

## CONVEX RISK MEASURES BASED ON DIVERGENCE

PAUL DOMMEL AND ALOIS PICHLER

**ABSTRACT.** Risk measures connect probability theory or statistics to optimization, particularly to convex optimization. They are nowadays standard in applications of finance and in insurance involving risk aversion.

This paper investigates a wide class of risk measures on Orlicz spaces. The characterizing function describes the decision maker's risk assessment towards increasing losses. We link the risk measures to a crucial formula developed by Rockafellar for the Average Value-at-Risk based on convex duality, which is fundamental in corresponding optimization problems. We characterize the dual and provide complementary representations.

### 1. INTRODUCTION

Risk measures are of fundamental importance in assessing risk, they have numerous applications in finance and in actuarial mathematics. A cornerstone is the Average Value-at-Risk, which has been considered in insurance first. Rockafellar and Uryasev [39, 40] develop its dual representation, which is an important tool when employing risk measures for concrete optimization. Even more, the Average Value-at-Risk is the major building block in what is known as the Kusuoka representation. The duality relations are also elaborated in Ogryczak and Ruszczyński [24, 25].

Risk measures are most typically considered on Lebesgue spaces as  $L^1$  or  $L^\infty$ , although these are not the most general Banach space to consider them. An important reason for choosing this domain is that risk measures are Lipschitz continuous on  $L^\infty$ .

A wide class of risk measures can be properly defined on function spaces as Orlicz spaces. These risk functionals get some attention in Bellini et al. [10], while Bellini and Rosazza Gianin [9], Cheridito and Li [14, 15] elaborate their general properties. Delbaen and Owari [17] investigate risk aversion on Orlicz spaces as well, but they consider a somewhat wider class of risk functionals, which is not necessarily law invariant.

Ahmadi-Javid [2] considers a specific risk measure — the Entropic Value-at-Risk — which is associated with Kullback–Leibler divergence. Delbaen [16] elaborates its Kusuoka representation and [4] present the natural domain. This paper as well

---

2020 *Mathematics Subject Classification.* 91G70, 94A17, 46E30, 49N1.

*Key words and phrases.* Risk measures, Orlicz spaces, duality.

Funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – SFB 1410.

notices possible extensions by involving a more general divergence. Rényi divergence is a specific extension of Kullback–Leibler divergence, which is the building block for the risk measures in [29]. Breuer and Csiszár [12, 13] realize that divergences are indeed essential in assessing risk. The divergence specifies a set of ambiguity, cf. Rockafellar and Royset [35].

This paper addresses general divergences and Fisher information. We derive the result that risk measures, which are built on divergence, are most naturally associated with a specific Orlicz space of random variables. For this reason we investigate them in depth here and identify its natural domain as well as its topological and convex dual.

Risk measures are not solely investigated to measure, to handle or to hedge risk. Rockafellar and Uryasev [38], Rockafellar et al. [41, 42], Krokmal et al. [21] develop a comprehensive theory involving risk measures in four different aspects, which are all interconnected. Their concept of *risk quadrangles* has become essential in understanding risk as well (cf. Rockafellar and Royset [36]).

Outline of the paper. The following section recalls essentials from generalized divergence and introduces the notation. Section 3 introduces the  $\varphi$ -divergence risk measure and Section 4 discusses its natural domain and the associated norm. In Section 5 we derive important representations, including the dual representation and the Kusuoka representation. We finally characterize the dual norm and exploit the convincing properties of the risk measure for concrete optimization problems. Section 7 concludes the paper with a closing discussion.

## 2. PRELIMINARIES

In what follows we repeat the definition of risk measures and divergence. The first subsection states the definition and interpretation of risk measures. We further provide some interpretations which cause their outstanding importance in economics.

**2.1. Risk measures.** A risk measure is a function  $\rho$  mapping random variables from some space  $L$  to the reals,  $\rho: L \rightarrow \mathbb{R} \cup \{\infty\}$ . The inherent interpretation is that the random variable  $X$  with random outcomes is associated with the risk  $\rho(X)$ . In insurance, the number  $\rho(X)$  is understood as premium for the insurance policy  $X$ .

Axioms for risk measures have been introduced by [5, 6]. A risk measure is called *coherent* if it satisfies the following axioms (cf. also [34]):

- A1. Monotonicity:  $\rho(X_1) \leq \rho(X_2)$  provided that  $X_1 \leq X_2$  almost surely.
- A2. Translation equivariance:  $\rho(X + c) = \rho(X) + c$  for any  $X \in L$  and  $c \in \mathbb{R}$ .
- A3. Subadditivity:  $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$  for all  $X_1, X_2 \in L$ .
- A4. Positive homogeneity:  $\rho(\lambda X) = \lambda \rho(X)$  for all  $X \in L$  and  $\lambda > 0$ .

