

Optimization of Convex Risk Functions

Andrzej Ruszczyński* Alexander Shapiro†

January 31, 2004

Abstract

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 involving risk functions.

Key words: Convex analysis, stochastic optimization, risk measures, mean-variance models, duality.

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 and that an uncertain outcome is represented by a function $X : \Omega \rightarrow \mathbb{R}$. 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 $\rho(X)$. In order 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

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

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA, e-mail: ashapiro@isye.gatech.edu

space equipped with σ -algebra \mathcal{F} of subsets of Ω , and that \mathcal{X} is a linear space of \mathcal{F} -measurable functions $X : \Omega \rightarrow \mathbb{R}$. Also, we consider risk functions which can take values in the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

Recently, Föllmer and Schied [6] have introduced several axioms of so-called convex risk functions¹. In our context, $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is a (*convex*) *risk function* if it satisfies the following conditions:

(A1) **Convexity:**

$$\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y) \text{ for all } X, Y \in \mathcal{X} \text{ and } \alpha \in [0, 1].$$

(A2) **Monotonicity:** If $X, Y \in \mathcal{X}$ and $Y \geq X$, then $\rho(Y) \geq \rho(X)$.

(A3) **Translation Equivariance:** If $a \in \mathbb{R}$ and $X \in \mathcal{X}$, then $\rho(X + a) = \rho(X) + a$.

These conditions were inspired by the axioms of *coherent risk measures* of Artzner, Delbaen, Eber and Heath [1], who postulated, in addition to (A1)–(A3), the positive homogeneity of $\rho(\cdot)$.

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 [11, 12], but many efforts have been made to use other deviation measures, like semideviation and deviations from quantiles (Ogryczak and Ruszczyński [13, 14, 15]). Recently, Rockafellar, Uryasev and Zabarankin [21] also developed an axiomatic approach to coherent risk 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 this paper.

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 \mathcal{Z} , and $F : \mathcal{Z} \rightarrow \mathcal{X}$. This creates the need to consider *composite risk functions*, of the form $\rho(F(z))$, and associated optimization problems:

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

where S is a convex subset of \mathcal{Z} .

Our plan is to exploit general results of convex analysis in topological vector spaces of measurable functions to derive properties of convex risk functions and of composite risk functions. In section 2 we generalize some dual-representation theorems given in [6] and [21]. Section 3 is devoted to the analysis of differentiability properties of

¹Föllmer and Schied use the name ‘risk measure’, but we reserve the term ‘measure’ for its classical meaning of a countably additive set function.

risk functions. In particular, we obtain representation of subgradients and directional derivatives of risk functions and composite risk functions. In section 4 we analyze risk functions resulting from several classical mean–risk models. In section 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 section 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 section 7 we introduce the concept of *risk value of perfect information*, for problem (1.1), and discuss its properties. Finally, in section 8 we develop a duality relation for optimization problems involving risk functions and nonanticipativity constraints.

2 Conjugate Duality of Risk Functions

Let (Ω, \mathcal{F}) be a measurable space and \mathbb{Y} be the (linear) space of all signed finite measures on (Ω, \mathcal{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 \mathcal{X} of \mathcal{F} -measurable functions $X : \Omega \rightarrow \mathbb{R}$. We use the cone

$$\mathcal{X}_+ := \{X \in \mathcal{X} : X(\omega) \geq 0, \forall \omega \in \Omega\} \quad (2.1)$$

to define the corresponding partial order on the space \mathcal{X} . That is, the relation $Y \geq X$ in condition (A2) means that $Y(\omega) \geq X(\omega)$ for all $\omega \in \Omega$. We associate with \mathcal{X} a linear space $\mathcal{Y} \subset \mathbb{Y}$ such that $\int_{\Omega} |X| d|\mu| < +\infty$ for every $X \in \mathcal{X}$ and $\mu \in \mathcal{Y}$, and define the scalar product

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

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

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

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

The above condition ensures that the cone \mathcal{Y}_+ is dual to \mathcal{X}_+ , i.e.,

$$\mathcal{Y}_+ = \{\mu \in \mathcal{Y} : \langle \mu, X \rangle \geq 0, \forall X \in \mathcal{X}_+\}.$$

We have that a measure μ is not nonnegative if $\mu(A) < 0$ for some $A \in \mathcal{F}$. Therefore, condition (C) holds, for example, if the space \mathcal{X} contains all functions $\mathbb{1}_A(\cdot)$, $A \in \mathcal{F}$, 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 \mathcal{X} satisfies condition (C).

Furthermore, we assume that \mathcal{X} and \mathcal{Y} are *paired locally convex* topological vector spaces. That is, \mathcal{X} and \mathcal{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 \mathcal{X} can be represented in the form $\langle \mu, \cdot \rangle$ for some $\mu \in \mathcal{Y}$, and every linear continuous functional on \mathcal{Y} can be represented in the form $\langle \cdot, X \rangle$ for some $X \in \mathcal{X}$. In particular, we can equip each space \mathcal{X} and \mathcal{Y} with its weak topology induced by its paired space. This will make \mathcal{X} and \mathcal{Y} paired locally convex topological vector spaces provided that for any $X \in \mathcal{X} \setminus \{0\}$ there exists $\mu \in \mathcal{Y}$ such that $\langle \mu, X \rangle \neq 0$, and for any $\mu \in \mathcal{Y} \setminus \{0\}$ there exists $X \in \mathcal{X}$ such that $\langle \mu, X \rangle \neq 0$.

A natural choice of \mathcal{X} is the space of all bounded \mathcal{F} -measurable functions $X : \Omega \rightarrow \mathbb{R}$. In that case we can take $\mathcal{Y} := \mathbb{Y}$. Another possible choice is $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ for some positive measure $\bar{\mu} \in \mathbb{Y}$ and $p \in [1, +\infty]$. In that case we can take \mathcal{Y} to be the linear space of measures $\nu \in \mathbb{Y}$ absolutely continuous with respect to $\bar{\mu}$ and with density (Radon–Nikodym derivative) $h = d\nu/d\bar{\mu}$ belonging to the space $\mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$, where $q \geq 1$ is such that $1/p + 1/q = 1$. In that case we identify \mathcal{Y} with $\mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$. Note that an element $h \in \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ is a class of functions which are equal each other for almost every (a.e.) $\omega \in \Omega$ with respect to the measure $\bar{\mu}$, and the corresponding cone \mathcal{X}_+ is formed by nonnegative almost everywhere $h \in \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$. The space $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ is a Banach space and, for $p \in [1, +\infty)$, $\mathcal{Y} := \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$ is its dual space of all continuous linear functionals on \mathcal{X} . When dealing with Banach spaces it is convenient to equip \mathcal{X} and $\mathcal{Y} := \mathcal{X}^*$ with the strong (norm) and weak* topologies, respectively. If \mathcal{X} is a reflexive Banach space, i.e., $\mathcal{X}^{**} = \mathcal{X}$, then \mathcal{X} and \mathcal{X}^* , both equipped with strong topologies, form paired spaces.

