \to \mathbb{R}$. We use the cone

$$\mathscr{X}_+ := \{X \in \mathscr{X}\colon X(\omega) \ge 0,\ \forall\, \omega \in \Omega\} \tag{2.1}$$

to define the corresponding partial order on the space $\mathscr{X}$. That is, the relation $Y \succeq X$ means that $Y - X \in \mathscr{X}_+$. We associate with $\mathscr{X}$ a linear space $\mathscr{Y} \subset \mathbb{Y}$ such that $\int_\Omega |X|\, d|\mu| < +\infty$ for every $X \in \mathscr{X}$ and $\mu \in \mathscr{Y}$, and we define the scalar product

$$\langle \mu, X\rangle := \int_\Omega X(\omega)\, d\mu(\omega). \tag{2.2}$$

By $\mathscr{Y}_+$, we denote the set of nonnegative measures $\mu \in \mathscr{Y}$, and by $\mathscr{P}$, the set of probability measures $\mu \in \mathscr{Y}$, i.e., $\mu \in \mathscr{P}$ if $\mu \in \mathscr{Y}_+$ and $\mu(\Omega) = 1$.

We also assume that the space $\mathscr{X}$ is sufficiently large so that the following condition holds true:

(C) If $\mu \notin \mathscr{Y}_+$, then there exists $X \in \mathscr{X}_+$ such that $\langle \mu, X\rangle < 0$.

The above condition ensures that the cone of nonnegative measures, $\mathscr{Y}_+$, is dual to the cone of nonnegative functions; i.e.,

$$\mathscr{Y}_+ = \{\mu \in \mathscr{Y}\colon \langle \mu, X\rangle \ge 0,\ \forall\, X \in \mathscr{X}_+\}.$$

Condition (C) is a very mild technical requirement on the pairing of $\mathscr{X}$ and $\mathscr{Y}$. We are using it in the key Theorem 2.2. Observe that a measure μ is not nonnegative if $\mu(A) < 0$ for some $A \in \mathscr{F}$. Therefore, condition (C) holds, for example, if the space $\mathscr{X}$ contains all indicator functions $\mathbb{1}_A(\cdot)$, $A \in \mathscr{F}$, where $\mathbb{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ for $\omega \notin A$. From now on, we shall always assume that the space $\mathscr{X}$ satisfies condition (C).

A natural choice of $\mathscr{X}$ is the space of all bounded $\mathscr{F}$-measurable functions $X\colon \Omega \to \mathbb{R}$. In that case, we can take $\mathscr{Y} := \mathbb{Y}$. Another possible choice is $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$ for some positive measure $\bar\mu \in \mathbb{Y}$ and $p \in [1, +\infty)$. The space $\mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$, equipped with the respective norm, is a Banach space. Its dual space $\mathscr{X}^*$ (of all continuous linear functionals on $\mathscr{X}$) is the space $\mathscr{X}^* = \mathscr{L}_q(\Omega, \mathscr{F}, \bar\mu)$, where $q \in (1, +\infty]$ is such that $1/p + 1/q = 1$. Note that an element $h \in \mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$ is a class of functions, which are equal to each other for almost every (a.e.) $\omega \in \Omega$ with respect to the measure $\bar\mu$. When dealing with $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$ and $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar\mu)$, we identify measure $\mu \in \mathscr{Y}$ with its density (Radon–Nikodym derivative) $\gamma = d\mu/d\bar\mu$ belonging to the space $\mathscr{L}_q(\Omega, \mathscr{F}, \bar\mu)$. The corresponding cones $\mathscr{X}_+ \subset \mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$ and $\mathscr{Y}_+ \subset \mathscr{L}_q(\Omega, \mathscr{F}, \bar\mu)$ are formed by nonnegative almost everywhere functions. Note that in the setting of $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar\mu)$ and $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar\mu)$, condition (C) always holds. Most of our considerations focus on the case when $\mathscr{X}$ is a Banach space and $\mathscr{Y} := \mathscr{X}^*$ is its topological dual.

The main theorem of this section is presented in the framework of *paired locally convex* topological vector spaces $\mathscr{X}$ and $\mathscr{Y}$. As compared with the Banach spaces setting, this general framework does not introduce any additional difficulties, while it enlarges the potential scope of applications. That is, unless stated otherwise, we assume that $\mathscr{X}$ and $\mathscr{Y}$ are equipped with respective topologies, which make them locally convex topological vector spaces and these topologies are compatible with the scalar product (2.2); i.e., every linear continuous functional on $\mathscr{X}$ can be represented in the form $\langle \mu, \cdot \rangle$ for some $\mu \in \mathscr{Y}$, and every linear continuous functional on $\mathscr{Y}$ can be represented in the form $\langle \cdot, X \rangle$ for some $X \in \mathscr{X}$. In particular, we can equip each space $\mathscr{X}$ and $\mathscr{Y}$ with its weak topology induced by its paired space. This will make $\mathscr{X}$ and $\mathscr{Y}$ paired locally convex topological vector spaces provided that for any $X \in \mathscr{X}\backslash\{0\}$, there exists $\mu \in \mathscr{Y}$ such that $\langle \mu, X \rangle \neq 0$, and for any $\mu \in \mathscr{Y}\backslash\{0\}$, there exists $X \in \mathscr{X}$ such that $\langle \mu, X \rangle \neq 0$. When dealing with Banach spaces, it is convenient to equip $\mathscr{X}$ and $\mathscr{Y} := \mathscr{X}^*$ with the strong (norm) and weak* topologies, respectively. If $\mathscr{X}$ is a reflexive Banach space, i.e., $\mathscr{X}^{**} = \mathscr{X}$, then $\mathscr{X}$ and $\mathscr{X}^*$, both equipped with strong topologies, form paired spaces.

Having defined the spaces $\mathscr{X}$ and $\mathscr{Y}$, we can return to the analysis of convex risk functions. We shall assume that every risk function ρ is *proper*, i.e., $\rho(X) > -\infty$ for all $X \in \mathscr{X}$ and its domain $\operatorname{dom}(\rho) := \{X \in \mathscr{X}: \rho(X) < +\infty\}$ is nonempty.

The conjugate $\rho^*: \mathscr{Y} \to \overline{\mathbb{R}}$ of a risk function ρ is defined as

$$\rho^*(\mu) := \sup_{X \in \mathscr{X}} \{\langle \mu, X \rangle - \rho(X)\}, \tag{2.3}$$

and the conjugate of ρ^* as

$$\rho^{**}(X) := \sup_{\mu \in \mathscr{Y}} \{\langle \mu, X \rangle - \rho^*(\mu)\}. \tag{2.4}$$

By $\operatorname{lsc}(\rho)$, we denote the lower semicontinuous hull of ρ taken with respect to the considered topology of $\mathscr{X}$. The following is the basic duality result of convex analysis (see, e.g., Rockafellar [22, Theorem 5], Aubin and Ekeland [2, Theorem 4.4.2] for a proof).

THEOREM 2.1 (FENCHEL–MOREAU). *Suppose that the function $\rho: \mathscr{X} \to \overline{\mathbb{R}}$ is convex and proper. Then, $\rho^{**} = \operatorname{lsc}(\rho)$.*

It follows that if ρ is convex, proper, and lower semicontinuous,[1] then the representation

$$\rho(X) = \sup_{\mu \in \mathscr{Y}} \{\langle \mu, X \rangle - \rho^*(\mu)\}, \quad X \in \mathscr{X} \tag{2.5}$$

holds true. Conversely, if (2.5) is satisfied for some function $\rho^*(\cdot)$, then ρ is lower semicontinuous and convex. Note also that if ρ is proper, lower semicontinuous, and convex, then its conjugate function ρ^* is proper. Let us also remark that if $\mathscr{X}$ is a Banach space and $\mathscr{Y} := \mathscr{X}^*$ is its dual (e.g., $\mathscr{X} = \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$ and $\mathscr{Y} = \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$) and ρ is convex, then ρ is lower semicontinuous in the weak topology if, and only if, (iff) it is lower semicontinuous in the strong (norm) topology. If the set Ω is finite, then the space $\mathscr{X}$ is finite dimensional. In that case, ρ is continuous (and hence lower semicontinuous) if it is real valued.

Clearly, we can write Equation (2.5) in the following equivalent form:

$$\rho(X) = \sup_{\mu \in \mathscr{A}} \{\langle \mu, X \rangle - \rho^*(\mu)\}, \quad \forall X \in \mathscr{X}, \tag{2.6}$$

where

$$\mathscr{A} := \operatorname{dom}(\rho^*) = \{\mu \in \mathscr{Y}: \rho^*(\mu) < +\infty\}$$

is the domain of the conjugate function $\rho^*: \mathscr{Y} \to \overline{\mathbb{R}}$.

The following theorem generalizes to the case of paired locally convex topological vector spaces, satisfying condition (C), the representation theorems obtained by Delbaen [7, Theorem 2.3], Föllmer and Schied [8, Theorems 5–6], and Cheridito et al. [6, Theorem 3.3].

THEOREM 2.2. *Suppose that $\rho: \mathscr{X} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous, and convex. Then, representation* (2.6) *holds with $\mathscr{A} := \operatorname{dom}(\rho^*)$. Moreover, we have that*

(i) *Condition* (A2) *holds iff every measure $\mu \in \mathscr{A}$ is nonnegative*;
(ii) *Condition* (A3) *holds iff $\mu(\Omega) = 1$ for every $\mu \in \mathscr{A}$*;
(iii) *Condition* (A4) *holds iff $\rho(\cdot)$ can be represented in the form*

$$\rho(X) = \sup_{\mu \in \mathscr{A}} \langle \mu, X \rangle, \quad \forall X \in \mathscr{X}. \tag{2.7}$$

[1] The Fatou property used in Artzner et al. [1], Cheridito et al. [6], Delbaen [7], and Föllmer and Schied [8] is equivalent to weak lower semicontinuity in this context.

Proof. If $\rho: \mathscr{X} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous, and convex, then representation (2.6) holds by the Fenchel–Moreau theorem (Theorem 2.1).

Now, suppose that condition (A2) holds true. It follows then that $\rho^*(\mu) = +\infty$ for any measure $\mu \in \mathscr{Y}$, which is not nonnegative. Indeed, if $\mu \notin \mathscr{Y}_+$, then we have by condition (C) that $\langle \mu, \overline{X} \rangle < 0$ for some $\overline{X} \in \mathscr{X}_+$. Take any X in the domain of ρ, i.e., such that $\rho(X)$ is finite, and consider $X_t := X - t\overline{X}$. Then, for $t \geq 0$, we have by assumption (A2) that $X \geq X_t$, and hence $\rho(X) \geq \rho(X_t)$. Consequently,

$$\rho^*(\mu) \geq \sup_{t \in \mathbb{R}_+} \{\langle \mu, X_t \rangle - \rho(X_t)\} \geq \sup_{t \in \mathbb{R}_+} \{\langle \mu, X \rangle - t\langle \mu, \overline{X} \rangle - \rho(X)\} = +\infty.$$

Conversely, suppose that every measure $\mu \in \mathscr{A}$ is nonnegative. Then, for every $\mu \in \mathscr{A}$ and $Y \succeq X$, we have that $\langle \mu, Y \rangle \geq \langle \mu, X \rangle$. By (2.6), this implies that $\rho(Y) \geq \rho(X)$. This completes the proof of assertion (i).

Suppose that condition (A3) holds. Then, for an $X \in \operatorname{dom}(\rho)$, we have

$$\rho^*(\mu) \geq \sup_{a \in \mathbb{R}} \{\langle \mu, X + a \rangle - \rho(X + a)\} = \sup_{a \in \mathbb{R}} \{a\mu(\Omega) - a + \langle \mu, X \rangle - \rho(X)\}.$$

It follows that $\rho^*(\mu) = +\infty$ for any $\mu \in \mathscr{Y}$ such that $\mu(\Omega) \neq 1$. Conversely, if $\mu(\Omega) = 1$, then $\langle \mu, X + a \rangle = \langle \mu, X \rangle + a$, and hence condition (A3) follows by (2.6). This completes the proof of (ii).

Clearly, if (2.7) holds, then ρ is positively homogeneous. Conversely, if ρ is positively homogeneous, then its conjugate function is the indicator function of a closed convex subset of $\mathscr{Y}$. Consequently, the representation (2.7) follows by (2.6). □

It follows from the above theorem that if a risk function ρ is proper, satisfies conditions (A1)–(A3), and is lower semicontinuous, then the representation (2.6) holds with $\mathscr{A}$ being a subset of the set $\mathscr{P}$ of probability measures. If, moreover, ρ is positively homogeneous (i.e., condition (A4) holds), then its conjugate ρ^* is the indicator function of a closed convex set $\mathscr{A} \subset \mathscr{P}$, which can be written in the form

$$\mathscr{A} = \{\mu \in \mathscr{P}: \langle \mu, X \rangle \leq \rho(X), \ \forall X \in \mathscr{X}\}. \tag{2.8}$$

This set $\mathscr{A}$ is called the *risk envelope* in Rockafellar et al. [26], where the above result has been developed in the space $\mathscr{X} := \mathscr{L}_2(\Omega, \mathscr{F}, \bar{\mu})$.

3. Continuity and differentiability.

3.1. Risk functions. In applications, it is usually straightforward to verify whether some (all) of the assumptions (A1)–(A4) hold. The assumption of the lower semicontinuity of ρ is more delicate. To verify continuity properties of ρ, it is technically advantageous to use the strong (rather than weak) topology of $\mathscr{X}$ if $\mathscr{X}$ is a Banach space. Therefore, when dealing with a Banach space $\mathscr{X}$, we equip it with its strong topology and use $\mathscr{Y} := \mathscr{X}^*$.

Suppose that ρ is proper and convex, and denote by $\operatorname{int}(\operatorname{dom} \rho)$ the interior of the domain of ρ. We have that if ρ is bounded from above on a neighborhood of some point $\overline{X} \in \mathscr{X}$, then ρ is continuous on $\operatorname{int}(\operatorname{dom} \rho)$ (e.g., Ioffe and Tihomirov [9, Theorem 1, p. 170]). Unfortunately, this result cannot be applied directly in many situations. Note, for example, that the cone of nonnegative-valued functions in space $\mathscr{L}_p(\mathbb{R}^m, \mathscr{B}, \mu)$, $p \in [1, +\infty)$, where $\mathscr{B}$ is the Borel sigma-algebra of $\mathbb{R}^m$ and μ is a continuous probability measure, has an empty interior. Therefore we use below a different approach.