The term risk measure is also used in the literature for the Axioms A1–A3, while the term *coherent* specifically refers to the Axiom A4.

The domain  $L$  of the risk functional is often not specified. In what follows we introduce  $\varphi$ -divergence and elaborate the natural domain, which is as large as possible, of the associated risk measures.

**2.2. Divergence.** Divergence is a concept originating from statistics. The divergence quantifies, how much a probability measure deviates from an other measure. We define divergence functions first to introduce the general  $\varphi$ -divergence.

**Definition 2.1** (Divergence function). A convex and lsc. function  $\varphi: \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$  is a *divergence function* if  $\varphi(1) = 0$ ,  $\text{dom}(\varphi) = [0, \infty)$  and

$$(2.1) \quad \lim_{x \rightarrow \infty} \frac{\varphi(x)}{x} = \infty.$$

**Remark 2.2** ( $\varphi$ -divergence). The term divergence function is inspired by  $\varphi$ -divergence. For a divergence function  $\varphi$ , the  $\varphi$ -divergence of a probability measure  $Q$  from  $P$  is given by

$$D_\varphi(Q \parallel P) := \int_\Omega \varphi \left( \frac{dQ}{dP} \right) dP$$

if  $Q \ll P$  and  $\infty$  otherwise. This divergence is an important concept of a non-symmetric distance between probability measures. Kullback–Leibler is the divergence obtained for  $\varphi(x) = x \log x$ . For a detailed discussion of the general  $\varphi$ -divergence we refer to [12, 13].

In what follows we assume that  $\varphi$  is a divergence function satisfying all conditions of Definition 2.1. Associated with  $\varphi$  is its convex conjugate  $\psi$  defined by  $\psi(y) := \sup_{z \in \mathbb{R}} yz - \varphi(z)$ . These two functions satisfy the Fenchel–Young inequality

$$(2.2) \quad xy \leq \varphi(x) + \psi(y), \quad x, y \text{ in } \mathbb{R},$$

and further properties, as stated in the following proposition.

**Proposition 2.3.** *Let  $\varphi$  be divergence function and  $\psi$  its convex conjugate. The following statements hold true:*

- (i)  $\varphi$  and  $\psi$  are continuous on  $(0, \infty)$  and  $(-\infty, \infty)$ , respectively.
- (ii)  $\psi$  is non-decreasing.
- (iii) It holds that  $y \leq \psi(y)$  for every  $y \in \mathbb{R}$ .

*Proof.* For the first assertion we recall [32, Theorem 10.4], which states that a convex function is continuous on the interior of its domain. Therefore continuity of  $\varphi$  is immediate. For continuity of  $\psi$  it is sufficient to demonstrate that  $\psi(y) < \infty$  holds for every  $y \in \mathbb{R}$ . By contraposition we assume there is a point  $y \in \mathbb{R}$  such that

$$\infty = \psi(y) = \sup_{z \in \mathbb{R}} yz - \varphi(z) = \sup_{z \in \text{dom}(\varphi)} yz - \varphi(z) = \sup_{z \geq 0} yz - \varphi(z).$$

The function  $\varphi$  is finite in its domain and thus the supremum can not be attained at some point  $z^* \geq 0$ . We thus have

$$\infty = \psi(y) = \lim_{z \rightarrow \infty} yz - \varphi(z) = \lim_{z \rightarrow \infty} z \left( y - \frac{\varphi(z)}{z} \right)$$

and consequently  $\lim_{z \rightarrow \infty} \left( y - \frac{\varphi(z)}{z} \right) \geq 0$ . This contradicts assumption (2.1), i.e.,  $\frac{\varphi(z)}{z}$  tends to  $\infty$  for  $z \rightarrow \infty$ .

The second assertion (ii) follows from

$$\psi(y_1) = \sup_{z \in \mathbb{R}} y_1 z - \varphi(z) = \sup_{z \geq 0} y_1 z - \varphi(z) \leq \sup_{z \geq 0} y_2 z - \varphi(z) = \psi(y_2)$$

for  $y_1 \leq y_2$ . We finally have that

$$\psi(y) = \sup_{z \in \mathbb{R}} yz - \varphi(z) \geq y \cdot 1 - \varphi(1) = y, \quad y \in \mathbb{R},$$

which completes the proof.  $\square$

### 3. $\varphi$ -DIVERGENCE RISK MEASURES

Ahmadi-Javid [2, 3] introduces the Entropic Value-at-Risk based on Kullback–Leibler divergence and briefly mentions a possible generalization. We pick up and advance this idea and demonstrate that  $\varphi$ -divergence risk measures are indeed coherent risk measures as specified by the Axioms A1–A4 above.