Having defined the spaces \mathcal{X} and \mathcal{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 $X \in \mathcal{X}$ and its domain $\text{dom}(\rho) := \{X \in \mathcal{X} : \rho(X) < +\infty\}$ is nonempty.

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

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

and the conjugate of ρ^* as

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

By $\text{lsc}(\rho)$ we denote the lower semicontinuous hull of ρ taken with respect to the considered topology of \mathcal{X} . The following is the basic duality result of convex analysis (see, e.g., [17, Theorem 5] and [2, Theorem 4.4.2] for a proof).

Theorem 1 (Fenchel-Moreau) *Suppose that the function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is convex and proper. Then $\rho^{**} = \text{lsc}(\rho)$.*

It follows that if ρ is convex and proper, then the representation

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

holds if ρ is lower semicontinuous. Conversely, if (2.5) holds 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 \mathcal{X} is a Banach space and $\mathcal{Y} := \mathcal{X}^*$ is its dual (e.g., $\mathcal{X} = \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ and $\mathcal{Y} = \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$) and ρ is convex, then ρ is lower semicontinuous in the weak topology iff it is lower semicontinuous in the strong (norm) topology. If the set Ω is finite, then the space \mathcal{X} is finite dimensional. In that case ρ is continuous (and hence lower semicontinuous) if it is real valued.

Theorem 2 *If assumptions (A1)–(A3) hold true and the function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is lower semicontinuous, then*

$$\rho(X) = \sup_{\mu \in \mathcal{P}} \{ \langle \mu, X \rangle - \rho^*(\mu) \}, \quad \forall X \in \mathcal{X}. \quad (2.6)$$

Conversely, if ρ can be represented in the form (2.6) for some function $\rho^ : \mathcal{Y} \rightarrow \overline{\mathbb{R}}$, then ρ is lower semicontinuous and assumptions (A1)–(A3) are satisfied.*

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

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

Suppose that assumption (A3) holds. Then, for an $X \in \text{dom}(\rho)$, 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 $\mu \in \mathcal{Y}$ such that $\mu(\Omega) \neq 1$. This shows that, under the specified assumptions, it suffices to take the supremum in (2.4) with respect to the set $\mathcal{P} \subset \mathcal{Y}$ of probability measures, and hence (2.6) follows by the Fenchel-Moreau theorem.

Conversely, suppose that representation (2.6) holds. Then ρ is given by the supremum of a family of continuous affine functions, and hence is convex and lower semicontinuous. Now if $Y \geq X$, then $\langle \mu, Y \rangle - \langle \mu, X \rangle = \langle \mu, Y - X \rangle \geq 0$ for any $\mu \in \mathcal{P}$. Consequently assumption (A2) follows from (2.6). Finally, we have that for any

$\mu \in \mathcal{P}$, $\langle \mu, X+a \rangle = \langle \mu, X \rangle + a$, and hence it follows by (2.6) that $\rho(X+a) = \rho(X) + a$. ■

The above theorem generalizes the results of Föllmer and Schied [6, Theorems 5 and 6].

It is said that the function ρ is *positively homogeneous* if $\rho(tX) = t\rho(X)$ for any $t > 0$ and $X \in \mathcal{X}$. If ρ is convex and positively homogeneous, then its conjugate ρ^* is the indicator function of a closed convex set $\mathcal{A} \subset \mathcal{Y}$ which is formed by such $\mu \in \mathcal{Y}$ that $\langle \mu, X \rangle \leq \rho(X)$ for all $X \in \mathcal{X}$. Moreover, under assumptions (A2)–(A3), it suffices to consider $\mu \in \mathcal{P}$, i.e.,

$$\mathcal{A} = \{\mu \in \mathcal{P} : \langle \mu, X \rangle \leq \rho(X), \quad \forall X \in \mathcal{X}\}. \quad (2.7)$$

Therefore Theorem 2 implies the following result.

Corollary 1 *Suppose that assumptions (A1)–(A3) hold and the function ρ is lower semicontinuous and positively homogeneous. Then*

$$\rho(X) = \sup_{\mu \in \mathcal{A}} \langle \mu, X \rangle, \quad \forall X \in \mathcal{X}, \quad (2.8)$$

where the set \mathcal{A} is defined in (2.7).

The set \mathcal{A} is called the *risk envelope* in Rockafellar, Uryasev and Zabarankin [21], where the above result has been developed in the space $\mathcal{X} := \mathcal{L}_2(\Omega, \mathcal{F}, \bar{\mu})$.

3 Continuity and Differentiability Properties of Risk Functions

In applications it is usually straightforward to verify assumptions (A1)–(A3). The assumption of the lower semicontinuity of ρ is more delicate. Suppose that ρ is proper and convex, and denote by $\text{int}(\text{dom } \rho)$ the interior of the domain of ρ . We have that if ρ is bounded from above on a neighborhood of some point $\bar{X} \in \mathcal{X}$, then ρ is continuous on $\text{int}(\text{dom } \rho)$ (e.g., [7, p. 170, Theorem 1]). In order to verify continuity properties of ρ it is technically advantageous to use the strong (rather than weak) topology of \mathcal{X} , if \mathcal{X} is a Banach space. Therefore when dealing with a Banach space \mathcal{X} we equip it with its strong topology and use $\mathcal{Y} := \mathcal{X}^*$.

A linear functional $\ell : \mathcal{X} \rightarrow \mathbb{R}$ is called an *algebraic subgradient* of ρ at $\bar{X} \in \text{dom } \rho$ if

$$\rho(X) \geq \rho(\bar{X}) + \ell(X - \bar{X}), \quad \forall X \in \mathcal{X}. \quad (3.1)$$

Note that the algebraic subgradient functional ℓ is not required to be continuous. If, moreover, $\ell \in \mathcal{Y}$, then we say that ℓ is a subgradient of ρ at \bar{X} . The set of

all subgradients $\ell \in \mathcal{Y}$, satisfying (3.1), is called the subdifferential of ρ at \bar{X} , and denoted $\partial\rho(\bar{X})$. It is said that ρ is subdifferentiable at \bar{X} if $\partial\rho(\bar{X})$ is nonempty.