A linear functional $l: \mathscr{X} \to \mathbb{R}$ is called an *algebraic subgradient* of ρ at $\overline{X} \in \operatorname{dom} \rho$ if

$$\rho(X) \geq \rho(\overline{X}) + l(X - \overline{X}), \quad \forall X \in \mathscr{X}. \tag{3.1}$$

Note that the algebraic subgradient functional l is not required to be continuous. If, moreover, $l \in \mathscr{Y}$, then we say that l is a subgradient of ρ at $\overline{X}$. The set of all subgradients $l \in \mathscr{Y}$, satisfying (3.1), is called the subdifferential of ρ at $\overline{X}$, and denoted $\partial\rho(\overline{X})$. It is said that ρ is subdifferentiable at $\overline{X}$ if $\partial\rho(\overline{X})$ is nonempty.

Let us observe that ρ always possesses an algebraic subgradient at any point $\overline{X} \in \operatorname{int}(\operatorname{dom} \rho)$ (cf. Levin [11, Lemma 1.1]). Indeed, consider the directional derivative function $\delta(\cdot) := \rho'(\overline{X}, \cdot)$, where

$$\rho'(\overline{X}, X) := \lim_{t \downarrow 0} \frac{\rho(\overline{X} + tX) - \rho(\overline{X})}{t}.$$

The function $\delta(\cdot)$ is positively homogeneous. By the convexity of ρ, it is convex and satisfies for all X the inequality $\rho(X) \geq \rho(\bar{X}) + \delta(X - \bar{X})$. Moreover, if $\bar{X} \in \operatorname{int}(\operatorname{dom}\rho)$, then $\delta(\cdot)$ is finite valued. By the Hahn-Banach theorem, we have that there exists a linear functional $l\colon \mathscr{X} \to \mathbb{R}$ such that $\delta(\cdot) \geq l(\cdot)$. It follows that l satisfies (3.1).

We show now that lower semicontinuity of ρ is implied by assumptions (A1)–(A2) if $\mathscr{X}$ has the structure of a Banach lattice. Recall that $\mathscr{X}$ is a *lattice* (with respect to the cone $\mathscr{X}_+$) if for any $X_1, X_2 \in \mathscr{X}$, the element $X_1 \vee X_2$, defined as

$$[X_1 \vee X_2](\omega) := \max\{X_1(\omega), X_2(\omega)\}, \quad \omega \in \Omega,$$

belongs to $\mathscr{X}$. For every $X \in \mathscr{X}$, we can then define $|X| \in \mathscr{X}$ in a natural way; i.e., $|X|(\omega) = |X(\omega)|$, $\omega \in \Omega$. The space $\mathscr{X}$ is a *Banach lattice* if it is a Banach space and $|X_1| \leq |X_2|$ implies $\|X_1\| \leq \|X_2\|$. For example, every space $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$, $p \in [1, +\infty]$, is a Banach lattice.

Proposition 3.1. *Suppose that $\mathscr{X}$ is a Banach lattice and $\rho\colon \mathscr{X} \to \overline{\mathbb{R}}$ is proper and satisfies assumptions* (A1) *and* (A2). *Then, $\rho(\cdot)$ is continuous and subdifferentiable on the interior of its domain.*

Proof. Let $\bar{X} \in \operatorname{int}(\operatorname{dom}\rho)$. By the discussion preceding the statement of the theorem, ρ possesses an algebraic subgradient, denoted l, at $\bar{X}$. It follows from the monotonicity of $\rho(\cdot)$ that l is positive in the sense that $l(X) \geq 0$ for all $X \in \mathscr{X}_+$. Indeed, if $l(Y) < 0$ for some $Y \in \mathscr{X}_+$, then it follows from (3.1) that $\rho(\bar{X} - Y) > \rho(\bar{X})$, which contradicts (A2). Now, by Levin [11, Theorem 0.12], we have that any positive linear functional on the Banach lattice $\mathscr{X}$ is continuous. Consequently, l is continuous, and hence $l \in \partial\rho(\bar{X})$. It follows then from (3.1) that ρ is lower semicontinuous at $\bar{X}$. Since $\bar{X}$ was an arbitrary point of $\operatorname{int}(\operatorname{dom}\rho)$, we obtain that $\rho(\cdot)$ is lower semicontinuous on the interior of its domain. This, combined with the fact that $\mathscr{X}$ is a Banach space, implies the continuity of $\rho(\cdot)$ on $\operatorname{int}(\operatorname{dom}\rho)$ (see, e.g., Phelps [19, Theorem 3.3]). □

We obtain that under the assumptions of the above proposition, if $\rho(X)$ is real valued for all $X \in \mathscr{X}$, then $\rho(\cdot)$ is continuous and subdifferentiable on $\mathscr{X}$. Proposition 3.1 can be applied, for example, to every space $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$ with $p \in [1, +\infty)$. We also can apply this framework to the space $\mathscr{X} := \mathscr{L}_\infty(\Omega, \mathscr{F}, \bar{\mu})$ if we equip it with its strong topology. This, however, will require us to pair $\mathscr{X}$ with its dual space $\mathscr{Y} := \mathscr{L}_\infty(\Omega, \mathscr{F}, \bar{\mu})^*$, which is larger than $\mathscr{L}_1(\Omega, \mathscr{F}, \bar{\mu})$.

Corollary 3.1. *Let $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$, with $p \in [1, +\infty)$, be paired with its dual space $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$. Suppose that $\rho\colon \mathscr{X} \to \overline{\mathbb{R}}$ is proper and satisfies assumptions* (A1) *and* (A2). *Then, $\rho(\cdot)$ is continuous and subdifferentiable on the interior of its domain.*

Consider a point $\bar{X} \in \operatorname{dom}(\rho)$. It immediately follows from the definitions that

$$\mu \in \partial\rho(\bar{X}) \quad \text{iff } \rho^*(\mu) = \langle \mu, \bar{X} \rangle - \rho(\bar{X}). \tag{3.2}$$

By applying this to the function ρ^{**}, instead of ρ, and using the identity $\rho^{***} = \rho^*$, which follows from the Fenchel–Moreau theorem, we obtain that (cf. Rockafellar [22, p. 35])

$$\partial\rho^{**}(\bar{X}) = \operatorname*{arg\,max}_{\mu \in \mathscr{A}} \{\langle \mu, \bar{X} \rangle - \rho^*(\mu)\} \tag{3.3}$$

(recall that $\mathscr{A} = \operatorname{dom}(\rho^*)$). We also have that if ρ is subdifferentiable at $\bar{X}$, then $\partial\rho^{**}(\bar{X}) = \partial\rho(\bar{X})$. It follows that if ρ is subdifferentiable at $\bar{X}$, then $\partial\rho(\bar{X})$ is equal to the right-hand side of (3.3) and, moreover, if assumptions (A1)–(A3) hold, then $\partial\rho(\bar{X}) \subset \mathscr{P}$. In particular, we obtain that if conditions (A1)–(A4) hold and ρ is lower semicontinuous, then the representation (2.7) holds, with $\mathscr{A} = \partial\rho(0) \subset \mathscr{P}$ and

$$\partial\rho(\bar{X}) = \operatorname*{arg\,max}_{\mu \in \mathscr{A}} \langle \mu, \bar{X} \rangle. \tag{3.4}$$

There is a duality relation between the subdifferential $\partial\rho(\bar{X})$ and the directional derivative function $\rho'(\bar{X}, \cdot)$. That is, if ρ is subdifferentiable at $\bar{X}$ and $\rho'(\bar{X}, \cdot)$ is lower semicontinuous at $0 \in \mathscr{X}$, then

$$\rho'(\bar{X}, X) = \sup_{\mu \in \partial\rho(\bar{X})} \langle \mu, X \rangle, \quad X \in \mathscr{X}. \tag{3.5}$$

In particular, if $\mathscr{X}$ is a Banach space and ρ is continuous at $\bar{X}$, then (3.5) holds and ρ is directionally differentiable at $\bar{X}$ in the Hadamard sense; i.e.,

$$\rho'(\bar{X}, X) = \lim_{\substack{X' \to X \\ t \downarrow 0}} \frac{\rho(\bar{X} + tX') - \rho(\bar{X})}{t}$$

(see, e.g., Bonnans and Shapiro [4, §2.2.1] for a discussion of Hadamard directional derivatives). Hadamard directional differentiability implies continuity of the directional derivative function $\rho'(\overline{X}, \cdot)$.

Recall that ρ is said to be Gâteaux differentiable at $\overline{X}$ if ρ is directionally differentiable at $\overline{X}$ and $\rho'(\overline{X}, \cdot)$ is linear and continuous, i.e., there exists an element $\nabla\rho(\overline{X}) \in \mathcal{Y}$ such that

$$\rho'(\overline{X}, \cdot) = \langle \nabla\rho(\overline{X}), X\rangle, \quad \forall X \in \mathcal{X}.$$

It follows that if $\mathcal{X}$ is a Banach space and ρ is continuous at $\overline{X}$, then ρ is Gâteaux (Hadamard) differentiable at $\overline{X}$ iff $\partial\rho(\overline{X})$ is a singleton, in which case $\partial\rho(\overline{X}) = \{\nabla\rho(\overline{X})\}$.

3.2. Composite risk functions. In the subsequent analysis, we shall deal with composite functions $\psi\colon \mathcal{Z} \to \overline{\mathbb{R}}$ of the form $\psi(\cdot) := \rho(F(\cdot))$. Here, $\mathcal{Z}$ is a vector space and $F\colon \mathcal{Z} \to \mathcal{X}$ is a mapping. We write $f(z, \omega)$ or $f_\omega(z)$ for $[F(z)](\omega)$, and view $f(z, \omega)$ as a random function defined on the measurable space $(\Omega, \mathcal{F})$. We say that the mapping F is *convex* if the function $f(\cdot, \omega)$ is convex for every $\omega \in \Omega$.

Lemma 3.1. *If the mapping $F\colon \mathcal{Z} \to \mathcal{X}$ is convex and $\rho\colon \mathcal{X} \to \overline{\mathbb{R}}$ satisfies assumptions* (A1)–(A2), *then the composite function $\psi(\cdot) := \rho(F(\cdot))$ is convex.*

Proof. For any $z_1, z_2 \in \mathcal{Z}$, and $\alpha \in [0, 1]$, we have by convexity of $F(\cdot)$ and monotonicity of $\rho(\cdot)$ that

$$\rho(F(\alpha z_1 + (1-\alpha)z_2)) \le \rho(\alpha F(z_1) + (1-\alpha)F(z_2)).$$

Hence convexity of $\rho(\cdot)$ implies that

$$\rho(F(\alpha z_1 + (1-\alpha)z_2)) \le \alpha\rho(F(z_1)) + (1-\alpha)\rho(F(z_2)).$$

This proves the convexity of $\rho(F(\cdot))$. □

Let us discuss differentiability properties of the composite function $\psi(\cdot)$ at a point $\bar{z} \in \mathcal{Z}$.

Proposition 3.2. *Suppose that $\mathcal{X}$ is a Banach space, the mapping F is convex, the function ρ is convex, finite valued and continuous at $\overline{X} := F(\bar{z})$. Then, the composite function ψ is directionally differentiable at $\bar{z}$, $\psi'(\bar{z}, z)$ is finite valued for every $z \in \mathcal{Z}$ and*

$$\psi'(\bar{z}, z) = \sup_{\mu \in \partial\rho(\overline{X})} \int_\Omega f'_\omega(\bar{z}, z)\, d\mu(\omega). \tag{3.6}$$

Proof. We have here that ρ is subdifferentiable and Hadamard directionally differentiable at $\overline{X} := F(\bar{z})$ and formula (3.5) holds. By the convexity of F, we also have that F is directionally differentiable at $\bar{z}$ with $[F'(\bar{z}, z)](\omega) = f'_\omega(\bar{z}, z)$. Because of the Hadamard directional differentiability of ρ, we can apply the chain rule to conclude that $\psi(\cdot)$ is directionally differentiable at $\bar{z}$, $\psi'(\bar{z}, z)$ is finite valued, and

$$\psi'(\bar{z}, z) = \rho'(\overline{X}, F'(\bar{z}, z)).$$

Together with (3.5), the above formula implies (3.6). □

Corollary 3.2. *Suppose that $\mathcal{X}$ and $\mathcal{Z}$ are Banach spaces, the mapping F is convex and continuous at $\bar{z}$, the function ρ satisfies assumptions* (A1) *and* (A2), *is continuous at $\overline{X} := F(\bar{z})$, and $\partial\rho(\overline{X}) = \{\bar{\mu}\}$ is a singleton. Then, the composite function ψ is Hadamard differentiable at $\bar{z}$ iff $f'_\omega(\bar{z}, \cdot)$ is linear for $\bar{\mu}$-almost every $\omega \in \Omega$.*

Proof. By Proposition 3.2, we have here that

$$\psi'(\bar{z}, z) = \int_\Omega f'_\omega(\bar{z}, z)\, d\bar{\mu}(\omega). \tag{3.7}$$

The function $\int_\Omega f'_\omega(\bar{z}, \cdot)\, d\bar{\mu}(\omega)$ is real valued, convex, and positively homogeneous. It is linear iff $f'_\omega(\bar{z}, \cdot)$ is linear for μ-almost every $\omega \in \Omega$. Therefore, $\psi'(\bar{z}, \cdot)$ is linear iff $f'_\omega(\bar{z}, \cdot)$ is linear for $\bar{\mu}$-almost every $\omega \in \Omega$. We also have that ψ is continuous at $\bar{z}$, and by Lemma 3.1, ψ is convex. It follows that if $\psi'(\bar{z}, \cdot)$ is linear, then $\partial\psi(\bar{z})$ is a singleton, and hence ψ is Hadamard differentiable at $\bar{z}$. □

It is also possible to write formula (3.6) in terms of the corresponding subdifferentials. Suppose that $\mathcal{Z}$ is a separable Banach space, $\mathcal{Z}^*$ is its dual space of all continuous linear functionals on $\mathcal{Z}$, $F: \mathcal{Z} \to \mathcal{X}$ is convex, and consider the integral function

$$\varphi_\mu(z) := \int_\Omega f_\omega(z)\, d\mu(\omega)$$

associated with nonnegative measure $\mu \in \mathcal{Y}_+$. Suppose, further, that functions $f_\omega(\cdot)$, $\omega \in \Omega$, and $\varphi_\mu(\cdot)$ are continuous at a point $\bar{z} \in \mathcal{Z}$. Then, by Strassen's theorem (in the general form provided in Levin [11, Theorem 1.1], see also Castaing and Valadier [5], Strassen [29]), we have that

$$\partial \varphi_\mu(\bar{z}) = \int_\Omega \partial f_\omega(\bar{z})\, d\mu(\omega). \tag{3.8}$$

The integral at the right-hand side of (3.8) is understood as the set of elements of $\mathcal{Z}^*$ of the form $\int_\Omega z^*(\omega)\, d\mu(\omega)$, where $z^*(\omega) \in \partial f_\omega(\bar{z}) \subset \mathcal{Z}^*$ for every $\omega \in \Omega$, and $z^*(\cdot)$ is weakly* μ-integrable; that is, $\langle z^*(\cdot), v\rangle$ is μ-integrable for every $v \in \mathcal{Z}$.