In what follows we deduce further properties of these risk measures, which are of importance in subsequent investigations.

**Definition 3.1** ( $\varphi$ -divergence risk measure). Let  $\varphi$  be a divergence function with convex conjugate  $\psi$ . The  $\varphi$ -divergence risk measure  $\rho_{\varphi, \beta}: L^1 \rightarrow \mathbb{R} \cup \{\infty\}$  is

$$(3.1) \quad \rho_{\varphi, \beta}(X) := \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right),$$

where the coefficient  $\beta > 0$  indicates risk aversion.

**Remark 3.2** (Interpretation and motivation). The divergence function  $\varphi$  characterizes the shape of risk aversion for increasing risk, while the risk aversion coefficient  $\beta$  describes the tendency of an investor to avoid risk.

The risk measure in (3.1) above is well defined for  $X \in L^1$ , as

$$(3.2) \quad \mathbb{E} X \leq \rho_{\varphi, \beta}(X)$$

by Proposition 2.3 (iii). Note, however, that the risk measure may be unbounded, i.e.,  $\rho_{\varphi, \beta}(X) = \infty$ . Further observe that  $\rho_{\varphi, \beta}$  only depends on the expectation and is therefore law invariant, i.e., the risk measure evaluates random variables  $X$  and  $X'$  equally, provided that  $P(X \leq x) = P(X' \leq x)$  for all  $x \in \mathbb{R}$ .

The following proposition demonstrates that  $\rho_{\varphi, \beta}$  is indeed a coherent risk measure.

**Proposition 3.3.** *The functional  $\rho_{\varphi, \beta}$  is a coherent risk measure, it satisfies all Axioms A1–A4 above.*

*Proof.* To demonstrate translation equivariance let  $c \in \mathbb{R}$  be given. Employing the substitution  $\tilde{\mu} := \mu - \frac{c}{t}$  we have that

$$\begin{aligned} \rho_{\varphi, \beta}(X + c) &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X + c}{t} - \mu \right) \right) \\ &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \tilde{\mu} + \frac{c}{t} + \mathbb{E} \psi \left( \frac{X}{t} - \tilde{\mu} \right) \right) = \rho_{\varphi, \beta}(X) + c, \end{aligned}$$

which is translation equivariance, A2. As for positive homogeneity observe that

$$\begin{aligned}\rho_{\varphi,\beta}(\lambda X) &= \inf_{\substack{\mu \in \mathbb{R} \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{\lambda X}{t} - \mu \right) \right) \\ &= \inf_{\substack{\mu \in \mathbb{R} \\ t > 0}} \lambda \tilde{t} \left( \beta + \mu + \mathbb{E} \psi \left( \frac{\lambda X}{\lambda \tilde{t}} - \mu \right) \right) = \lambda \rho_{\varphi,\beta}(X),\end{aligned}$$

where we have substituted  $\tilde{t} := \frac{t}{\lambda}$ .

Monotonicity follows directly from monotonicity of  $\psi$  (Proposition 2.3 (ii)). Indeed, provided that  $X_1 \leq X_2$  we have that

$$\mathbb{E} \psi \left( \frac{X_1}{t} - \mu \right) \leq \mathbb{E} \psi \left( \frac{X_2}{t} - \mu \right),$$

which implies  $\rho_{\varphi,\beta}(X_1) \leq \rho_{\varphi,\beta}(X_2)$ .

As for subadditivity let  $X, Y \in L^1$  be given. Applying Jensen's inequality for the weights  $\frac{t_1}{t_1+t_2}$  and  $\frac{t_2}{t_1+t_2}$  gives

$$\begin{aligned}\rho_{\varphi,\beta}(X) + \rho_{\varphi,\beta}(Y) &\geq \inf_{\substack{\mu_1, \mu_2 \in \mathbb{R} \\ t_1, t_2 > 0}} (t_1 + t_2) \left( \beta + \frac{t_1 \mu_1 + t_2 \mu_2}{t_1 + t_2} + \mathbb{E} \left( \psi \left( \frac{X + Y}{t_1 + t_2} - \frac{t_1 \mu_1 + t_2 \mu_2}{t_1 + t_2} \right) \right) \right) \\ &= \rho_{\varphi,\beta}(X + Y),\end{aligned}$$

as  $t_1 + t_2 > 0$  and  $\frac{t_1 \mu_1 + t_2 \mu_2}{t_1 + t_2} \in \mathbb{R}$ . This proves A3 (subadditivity).  $\square$

**Remark 3.4.** This proof of coherence of  $\rho_{\varphi,\beta}$  does not involve all conditions imposed on  $\varphi$  above. However, the particular condition (2.1) turns out to be of importance for the proper domain of these risk measures, as Section 4 outlines below.