Let us observe that ρ always possesses an algebraic subgradient at any point $\bar{X} \in \text{int}(\text{dom } \rho)$ (cf., [9, Lemma 1.1]). Indeed, consider the directional derivative function $\delta(\cdot) := \rho'(\bar{X}, \cdot)$, where

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

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 $\bar{X} \in \text{int}(\text{dom } \rho)$, then $\delta(\cdot)$ is finite valued. By the Hahn-Banach Theorem we have that there exists a linear functional $\ell : \mathcal{X} \rightarrow \mathbb{R}$ such that $\delta(\cdot) \geq \ell(\cdot)$. It follows that ℓ satisfies (3.1).

We show now that the lower semicontinuity of ρ is implied by assumptions (A1)–(A2), if \mathcal{X} has the structure of a Banach lattice. Recall that \mathcal{X} is a *lattice* (with respect to the cone \mathcal{X}_+) if for any $X_1, X_2 \in \mathcal{X}$ 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 \mathcal{X} . For every $X \in \mathcal{X}$ we can then define $|X| \in \mathcal{X}$ in a natural way, i.e., $|X|(\omega) = |X(\omega)|$, $\omega \in \Omega$. The space \mathcal{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 $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$, $p \in [1, +\infty]$, is a Banach lattice.

Proposition 1 *Suppose that \mathcal{X} is a Banach lattice and $\rho : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ satisfies assumptions (A1) and (A2). Then $\rho(\cdot)$ is continuous and subdifferentiable on the interior of its domain.*

Proof. Let $\bar{X} \in \text{int}(\text{dom } \rho)$. By the above discussion, 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 $\ell(X) \geq 0$ for all $X \in \mathcal{X}_+$. Indeed, if $\ell(Y) < 0$ for some $Y \in \mathcal{X}_+$, then it follows from (3.1) that $\rho(\bar{X} - Y) > \rho(\bar{X})$, which contradicts (A2). Now by [9, Theorem 0.12] we have that any positive linear functional on the Banach lattice \mathcal{X} is continuous. Consequently ℓ is continuous, and hence $\ell \in \partial\rho(\bar{X})$. It follows then from (3.1) that ρ is lower semicontinuous at \bar{X} . Since \bar{X} was an arbitrary point of $\text{int}(\text{dom } \rho)$, we obtain that $\rho(\cdot)$ is lower semicontinuous on the interior of its domain. This, combined with the fact that \mathcal{X} is a Banach space, implies the continuity of $\rho(\cdot)$ on $\text{int}(\text{dom } \rho)$ (see, e.g., [16, Theorem 3.3]). ■

We obtain that, under the assumptions of the above proposition, if $\rho(X)$ is real valued for all $X \in \mathcal{X}$, then $\rho(\cdot)$ is continuous and subdifferentiable on \mathcal{X} . Proposition 1 can be applied, for example, to every space $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ with $p \in [1, +\infty)$. We also can apply this framework to the space $\mathcal{X} := \mathcal{L}_\infty(\Omega, \mathcal{F}, \bar{\mu})$ if we equip it

with its strong topology. This, however, will require to pair \mathcal{X} with its dual space $\mathcal{Y} := \mathcal{L}_\infty(\Omega, \mathcal{F}, \bar{\mu})^*$ which is larger than $\mathcal{L}_1(\Omega, \mathcal{F}, \bar{\mu})$.

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

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

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

$$\partial\rho^{**}(\bar{X}) = \arg \max_{\mu \in \mathcal{Y}} \{ \langle \mu, \bar{X} \rangle - \rho^*(\mu) \} \quad (3.3)$$

(cf., [17, p.35]). 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

$$\partial\rho(\bar{X}) \subset \mathcal{P}.$$

In particular, we obtain that under the assumptions of Corollary 1, the representation (2.8) holds with the set $\mathcal{A} = \partial\rho(0)$, and

$$\partial\rho(\bar{X}) = \arg \max_{\mu \in \mathcal{A}} \langle \mu, \bar{X} \rangle. \quad (3.4)$$

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 $0 \in \mathcal{X}$, then

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

In particular, if \mathcal{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.,

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

(see, e.g., [4, section 2.2.1] for a discussion of Hadamard directional derivatives). Hadamard directional differentiability implies continuity of the directional derivative function $\rho'(\bar{X}, \cdot)$. Recall that ρ is said to be Gâteaux differentiable at \bar{X} if $\rho'(\bar{X}, \cdot)$ is linear and continuous. Therefore, if \mathcal{X} is a Banach space and ρ is continuous at \bar{X} , then ρ is Gâteaux (Hadamard) differentiable at \bar{X} iff $\partial\rho(\bar{X})$ is a singleton.

In the subsequent analysis we shall deal with composite functions $\psi : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$ of the form $\psi(\cdot) := \rho(F(\cdot))$. Here \mathcal{Z} is a vector space and $F : \mathcal{Z} \rightarrow \mathcal{X}$ 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 (Ω, \mathcal{F}) . We say that the mapping F is *convex* if the function $f(\cdot, \omega)$ is convex for every $\omega \in \Omega$.

Proposition 2 *If the mapping $F : \mathcal{Z} \rightarrow \mathcal{X}$ is convex and $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ 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

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

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

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

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

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

Proposition 3 *Suppose that \mathcal{X} is a Banach space, the mapping F is convex, the function ρ is convex, finite valued and continuous at $\bar{X} := F(\bar{z})$. 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*

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

Proof. We have here that ρ is subdifferentiable and Hadamard directionally differentiable at $\bar{X} := F(\bar{z})$ 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

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

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

Corollary 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 $\bar{X} := F(\bar{z})$ and $\partial \rho(\bar{X}) = \{\bar{\mu}\}$ is a singleton. Then the composite function ψ is Hadamard differentiable at \bar{z} if and only if $f'_{\omega}(\bar{z}, \cdot)$ is linear for $\bar{\mu}$ -almost every $\omega \in \Omega$.*

Proof. By Proposition 3 we have here that

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

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 Proposition 2, 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 locally convex topological vector space, \mathcal{Z}^* is its dual space of all continuous linear functionals on \mathcal{Z} , $F : \mathcal{Z} \rightarrow \mathcal{X}$ is convex, and consider the integral function

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

associated with measure $\mu \in \mathcal{P}$. 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 [9, Theorem 1.1], see also [23, 5]) we have that

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

The integral in 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 $v \in \mathcal{Z}$.

We also have that $\partial\psi(\bar{z})$ is equal to the subdifferential of the directional derivative function $\psi'(\bar{z}, \cdot)$ at $0 \in \mathcal{Z}$. Therefore, if formula (3.6) holds, then

$$\partial\psi(\bar{z}) := \overline{\text{conv}} \left(\bigcup_{\mu \in \partial\rho(\bar{X})} \int_{\Omega} \partial f_{\omega}(\bar{z}) d\mu(\omega) \right), \quad (3.9)$$

where $\overline{\text{conv}}(S)$ denotes the closure (in the weak* topology) of the convex hull of the set $S \subset \mathcal{Z}^*$.