By $\mathrm{cl}(S)$, we denote the closure (in the weak* topology) of set $S \subset \mathcal{Z}^*$.

Proposition 3.3. *Suppose that $\mathcal{X}$ is a Banach space, $\mathcal{Z}$ is a separable Banach space, the mapping F is convex and continuous at $\bar{z}$, the function ρ satisfies assumptions* (A1) *and* (A2), *and is finite valued and continuous at $\overline{X} := F(\bar{z})$. Then, the composite function ψ is subdifferentiable at $\bar{z}$ and*

$$\partial \psi(\bar{z}) = \mathrm{cl}\left(\bigcup_{\mu \in \partial \rho(\overline{X})} \int_\Omega \partial f_\omega(\bar{z})\, d\mu(\omega) \right). \tag{3.9}$$

Proof. Since ρ is convex, finite valued, and continuous at $\overline{X}$, it is subdifferentiable at $\overline{X}$. Moreover, by condition (A2), we have that $\partial\rho(\overline{X}) \subset \mathcal{Y}_+$ (see Theorem 2.2(i)). Because $\partial\rho(\overline{X})$ is convex and the right-hand side of (3.8) linear in μ, the set inside the parentheses at the right-hand side of (3.9) is convex. Now, under the specified assumptions, the composite function ψ is continuous at $\bar{z}$ and formula (3.6) holds. It follows that $\psi'(\bar{z}, \cdot)$ is convex, continuous, positively homogeneous, and $\partial\psi(\bar{z})$ is equal to the subdifferential of the function $\psi'(\bar{z}, \cdot)$ at $0 \in \mathcal{Z}$. Moreover, $\psi'(\bar{z}, \cdot) = \sup_{\mu \in \partial\rho(\overline{X})} \eta_\mu(\cdot)$, where functions

$$\eta_\mu(\cdot) := \int_\Omega f'_\omega(\bar{z}, \cdot)\, d\mu(\omega) \tag{3.10}$$

are convex continuous and positively homogeneous. Consequently, $\partial\psi(\bar{z})$ is equal to the topological closure of the convex hull of the union of $\partial\eta_\mu(0)$, over $\mu \in \partial\rho(\overline{X})$ (e.g., Bonnans and Shapiro [4, Proposition 2.116(ii)]). Since $\partial\eta_\mu(0) = \partial\varphi_\mu(\bar{z})$, applying (3.8), we obtain (3.9). This completes the proof. □

It is possible to derive from formula (3.9) various conditions ensuring differentiability of the composite function.

Corollary 3.3. *Suppose that the assumptions of Proposition 3.3 are satisfied and, moreover, $\partial f_\omega(\bar{z}) = \{\nabla f_\omega(\bar{z})\}$ is a singleton for μ-almost every $\omega \in \Omega$, for all $\mu \in \partial\rho(\overline{X})$. Then, ψ is Hadamard differentiable at $\bar{z}$ iff $\partial\rho(\overline{X}) = \{\bar{\mu}\}$ is a singleton, in which case*

$$\nabla\psi(\bar{z}) = \int_\Omega \nabla f_\omega(\bar{z})\, d\bar{\mu}(\omega). \tag{3.11}$$

Of course, if the set inside the parentheses in the right-hand side of (3.9) is closed, then formula (3.9) takes the form

$$\partial\psi(\bar{z}) = \bigcup_{\mu \in \partial\rho(\overline{X})} \int_\Omega \partial f_\omega(\bar{z})\, d\mu(\omega). \tag{3.12}$$

This is always true if the assumptions of Proposition 3.3 are satisfied and either the subdifferential $\partial\rho(\overline{X})$ is a singleton or the space $\mathcal{Z}$ is finite dimensional and Ω is finite. Another case where this holds is the following.

Proposition 3.4. *Suppose that the assumptions of Proposition 3.3 are satisfied, the space $\mathcal{Z}$ is finite dimensional and $\partial f_\omega(\bar{z}) = \{\nabla f_\omega(\bar{z})\}$ is a singleton for μ-almost every $\omega \in \Omega$, for all $\mu \in \partial\rho(\overline{X})$. Then,*

$$\partial\psi(\bar{z}) = \bigcup_{\mu \in \partial\rho(\overline{X})} \int_\Omega \nabla f_\omega(\bar{z})\, d\mu(\omega). \tag{3.13}$$

Proof. In view of Proposition 3.3, we only need to show that the set in the right-hand side of (3.13) is closed in the standard topology of $\mathscr{Z}$ (recall that $\mathscr{Z}$ is assumed to be finite dimensional). Now, since ρ is continuous at $\overline{X}$, we have that the set $\partial\rho(\overline{X})$ is compact in the weak* topology of $\mathscr{X}^*$ (e.g., Ioffe and Tihomirov [9, Proposition 3, p. 199]). Also, the mapping $\mu \mapsto \int_\Omega \nabla f_\omega(\bar{z})\, d\mu(\omega)$ is continuous with respect to the weak* topology of $\mathscr{X}^*$ and the standard topology of $\mathscr{Z}$. It follows that the image of the set $\partial\rho(\overline{X})$ by this mapping is compact, and hence closed. We obtain that the set at the right-hand side of (3.13) is closed. □

4. Examples of risk functions.

In this section, we investigate several examples of risk models, which are discussed in the literature.

Example 4.1 (Mean-Deviation Risk Function). Let $\bar{\mu}$ be a probability measure on $(\Omega, \mathscr{F})$ and consider the space $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$ for some $p \in [1, +\infty)$. Define

$$\rho(X) := \langle \bar{\mu}, X\rangle + c\psi_p(X), \tag{4.1}$$

where $c \geq 0$ is a constant and

$$\psi_p(X) := \|X - \langle \bar{\mu}, X\rangle\|_p = \left(\int_\Omega |X(\omega) - \langle \bar{\mu}, X\rangle|^p\, d\bar{\mu}(\omega)\right)^{1/p}. \tag{4.2}$$

Note that for $p = 2$, the function $\rho(\cdot)$ corresponds to the classical mean-variance model of Markowitz [13], but with the standard deviation instead of the variance.

The functions ψ_p, ρ: $\mathscr{X} \to \mathbb{R}$ are convex, positively homogeneous, and continuous in the strong (norm) topology of $\mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$. Consider the set

$$\mathscr{M}_p := \{\nu \in \mathbb{Y}\colon \langle \nu, X\rangle \leq \psi_p(X),\ X \in \mathscr{X}\} \tag{4.3}$$

and $\nu \in \mathscr{M}_p$. For a set $A \in \mathscr{F}$, let $A = A^+ \cup A^-$ be the Jordan decomposition of A with respect to ν; i.e., $A^+ \cap A^- = \varnothing$ and $|\nu|(A) = \nu(A^+) - \nu(A^-)$. Let $X(\cdot) := \mathbb{1}_{A^+}(\cdot) - \mathbb{1}_{A^-}(\cdot)$. Then, $|\nu|(A) = \langle \nu, X\rangle$, and if $\bar{\mu}(A) = 0$, then $\psi_p(X) = 0$. It follows, by the definition of the set $\mathscr{M}_p$, that if $\bar{\mu}(A) = 0$, then $|\nu|(A) = 0$, and hence ν is absolutely continuous with respect to $\bar{\mu}$. Consider the Radon–Nikodym derivative $h = d\nu/d\bar{\mu}$. It is natural then to embed the set $\mathscr{M}_p$ into the space of absolutely continuous measures (with respect to $\bar{\mu}$) having a density $h \in \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$. With some abuse of the notation, we take $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$ and write

$$\mathscr{M}_p = \left\{h \in \mathscr{Y}\colon \int_\Omega X(\omega)h(\omega)\, d\bar{\mu}(\omega) \leq \psi_p(X),\ X \in \mathscr{X}\right\}. \tag{4.4}$$

We have that $\mathscr{M}_p$ is equal to the subdifferential $\partial\psi_p(X)$ at $X = 0$. Also, recall that the subdifferential of the norm $\|X\|_p$, at $X = 0$, is equal to the unit ball $B_q := \{h \in \mathscr{Y}\colon \|h\|_q \leq 1\}$ in the dual space $\mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$. Consider the (linear) operator $A(X) := X - \langle \bar{\mu}, X\rangle$. By the Moreau–Rockafellar theorem, we have that $\partial\psi_p(0) = A^*(B_q)$, where A^*: $\mathscr{Y} \to \mathscr{Y}$ is the adjoint of the operator A. By a straightforward calculation, we have that $A^*(h) = h - \int_\Omega h\, d\bar{\mu}$. Consequently,

$$\mathscr{M}_p = \left\{h - \int_\Omega h\, d\bar{\mu}\colon h \in B_q\right\}. \tag{4.5}$$

It follows that

$$\rho(X) = \sup_{\mu \in \mathscr{A}_p} \langle \mu, X\rangle, \tag{4.6}$$

where the set $\mathscr{A}_p := 1 + c\mathscr{M}_p$ can be written in the form

$$\mathscr{A}_p = \left\{g \in \mathscr{Y}\colon g = 1 + h - \int_\Omega h\, d\bar{\mu},\ \|h\|_q \leq c\right\}. \tag{4.7}$$

Suppose now that $p = 1$. Then, $q = +\infty$; i.e., the dual norm $\|h\|_\infty$ is given by the essential maximum of $|h(\omega)|$, $\omega \in \Omega$. In that case, all functions $g \in \mathscr{A}_p$ are almost everywhere nonnegative valued, and hence $\mathscr{A}_p$ is a set of probability measures if $c \leq 1/2$. Indeed, it is clear that if $\|h\|_\infty \leq c$, then

$$1 + h(\omega) - \int_\Omega h\, d\bar{\mu} \geq 1 - |h(\omega)| - \int_\Omega |h|\, d\bar{\mu} \geq 1 - 2c$$

for a.e. $w \in \Omega$. Moreover, $\mathscr{A}_p$ is a set of probability measures iff $c \leq 1/2$ provided that the following condition holds.

(∗) For every $\varepsilon > 0$, there exists $A \in \mathcal{F}$ such that $0 < \bar{\mu}(A) < \varepsilon$.
To show the "only if" part, take

$$h(\cdot) := c[-\mathbb{1}_A(\cdot) + \mathbb{1}_{\Omega\setminus A}(\cdot)].$$

Then, $\|h\|_\infty = c$, $\int_\Omega h\, d\bar{\mu} = c[1 - 2\bar{\mu}(A)]$, and hence

$$\inf_{\omega\in\Omega}\left\{1 + h(\omega) - \int_\Omega h\, d\bar{\mu}\right\} = 1 - 2c + 2c\bar{\mu}(A).$$

Consequently, if $c > 1/2$, then for $A \in \mathcal{F}$ such that $\bar{\mu}(A) > 0$ is small enough, the right-hand side of the above equation is negative.

For $p > 1$, the situation is different. Suppose for the moment that Ω is finite; say, $\Omega = \{\omega_1, \dots, \omega_K\}$ with respective (positive) probabilities $p_1, \dots, p_K$. In that case, a necessary condition for $\mathcal{A}_p$ to be a set of probability measures is that the following inequality should hold:

$$c \le \min_{1\le i\le K}\left[p_i^{-1/q} - 1\right]^{-1}. \tag{4.8}$$

The right-hand side of the above inequality is less than or equal to $1/(K^{1/q} - 1)$, with the equality for $p_1 = \cdots = p_K = 1/K$. Therefore, for large K, the allowable values of c (for which $\mathcal{A}_p$ is a set of probability measures) are very small. If the measure $\bar{\mu}$ is such that the above property (∗) holds, then for $p > 1$, the set $\mathcal{A}_p$ is not a set of probability measures, no matter what the value of $c > 0$ is.

Remark 4.1. It might be worth mentioning that $\psi_p(X)$ satisfies all axioms of a deviation measure specified in Rockafellar et al. [26]. Note, however, that for $p > 1$ (and, in particular, for $p = 2$), the resulting mean-deviation model (4.1) may violate the monotonicity property (A2). In fact, the mean-deviation model (4.1) violates the monotonicity property for *any* $c > 0$ if the measure $\bar{\mu}$ satisfies the above specified property (∗).

Example 4.2 (Mean-Semideviation Risk Function). Let, as in Example 4.1, $\bar{\mu}$ be a probability measure on $(\Omega, \mathcal{F})$ and $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ for some $p \in [1, +\infty)$. Consider now the function

$$\rho(X) := \langle\bar{\mu}, X\rangle + c\sigma_p(X), \tag{4.9}$$

where $c \ge 0$ and

$$\sigma_p(X) := \|[X - \langle\bar{\mu}, X\rangle]_+\|_p = \left(\int_\Omega [X(\omega) - \langle\bar{\mu}, X\rangle]_+^p\, d\bar{\mu}(\omega)\right)^{1/p} \tag{4.10}$$

is the upper semideviation of X of order p with $p \ge 1$. Note that $[a]_+^p := ([a]_+)^p$. The risk function (4.2) represents the mean-semideviation models analyzed in Ogryczak and Ruszczyński [16, 17].