**Remark 3.5** (Bounds). The general inequality

$$0 \leq \rho_{\varphi,\beta}(0) = \inf_{\substack{\mu \in \mathbb{R} \\ t > 0}} t(\mu + \psi(-\mu + \beta)) \leq 0$$

follows from (3.2) for the constant random variable  $X = 0$  and by letting  $t \rightarrow 0$ . The general bounds

$$(3.3) \quad \mathbb{E} X \leq \rho_{\varphi,\beta}(X) \leq \text{ess sup}(X)$$

follow from translation equivariance.

The following proposition exposes the parameter of risk aversion  $\beta$ . We demonstrate that a larger parameter of risk aversion increases the risk assessment for every random variable.

**Proposition 3.6.** *Suppose that  $0 < \beta_1 \leq \beta_2$ . It holds that*

$$\rho_{\varphi,\beta_1}(X) \leq \rho_{\varphi,\beta_2}(X)$$

*for every  $X \in L^1$ . Conversely, for any non-negative random variable  $X \geq 0$  we have that*

$$\rho_{\varphi,\beta_2}(X) \leq \frac{\beta_2}{\beta_1} \rho_{\varphi,\beta_1}(X).$$

*Proof.* It is immediate that

$$t \left( \beta_1 + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right) \leq t \left( \beta_2 + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right), \quad t > 0, \mu \in \mathbb{R},$$

and hence the first assertion.

As for the second inequality assume that  $X$  is non-negative. The inequality

$$\rho_{\varphi, \beta}(X) = t^* \left( \beta_1 + \mu^* + \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right) \right) \geq t^* \beta_1 + \mathbb{E} X \geq t^* \beta_1$$

follows with (3.2), where  $t^*$  denotes the optimal value inside of (3.1) (and 0, if the infimum is not attained). In other words, the set of possible optimal values of  $t$  is bounded by  $\frac{\rho_{\varphi, \beta_1}(X)}{\beta_1}$ . Consequently we have

$$\begin{aligned} \frac{\beta_2}{\beta_1} \rho_{\varphi, \beta_1}(X) &= \frac{\beta_2 - \beta_1}{\beta_1} \rho_{\varphi, \beta_1}(X) + \inf_{\substack{\mu \in \mathbb{R}, \\ 0 < t \leq \frac{\rho_{\varphi, \beta_1}(X)}{\beta_1}}} t \left( \beta_1 + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right) \\ &\geq \inf_{\substack{\mu \in \mathbb{R}, \\ 0 < t \leq \frac{\rho_{\varphi, \beta_1}(X)}{\beta_1}}} t (\beta_2 - \beta_1) + t \left( \beta_1 + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right) \\ &\geq \rho_{\varphi, \beta_2}(X), \end{aligned}$$

the assertion.  $\square$

#### 4. NORMS AND DOMAINS

This section demonstrates that the largest vector space on which  $\varphi$ -divergence risk measures are finite, are specific Orlicz spaces. We further show that  $\varphi$ -divergence norms, which are based on  $\varphi$ -divergence risk measures, are equivalent to certain Orlicz norms on these spaces.

**4.1. Norms associated with risk functionals.** Coherent risk measures induce semi-norms, cf. [19, 27, 28]. Following this setting we introduce  $\varphi$ -divergence norms by

$$(4.1) \quad \|X\|_{\varphi, \beta} := \rho_{\varphi, \beta}(|X|).$$

This is indeed a norm, as  $\|X\|_{\varphi, \beta} = 0$  if and only if  $X = 0$ , as follows from (3.3).