4 Examples of Risk Functions

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

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

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

where $c \geq 0$ is a constant and

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

Note that for $p = 2$ the function $\rho(\cdot)$ corresponds to the classical mean–variance model of Markowitz [11], but with the standard deviation instead of the variance.

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

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

and $\nu \in \mathcal{M}_p$. For a set $A \in \mathcal{F}$ let $A = A^+ \cup A^-$ be the Jordan decomposition of A with respect to ν , i.e., $A^+ \cap A^- = \emptyset$ 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 \mathcal{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 \mathcal{M}_p into the space of absolutely continuous, with respect to $\bar{\mu}$, measures with density $h \in \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$. With some abuse of the notation we take $\mathcal{Y} := \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$ and write

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

We have that \mathcal{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 $B_q := \{h \in \mathcal{Y} : \|h\|_q \leq 1\}$ in the dual space $\mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$. 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 $A^* : \mathcal{Y} \rightarrow \mathcal{Y}$ is the adjoint of the operator A . By a straightforward calculation we have that $A^*(h) = h - \int_{\Omega} h d\bar{\mu}$. Consequently,

$$\mathcal{M}_p = \{h - \int_{\Omega} h d\bar{\mu} : h \in B_q\}. \quad (4.5)$$

It follows that

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

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

$$\mathcal{A}_p = \{g \in \mathcal{Y} : g = 1 + h - \int_{\Omega} h d\bar{\mu}, \|h\|_q \leq c\}. \quad (4.7)$$

Now if $p = 1$, then $q = +\infty$, i.e., the corresponding norm $\|h\|_q$ is given by the essential maximum of $|h(\omega)|$, $\omega \in \Omega$. In that case all functions $g \in \mathcal{A}_p$ are almost everywhere nonnegative valued, and hence \mathcal{A}_p is a set of probability measures, if $c \leq 1/2$. In fact, if the measure $\bar{\mu}$ is such that for every $\varepsilon > 0$ there exists $A \in \mathcal{F}$ such that $0 < \bar{\mu}(A) < \varepsilon$, then \mathcal{A}_p is a set of probability measures if and only if $c \leq 1/2$. Indeed, it is clear that if $\|h\|_q \leq c$, 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. $w \in \Omega$. Conversely, take

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

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

$$c \leq \min_{1 \leq i \leq K} [p_i^{-1/q} - 1]^{-1}. \quad (4.8)$$

The right hand side of the above inequality is less than or equal to $1/(K^{1/q} - 1)$, with the equality for $p_1 = \dots = p_K = 1/K$. 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 property: (i) “for every $\varepsilon > 0$ there exists $A \in \mathcal{F}$ such that $0 < \bar{\mu}(A) < \varepsilon$ ” 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 1 It might be worth mentioning that $\psi_p(X)$ satisfies all axioms of a deviation measure specified in [21]. 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 specified above property (i).

Example 2 Let, as in Example 1, $\bar{\mu}$ be a probability measure on (Ω, \mathcal{F}) and $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ for some $p \in [1, +\infty)$. Consider now the function

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

where $c \geq 0$ and

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

is the upper semi-deviation of X of order p with $p \geq 1$. Note that $[a]_+^p := ([a]_+)^p$. The risk function (4.2) represents the mean-semideviation models analyzed in [13, 14].

Again, the functions $\sigma_p(\cdot)$ and $\rho(\cdot)$ are a convex, positively homogeneous and continuous in the strong topology of $\mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$. Similarly to the analysis of Example

1, we need to consider only measures which 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 $\{h \in B_q : h \succeq 0\}$, where the notation $h \succeq 0$ means that $h(\omega) \geq 0$ for $\bar{\mu}$ -almost every $\omega \in \Omega$. Consequently, in a way similar to the derivations of Example 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

$$\mathcal{A}_p := \{g \in \mathcal{Y} : g = 1 + h - \int_{\Omega} h d\bar{\mu}, \|h\|_q \leq c, h \succeq 0\}. \quad (4.11)$$

Since $|\int_{\Omega} h d\bar{\mu}| \leq \|h\|_q$ 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 if $\bar{\mu}$ is not a trivial measure, only if) $c \in [0, 1]$.

Since 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, $\partial\rho(X) = \{1 + h - \int_{\Omega} h d\bar{\mu} : h \in \mathcal{D}_X\}$, where

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

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 $A_X := \{\omega \in \Omega : Y(\omega) > 0\}$. 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 $\omega \in \Omega \setminus A_X$. It follows that $\mathcal{D}_X = \{ch_X^*\}$. We obtain that \mathcal{D}_X is a singleton, and hence ρ is Hadamard differentiable at X , for every nonconstant $X \in \mathcal{X}$.

Suppose now that $p = 1$. Then B_q , with $q = +\infty$, is formed by $h \in \mathcal{Y}$ such that $|h(\omega)| \leq 1$ for $\bar{\mu}$ -almost every $\omega \in \Omega$. In that case we have that if $h \in \mathcal{D}_X$, 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 $\omega \in \Omega \setminus A_X$. We obtain that

$$\mathcal{D}_X = \{h \in cB_q : h(\omega) = c \text{ if } Y(\omega) > 0, h(\omega) = 0 \text{ if } Y(\omega) < 0\}. \quad (4.13)$$

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

Example 3 Let $\bar{\mu}$ be a probability measure on (Ω, \mathcal{F}) and consider spaces $\mathcal{X} := \mathcal{L}_1(\Omega, \mathcal{F}, \bar{\mu})$ and $\mathcal{Y} := \mathcal{L}_{\infty}(\Omega, \mathcal{F}, \bar{\mu})$. For constants $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, consider the function $\rho(X) := \langle \bar{\mu}, X \rangle + \phi(X)$, where

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

and $G(x) := \bar{\mu}(\{\omega : X(\omega) \leq x\})$ is the cumulative distribution function of $X(\omega)$ with respect to the probability measure $\bar{\mu}$. It can be noted that the infimum in 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

$$\rho(X) = (1 - \varepsilon_1)\langle \bar{\mu}, X \rangle + \varepsilon_1 \text{CV@R}_p[X], \quad (4.15)$$

where

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

The quantity (4.16) was called the Conditional Value at Risk in Rockafellar and Uryasev [20]. It is the financial counterpart of the function of the *integrated chance constraint* introduced in Klein Haneveld [8]. Both are special cases of the classical concept of the absolute Lorenz curve, evaluated at point p , [10, 15]. A risk envelope representation of CV@R has been developed in [21] and [22].