The functions $\sigma_p(\cdot)$ and $\rho(\cdot)$ are convex, positively homogeneous, and continuous in the strong topology of $\mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$. Similarly to the analysis of Example 4.1, we need to consider only measures that are absolutely continuous with respect to $\bar{\mu}$, and can take $\mathcal{Y} := \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$. Moreover, the subdifferential of $\|[X]_+\|_p$, at $X = 0$, is equal to $\{h \in B_q\colon h \succeq 0\}$, where the notation $h \succeq 0$ means that $h(\omega) \ge 0$ for $\bar{\mu}$-almost every $\omega \in \Omega$. Consequently, in a way similar to the derivations of Example 4.1, it can be shown that the representation (4.6), for the function ρ, holds with the set $\mathcal{A}_p$, which can be written in the form

$$\mathcal{A}_p := \left\{g \in \mathcal{Y}\colon g = 1 + h - \int_\Omega h\, d\bar{\mu},\ \|h\|_q \le c,\ h \succeq 0\right\}. \tag{4.11}$$

Because $|\int_\Omega h\, d\bar{\mu}| \le \|h\|_q$ for any $h \in \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$, we have here that $\mathcal{A}_p$ is a set of probability measures if (and only if provided that condition (∗) holds) $c \in [0, 1]$.

Because here ρ is convex, positively homogeneous, and continuous, we have that for any $X \in \mathcal{X}$, the subdifferential $\partial\rho(X)$ is nonempty and is given by formula (3.4). That is, $\partial\rho(X) = \{1 + h - \int_\Omega h\, d\bar{\mu}\colon h \in \mathcal{D}_X\}$, where

$$\mathcal{D}_X := \arg\max_{h\in\mathcal{Y}}\left\{\int_\Omega \left(X - \int_\Omega X\, d\bar{\mu}\right) h\, d\bar{\mu}\colon \|h\|_q \le c,\ h \succeq 0\right\}. \tag{4.12}$$

The set $\mathcal{D}_X$ can be described as follows. Consider the functions $Y(\cdot) := X(\cdot) - \int_\Omega X\, d\bar{\mu}$ and $Y_+(\cdot) := \max\{Y(\cdot), 0\}$ and the set $A_X := \{\omega \in \Omega\colon Y(\omega) > 0\}$. With $Y_+ \in \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$, we associate a (dual) point $h_X^* \in B_q$ such that $\|Y_+\|_p = \langle Y_+, h_X^*\rangle$. The point h_X^* is a maximizer of $\langle Y_+, h\rangle$ over $h \in B_q$, and hence $\|h_X^*\| = 1$ unless $Y_+ = 0$. If the function $X(\cdot)$ is constant, then $Y(\cdot) \equiv 0$ and $\partial\rho(X) = \mathcal{A}_p$. So suppose that $X(\cdot)$ is not constant (this and similar subsequent statements should be understood, of course, up to a set of $\bar{\mu}$-measure zero), and hence the set A_X has a positive $\bar{\mu}$-measure.

Consider the case of $1 < p < +\infty$. In that case, the dual point h_X^* is unique, $h_X^* \succeq 0$ and $h_X^*(\omega) = 0$ for all $\omega \in \Omega \backslash A_X$. It follows that $\mathscr{D}_X = \{ch_X^*\}$. We obtain that $\mathscr{D}_X$ is a singleton, and hence ρ is Hadamard differentiable at X, for every nonconstant $X \in \mathscr{X}$.

Suppose now that $p = 1$. Then, B_q with $q = +\infty$ is formed by $h \in \mathscr{Y}$ such that $|h(\omega)| \leq 1$ for $\bar{\mu}$-almost every $\omega \in \Omega$. In that case, we have that if $h \in \mathscr{D}_X$, then $h(\omega) = 0$ for every $\omega \in \Omega$ such that $Y(\omega) < 0$. Also, h_X^* is a dual point of Y_+ iff $h_X^*(\omega) = 1$ for $\omega \in A_X$ and $|h_X^*(\omega)| \leq 1$ for $\omega \in \Omega \backslash A_X$. We obtain that

$$\mathscr{D}_X = \big\{h \in cB_q\colon\ h(\omega) = c \text{ if } Y(\omega) > 0,\ h(\omega) = 0 \text{ if } Y(\omega) < 0\big\}. \tag{4.13}$$

It follows that $\mathscr{D}_X$ is a singleton iff $Y(\omega) \neq 0$ for $\bar{\mu}$-almost every $\omega \in \Omega$.

Example 4.3 (Conditional Value at Risk). Let $\bar{\mu}$ be a probability measure on $(\Omega, \mathscr{F})$ and consider spaces $\mathscr{X} := \mathscr{L}_1(\Omega, \mathscr{F}, \bar{\mu})$ and $\mathscr{Y} := \mathscr{L}_\infty(\Omega, \mathscr{F}, \bar{\mu})$. For constants $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, consider the function $\rho(X) := \langle \bar{\mu}, X\rangle + \phi(X)$, where

$$\begin{aligned}\phi(X) &:= \inf_{z\in\mathbb{R}} \int_\Omega \{\varepsilon_1[z - X(\omega)]_+ + \varepsilon_2[X(\omega) - z]_+\}\, d\bar{\mu}(\omega)\\ &= \inf_{z\in\mathbb{R}} \int_{-\infty}^{+\infty} \{\varepsilon_1[z - x]_+ + \varepsilon_2[x - z]_+\}\, dG(x),\end{aligned} \tag{4.14}$$

and $G(x) := \bar{\mu}(\{\omega\colon X(\omega) \leq x\})$ is the cumulative distribution function of $X(\omega)$ with respect to the probability measure $\bar{\mu}$. It can be noted that the infimum at the right-hand side of (4.14) is attained at any $\bar{z}$ such that $\bar{\mu}[X \leq \bar{z}] \geq p$ and $\bar{\mu}[X \geq \bar{z}] \geq 1 - p$, where

$$p := \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} = 1 - \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2};$$

the point $\bar{z}$ is called a p-quantile of the cdf $G(x)$. Note also that

$$\varepsilon_1[z - X]_+ + \varepsilon_2[X - z]_+ = \varepsilon_1(z + (1-p)^{-1}[X - z]_+ - X).$$

Consequently,

$$\rho(X) = (1 - \varepsilon_1)\langle \bar{\mu}, X\rangle + \varepsilon_1 \mathrm{CVaR}_p[X], \tag{4.15}$$

where

$$\mathrm{CVaR}_p[X] := \inf_{z\in\mathbb{R}} \left\{ z + \frac{1}{1-p} \int_{-\infty}^{+\infty} [x - z]_+\, dG(x) \right\}. \tag{4.16}$$

The quantity (4.16) was called the conditional value at risk in Rockafellar and Uryasev [23]. It is the financial counterpart of the function of the *integrated chance constraint* introduced in Klein Haneveld [10]. Both are special cases of the classical concept of the absolute Lorenz curve, evaluated at point p, Lorenz [12], Ogryczak and Ruszczyński [18]. A risk envelope representation of CVaR has been developed in Rockafellar et al. [26] and Shapiro and Ahmed [28].

We have that

$$\rho(X) = \sup_{\mu\in\mathscr{A}} \langle \mu, X\rangle, \tag{4.17}$$

where $\gamma_1 := 1 - \varepsilon_1$ and $\gamma_2 := 1 + \varepsilon_2$, and

$$\mathscr{A} := \left\{ h \in \mathscr{Y}\colon\ \gamma_1 \leq h(\omega) \leq \gamma_2,\ \text{a.e. } \omega \in \Omega,\ \int_\Omega h\, d\bar{\mu} = 1 \right\}. \tag{4.18}$$

Let us observe that the set $\mathscr{A}$ is a set of probability measures if $\varepsilon_1 \leq 1$. This shows that for $\varepsilon_1 \in (0, 1]$ and $\varepsilon_2 > 0$, the function ρ is a risk function.

Similarly to the previous example, ρ is subdifferentiable at every $X \in \mathscr{X}$ and

$$\partial\rho(X) = \operatorname*{arg\,max}_{h\in\mathscr{Y}} \left\{ \int_\Omega Xh\, d\bar{\mu}\colon\ \gamma_1 \leq h(\omega) \leq \gamma_2,\ \text{a.e. } \omega \in \Omega,\ \int_\Omega h\, d\bar{\mu} = 1 \right\}. \tag{4.19}$$

Moreover, ρ is Hadamard differentiable at X iff the "arg max" set in the right-hand side of (4.19) is a singleton.

Let us consider the maximization problem in the right-hand side of (4.19). We can write it in the max-min form

$$\max_{\gamma_1 \leq h(\cdot) \leq \gamma_2}\ \inf_{\lambda\in\mathbb{R}} \left\{ \int_\Omega (X - \lambda) h\, d\bar{\mu} + \lambda \right\}.$$

Because $0 < \gamma_1 < \gamma_2$, by interchanging the "min" and "max" operators in the last problem, we obtain that it is equivalent to

$$\operatorname*{Min}_{\lambda\in\mathbb{R}}\left\{\int_\Omega \max[\gamma_1(X-\lambda),\gamma_2(X-\lambda)]\,d\bar{\mu}+\lambda\right\}. \tag{4.20}$$

Let $\bar{\lambda}$ be an optimal solution of (4.20). Considering the left- and right-side derivatives, at $\bar{\lambda}$, of the objective function in (4.20), we obtain that

$$1-\gamma_1\bar{\mu}\{X<\bar{\lambda}\}-\gamma_2\bar{\mu}\{X\geq\bar{\lambda}\}\leq 0\leq 1-\gamma_1\bar{\mu}\{X\leq\bar{\lambda}\}-\gamma_2\bar{\mu}\{X>\bar{\lambda}\}.$$

This can be rewritten as follows:

$$\varepsilon_1\bar{\mu}\{X<\bar{\lambda}\}-\varepsilon_2\bar{\mu}\{X\geq\bar{\lambda}\}\leq 0\leq \varepsilon_1\bar{\mu}\{X\leq\bar{\lambda}\}-\varepsilon_2\bar{\mu}\{X>\bar{\lambda}\}.$$

Recalling that $p=\varepsilon_2/(\varepsilon_1+\varepsilon_2)$, we conclude that the set of optimal solutions of (4.20) is the set of p-quantiles of the cdf $G(\cdot)$. Suppose for simplicity that the p-quantile $\bar{\lambda}$ is defined uniquely. Then, the "arg max" set in (4.19) is given by such $h(\omega)$ that

$$\begin{aligned} &h(\omega)=\gamma_2 \quad \text{if } X(\omega)>\bar{\lambda},\\ &h(\omega)=\gamma_1 \quad \text{if } X(\omega)<\bar{\lambda},\\ &h(\omega)\in[\gamma_1,\gamma_2] \quad \text{if } X(\omega)=\bar{\lambda}, \qquad \text{and}\\ &\int_\Omega h\,d\bar{\mu}=1. \end{aligned} \tag{4.21}$$

It follows that the "arg max" set in (4.19) is a singleton, and ρ is Hadamard differentiable at X, iff the system (4.21) has a unique solution h. This is equivalent to the following statement:

$$\bar{\mu}\{X<\bar{\lambda}\}=p \qquad \text{or} \qquad \bar{\mu}\{X>\bar{\lambda}\}=1-p. \tag{4.22}$$

If the quantile $\bar{\lambda}$ is not unique, then the set of p-quantiles has $\bar{\mu}$-measure zero, and thus the differentiability condition (4.22) can be understood as holding for any (or for all) p-quantiles. In summary, ρ is Hadamard differentiable at X iff condition (4.22) holds for a p-quantile $\bar{\lambda}$. Note that condition (4.22) always holds true if the set $\{\omega\in\Omega\colon X(\omega)=\bar{\lambda}\}$ has $\bar{\mu}$-measure zero, but may also hold when this set has a positive $\bar{\mu}$-measure.

In particular, for $\varepsilon_1=1$, we have that $\rho(\cdot)=\mathrm{CVaR}_p[\cdot]$. Therefore $\mathrm{CVaR}_p[X]$ is equal to the right-hand side of (4.17) for

$$\mathcal{A}:=\left\{h\in\mathcal{Y}\colon 0\leq h(\omega)\leq(1-p)^{-1},\ \text{a.e. }\omega\in\Omega,\ \int_\Omega h\,d\bar{\mu}=1\right\}. \tag{4.23}$$

The dual representation and formulas for the set $\mathcal{A}$ and the subdifferential of $\mathrm{CVaR}_p[X]$ were derived in Rockafellar et al. [26, Examples 12 and 20] in the space $\mathcal{X}:=\mathcal{L}_2(\Omega,\mathcal{F},\bar{\mu})$.

Many other examples of risk functions, including examples of nonhomogeneous risk functions, can be found in Ruszczyński and Shapiro [27].

5. Risk-averse functions.

5.1. General properties. Let $\bar{\mu}$ be a (reference) probability measure on $(\Omega,\mathcal{F})$, $\mathcal{X}$ be a linear space of $\bar{\mu}$-integrable functions, and $\mathcal{Y}\subset\mathbb{Y}$ be a dual space of measures. Unless stated otherwise, we assume in this section, that all expectations (conditional expectations) are taken with respect to the reference measure $\bar{\mu}$. We also assume that if $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$ and $X\in\mathcal{X}$, then $\mathbb{E}[X\mid\mathcal{G}]\in\mathcal{X}$. Then, for a σ-algebra $\mathcal{G}\subset\mathcal{F}$, we can consider the mapping

$$P_{\mathcal{G}}(\cdot):=\mathbb{E}[\cdot\mid\mathcal{G}]\colon \mathcal{X}\to\mathcal{X}. \tag{5.1}$$

Note that $P_{\mathcal{G}}$ is a projection onto the subspace of $\mathcal{X}$ formed by $\mathcal{G}$-measurable functions. Note also that the conditional expectation $\mathbb{E}[X\mid\mathcal{G}]$ is defined up to a set of $\bar{\mu}$-measure zero. That is, any two versions of $\mathbb{E}[X\mid\mathcal{G}](\omega)$ are equal for almost every $\omega\in\Omega$. Unless stated otherwise, we assume in the subsequent analysis that a considered property holds for every version of $\mathbb{E}[X\mid\mathcal{G}]$.