It is a consequence of A1–A4 and the vector space axioms that  $\|\cdot\|_{\varphi, \beta}$  is finite, iff  $\rho_{\varphi, \beta}(\cdot)$  is finite. We therefore consider the risk measure on the set

$$(4.2) \quad \{X \in L^0 : \|X\|_{\varphi, \beta} < \infty\}.$$

**Remark 4.1.** By Proposition 3.6 it follows for  $\beta_1 < \beta_2$  that

$$(4.3) \quad \|X\|_{\varphi, \beta_1} \leq \|X\|_{\varphi, \beta_2} \leq \frac{\beta_2}{\beta_1} \|X\|_{\varphi, \beta_1}.$$

The norms associated with risk functionals are thus equivalent for varying risk aversion parameters  $\beta > 0$ .

**4.2. Orlicz spaces.** In what follows we discuss the spaces (4.2) endowed with norm (4.1). To this end we introduce the Orlicz class with their associated norms first.

**Definition 4.2** (Orlicz norms and spaces). A convex function  $\Phi: [0, \infty) \rightarrow [0, \infty)$  with  $\Phi(0) = 0$ ,

$$\lim_{x \rightarrow 0} \frac{\Phi(x)}{x} = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{\Phi(x)}{x} = \infty$$

and its convex conjugate  $\Psi$  are called a *pair of complementary Young-functions*. Given a pair of complementary Young-functions  $\Phi$  and  $\Psi$ , the norms

$$(4.4) \quad \|X\|_{\Phi} := \sup_{\mathbb{E} \Psi(|Z|) \leq 1} \mathbb{E} X Z \quad \text{and}$$

$$(4.5) \quad \|X\|_{(\Phi)} := \inf \left\{ \lambda > 0: \mathbb{E} \Phi \left( \frac{|X|}{\lambda} \right) \leq 1 \right\}$$

are called *Orlicz norm* and *Luxemburg norm*, respectively. Further, the spaces

$$(4.6) \quad M^{\Phi} := \{X \in L^0: \mathbb{E} \Phi(t|X|) < \infty \text{ for all } t > 0\} \quad \text{and}$$

$$(4.7) \quad L^{\Phi} := \{X \in L^0: \mathbb{E} \Phi(t|X|) < \infty \text{ for some } t > 0\}$$

are called *Orlicz heart* and *Orlicz opace*, respectively.

**Remark 4.3.** The Orlicz norm  $\|\cdot\|_{\Phi}$  and the Luxemburg norm  $\|\cdot\|_{(\Phi)}$  are topologically equivalent. More specifically, it holds that

$$\|X\|_{(\Phi)} \leq \|X\|_{\Phi} \leq 2 \|X\|_{(\Phi)}$$

on  $L^{\Phi}$  (see [31, Theorem 4.8.5]).

The next Lemma relates divergence functions and Young functions.

**Lemma 4.4.** Let  $\varphi$  be a divergence function (cf. Definition 2.1). The function

$$(4.8) \quad \Phi(x) := \begin{cases} 0 & \text{if } x \in [0, 1] \\ \max\{0, \varphi(x)\} & \text{else} \end{cases}$$

is an Young-function (cf. Definition 4.2) and a divergence function (Definition 2.1). Further, for every  $X \in L^1$ , it holds that  $\|X\|_{\varphi, \beta} < \infty$  if and only if  $\|X\|_{\Phi, \beta} < \infty$  and

$$\frac{\beta}{\beta + d} \|X\|_{\varphi, \beta} \leq \|X\|_{\Phi, \beta} \leq \frac{\beta + d}{\beta} \|X\|_{\varphi, \beta},$$

where  $d := \|\varphi - \Phi\|_{L^{\infty}} = \sup_{x \geq 0} |\varphi(x) - \Phi(x)|$ .

*Proof.* For the first assertion it is sufficient to show that  $\Phi$  is convex, as the other properties are evident by the definition of  $\varphi$  and  $\Phi$ . Let  $0 \leq x \leq y$  and  $\lambda \in (0, 1)$  be given. As  $\max\{0, \varphi\}$  is still convex, we may assume  $x \in [0, 1]$  and  $y > 1$ . By employing  $\varphi(1) = 0$ ,  $\max\{0, \varphi(x)\} \geq 0$  and the convexity of  $\max\{0, \varphi\}$ , it follows that  $\Phi$  is non-decreasing on  $[1, \infty)$  and thus on  $[0, \infty)$ . We therefore have

$$\begin{aligned} \Phi(\lambda x + (1 - \lambda)y) &\leq \Phi(\lambda + (1 - \lambda)y) \\ &\leq \lambda \Phi(1) + (1 - \lambda) \Phi(y) = \lambda \Phi(x) + (1 - \lambda) \Phi(y) \end{aligned}$$

and hence the first assertion.