We have that the function inside the integral in the right hand side of (4.14) is convex jointly in X and z . It follows that the functions $\phi(X)$ and $\rho(X)$ are convex. It is also not difficult to see that $\phi(X + a) = \phi(X)$, for any $X \in \mathcal{X}$ and $a \in \mathbb{R}$, and hence the assumption (A3) holds here, and that ϕ is positively homogeneous. Similarly the assumptions (A1) and (A3) hold for the function $\text{CV@R}_p[\cdot]$ and it is positively homogeneous. It is also not difficult to see that the function $\text{CV@R}_p[\cdot]$ satisfies assumption (A2). It follows then by (4.15) that the function $\rho(X)$ satisfies assumption (A2) if $\varepsilon_1 \in (0, 1]$.

It is straightforward to verify that

$$\int_{\Omega} \{\varepsilon_1[z - X(\omega)]_+ + \varepsilon_2[X(\omega) - z]_+\} d\bar{\mu}(\omega) = \sup_{h \in \mathcal{M}} \langle h, X - z \rangle,$$

where²

$$\mathcal{M} := \{h \in \mathcal{Y} : -\varepsilon_1 \leq h(\omega) \leq \varepsilon_2, \text{ a.e. } \omega \in \Omega\}.$$

²Recall that $\mathcal{Y} := \mathcal{L}_{\infty}(\Omega, \mathcal{F}, \bar{\mu})$ and ‘‘a.e. $\omega \in \Omega$ ’’ means here that the corresponding property holds for almost every $\omega \in \Omega$ with respect to the measure $\bar{\mu}$.

Consequently

$$\phi(X) = \inf_{z \in \mathbb{R}} \sup_{h \in \mathcal{M}} \left\{ \langle h, X \rangle - z \int_{\Omega} h d\bar{\mu} \right\}. \quad (4.17)$$

Since the set of $z \in \mathbb{R}$, which minimize the function in the right hand side of (4.17), is nonempty and bounded, the “inf” and “sup” operators in (4.17) can be interchanged. Hence

$$\phi(X) = \sup \left\{ \langle h, X \rangle : -\varepsilon_1 \leq h(\omega) \leq \varepsilon_2, \text{ a.e. } \omega \in \Omega, \int_{\Omega} h d\bar{\mu} = 0 \right\}. \quad (4.18)$$

It follows that

$$\rho(X) = \sup_{\mu \in \mathcal{A}} \langle \mu, X \rangle, \quad (4.19)$$

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

$$\mathcal{A} := \left\{ h \in \mathcal{Y} : \gamma_1 \leq h(\omega) \leq \gamma_2, \text{ a.e. } \omega \in \Omega, \int_{\Omega} h d\bar{\mu} = 1 \right\}. \quad (4.20)$$

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

Similarly to the previous example, we have here that ρ is subdifferentiable at every $X \in \mathcal{X}$ and

$$\partial\rho(X) = \arg \max_{h \in \mathcal{Y}} \left\{ \int_{\Omega} X h d\bar{\mu} : \gamma_1 \leq h(\omega) \leq \gamma_2, \text{ a.e. } \omega \in \Omega, \int_{\Omega} h d\bar{\mu} = 1 \right\}. \quad (4.21)$$

Moreover, ρ is Hadamard differentiable at X iff the ‘argmax’ set in the right hand side of (4.21) is a singleton.

Let us consider the maximization problem in the right hand side of (4.21). We can write it in the max-min form:

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

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

$$\text{Min}_{\lambda \in \mathbb{R}} \left\{ \int_{\Omega} \max [\gamma_1(X - \lambda), \gamma_2(X - \lambda)] d\bar{\mu} + \lambda \right\}. \quad (4.22)$$

Let $\bar{\lambda}$ be an optimal solution of (4.22). Considering the left and right side derivatives, at $\bar{\lambda}$, of the objective function in (4.22) 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 (4.22) 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 ‘argmax’ set in (4.21) is given by such $h(\omega)$ that

$$\begin{aligned} h(\omega) &= \gamma_2, & \text{if } X(\omega) > \bar{\lambda}, \\ h(\omega) &= \gamma_1, & \text{if } X(\omega) < \bar{\lambda}, \\ h(\omega) &\in [\gamma_1, \gamma_2], & \text{if } X(\omega) = \bar{\lambda}, \\ &\text{and } \int_{\Omega} h d\bar{\mu} = 1. \end{aligned} \tag{4.23}$$

It follows that the ‘argmax’ set in (4.21) is a singleton, and ρ is Hadamard differentiable at X , iff the system (4.23) has a unique solution h . This is equivalent to the following statement:

$$\bar{\mu}\{X < \bar{\lambda}\} = p \quad \text{or} \quad \bar{\mu}\{X > \bar{\lambda}\} = 1 - p. \tag{4.24}$$

If the quantile $\bar{\lambda}$ is not unique then the set of p -quantiles has $\bar{\mu}$ -measure zero, and thus the differentiability condition (4.24) can be understood as holding for any (or for all) p -quantiles. In summary, ρ is Hadamard differentiable at X iff condition (4.24) holds for a p -quantile $\bar{\lambda}$. Note that condition (4.24) always holds true if the set $\{\omega \in \Omega : X(\omega) = \bar{\lambda}\}$ 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 $\rho(\cdot) = \text{CV@R}_p[\cdot]$. Therefore $\text{CV@R}_p[X]$ is equal to the right hand side of (4.19) for

$$\mathcal{A} := \left\{ h \in \mathcal{Y} : 0 \leq h(\omega) \leq (1 - p)^{-1}, \text{ a.e. } \omega \in \Omega, \int_{\Omega} h d\bar{\mu} = 1 \right\}. \tag{4.25}$$

The dual representation and formulas for the set \mathcal{A} and the subdifferential of $\text{CV@R}_p[X]$ were derived in [21, examples 12 and 20] in the space $\mathcal{X} := \mathcal{L}_2(\Omega, \mathcal{F}, \bar{\mu})$.

5 Risk Averse Functions

Let $\bar{\mu}$ be a (reference) probability measure on (Ω, \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 expectation) are taken with respect to the reference measure $\bar{\mu}$. For a σ -algebra $\mathcal{G} \subset \mathcal{F}$ consider the mapping

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

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 $\mathbb{E}[X | \mathcal{G}]$ is defined up to a set of $\bar{\mu}$ -measure zero. That is, any two versions of $\mathbb{E}[X | \mathcal{G}](\omega)$ 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 $\mathbb{E}[X | \mathcal{G}]$.