Definition 5.1. We say that a risk function $\rho\colon\mathcal{X}\to\overline{\mathbb{R}}$ is *risk averse* (with respect to $\bar{\mu}$) if for every σ-algebra $\mathcal{G}\subset\mathcal{F}$, the following inequality holds:

$$\rho(X)\geq\rho(P_{\mathcal{G}}(X)) \quad \text{for all } X\in\mathcal{X}. \tag{5.2}$$

With every risk function $\rho\colon \mathscr{X} \to \overline{\mathbb{R}}$ is associated its conjugate function $\rho^*\colon \mathscr{Y} \to \overline{\mathbb{R}}$. We also say that ρ^* is *risk averse*, if for every σ-algebra $\mathscr{G} \subset \mathscr{F}$, the following inequality holds:

$$\rho^*(\mu) \geq \rho^*(P^*_{\mathscr{G}}(\mu)) \quad \text{for all } \mu \in \mathscr{Y}. \tag{5.3}$$

Here, $P^*_{\mathscr{G}}\colon \mathscr{Y} \to \mathscr{Y}$ denotes the adjoint of the operator $P_{\mathscr{G}}$. Recall that $P^*_{\mathscr{G}}$ is defined by the equation $\langle \mu, P_{\mathscr{G}}(X)\rangle = \langle P^*_{\mathscr{G}}(\mu), X\rangle$ for all $X \in \mathscr{X}$ and $\mu \in \mathscr{Y}$. In particular, if $\mathscr{G} = \{\varnothing, \Omega\}$, then $P_{\mathscr{G}}(\cdot) = \mathbb{E}[\cdot]$, and hence $P^*_{\mathscr{G}}(\mu) = a_\mu \bar{\mu}$, where $a_\mu := \mu(\Omega)$.

Suppose for the moment that every measure $\mu \in \mathscr{Y}$ is absolutely continuous with respect to the reference measure $\bar{\mu}$, i.e., $d\mu = h\,d\bar{\mu}$, and the corresponding density $h(\omega)$ is $\bar{\mu}$-integrable. In that case, we can identify $\mathscr{Y}$ with the corresponding linear space of $\bar{\mu}$-integrable functions. Take, for example, $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$ and $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$ for some $p \in [1, +\infty)$ and $1/p + 1/q = 1$. Note that ρ^* does not satisfy the monotonicity property, and also $\rho^*(h + a) = \rho^*(h)$ for $h \in \mathscr{Y}$ and $a \in \mathbb{R}$. In this way, the conjugate risk function is a deviation measure of Rockafellar et al. [26].

We have that for $X \in \mathscr{X}$ and $h \in \mathscr{Y}$,

$$\begin{aligned}\langle h, P_{\mathscr{G}}(X)\rangle &= \int_\Omega P_{\mathscr{G}}(X) h\, d\bar{\mu} = \mathbb{E}[hP_{\mathscr{G}}(X)] = \mathbb{E}[\mathbb{E}[hP_{\mathscr{G}}(X) \mid \mathscr{G}]] \\ &= \mathbb{E}[P_{\mathscr{G}}(X)\mathbb{E}[h \mid \mathscr{G}]] = \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]\mathbb{E}[h \mid \mathscr{G}]] = \mathbb{E}[\mathbb{E}[X\mathbb{E}[h \mid \mathscr{G}] \mid \mathscr{G}]] \\ &= \mathbb{E}[X\mathbb{E}[h \mid \mathscr{G}]] = \langle \mathbb{E}[h \mid \mathscr{G}], X\rangle.\end{aligned}$$

It follows that

$$P^*_{\mathscr{G}}(\cdot) = \mathbb{E}[\cdot \mid \mathscr{G}]. \tag{5.4}$$

Proposition 5.1. *Let ρ be a risk function satisfying assumptions* (A1)–(A3). *Suppose that ρ is lower semicontinuous. Then, ρ is risk averse iff ρ^* is risk averse.*

Proof. Consider a σ-algebra $\mathscr{G} \subset \mathscr{F}$. Suppose that ρ^* is risk averse. By Theorem 2.1, we have

$$\rho(P_{\mathscr{G}}(X)) = \sup_{\mu \in \mathscr{Y}} \{\langle \mu, P_{\mathscr{G}}(X)\rangle - \rho^*(\mu)\}.$$

Because $\langle \mu, P_{\mathscr{G}}(X)\rangle = \langle P^*_{\mathscr{G}}(\mu), X\rangle$ and because of (5.3), it follows that

$$\rho(P_{\mathscr{G}}(X)) \leq \sup_{\mu \in \mathscr{Y}} \{\langle P^*_{\mathscr{G}}(\mu), X\rangle - \rho^*(P^*_{\mathscr{G}}(\mu))\}.$$

By making a change of variables $\nu = P^*_{\mathscr{G}}(\mu)$, we obtain

$$\rho(P_{\mathscr{G}}(X)) \leq \sup_{\nu \in \mathscr{Y}} \{\langle \nu, X\rangle - \rho^*(\nu)\} = \rho(X).$$

The converse implication can be proved similarly. □

If the risk function ρ is positively homogeneous, then its conjugate function ρ^* is the indicator function of a set $\mathscr{A} \subset \mathscr{P}$, which can be written in the form (2.8). In that case, we have that ρ^*, and hence ρ, is risk averse iff for every σ-algebra $\mathscr{G} \subset \mathscr{F}$, the following relation holds true:

$$P^*_{\mathscr{G}}(\mathscr{A}) \subseteq \mathscr{A}. \tag{5.5}$$

In particular, for $\mathscr{G} = \{\varnothing, \Omega\}$ and $\mu \in \mathscr{A}$, we have that $P^*_{\mathscr{G}}(\mu) = \bar{\mu}$, and hence it follows from (5.5) that $\bar{\mu}$ should be an element of $\mathscr{A}$. Recall that if $\mathscr{X} := \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$ and $\mathscr{Y} := \mathscr{L}_q(\Omega, \mathscr{F}, \bar{\mu})$, then $P^*_{\mathscr{G}}(\cdot) = \mathbb{E}[\cdot \mid \mathscr{G}]$. Therefore, in that case, ρ^* and ρ are risk averse iff for any σ-algebra $\mathscr{G} \subseteq \mathscr{F}$, the following holds:

$$\mathbb{E}[h \mid \mathscr{G}] \in \mathscr{A}, \quad \forall\, h \in \mathscr{A}. \tag{5.6}$$

It follows that the function $h(\cdot) \equiv 1$ must be an element of $\mathscr{A}$.

Let us return to Example 4.2 and consider the function ρ defined in (4.9). We know that this function can be represented as

$$\rho(X) = \sup_{g \in \mathscr{A}_p} \langle g, X\rangle,$$

with the set $\mathcal{A}_p$ given in (4.11). Consider an element $g \in \mathcal{A}_p$. By (4.11), we have that $g = 1 + h - \mathbb{E}[h]$ for some $h \in \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$ such that $\|h\|_q \le c$ and $h(\omega) \ge 0$ for a.e. $\omega \in \Omega$. Because $\mathbb{E}[h] = \mathbb{E}[\mathbb{E}[h \mid \mathcal{G}]]$, it follows that

$$P^*_{\mathcal{G}}(g) = \mathbb{E}[g \mid \mathcal{G}] = 1 + \mathbb{E}[h \mid \mathcal{G}] - \mathbb{E}[\mathbb{E}[h \mid \mathcal{G}]].$$

Moreover, $\|\mathbb{E}[h \mid \mathcal{G}]\|_q \le \|h\|_q$ and $\mathbb{E}[h \mid \mathcal{G}](\omega) \ge 0$ for a.e. $\omega \in \Omega$. Thus condition (5.6) is satisfied, and hence ρ is risk averse. Similar considerations apply to Example 4.1.

Consider now the risk function $\rho(\cdot) := \mathrm{CVaR}_p[\cdot]\colon \mathcal{X} \to \mathbb{R}$ discussed in Example 4.3. Here, $\mathcal{X} := \mathcal{L}_1(\Omega, \mathcal{F}, \bar{\mu})$ and $p \in (0, 1)$. It immediately follows from the description (4.23) of the corresponding set $\mathcal{A}$ that condition (5.6) is satisfied, and hence ρ is risk averse. It follows then that the function ρ defined in (4.15) is also risk averse for any $\varepsilon_1 \in [0, 1]$.

By using (A3) and setting $\mathcal{G} = \{\varnothing, \Omega\}$, we obtain that a risk-averse function ρ satisfies the inequality $\rho(X) \ge \langle \bar{\mu}, X \rangle$ for all $X \in \mathcal{X}$. This property of risk aversion is related to the classical Jensen's inequality for the expected value of a convex function, but it is not implied by the convexity of the risk function. For example, relation (5.5) is not implied by the convexity of the set $\mathcal{A}$.

5.2. Consistency with stochastic orders. In all examples considered in §4, the space $\mathcal{X}$ was given by $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $\mathcal{X}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$ and, moreover, the risk functions $\rho(X)$ discussed there were dependent only on the distribution of X. That is, each risk function $\rho(\cdot)$ considered in §4, could be formulated in terms of the cumulative distribution function (cdf) $F_X(t) := P(X \le t)$ associated with $X \in \mathcal{X}$. We call such risk functions distribution invariant.

Definition 5.2. We say that risk function $\rho\colon \mathcal{X} \to \mathbb{R}$ is *distribution invariant*, with respect to reference measure P, if the following condition holds: if $X_1, X_2 \in \mathcal{Z}$ are such that $P(X_1 \le t) = P(X_2 \le t)$ for all $t \in \mathbb{R}$, then $\rho(X_1) = \rho(X_2)$.

For distribution-invariant risk functions, it makes sense to discuss their monotonicity properties with respect to various stochastic orders defined for (real-valued) random variables.

Many stochastic orders can be characterized by a class $\mathcal{U}$ of functions $u\colon \mathbb{R} \to \mathbb{R}$ as follows. For (real-valued) random variables X_1 and X_2, it is said that X_2 dominates X_1, denoted $X_2 \succeq_{\mathcal{U}} X_1$, if $\mathbb{E}[u(X_2)] \ge \mathbb{E}[u(X_1)]$ for all $u \in \mathcal{U}$ for which the corresponding expectations do exist. This stochastic order is called the *integral stochastic order* with *generator* $\mathcal{U}$. We refer to Müller and Stoyan [15, Chapter 2] for a thorough discussion of this concept. For example, the *usual stochastic order*, written $X_2 \succeq_{\mathrm{st}} X_1$, corresponds to the generator $\mathcal{U}$ formed by all nondecreasing functions $u\colon \mathbb{R} \to \mathbb{R}$. It is possible to show that $X_2 \succeq_{\mathrm{st}} X_1$ iff $F_{X_2}(t) \le F_{X_1}(t)$ for all $t \in \mathbb{R}$ (e.g., Müller and Stoyan [15, Theorem 1.2.8]). We say that the integral stochastic order is *increasing* if all functions in the set $\mathcal{U}$ are nondecreasing. The usual stochastic order is an example of increasing integral stochastic order.

We say that a distribution-invariant risk function ρ is *consistent* with the integral stochastic order if $X_2 \succeq_{\mathcal{U}} X_1$ implies $\rho(X_2) \ge \rho(X_1)$ for all $X_1, X_2 \in \mathcal{X}$, i.e., ρ is monotone with respect to $\succeq_{\mathcal{U}}$. For an increasing integral stochastic order, we have that if $X_2(\omega) \ge X_1(\omega)$ for a.e. $\omega \in \Omega$, then $u(X_2(\omega)) \ge u(X_1(\omega))$ for any $u \in \mathcal{U}$ and a.e. $\omega \in \Omega$, and hence $\mathbb{E}[u(X_2)] \ge \mathbb{E}[u(X_1)]$. That is, if $X_2 \succeq X_1$ in the almost sure sense, then $X_2 \succeq_{\mathcal{U}} X_1$. It follows that if ρ is distribution invariant and consistent with respect to an increasing integral stochastic order, then it satisfies the monotonicity condition (A2). In other words, if ρ does not satisfy condition (A2), then it cannot be consistent with any increasing integral stochastic order. In particular, for $p > 1$ the mean-deviation risk function, defined in (4.1)–(4.2), is not consistent with any increasing integral stochastic order, provided that condition $(*)$, introduced in Example 4.1, holds true.

A general way of proving consistency of distribution-invariant risk functions with stochastic orders can be obtained via the following construction. For a given pair of random variables X_1 and X_2 in $\mathcal{X}$, consider another pair of random variables, $\widehat{X}_1$ and $\widehat{X}_2$, which have identical distributions as the original pair, denoted $\widehat{X}_1 \stackrel{\mathrm{D}}{\sim} X_1$ and $\widehat{X}_2 \stackrel{\mathrm{D}}{\sim} X_2$. The construction is such that the postulated consistency result becomes evident. For this method to be applicable, the probability space $(\Omega, \mathcal{F}, P)$ must be sufficiently rich. Therefore we assume in the remainder of this section that a uniform random variable exists on the space $(\Omega, \mathcal{F}, P)$; i.e., there exists a measurable function $U\colon \Omega \to \mathbb{R}$ such that $P(U \le t) = t$ for all $t \in [0, 1]$. For a cdf $F(t)$, its inverse is defined as

$$F^{-1}(t) := \inf\{s\colon F(s) \ge t\}.$$

We also assume that $\mathcal{X} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for some $p \in [1, \infty)$.

Lemma 5.1. *If the risk function ρ is distribution invariant, then it is consistent with the usual stochastic order iff it satisfies condition* (A2).

Proof. By the discussion preceding the lemma, it is sufficient to prove that (A2) implies consistency with the usual stochastic order.