For the second observe that  $d < \infty$  by (2.1) and convexity of  $\varphi$ . Employing the obvious inequality  $\varphi(x) - d \leq \Phi(x)$  we get that

$$\begin{aligned}\Psi(y) &= \sup_{z \in \mathbb{R}} yz - \Phi(z) \leq \sup_{z \in \mathbb{R}} yz - (\varphi(z) - d) \\ &\leq \sup_{z \in \mathbb{R}} yz - \varphi(z) + d = \psi(y) + d\end{aligned}$$

for all  $y \in \mathbb{R}$ . Inserting this into (3.1) it follows  $\|X\|_{\Phi, \beta} < \infty$  if  $\|X\|_{\varphi, \beta} < \infty$  and

$$\begin{aligned}\|X\|_{\Phi, \beta} &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \\ &\leq \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + d + \mu + \mathbb{E} \psi \left( \frac{|X|}{t} - \mu \right) \right) \\ &= \|X\|_{\varphi, \beta+d} \leq \frac{\beta+d}{\beta} \|X\|_{\varphi, \beta}\end{aligned}$$

by (4.3). The proof of the converse statement is analogous.  $\square$

The following two theorems, which are the main results of this section, establish that the domains of divergence risk measures are specific Orlicz spaces.

**Theorem 4.5** (Equivalence of norms). *Let  $\varphi$  be a divergence function and the associated Young-function  $\Phi$  be given from (4.8). It holds that  $\|X\|_{\varphi, \beta} < \infty$  and  $\|X\|_{\Phi, \beta} < \infty$  if and only if  $X \in L^\Psi$ . Furthermore, the norms*

$$\|\cdot\|_{\varphi, \beta}, \quad \|\cdot\|_{\Phi, \beta} \quad \text{and} \quad \|\cdot\|_{\Phi}$$

*are equivalent on  $L^\Psi$ . In particular we have the inequality*

$$(4.9) \quad \frac{1}{\max\{1, \beta\}} \|X\|_{\Phi, \beta} \leq \|X\|_{\Phi} \leq \frac{\Psi(1) + 1}{\min\{1, \beta\}} \|X\|_{\Phi, \beta}$$

*for all  $X \in L^\Psi$ .*

*Proof.* Let be  $X \in L^\Psi$ . By employing (4.3) with  $\beta = 1$  it follows that

$$\frac{1}{\max\{1, \beta\}} \|X\|_{\Phi, \beta} \leq \|X\|_{\Phi, 1} \leq \frac{1}{\min\{1, \beta\}} \|X\|_{\Phi, \beta}$$

and it is thus sufficient to show (4.9) for  $\beta = 1$ . We have that

$$\|X\|_{\Phi, 1} = \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( 1 + \mu + \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \leq \inf_{t > 0} t \left( 1 + \mathbb{E} \Psi \left( \frac{|X|}{t} \right) \right),$$

where the last term is an equivalent expression of the Orlicz norm in (4.4) (see [20, Theorem 10.5]). Therefore, the inequality

$$\frac{1}{\max\{1, \beta\}} \|X\|_{\Phi, \beta} \leq \inf_{t > 0} t \left( 1 + \mathbb{E} \Psi \left( \frac{|X|}{t} \right) \right) = \|X\|_{\Phi} < \infty$$

holds true.

To prove the converse inequality assume  $\|X\|_{\Phi, 1} < \infty$ . By the definition of  $\Psi$  and Proposition 2.3 (iii) we have that  $\Psi(0) = -\inf_{z \in \mathbb{R}} \Phi(z) = 0$  and  $-y + \Psi(y) \geq 0$



for all  $y \in \mathbb{R}$ . Therefore, as  $-y + \Psi(y)$  is a non-negative, convex function which is 0 in the origin, it is non-decreasing on  $[0, \infty)$ . Hence the infimum in (3.1) is *not* attained for  $\mu < 0$  and it follows that

$$\begin{aligned} \|X\|_{\Phi,1} &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( 1 + \mu + \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \\ &= \inf_{\substack{\mu \geq 0, \\ t > 0}} t \left( 1 + \mu + \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right). \end{aligned}$$