Definition 1 We say that a risk function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is risk averse (with respect to $\bar{\mu}$) if for every σ -algebra $\mathcal{G} \subset \mathcal{F}$ the following inequality holds

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

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

$$\rho^*(\mu) \geq \rho^*(P_{\mathcal{G}}^*(\mu)) \quad \text{for all } \mu \in \mathcal{Y}. \quad (5.3)$$

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

Suppose for the moment that every measure $\mu \in \mathcal{Y}$ 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 \mathcal{Y} with the corresponding linear space of $\bar{\mu}$ -integrable functions. Take, for example, $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ and $\mathcal{Y} := \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$ for some $p \in [1, +\infty)$ and $1/p + 1/q = 1$. Note then even so, ρ^* is not a risk function since it does not satisfy the monotonicity property, and $\rho^*(h + a) = \rho^*(h)$ for $h \in \mathcal{Y}$ and $a \in \mathbb{R}$. We have that for $X \in \mathcal{X}$ and $h \in \mathcal{Y}$,

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

It follows that $P_{\mathcal{G}}^*(\cdot) = \mathbb{E}[\cdot | \mathcal{G}]$.

Proposition 4 Let ρ be a risk function satisfying assumptions (A1)–(A3). Suppose that ρ is lower semicontinuous. Then ρ is risk averse if and only if ρ^* is risk averse.

Proof. Consider a σ -algebra $\mathcal{G} \subset \mathcal{F}$. Suppose that ρ^* is risk averse. By Theorem 1 we have

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

Since $\langle \mu, P_{\mathcal{G}}(X) \rangle = \langle P_{\mathcal{G}}^*(\mu), X \rangle$ and because of (5.3), it follows that

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

By making change of variables $\nu = P_{\mathcal{G}}^*(\mu)$, we obtain

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

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 $\mathcal{A} \subset \mathcal{P}$, which can be written in form (2.7). In that case we have that ρ^* , and hence ρ , is risk averse iff for every σ -algebra $\mathcal{G} \subseteq \mathcal{F}$ the following relation holds true:

$$P_{\mathcal{G}}^*(\mathcal{A}) \subseteq \mathcal{A}. \quad (5.4)$$

In particular, for $\mathcal{G} = \{\emptyset, \Omega\}$ and $\mu \in \mathcal{A}$ we have that $P_{\mathcal{G}}^*(\mu) = \bar{\mu}$, and hence it follows from (5.4) that $\bar{\mu}$ should be an element of \mathcal{A} . Recall that if $\mathcal{X} := \mathcal{L}_p(\Omega, \mathcal{F}, \bar{\mu})$ and $\mathcal{Y} := \mathcal{L}_q(\Omega, \mathcal{F}, \bar{\mu})$, then $P_{\mathcal{G}}^*(\cdot) = \mathbb{E}[\cdot|\mathcal{G}]$. Therefore in that case, ρ^* and ρ are risk averse iff for any σ -algebra $\mathcal{G} \subseteq \mathcal{F}$ the following holds

$$\mathbb{E}[h|\mathcal{G}] \in \mathcal{A}, \quad \forall h \in \mathcal{A}. \quad (5.5)$$

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

Let us return to Example 2 and consider the function ρ defined in (4.9). We know that this function can be represented as

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

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 $\|h\|_q \leq c$ and $h(\omega) \geq 0$ for a.e. $\omega \in \Omega$. Since $\mathbb{E}[h] = \mathbb{E}[\mathbb{E}[h|\mathcal{G}]]$, it follows that

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

Moreover, $\|\mathbb{E}[h|\mathcal{G}]\|_q \leq \|h\|_q$ and $\mathbb{E}[h|\mathcal{G}](\omega) \geq 0$ for a.e. $\omega \in \Omega$. Thus condition (5.5) is satisfied, and hence ρ is risk averse. Similar considerations apply to Example 1.

Consider now the risk function $\rho(\cdot) := \text{CV@R}_p[\cdot] : \mathcal{X} \rightarrow \mathbb{R}$ discussed in Example 3. Here $\mathcal{X} := \mathcal{L}_1(\Omega, \mathcal{F}, \bar{\mu})$ and $p \in (0, 1)$. It immediately follows from the description (4.25) of the corresponding set \mathcal{A} that condition (5.5) 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 $\varepsilon_1 \in [0, 1]$.

Using (A3) and setting $\mathcal{G} = \{\emptyset, \Omega\}$, we obtain that a risk averse function ρ satisfies the inequality $\rho(X) \geq \langle \bar{\mu}, X \rangle$ 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 (5.4) is not implied by the convexity of the set \mathcal{A} .

6 Optimization of Risk Functions

Let \mathcal{Z} be a vector space and consider a mapping $F : \mathcal{Z} \rightarrow \mathcal{X}$. As in section 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 (Ω, \mathcal{F}) . Consider the problem

$$\text{Min}_{z \in S} \{ \psi(z) := \rho(F(z)) \}, \quad (6.1)$$

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

Suppose that the mapping $F : \mathcal{Z} \rightarrow \mathcal{X}$ is convex and the function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ is proper and lower semicontinuous, and that assumptions (A1)–(A3) are satisfied. By Theorem 2 we can use representation (2.6) to write problem (6.1) in the form

$$\text{Min sup}_{z \in S \quad \mu \in \mathcal{P}} \Xi(z, \mu), \quad (6.2)$$

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

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

As we mentioned earlier, under the above assumptions, the function $\rho^*(\cdot)$ is also proper. We have here that 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, and 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 5 *Suppose that $\mathcal{Z} = \mathbb{R}^n$, the mapping $F : \mathbb{R}^n \rightarrow \mathcal{X}$ is convex and the function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ 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*

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

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 *Suppose that \mathcal{X} is a Banach space, $\mathcal{Y} = \mathcal{X}^*$, the mapping $F : \mathbb{R}^n \rightarrow \mathcal{X}$ is convex, the function $\rho : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ 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 and weakly* closed in \mathcal{X}^* . By 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., [2, Theorem 6.2.7]). ■

Let $\Xi(z, \mu)$ be the function defined in (6.3). Suppose that the assumptions of Proposition 5 or 6 hold true. Consider elements $\hat{z} \in S$ and $\hat{\mu} \in \mathcal{P}$. Since 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 $\Xi(z, \mu)$. Note that, since $\hat{\mu}$ 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), then \hat{z} is an optimal solution of the problem

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

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

$$\text{Min sup}_{z \in S \quad \mu \in \mathcal{A}} \mathbb{E}_{\mu}[F(z)], \quad (6.6)$$

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

If we cannot use Proposition 6, to ensure existence of an optimal solution of problem (6.4), we need additional conditions.