Suppose that the risk function ρ is distribution invariant and satisfies the monotonicity condition (A2). Recall that $X_2 \succeq_{\rm st} X_1$ iff $F_{X_2}(t) \le F_{X_1}(t)$ for all $t \in \mathbb{R}$. For a uniform random variable $U(\omega)$, consider the random variables $\hat{X}_1 := F_{X_1}^{-1}(U)$ and $\hat{X}_2 := F_{X_2}^{-1}(U)$. We obtain that if $X_2 \succeq_{\rm st} X_1$, then $\hat{X}_2(\omega) \ge \hat{X}_1(\omega)$ for all $\omega \in \Omega$, and hence by virtue of (A2), $\rho(\hat{X}_2) \ge \rho(\hat{X}_1)$. By construction, $\hat{X}_1 \stackrel{\rm D}{\sim} X_1$ and $\hat{X}_2 \stackrel{\rm D}{\sim} X_2$. Because the risk function is distribution invariant, we conclude that $\rho(X_2) \ge \rho(X_1)$. Consequently, the risk function ρ is consistent with the usual stochastic order. □

It is said that X_2 is bigger than X_1 in the *increasing convex order*, written $X_2 \succeq_{\rm icx} X_1$, if $\mathbb{E}[u(X_2)] \ge \mathbb{E}[u(X_1)]$ for all increasing convex functions u: $\mathbb{R} \to \mathbb{R}$ such that the expectations exist. Clearly, this is an integral stochastic order with the corresponding generator given by the set of increasing convex functions. It is the counterpart of the classical stochastic dominance relation, which is the increasing concave order (recall that we are dealing here with minimization rather than maximization problems).

Theorem 5.1. *If the risk function ρ is distribution invariant, then it is consistent with the increasing convex order iff it satisfies condition* (A2) *and is risk averse.*

Proof. Suppose that the risk function ρ is consistent with the increasing convex order. Then, it is consistent with respect to the usual stochastic order. By Lemma 5.1, it satisfies (A2). Next, for every random variable X and every σ-algebra $\mathcal{G} \subset \mathcal{F}$, we have by Jensen's inequality that

$$\mathbb{E}[u(\mathbb{E}[X \mid \mathcal{G}])] \le \mathbb{E}[\mathbb{E}[u(X) \mid \mathcal{G}]] = \mathbb{E}[u(X)],$$

for any convex function $u(\cdot)$. Consequently, $\mathbb{E}[X \mid \mathcal{G}] \preceq_{\rm icx} X$. Because ρ is consistent with the increasing convex order, $\rho(X) \ge \rho(P_{\mathcal{G}}(X))$. It follows that the risk function ρ is risk averse.

Assume now that ρ is distribution invariant, risk averse, and satisfies (A2). If $X_1 \preceq_{\rm icx} X_2$, by Müller and Stoyan [15, Corollary 1.5.21] (see also Blackwell [3], Strassen [29]), we can find $\hat{X}_1 \stackrel{\rm D}{\sim} X_1$ and $\hat{X}_2 \stackrel{\rm D}{\sim} X_2$ such that

$$\hat{X}_1 \le \mathbb{E}[\hat{X}_2 \mid \hat{X}_1] \quad \text{a.s.}$$

By using the monotonicity and risk aversion of ρ, we conclude that

$$\rho(X_1) = \rho(\hat{X}_1) \le \rho(\mathbb{E}[\hat{X}_2 \mid \hat{X}_1]) \le \rho(\hat{X}_2) = \rho(X_2).$$

This means that ρ is consistent with the increasing convex order. □

The risk functions of Examples 4.2 and 4.3, with $c \in [0, 1]$, are monotone and risk averse, and therefore they are consistent with the increasing convex order. The consistency results of Ogryczak and Ruszczyński [16, 17, 18] and Pflug [20] for specific mean-risk models are special cases of Theorem 5.1.

6. Optimality conditions for optimization problems with risk functions. Let $\mathcal{Z}$ be a vector space and consider a mapping F: $\mathcal{Z} \to \mathcal{X}$. As in §3, we write $f(z, \omega)$ or $f_\omega(z)$ for $[F(z)](\omega)$, and view $f(z, \omega)$ as a random function defined on the measurable space $(\Omega, \mathcal{F})$. Consider the problem

$$\operatorname*{Min}_{z \in S} \{\psi(z) := \rho(F(z))\}, \tag{6.1}$$

where S is a nonempty convex subset of $\mathcal{Z}$ and ρ: $\mathcal{X} \to \overline{\mathbb{R}}$ is a risk function.

Suppose that the mapping F: $\mathcal{Z} \to \mathcal{X}$ is convex and the function ρ: $\mathcal{X} \to \overline{\mathbb{R}}$ is proper and lower semicontinuous, and that assumptions (A1)–(A3) are satisfied. By Theorem 2.2, we can use representation (2.6) to write problem (6.1) in the form

$$\operatorname*{Min}_{z \in S} \sup_{\mu \in \mathcal{P}} \Xi(z, \mu), \tag{6.2}$$

where the function Ξ: $\mathcal{Z} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is defined by

$$\Xi(z, \mu) := \int_\Omega f(z, \omega)\, d\mu(\omega) - \rho^*(\mu). \tag{6.3}$$

As we mentioned earlier, under the above assumptions, the function $\rho^*(\cdot)$ is also proper. For every $\mu \in \mathcal{P}$, the function $\Xi(\cdot, \mu)$ is convex and if, moreover, μ is in the domain of $\rho^*(\cdot)$, then $\Xi(\cdot, \mu)$ is real valued. For every $z \in S$, the function $\Xi(z, \cdot)$ is concave. Therefore, under various regularity conditions, the "min" and "sup" operators in (6.2) can be interchanged. When $\mathcal{Z} = \mathbb{R}^n$, a sufficient condition for such interchangeability is that problem (6.1) has a nonempty and bounded set of optimal solutions. We obtain the following result.

Proposition 6.1. *Suppose that $\mathcal{Z}=\mathbb{R}^n$, the mapping $F\colon \mathbb{R}^n\to\mathcal{X}$ is convex, and the function $\rho\colon \mathcal{X}\to\overline{\mathbb{R}}$ is proper, lower semicontinuous, and assumptions* (A1)–(A3) *are satisfied. Suppose, further, that problem* (6.1) *has a nonempty and bounded set of optimal solutions. Then, the optimal value of problem* (6.1) *is equal to the optimal value of the problem*

$$\operatorname*{Max}_{\mu\in\mathcal{P}}\inf_{z\in S}\left\{\int_\Omega f(z,\omega)\,d\mu(\omega)-\rho^*(\mu)\right\}. \tag{6.4}$$

If $\mathcal{X}$ is a Banach space and $\mathcal{Y}=\mathcal{X}^*$, a similar statement can be obtained for a general vector space $\mathcal{Z}$.

Proposition 6.2. *Suppose that $\mathcal{X}$ is a Banach space, $\mathcal{Y}=\mathcal{X}^*$, the mapping $F\colon\mathcal{Z}\to\mathcal{X}$ is convex, the function $\rho\colon\mathcal{X}\to\overline{\mathbb{R}}$ is proper, lower semicontinuous, and assumptions* (A1)–(A3) *are satisfied. Then, the optimal value of problem* (6.1) *is equal to the optimal value of problem* (6.4). *Moreover, problem* (6.4) *has an optimal solution.*

Proof. The set $\mathcal{P}$ is bounded, convex, and weakly* closed in $\mathcal{X}^*$. By the Banach–Alaoglu theorem, it is weakly* compact in $\mathcal{X}^*$. The function $\Xi(z,\cdot)$ is concave and weakly* continuous, and the function $\Xi(\cdot,\mu)$ is convex. Our assertion then follows from the asymmetric min-max theorem (see, e.g., Aubin and Ekeland [2, Theorem 6.2.7]). □

Let $\Xi(z,\mu)$ be the function defined in (6.3). Suppose that the assumptions of Propositions 6.1 or 6.2 hold true. Consider elements $\hat z\in S$ and $\hat\mu\in\mathcal{P}$. Because the optimal values of (6.1) and (6.4) are equal, we have that $\hat z$ is an optimal solution of (6.1) and $\hat\mu$ is an optimal solution of problem (6.4) iff $(\hat z,\hat\mu)$ is a saddle point of $\Xi(z,\mu)$. Note that because $\hat\mu$ is a probability measure, the integral $\int_\Omega f(z,\omega)\,d\hat\mu(\omega)$ can be written as the expectation $\mathbb{E}_{\hat\mu}[F(z)]$. It follows that if $\hat z$ is an optimal solution of problem (6.1) and $\hat\mu$ is an optimal solution of problem (6.4), then $\hat z$ is an optimal solution of the problem

$$\operatorname*{Min}_{z\in S}\{\mathbb{E}_{\hat\mu}[F(z)]-\rho^*(\hat\mu)\}. \tag{6.5}$$

That is, problem (6.1) is "almost" equivalent to the optimization problem (6.5) in the sense that the set of optimal solutions of problem (6.5) contains the set of optimal solutions of problem (6.1) and their optimal values are equal to each other. If, moreover, ρ is positively homogeneous, then ρ^* is the indicator function of a set $\mathcal{A}\subset\mathcal{P}$. In that case, problem (6.2) takes the form

$$\operatorname*{Min}_{z\in S}\sup_{\mu\in\mathcal{A}}\mathbb{E}_\mu[F(z)], \tag{6.6}$$

and $\rho^*(\hat\mu)=0$ in problem (6.5).

If we cannot use Proposition 6.2 to ensure existence of an optimal solution of problem (6.4), we need additional conditions.

Proposition 6.3. *Let $F\colon\mathcal{Z}\to\mathcal{X}$ be convex and let $\rho\colon\mathcal{X}\to\overline{\mathbb{R}}$ satisfy conditions* (A1)–(A3)*. Suppose that a point $\hat z\in S$ is an optimal solution of problem* (6.1) *and that $\rho(\cdot)$ is subdifferentiable at $\widehat X:=F(\hat z)$. Then, there exists a measure $\hat\mu\in\partial\rho(\widehat X)$ such that $\hat z$ is an optimal solution of problem* (6.5).

Proof. Since ρ is subdifferentiable at $\widehat X$, we have that $\partial\rho^{**}(\widehat X)=\partial\rho(\widehat X)$, and hence it follows by (3.3) that

$$\partial\rho(\widehat X)=\arg\max_{\mu\in\mathcal{Y}}\{\langle\mu,\widehat X\rangle-\rho^*(\mu)\}. \tag{6.7}$$

Of course, the maximum in (6.7) over $\mu\in\mathcal{Y}$ can be replaced by the maximum over $\mu\in\operatorname{dom}(\rho^*)$, and we have that $\operatorname{dom}(\rho^*)\subset\mathcal{P}$. Therefore

$$\partial\rho(\widehat X)=\arg\max_{\mu\in\mathcal{P}}\Xi(\hat z,\mu). \tag{6.8}$$

By the monotonicity of $\rho(\cdot)$, problem (6.1) is equivalent to the problem

$$\min_{(z,X)\in U}\rho(X) \tag{6.9}$$

with

$$U:=\{(z,X)\in\mathcal{Z}\times\mathcal{X}\colon X\succeq F(z),\ z\in S\}.$$

Since $\hat z$ is an optimal solution of (6.1), the pair $(\hat z,\widehat X)$ constitutes an optimal solution of (6.9). By the convexity of F, the set U is convex. The optimality of $(\hat z,\widehat X)$ and the subdifferentiability of $\rho(\cdot)$ imply that there exists a subgradient $\hat\mu\in\partial\rho(\widehat X)$ such that

$$\langle\hat\mu,X-\widehat X\rangle\ge 0\quad\text{for all }(z,X)\in U.$$

In particular, setting $X = F(z)$, we obtain that

$$\langle \hat{\mu}, F(z) - F(\hat{z}) \rangle \geq 0 \quad \text{for all } z \in S.$$

Thus $\hat{z}$ is a solution of problem (6.5). □

The condition of the subdifferentiability of $\rho(\cdot)$ at $F(\hat{z})$ has to be verified in each application by special methods. For example, if $\rho(\cdot)$ is continuous at $\overline{X} \in \mathcal{X}$, then it is subdifferentiable at $\overline{X}$. The risk functions in Examples 4.1, 4.2, and 4.3 are continuous, and therefore they are subdifferentiable everywhere.

Finally, we can formulate an equivalent condition to the saddle point conditions of Propositions 6.2 or 6.3. We use the notation

$$N_S(\bar{z}) := \{z^* \in \mathcal{Z}^*: \langle z - \bar{z}, z^* \rangle \leq 0, \forall z \in S\}$$

for the normal cone to the convex set S at $z \in S$.

Proposition 6.4. *Suppose that $\mathcal{Z}$ is a Banach space, the risk function ρ satisfies conditions* (A1)–(A3), *the set S and the mapping F are convex, and $\bar{z} \in X$ and $\bar{\mu} \in \mathcal{P}$ are such that $\mathbb{E}_{\bar{\mu}}[F(\cdot)]$ is continuous $\bar{z}$. Denote $\overline{X} := F(\bar{z})$. Then, $(\bar{z}, \bar{\mu})$ is a saddle point of $\Xi(z, \mu)$ iff*

$$0 \in N_S(\bar{z}) + \mathbb{E}_{\bar{\mu}}[\partial f_\omega(\bar{z})] \qquad \textit{and} \qquad \bar{\mu} \in \partial \rho(\overline{X}). \tag{6.10}$$

Proof. By the definition, $(\bar{z}, \bar{\mu})$ is a saddle point of $\Xi(z, \mu)$ iff

$$\bar{z} \in \arg\min_{z \in S} \Xi(z, \bar{\mu}) \qquad \text{and} \qquad \bar{\mu} \in \arg\max_{\mu \in \mathcal{P}} \Xi(\bar{z}, \mu). \tag{6.11}$$

The first of the above conditions means that $\bar{z} \in \arg\min_{z \in S} \varphi(z)$, where $\varphi(z) := \int_\Omega f(z, \omega)\, d\bar{\mu}(\omega)$. Because we deal here with a convex problem and $\varphi(\cdot)$ is continuous at $\bar{z}$, by the standard optimality conditions, this holds iff $0 \in N_S(\bar{z}) + \partial \varphi(\bar{z})$. Moreover, by Strassen's [29] theorem, we have $\partial \varphi(\bar{z}) = \mathbb{E}_{\bar{\mu}}[\partial f_\omega(\bar{z})]$. Therefore the first conditions in (6.10) and (6.11), respectively, are equivalent. The respective second conditions in (6.10) and (6.11) are equivalent by Equation (6.8). □

Under the assumptions of Proposition 6.4, conditions (6.10) can be viewed as optimality conditions for a point $\bar{z} \in S$ to be an optimal solution of problem (6.1). That is, if there exists a probability measure $\bar{\mu} \in \partial \rho(\overline{X})$ such that the first condition of (6.10) holds, then $\bar{z}$ is an optimal solution of problem (6.1); i.e., (6.10) are sufficient conditions for optimality. Moreover, under the assumptions of Proposition 6.2 or 6.3, the existence of such a probability measure $\bar{\mu}$ is a necessary condition for optimality of $\bar{z}$.