Moreover, as  $t$  and  $\Psi$  are non-negative, we get from  $1 + \Psi(1) \geq 1$  that

$$\begin{aligned} &\inf_{\substack{\mu \geq 0, \\ t > 0}} t \left( 1 + \mu + \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \\ &\geq \frac{1}{1 + \Psi(1)} \inf_{\substack{\mu \geq 0, \\ t > 0}} t + \mu t (1 + \Psi(1)) + t \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \\ &= \inf_{\substack{\mu \geq 0, \\ t > 0}} \frac{t + \mu t}{\Psi(1) + 1} \left( 1 + \frac{\mu t}{t + \mu t} \Psi(1) + \frac{t}{t + \mu t} \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \end{aligned}$$

and therefore, by applying Jensen's inequality,

$$\begin{aligned} \infty > \|X\|_{\Phi,1} &\geq \inf_{\substack{\mu \geq 0, \\ t > 0}} \frac{t + \mu t}{\Psi(1) + 1} \left( 1 + \frac{\mu t}{t + \mu t} \Psi(1) + \frac{t}{t + \mu t} \mathbb{E} \Psi \left( \frac{|X|}{t} - \mu \right) \right) \\ &\geq \inf_{\substack{\mu \geq 0, \\ t > 0}} \frac{t + \mu t}{\Psi(1) + 1} \left( 1 + \mathbb{E} \Psi \left( \frac{|X|}{t + \mu t} \right) \right) \geq \frac{1}{\Psi(1) + 1} \|X\|_{\Phi}. \end{aligned}$$

This establishes  $\|X\|_{\Phi,\beta} < \infty \iff X \in L^\Psi$  as well as (4.9). The remaining statement is immediate by Lemma 4.4. This yields the claim.  $\square$

**Theorem 4.6** (Equivalence of spaces). *Let  $\psi$  be the convex conjugate of a divergence function  $\varphi$ .<sup>1</sup> It holds that  $\|X\|_{\varphi,\beta} < \infty$  if and only if  $X \in L^\psi$  and  $(M^\psi, \|\cdot\|_{\varphi,\beta}) \cong (M^\Psi, \|\cdot\|_{\Phi})$  as well as  $(L^\psi, \|\cdot\|_{\varphi,\beta}) \cong (L^\Psi, \|\cdot\|_{\Phi})$  (here,  $\cong$  indicates a continuous isomorphism).*

*Proof.* We have  $\psi(y) - d < \Psi(y) < \psi(y) + d$  as shown in the proof of Lemma 4.4 and hence the setwise identities  $M^\Psi = M^\psi$  and  $L^\Psi = L^\psi$ . The remaining assertion follows from Theorem 4.5.  $\square$

To emphasize the strength of the previous result we provide some propositions which are consequences of Theorem 4.6 and general results on Orlicz space theory.

**Proposition 4.7.** *The pairs  $(M^\psi, \|\cdot\|_{\varphi,\beta})$  and  $(L^\psi, \|\cdot\|_{\varphi,\beta})$  are Banach spaces.*

**Proposition 4.8.** *The simple functions are dense in  $(M^\psi, \|\cdot\|_{\varphi,\beta})$ .*

*Proof.* Cf. [31, Theorem 4.9.1, Theorem 4.12.8].  $\square$

**Proposition 4.9.** *The following duality relations hold true:*

<sup>1</sup>The sets  $M^\psi$  ( $L^\psi$ , resp.) are defined as in (4.6) (in (4.7), resp.).

(i)  $(M^\psi, \|\cdot\|_{\varphi,\beta})^* \cong (L^\varphi, \|\cdot\|_{\varphi,\beta}^*)$ , where  $*$  indicates the dual space (the dual norm, resp.).

(ii) Assume  $\varphi$  satisfies the  $\Delta_2$ -condition, i.e., there exist numbers  $T, k \geq 0$  such that

$$(4.10) \quad \varphi(2x) \leq k\varphi(x) \quad \text{for all } T < x.$$

Then  $(M^\varphi, \|\cdot\|_{\varphi,\beta}) = (L^\varphi, \|\cdot\|_{\varphi,\beta})$  and  $(L^\psi, \|\cdot\|_{\varphi,\beta}) \cong (M^\psi, \|\cdot\|_{\varphi,\beta})^{**}$ .

(iii)  $(M^\psi, \|\cdot\|_{\varphi,\beta})$  is reflexive if and only if  $\varphi$  and  $\psi$  satisfy the  $\Delta_2$ -condition.

*Proof.* Cf. [31, Theorem 4.13.6, Remark 4.13.8 and Theorem 4.13.9].  $\square$

## 5. REPRESENTATIONS

This section establishes the dual representation of  $\varphi$ -divergence risk measures. We further deduce a simple criterion to ensure that the infimum in (3.1) is attained. The Kusuoka's representation relates the  $\varphi$ -divergence risk measures with distortion risk measures, which are of practical importance.