Theorem 3 *Let $F : \mathcal{Z} \rightarrow \mathcal{X}$ be convex. Suppose that a point $\hat{z} \in S$ is an optimal solution of problem (6.1) and $\rho(\cdot)$ is subdifferentiable at $\hat{X} := F(\hat{z})$. Then there exists a measure $\hat{\mu} \in \partial\rho(\hat{X})$ such that \hat{z} is an optimal solution of problem (6.5).*

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

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

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

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

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

$$\min_{(z, X) \in U} \rho(X) \tag{6.8}$$

with

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

Since \hat{z} is an optimal solution of (6.1), the pair (\hat{z}, \hat{X}) constitutes an optimal solution of (6.8). By the convexity of F , the set U is convex. The optimality of (\hat{z}, \hat{X}) and the subdifferentiability of $\rho(\cdot)$ imply that there exists a subgradient $\hat{\mu} \in \partial\rho(\hat{X})$ such that

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

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

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

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 $\bar{X} \in \mathcal{X}$, then it is subdifferentiable at \bar{X} . The risk functions in Examples 1, 2 and 3 are continuous, and therefore they are subdifferentiable everywhere.

7 Value of Perfect Information

Let us consider, as in section 6, a mapping $F : \mathcal{Z} \rightarrow \mathcal{X}$ and a set $S \subset \mathcal{Z}$. In order to avoid technical difficulties, we assume in this section that $\mathcal{Z} = \mathbb{R}^n$. Moreover, we assume that the space \mathcal{X} is *solid* in the sense that for every two elements $X_1, X_2 \in \mathcal{X}$ and every \mathcal{F} -measurable function X satisfying $X_1(\cdot) \leq X(\cdot) \leq X_2(\cdot)$, the function X is an element of \mathcal{X} .

Let us define the operation “inf” on the family $\{F(z) : z \in S\}$ of elements of \mathcal{X} as the pointwise infimum:

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

Assume that there exists $\phi \in \mathcal{X}$ such that $f(z, \omega) \geq \phi(\omega)$ for all $z \in S$ and all $\omega \in \Omega$. Then for any $z_0 \in S$ we have that

$$\phi(\omega) \leq \inf_{z \in S} f(z, \omega) \leq f(z_0, \omega), \quad \omega \in \Omega.$$

Moreover, the function $\inf_{z \in S} f(z, \omega)$ depends measurably on ω (e.g., [19, Theorem 14.37]), and therefore $\inf_{z \in S} F(z) \in \mathcal{X}$.

By the monotonicity assumption (A2) we have that for any $z \in S$, the inequality

$$\rho(F(z)) \geq \rho\left(\inf_{z \in S} F(z)\right)$$

holds true. Consequently

$$\inf_{z \in S} \rho(F(z)) \geq \rho\left(\inf_{z \in S} F(z)\right).$$

The difference

$$\text{RVPI}_\rho := \inf_{z \in S} \rho(F(z)) - \rho\left(\inf_{z \in S} F(z)\right) \quad (7.1)$$

can be called the *risk value of perfect information*. Suppose that $\rho(\cdot)$ is lower semicontinuous and positively homogeneous, and assumptions (A1)–(A3) are satisfied. Then it follows from Theorem 2 that representation (2.8) of ρ holds true for some convex set $\mathcal{A} \subset \mathcal{P}$, and hence we can write

$$\text{RVPI}_\rho = \inf_{z \in S} \sup_{\mu \in \mathcal{A}} \mathbb{E}_\mu[F(z)] - \sup_{\mu \in \mathcal{A}} \mathbb{E}_\mu\left[\inf_{z \in S} F(z)\right]. \quad (7.2)$$

Moreover, under the assumptions of Proposition 5 or 6, 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 (7.2) we obtain

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

Therefore RVPI_ρ can be estimated from below and above as follows

$$\inf_{\mu \in \mathcal{A}} \text{EVPI}_\mu \leq \text{RVPI}_\rho \leq \sup_{\mu \in \mathcal{A}} \text{EVPI}_\mu, \quad (7.3)$$

where

$$\text{EVPI}_\mu := \inf_{z \in S} \mathbb{E}_\mu[F(z)] - \mathbb{E}_\mu\left[\inf_{z \in S} F(z)\right] \quad (7.4)$$

is the expected value of perfect information with respect to the probability measure μ .

It is said that a linear space \mathcal{M} of \mathcal{F} -measurable functions $Z : \Omega \rightarrow \mathbb{R}^n$ is *decomposable* if for every $Z \in \mathcal{M}$ and $B \in \mathcal{F}$, and every bounded and \mathcal{F} -measurable function $W : \Omega \rightarrow \mathbb{R}^n$, 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 [19, p.676]). If \mathcal{M} is a decomposable linear space of \mathcal{F} -measurable functions and $g : \mathbb{R}^n \times \Omega \rightarrow \overline{\mathbb{R}}$ is a random lower semicontinuous function³, then for any probability measure μ on (Ω, \mathcal{F}) the following interchangeability formula holds

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

(Rockafellar and Wets [19, Theorem 14.60]). It is possible to extend this result to risk functions as follows. With $F : \mathbb{R}^n \rightarrow \mathcal{X}$ and $Z \in \mathcal{M}$ we associate an element $F_Z \in \mathcal{X}$ defined as follows

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

Theorem 4 *Let \mathcal{M} be a decomposable space, ρ be a risk function, and let $F : \mathbb{R}^n \rightarrow \mathcal{X}$ be such that the function $f(z, \omega) := [F(z)](\omega)$ is random lower semicontinuous. Then*

$$\rho\left(\inf_{z \in S} F(z)\right) = \inf_{Z \in \mathcal{M}_S} \rho(F_Z), \quad (7.6)$$

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

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

$$\left[\inf_{z \in S} F(z) \right] (\omega) \leq F_Z(\omega)$$

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

$$\rho\left(\inf_{z \in S} F(z)\right) \leq \rho(F_Z),$$

and hence

$$\rho\left(\inf_{z \in S} F(z)\right) \leq \inf_{Z \in \mathcal{M}_S} \rho(F_Z). \quad (7.7)$$

The opposite of inequality (7.7) can be proved in the same way as in the proof of Theorem 14.60 in [19]. ■

8 Dualization of Nonanticipativity Constraints

Consider the framework of section 6 with $\mathcal{Z} = \mathbb{R}^n$ and convex mapping $F : \mathcal{Z} \rightarrow \mathcal{X}$. We can write the optimization problem (6.1) in the following equivalent form

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

³Random lower semicontinuous functions are also called *normal integrands* (see Definition 14.27 in [19, p.676]).