7. Value of perfect information. Let us consider, as in §6, a mapping $F: \mathcal{Z} \to \mathcal{X}$ and a set $S \subset \mathcal{Z}$. In the formulation of problem (6.1), we assume that the choice of $z \in S$ has to be made *before* the elementary event ω becomes known. It is of interest to consider also the opposite situation, when $z \in S$ is chosen *after* ω is known. The best value of $f(z, \omega)$, in this case, is equal to $\inf_{z \in S} f(z, \omega)$. We thus define the operation "inf" on the family $\{F(z): z \in S\}$ of elements of $\mathcal{X}$ as the pointwise infimum

$$\Big[\inf_{z \in S} F(z)\Big](\omega) := \inf\{f(z, \omega): z \in S\}, \quad \omega \in \Omega.$$

Suppose for a moment that $\inf_{z \in S} F(z)$ is an element of $\mathcal{X}$. Then, we can consider the value of the risk function ρ on this element. We define the *risk value of perfect information* as the difference

$$\text{RVPI}_\rho := \inf_{z \in S} \rho(F(z)) - \rho\Big(\inf_{z \in S} F(z)\Big). \tag{7.1}$$

Observe that for a risk function ρ satisfying condition (A2), RVPI_ρ is always nonnegative. Indeed, for every point $z' \in S$, we have that

$$f(z', \omega) \geq \inf_{z \in S} f(z, \omega), \quad \omega \in \Omega.$$

Using (A2), we obtain that

$$\rho(F(z')) \geq \rho\Big(\inf_{z \in S} F(z)\Big) \quad \text{for all } z' \in S.$$

Taking the infimum over $z' \in S$ yields $\text{RVPI}_\rho \geq 0$. The intuitive meaning of RVPI_ρ is the loss of the value of $\rho(\cdot)$ because of the fact that z cannot be made dependent on ω. Our intention is to estimate this quantity.

To avoid technical difficulties, we assume in this section, that $\mathcal{Z} := \mathbb{R}^n$ and $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$, with some $p \in [1, \infty]$. Assume that there exists $\phi \in \mathcal{X}$ such that $f(z, \omega) \ge \phi(\omega)$ for all $z \in S$ and all $\omega \in \Omega$. Then, for any $z' \in S$, we have that

$$\phi(\omega) \le \inf_{z \in S} f(z, \omega) \le f(z', \omega), \quad \omega \in \Omega.$$

Since the function $\inf_{z\in S} f(z, \omega)$ depends measurably on ω (e.g., Rockafellar and Wets [25, Theorem 14.37]), we conclude that $\inf_{z\in S} F(z) \in \mathcal{X}$. Thus (7.1) is well defined.

Together with (7.1), we consider the classical notion of the *expected value of perfect information* with respect to the probability measure μ:

$$\mathrm{EVPI}_\mu := \inf_{z\in S} \mathbb{E}_\mu[F(z)] - \mathbb{E}_\mu\Big[\inf_{z\in S} F(z)\Big]. \tag{7.2}$$

For the background on EVPI_μ, see Raiffa and Schlaifer [21].

Proposition 7.1. *Suppose that $\rho(\cdot)$ is lower semicontinuous and satisfies assumptions* (A1)–(A4). *Let $\mathcal{A} \subset \mathcal{P}$ be the convex set for which representation* (2.7) *of ρ holds true. If assumptions of Propositions 6.1 or 6.2 are satisfied, then*

$$\inf_{\mu\in\mathcal{A}} EVPI_\mu \le RVPI_\rho \le \sup_{\mu\in\mathcal{A}} EVPI_\mu. \tag{7.3}$$

Proof. It follows from Theorem 2.2 that

$$\mathrm{RVPI}_\rho = \inf_{z\in S} \sup_{\mu\in\mathcal{A}} \mathbb{E}_\mu[F(z)] - \sup_{\mu\in\mathcal{A}} \mathbb{E}_\mu\Big[\inf_{z\in S} F(z)\Big]. \tag{7.4}$$

Moreover, under the assumptions of Propositions 6.1 or 6.2, we have that

$$\inf_{z\in S} \sup_{\mu\in\mathcal{A}} \mathbb{E}_\mu[F(z)] = \sup_{\mu\in\mathcal{A}} \inf_{z\in S} \mathbb{E}_\mu[F(z)].$$

Substituting this into (7.4), we obtain

$$\mathrm{RVPI}_\rho = \sup_{\mu\in\mathcal{A}} \inf_{z\in S} \mathbb{E}_\mu[F(z)] - \sup_{\mu\in\mathcal{A}} \mathbb{E}_\mu\Big[\inf_{z\in S} F(z)\Big].$$

Therefore RVPI_ρ can be estimated from below and above as in (7.3). □

Another way of interpreting the risk value of perfect information is to view the hypothetical decision z, chosen after the elementary event ω is known, as a value of a function $Z\colon \Omega \to \mathcal{Z}$. To adapt this viewpoint, we need to assume a little more about the class of functions Z considered.

It is said that a linear space $\mathcal{M}$ of $\mathcal{F}$-measurable functions $Z\colon \Omega \to \mathbb{R}^n$ is *decomposable* if for every $Z \in \mathcal{M}$ and $B \in \mathcal{F}$, and every bounded and $\mathcal{F}$-measurable function $W\colon \Omega \to \mathbb{R}^n$, the space $\mathcal{M}$ also contains the function $V(\cdot) := \mathbb{1}_{\Omega\setminus B}(\cdot)Z(\cdot) + \mathbb{1}_B(\cdot)W(\cdot)$ (Rockafellar and Wets [25, p. 676]). Obviously, the spaces $\mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu}, \mathbb{R}^n)$, $p \in [1, \infty]$, of measurable functions $Z\colon \Omega \to \mathbb{R}^n$ are decomposable.

If $\mathcal{M}$ is a decomposable linear space of $\mathcal{F}$-measurable functions and $g\colon \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is a random lower semicontinuous function,[2] then for any probability measure μ on $(\Omega, \mathcal{F})$, the following interchangeability formula holds:

$$\int_\Omega \inf_{z\in\mathbb{R}^n} g(z, \omega)\, d\mu(\omega) = \inf_{Z\in\mathcal{M}} \int_\Omega g(Z(\omega), \omega)\, d\mu(\omega) \tag{7.5}$$

(Rockafellar and Wets [25, Theorem 14.60]). Thus the optimization after ω is known, and the optimization with respect to decision rules $Z \in \mathcal{M}$ are equivalent.

It is possible to extend this result to risk functions as follows. With $F\colon \mathbb{R}^n \to \mathcal{X}$ and $Z \in \mathcal{M}$, we associate an element $F_Z \in \mathcal{X}$ defined as follows:

$$F_Z(\omega) := [F(Z(\omega))](\omega) = f(Z(\omega), \omega).$$

We can then consider the problem

$$\operatorname*{Min}_{Z\in\mathcal{M}_S} \rho(F_Z), \tag{7.6}$$

where $\mathcal{M}_S := \{Z \in \mathcal{M}\colon Z(\omega) \in S,\ \forall\, \omega \in \Omega\}$.

[2] Random lower semicontinuous functions are also called *normal integrands* (see Rockafellar and Wets [25, Definition 14.27, p. 676]).

THEOREM 7.1. *Let $\mathcal{M}$ be a decomposable space, ρ be a risk function satisfying* (A2), *and $F: \mathbb{R}^n \to \mathcal{X}$ be such that the function $f(z,\omega) := [F(z)](\omega)$ is random lower semicontinuous. Then,*

$$\rho\Big(\inf_{z\in S} F(z)\Big) = \inf_{Z\in\mathcal{M}_S} \rho(F_Z). \tag{7.7}$$

PROOF. For any $Z \in \mathcal{M}_S$, we have that $Z(\omega) \in S$, and hence the inequality

$$\Big[\inf_{z\in S} F(z)\Big](\omega) \le F_Z(\omega)$$

holds for all $\omega \in \Omega$. By the monotonicity of ρ, this implies that

$$\rho\Big(\inf_{z\in S} F(z)\Big) \le \rho(F_Z),$$

and hence

$$\rho\Big(\inf_{z\in S} F(z)\Big) \le \inf_{Z\in\mathcal{M}_S} \rho(F_Z). \tag{7.8}$$

The opposite of inequality (7.8) can be proved in the same way as in the proof of Theorem 14.60 in Rockafellar and Wets [25]. □

In view of this result, we can interpret (7.1) as the difference between optimal values of two problems: problem (6.1) and problem (7.6).

8. Dualization of nonanticipativity constraints. Consider the framework of §6 with a convex mapping $F: \mathcal{Z} \to \mathcal{X}$, where $\mathcal{Z} := \mathbb{R}^n$, $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar\mu)$, $p \in [1, +\infty)$, and $\bar\mu$ is a probability measure on $(\Omega, \mathcal{F})$. We can write the optimization problem (6.1) in the following equivalent form:

$$\underset{Z\in\mathcal{M}_S,\, v\in\mathbb{R}^n}{\mathrm{Min}}\ \rho(F_Z) \quad \text{subject to } Z(\omega) = v, \quad \forall\, \omega \in \Omega. \tag{8.1}$$

The constraints $Z(\omega) = v$, $\omega \in \Omega$ in the above problem are called the *nonanticipativity* constraints. The equivalence of (6.1) and (8.1) is evident. The advantage of formulation (8.1) over (6.1) is the possibility of developing a duality framework associated with the nonanticipativity constraint. This, in turn, allows in many cases, for the decomposition of problem (8.1) into simpler problems for each $\omega \in \Omega$. Study of nonanticipativity as a constraint in a dualization procedure was initiated in Rockafellar and Wets [24]. We show, in this section, that in the context of risk measures a corresponding duality relation can be obtained as well. We assume that the risk function ρ satisfies conditions (A1)–(A3).

Suppose that $\mathcal{M} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar\mu; \mathbb{R}^n)$ and let $\mathcal{M}^* := \mathcal{L}_q(\Omega, \mathcal{F}, \bar\mu; \mathbb{R}^n)$ be its dual space. Unless stated otherwise, all expectations and probabilistic statements in this section are made with respect to $\bar\mu$. Problem (8.1) is associated with the Lagrangian

$$L(Z, v, \lambda) := \rho(F_Z) + \mathbb{E}[\lambda^T(Z - v)], \quad \lambda \in \mathcal{M}^*.$$

Let us note that $\inf_{v\in\mathbb{R}^n} L(Z, v, \lambda)$ is equal to $-\infty$ if $\mathbb{E}[\lambda] \ne 0$ and to $L(Z, 0, \lambda)$ if $\mathbb{E}[\lambda] = 0$. Therefore the (Lagrangian) dual problem of problem (8.1) takes on the form

$$\underset{\lambda\in\mathcal{M}^*}{\mathrm{Max}}\Big\{\inf_{Z\in\mathcal{M}_S} L_0(Z, \lambda)\Big\} \quad \text{subject to } \mathbb{E}[\lambda] = 0, \tag{8.2}$$

where

$$L_0(Z, \lambda) := L(Z, 0, \lambda) = \rho(F_Z) + \mathbb{E}[\lambda^T Z]. \tag{8.3}$$

By the standard theory of Lagrangian duality, we have that the optimal value of the primal problem (8.1) is greater than or equal to the optimal value of the dual problem (8.2). Moreover, under some standard regularity conditions, there is no duality gap between problems (8.1) and (8.2), i.e., their optimal values are equal to each other (see, e.g., Rockafellar [22], Bonnans and Shapiro [4, §2.5]). In particular, there is no duality gap between problems (8.1) and (8.2) and $\bar Z$ and $\bar\lambda$ are optimal solutions of (8.1) and (8.2), respectively, iff $((\bar Z, \bar v), \bar\lambda)$ is a saddle point of $L(Z, v, \lambda)$ for some $\bar v \in \mathbb{R}^n$. Noting that $L(Z, v, \lambda)$ is linear with respect to λ and to v, we obtain that $((\bar Z, \bar v), \bar\lambda)$ is a saddle point of $L(Z, v, \lambda)$ iff the following conditions hold:

$$\bar Z(\omega) = \bar v, \quad \text{a.e. } \omega \in \Omega, \qquad \text{and} \qquad \mathbb{E}[\bar\lambda] = 0, \tag{8.4}$$

$$\bar Z \in \underset{Z\in\mathcal{M}_S}{\arg\min}\, L_0(Z, \bar\lambda). \tag{8.5}$$

Consider the function $\Phi(Z) := \rho(F_Z)$: $\mathscr{M} \to \mathbb{R}$. Because of convexity of F and assumptions (A1) and (A2), this function is convex. Its subdifferential $\partial\Phi(Z) \subset \mathscr{M}^*$ is defined in the usual way. By convexity, assuming that ρ is continuous at $\bar{X} := F(\bar{z})$, we can write the following optimality conditions for (8.5) to hold at $\bar{Z}(\omega) \equiv \bar{v}$:

$$-\bar{\lambda} \in N_S(\bar{v}) + \partial\Phi(\bar{Z}). \tag{8.6}$$

Therefore we obtain that if problem (6.1) possesses an optimal solution $\bar{z}$, then the Lagrangian $L(Z, v, \lambda)$ has a saddle point iff there exists $\bar{\lambda} \in \mathscr{M}^*$ satisfying condition (8.6) and such that $\mathbb{E}[\bar{\lambda}] = 0$. We show now existence of such $\bar{\lambda}$.