The constraints $Z(\omega) = v$, $\omega \in \Omega$, in the above problem, represent the so-called *nonanticipativity* constraints. The equivalence of (6.1) and (8.1) is evident. The study of nonanticipativity as a constraint was initiated in [18]. We show in this section that in the context of risk measures a corresponding duality relation can be obtained as well.

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, where $\bar{\mu}$ is a probability measure on (Ω, \mathcal{F}) and $p \in [1, +\infty)$. Unless stated otherwise, all expectations and probabilistic statements in this section are made with respect to $\bar{\mu}$. With problem (8.1) is associated 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 $\mathbb{E}[\lambda] \neq 0$, and to $L(Z, 0, \lambda)$ if $\mathbb{E}[\lambda] = 0$. Therefore, the (Lagrangian) dual problem of problem (8.1) takes on the form:

$$\text{Max}_{\lambda \in \mathcal{M}^*} \left\{ \inf_{Z \in \mathcal{M}_S} L(Z, \lambda) \right\} \quad \text{subject to } \mathbb{E}[\lambda] = 0, \quad (8.2)$$

where

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

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., [17],[4, section 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:

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

$$\bar{Z} \in \arg \min_{Z \in \mathcal{M}_S} L(Z, \bar{\lambda}). \quad (8.5)$$

We assume in the remainder of this section that

$$\inf_{Z \in \mathcal{M}_S} \sup_{\mathbb{E}[\lambda]=0} L(Z, \lambda) = \sup_{\mathbb{E}[\lambda]=0} \inf_{Z \in \mathcal{M}_S} L(Z, \lambda), \quad (8.6)$$

i.e., that there is no duality gap between problems (8.1) and (8.2). Suppose that $\rho(\cdot)$ satisfies (A1)–(A3) and is lower semicontinuous. It follows then by Theorem 2 that

$$L(Z, \lambda) = \sup_{h \in \mathcal{P}} \left\{ \mathbb{E}[hF_Z + \lambda^T Z] - \rho^*(h) \right\}, \quad (8.7)$$

with \mathcal{P} denoting the set of densities $h \in \mathcal{M}^*$ of probability measures. Assuming that the corresponding ‘min’ and ‘sup’ operations can be interchanged (as in Propositions 5 and 6) we obtain that

$$\begin{aligned} \inf_{Z \in \mathcal{M}_S} L(Z, \lambda) &= \inf_{Z \in \mathcal{M}_S} \sup_{h \in \mathcal{P}} \left\{ \mathbb{E}[hF_Z + \lambda^T Z] - \rho^*(h) \right\} \\ &= \sup_{h \in \mathcal{P}} \left\{ \inf_{Z \in \mathcal{M}_S} \mathbb{E}[hF_Z + \lambda^T Z] - \rho^*(h) \right\} \\ &= \sup_{h \in \mathcal{P}} \left\{ \mathbb{E} \left(\inf_{z \in S} [h(\omega)f(z, \omega) + \lambda^T(\omega)z] \right) - \rho^*(h) \right\}. \end{aligned}$$

The last equation follows by the decomposability of \mathcal{M} .

Therefore we obtain that, under the specified assumptions, the following duality relation holds

$$\inf_{z \in S} \rho(F(z)) = \sup_{\substack{\mathbb{E}[\lambda]=0 \\ h \in \mathcal{P}}} D(\lambda, h), \quad (8.8)$$

where the function

$$D(\lambda, h) := \mathbb{E} \left\{ \inf_{z \in S} [h(\omega)f(z, \omega) + \lambda^T(\omega)z] \right\} - \rho^*(h) \quad (8.9)$$

can be interpreted as the dual function associated with the composite-risk optimization problem (6.1). Recall that if the risk function $\rho(\cdot)$ is positively homogeneous, then $\rho^*(\cdot)$ is the indicator function of the set \mathcal{A} defined in (2.7). Note also that the calculation of the dual function $D(\lambda, h)$ decomposes into problems associated with elementary events $\omega \in \Omega$:

$$\text{Min}_{z \in S} \{h(\omega)f(z, \omega) + \lambda^T(\omega)z\},$$

and then taking the corresponding integral of the obtained function. This creates a potential for developing decomposition methods for optimizing risk functions.

References

- [1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, Coherent measures of risk, *Mathematical Finance* 9 (1999) 203–228.
- [2] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Wiley, New York, 1984.
- [3] P. Billingsley, *Probability and Measure*, Wiley, New York, 1995.
- [4] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [5] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Springer-Verlag, Berlin, 1977.

- [6] H. Föllmer and A. Schied, Convex measures of risk and trading constraints, *Finance and Stochastics* 6 (2002) 429–447.
- [7] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems*, North-Holland Publishing Company, Amsterdam, 1979.
- [8] W.K. Klein Haneveld, *Duality in Stochastic Linear and Dynamic Programming*, Lecture Notes in Economics and Mathematical Systems 274, Springer-Verlag, New York, 1986.
- [9] V.L. Levin, *Convex Analysis in Spaces of Measurable Functions and Its Applications in Economics*, Nauka, Moscow, 1985 (in Russian).
- [10] M.O. Lorenz, Methods of measuring concentration of wealth, *Journal of the American Statistical Association* 9 (1905) 209–219.
- [11] H.M. Markowitz, *Portfolio selection*, *Journal of Finance*, 7 (1952), 77–91.
- [12] H.M. Markowitz, *Mean–Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1987.
- [13] W. Ogryczak and A. Ruszczyński, From stochastic dominance to mean–risk models: semideviations as risk measures, *European Journal of Operational Research*, 116 (1999), 33–50.
- [14] W. Ogryczak and A. Ruszczyński, On consistency of stochastic dominance and mean–semideviation models, *Mathematical Programming*, 89 (2001), 217–232.
- [15] W. Ogryczak and A. Ruszczyński, Dual stochastic dominance and related mean–risk models, *SIAM Journal on Optimization*, 13 (2002), 60–78.
- [16] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1989.
- [17] R.T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1974.
- [18] R.T. Rockafellar and R.J-B. Wets, *Stochastic convex programming: basic duality*, *Pacific J. Math.* 62 (1976), 173–195.
- [19] R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [20] R.T. Rockafellar and S.P. Uryasev, Optimization of conditional value-at-risk, *The Journal of Risk*, 2 (2000), 21–41.
- [21] R.T. Rockafellar, S. Uryasev and M. Zabarankin, Deviation measures in risk analysis and optimization, Research Report 2002-7, Department of Industrial and Systems Engineering, University of Florida.
- [22] A. Shapiro and S. Ahmed, On a class of minimax stochastic programs, *SIAM Journal on Optimization*, accepted for publication.
- [23] V. Strassen, The existence of probability measures with given marginals, *Annals of Mathematical Statistics* 38 (1965) 423–439.