By Proposition 6.3, there exists a measure $\hat{\mu}$ such that the optimal solution $\hat{z}$ of (6.1) is also the optimal solution of problem (6.5). Let $\hat{\gamma} = d\hat{\mu}/d\bar{\mu}$ be the corresponding density of $\hat{\mu}$ in the considered space $\mathscr{X} = \mathscr{L}_p(\Omega, \mathscr{F}, \bar{\mu})$. In the sequel, we identify $\hat{\mu}$ with its density $\hat{\gamma}$. We have that $\hat{\gamma} \in \partial\rho(\widehat{X})$ with $\widehat{X} = F(\hat{z})$. Using the optimality conditions for (6.5) of Proposition 6.4, we conclude that

$$0 \in N_S(\hat{z}) + \int_\Omega \hat{\gamma}(\omega)\partial f_\omega(\hat{z})\, d\bar{\mu}(\omega).$$

This means that there exists a measurable selection $g(\omega) \in \partial f_\omega(\hat{z})$ such that

$$0 \in N_S(\hat{z}) + \int_\Omega \hat{\gamma}(\omega) g(\omega)\, d\bar{\mu}(\omega). \tag{8.7}$$

Let us now define

$$\bar{\lambda}(\omega) := \int_\Omega \hat{\gamma}(\omega) g(\omega)\, d\bar{\mu}(\omega) - \hat{\gamma}(\omega) g(\omega), \quad \omega \in \Omega.$$

By construction, $\mathbb{E}[\bar{\lambda}] = 0$. Furthermore, because $g(\omega) \in \partial f_\omega(\hat{z})$, we have that

$$f_\omega(Z(\omega)) \geq f_\omega(\hat{z}) + g(\omega)^T(Z(\omega) - \hat{z}), \quad \omega \in \Omega.$$

Because of $\hat{\gamma} \in \partial\rho(\widehat{X})$, this implies that

$$\rho(F_Z) \geq \rho(F(\hat{z})) + \int_\Omega \hat{\gamma}(\omega) g(\omega)^T (Z(\omega) - \hat{z})\, d\bar{\mu}(\omega).$$

We obtain that $\hat{\gamma} g \in \partial\Phi(\widehat{Z})$. This together with Equation (8.7) imply that $\bar{\lambda}$ satisfies condition (8.6) at $\widehat{Z}$. We have thus proved the following result.

Proposition 8.1. *Suppose that ρ: $\mathscr{X} \to \overline{\mathbb{R}}$ satisfies conditions* (A1)–(A3) *and F: $\mathscr{Z} \to \mathscr{X}$ is convex. Furthermore, suppose that problem* (6.1) *possesses an optimal solution $\hat{z}$ and ρ is subdifferentiable at $\widehat{X} = F(\hat{z})$. Then, there exists $\bar{\lambda}$ such that $((\widehat{Z}, \hat{z}), \bar{\lambda})$, where $\widehat{Z}(\omega) \equiv \hat{z}$ is a saddle point of the Lagrangian $L(Z, v, \lambda)$, and hence there is no duality gap between problems* (6.1) *and* (8.2) *and $(\widehat{Z}, \hat{z})$ and $\bar{\lambda}$ are optimal solutions of problems* (8.1) *and* (8.2), *respectively.*

Let us return to the question of decomposing problem (8.5). Suppose that ρ is real valued and conditions (A1)–(A3) are satisfied. Then, ρ is continuous by Proposition 3.1. By Theorem 2.2, the representation

$$\rho(X) = \sup_{\mu \in \mathscr{P}} \{\langle \mu, X \rangle - \rho^*(\mu)\}$$

holds true. Then,

$$\inf_{Z \in \mathscr{M}_S} L_0(Z, \lambda) = \inf_{Z \in \mathscr{M}_S} \sup_{\zeta \in \mathscr{P}} \{\mathbb{E}[\zeta F_Z + \lambda^T Z] - \rho^*(\zeta)\}. \tag{8.8}$$

Suppose, further, that the "inf" and "sup" operators at the right-hand side of the above Equation (8.8) can be interchanged (note that the function inside the parentheses in the right-hand side of (8.8) is convex in X and concave in ζ). Then,

$$\begin{aligned}\inf_{Z \in \mathscr{M}_S} L_0(Z, \lambda) &= \sup_{\zeta \in \mathscr{P}} \inf_{Z \in \mathscr{M}_S} \{\mathbb{E}[\zeta F_Z + \lambda^T Z] - \rho^*(\zeta)\} \\ &= \sup_{\zeta \in \mathscr{P}} \left\{ \mathbb{E}\Big(\inf_{z \in S} [\zeta(\omega) f(z, \omega) + \lambda(\omega)^T z] \Big) - \rho^*(\zeta) \right\},\end{aligned}$$

where the last equality follows by the interchangeability principle. Therefore, we obtain that under the specified assumptions, the optimal value of the dual problem (8.2) is equal to $\sup_{\mathbb{E}[\lambda]=0,\, \zeta\in\mathcal{P}} D(\lambda,\zeta)$, where

$$D(\lambda,\zeta) := \mathbb{E}\Big\{\inf_{z\in S}[\zeta(\omega)f(z,\omega)+\lambda(\omega)^T z]\Big\} - \rho^*(\zeta). \tag{8.9}$$

If, moreover, there is no duality gap between problems (8.1) and (8.2), then the following duality relation holds:

$$\inf_{z\in S}\rho[F(z)] = \sup_{\substack{\lambda\in\mathcal{M}^*,\,\zeta\in\mathcal{P}\\ \mathbb{E}[\lambda]=0}} D(\lambda,\zeta). \tag{8.10}$$

Note the separable structure of the right-hand side of (8.9). That is, in order to calculate $D(\lambda,\zeta)$, one needs to solve the minimization problem inside the parentheses at the right-hand side of (8.9) separately for every $\omega\in\Omega$, and then to take the expectation of the optimal values calculated.

Acknowledgments. The research of the first author was supported by National Science Foundation (NSF) Awards DMS-0303545 and DMI-0354678. The research of the second author was supported by NSF Award DMS-0510324.

References

[1] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.
[2] Aubin, J.-P., I. Ekeland. 1984. *Applied Nonlinear Analysis*. John Wiley and Sons, New York.
[3] Blackwell, D. 1953. Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.
[4] Bonnans, J. F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York.
[5] Castaing, C., M. Valadier. 1977. *Convex Analysis and Measurable Multifunctions*. Springer-Verlag, Berlin, Germany.
[6] Cheridito, P., F. Delbaen, M. Kupper. 2004. Coherent and convex risk measures for bounded càdlàg processes. *Stochastic Processes Their Appl.* **112**(1) 1–22.
[7] Delbaen, F. 2002. Coherent risk measures on general probability spaces. *Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, Germany.
[8] Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance Stochastics* **6** 429–447.
[9] Ioffe, A. D., V. M. Tihomirov. 1979. *Theory of Extremal Problems*. North-Holland Publishing Company, Amsterdam, The Netherlands.
[10] Klein Haneveld, W. K. 1986. Duality in stochastic linear and dynamic programming. *Lecture Notes in Economics and Mathematical Systems*, Vol. 274. Springer-Verlag, New York.
[11] Levin, V. L. 1985. *Convex Analysis in Spaces of Measurable Functions and Its Applications in Economics*. Nauka, Moscow, Russia.
[12] Lorenz, M. O. 1905. Methods of measuring concentration of wealth. *J. Amer. Statist. Association* **9** 209–219.
[13] Markowitz, H. M. 1952. Portfolio selection. *J. Finance* **7** 77–91.
[14] Markowitz, H. M. 1987. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Blackwell Publishing, Oxford, UK.
[15] Müller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley and Sons, Chichester, UK.
[16] Ogryczak, W., A. Ruszczyński. 1999. From stochastic dominance to mean-risk models: Semideviations as risk measures. *Eur. J. Oper. Res.* **116** 33–50.
[17] Ogryczak, W., A. Ruszczyński. 2001. On consistency of stochastic dominance and mean-semideviation models. *Math. Programming* **89** 217–232.
[18] Ogryczak, W., A. Ruszczyński. 2002. Dual stochastic dominance and related mean-risk models. *SIAM J. Optim.* **13** 60–78.
[19] Phelps, R. R. 1989. Convex functions, monotone operators and differentiability. *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, Germany.
[20] Pflug, G. C. 2000. Some remarks on the value-at-risk and the conditional value-at-risk. S. Uryasev, ed. *Probabilistic Constrained Optimization—Methodology and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 272–281.
[21] Raiffa, H., R. Schlaifer. 1961. *Applied Statistical Decision Theory*. Harvard University Press, Boston, MA.
[22] Rockafellar, R. T. 1974. Conjugate duality and optimization. *Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA.
[23] Rockafellar, R. T., S. P. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 21–41.
[24] Rockafellar, R. T., R. J.-B. Wets. 1976. Stochastic convex programming: Basic duality. *Pacific J. Math.* **62** 173–195.
[25] Rockafellar, R. T., R. J.-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin, Germany.
[26] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2006. Deviation measures in risk analysis and optimization. *Finance Stochastics* **10** 51–74.
[27] Ruszczyński, A., A. Shapiro. 2005. Optimization of risk measures. G. Calafiore, F. Dabbene, eds. *Probabilistic and Randomized Methods for Design Under Uncertainty*. Springer-Verlag, London, UK, 117–158.
[28] Shapiro, A., S. Ahmed. 2004. On a class of minimax stochastic programs. *SIAM J. Optim.* **14** 1237–1249.
[29] Strassen, V. 1965. The existence of probability measures with given marginals. *Ann. Math. Statist.* **38** 423–439.

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 32, No. 2, May 2007, p. 496
ISSN 0364-765X | EISSN 1526-5471 | 07 | 3202 | 0496

DOI 10.1287/moor.1070.0265

Corrigendum to: "Optimization of Convex Risk Functions," *Mathematics of Operations Research* **31** (2006) 433–452

Andrzej Ruszczyński
Department of Management Science and Information Systems, Rutgers University, Piscataway, New Jersey 08854,
rusz@business.rutgers.edu, http://www.rusz.rutgers.edu

Alexander Shapiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332,
ashapiro@isye.gatech.edu, http://www2.isye.gatech.edu/people/faculty/Alex_Shapiro/

Proposition 6.2 is not correct, because the set $\mathrm{dom}(\rho^*)$ may be not weakly* compact. The correct statement should read as follows:

PROPOSITION 6.2. *Suppose that $\mathscr{X}$ is a Banach space, $\mathscr{Y} = \mathscr{X}^*$, the mapping $F : \mathscr{Z} \to \mathscr{X}$ is convex, the function $\rho : \mathscr{X} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and assumptions* (A1)–(A4) *are satisfied. If ρ is continuous at* 0, *then the optimal value of problem* (6.1) *is equal to the optimal value of problem* (6.4). *Moreover, problem* (6.4) *has an optimal solution.*

PROOF. The supremum over $\mathscr{P}$ in (6.2) may be replaced by the supremum over $\mathscr{A} = \mathrm{dom}(\rho^*)$. Due to (A1), (A4), and the continuity of ρ at 0, $\mathscr{A} = \partial\rho(0)$ and is weakly* closed and bounded. Thus it is weakly* compact in $\mathscr{X}^*$. (This could not be guaranteed without the additional continuity assumption.) The function $\Xi(z, \cdot)$ is concave and weakly* continuous, and the function $\Xi(\cdot, \mu)$ is convex. Our assertion then follows from the asymmetric min-max theorem [2, Theorem 6.2.7].

In Proposition 6.3 we need the stronger condition that $\rho(\cdot)$ is continuous at $\hat{X}$, rather than just subdifferentiable. The correct statement reads as follows.

PROPOSITION 6.3. *Let $F: \mathscr{Z} \to \mathscr{X}$ be convex and let $\rho : \mathscr{X} \to \overline{\mathbb{R}}$ satisfy conditions* (A1)–(A3). *Suppose that a point $\hat{z} \in S$ is an optimal solution of problem* (6.1) *and that $\rho(\cdot)$ is continuous at $\hat{X} := F(\hat{z})$. Then there exists a measure $\hat{\mu} \in \partial\rho(\hat{X})$ such that that $\hat{z}$ is an optimal solution of problem* (6.5).

PROOF. As ρ is continuous, it is subdifferentiable at $\hat{X}$. Then $\partial\rho^{**}(\hat{X}) = \partial\rho(\hat{X})$, and it follows by (3.3) that $\partial\rho(\hat{X}) = \arg\max_{\mu\in\mathscr{Y}} \{\langle \mu, \hat{X}\rangle - \rho^*(\mu)\}$. Since $\mathrm{dom}(\rho^*) \subset \mathscr{P}$, $\partial\rho(\hat{X}) = \arg\max_{\mu\in\mathscr{P}} \Xi(\hat{z}, \mu)$. By monotonicity of $\rho(\cdot)$, (6.1) is equivalent to (6.9). As $\hat{z}$ is optimal in (6.1), the pair $(\hat{z}, \hat{X})$ is optimal in (6.9). By convexity of F, the set U is convex. Problem (6.9) can be equivalently written as: $\min \rho(X) + \delta_U(z, X)$. Define $\tilde{\rho}(z, X) = \rho(X)$. The optimality of $(\hat{z}, \hat{X})$ implies that $0 \in \partial(\tilde{\rho} + \delta_U)(\hat{z}, \hat{X})$. By continuity of $\rho(\cdot)$ at $\hat{X}$, we can use the Moreau–Rockafellar theorem: $\partial(\tilde{\rho} + \delta_U)(\hat{z}, \hat{X}) = \partial\tilde{\rho}(\hat{z}, \hat{X}) + N_U(\hat{z}, \hat{X})$, with N_U denoting the normal cone to U, see [19, Theorem 3.23]. Thus, there exists $\hat{\mu} \in \partial\rho(\hat{X})$ such that $\langle \hat{\mu}, X - \hat{X}\rangle \geq 0$ for all $(z, X) \in U$. Setting $X = F(z)$, we obtain $\langle \hat{\mu}, F(z) - F(\hat{z})\rangle \geq 0$ for all $z \in S$. Hence $\hat{z}$ is a solution of (6.5).

Under the assumptions of Proposition 6.3, the optimal values of (6.2) and (6.4) are equal, and (6.4) has an optimal solution. Proposition 6.2 is fully absorbed by Proposition 6.3. All references to Proposition 6.2 should be replaced by references to Proposition 6.3.

Finally, for (8.6) to hold true, we need to additionally assume that F is continuous at $\bar{z}$.

We apologize to the readers of *Mathematics of Operations Research* for possible confusion.