**5.1. Dual Representation.** The subsequent theorem provides the exact shape of the dual representation of the  $\varphi$ -divergence risk measure. [2] gives a similar result for  $L^\infty$ , but this space is *not* dense in  $L^\psi$  as [4, Theorem 3.2] elaborate for the Entropic Value-at-Risk.

**Theorem 5.1** (Dual representation). *For every  $X \in L^\psi$ , the  $\varphi$ -divergence risk measure has the representation*

$$(5.1) \quad \rho_{\varphi,\beta}(X) = \sup_{Z \in M_{\varphi,\beta}} \mathbb{E} X Z,$$

where

$$(5.2) \quad M_{\varphi,\beta} := \{Z \in L^1 : Z \geq 0, \mathbb{E} Z = 1, \mathbb{E} \varphi(Z) \leq \beta\}.$$

In order to prove the dual representation we need to recall a result on so-called normal convex integrands. A function  $g: \Omega \times \mathbb{R} \rightarrow (-\infty, \infty]$  is said to be a *normal convex integrand*, if (i)  $\omega \mapsto g(\omega, x)$  is measurable for every fixed  $x$  and (ii) if  $x \mapsto g(\omega, x)$  is convex, lower semicontinuous and  $\text{int dom}(g(\omega, \cdot)) = \emptyset$  for almost all  $\omega \in \Omega$ . The following theorem is a special case of [33, p. 185, Theorem 3A]. It states that the supremum and expectation can be interchanged for normal convex integrands, if certain conditions are satisfied (the space  $L^1$  is notably decomposable).

**Theorem 5.2** (Interchangeability principle). *Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $g: \Omega \times \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$  a normal convex integrand. Then*

$$\sup_{X \in L^1(\Omega, \mathcal{F}, P)} \int_{\Omega} g(\omega, X(\omega)) P(d\omega) = \int_{\Omega} \sup_{x \in \mathbb{R}} g(\omega, x) P(d\omega)$$

*holds if the left supremum is finite.*

We now establish the dual representation (5.1) of the divergence risk measure.

*Proof of Theorem 5.1.* Let  $X \in L^\psi$  and  $Z \in M_{\varphi,\beta}$  be given. By applying the Fenchel–Young inequality (2.2) inside of the objective function in (3.1) we get for  $Z \in M_{\varphi,\beta}$  that

$$\begin{aligned} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right) &\geq t \left( \beta + \mu + \mathbb{E} \left( \frac{X}{t} - \mu \right) Z - \varphi(Z) \right) \\ &\geq t \mu + t \beta - t \mu \mathbb{E} Z - t \mathbb{E} \varphi(Z) + \mathbb{E} X Z \\ &\geq \mathbb{E} X Z, \end{aligned}$$

provided that  $t > 0$  and  $\mu \in \mathbb{R}$ . Taking the infimum among all  $t > 0$ ,  $\mu \in \mathbb{R}$  on the left hand side and the supremum for all  $Z \in M_{\varphi,\beta}$  on the right hand side it follows that

$$(5.3) \quad \infty > \rho_{\varphi,\beta}(X) \geq \sup_{Z \in M_{\varphi,\beta}} \mathbb{E} X Z.$$

This is the first inequality required (5.1).

As for the converse observe that the constant random variable  $Z \equiv 1$  is feasible and satisfies  $\mathbb{E} \varphi(Z) = \varphi(1) = 0 < \beta$ . This is, as stated in [23, p. 236 Problem 7], a sufficient condition for strong duality for the right problem in (5.1), i.e., there exist Lagrange multipliers  $\mu^* \in \mathbb{R}$  and  $t^* \geq 0$  such that

$$(5.4) \quad \sup_{Z \in M_{\varphi,\beta}} \mathbb{E} X Z = \sup_{Z \geq 0} \mathbb{E} X Z - \mu^* (\mathbb{E} Z - 1) - t^* (\mathbb{E} \varphi(Z) - \beta).$$

Further, by employing  $\inf_{x \geq 0} \varphi(x) > -\infty$  and substituting  $t \bar{\mu} = \mu$  we have that

$$\begin{aligned} &\sup_{Z \geq 0} \mathbb{E} X Z - \mu^* (\mathbb{E} Z - 1) - t^* (\mathbb{E} \varphi(Z) - \beta) \\ &\geq \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} \sup_{Z \geq 0} \mathbb{E} X Z - \mu (\mathbb{E} Z - 1) - t \inf_{x \geq 0} \varphi(x) + t \beta \\ &\geq \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} \sup_{Z \geq 0} \mathbb{E} X Z - \mu (\mathbb{E} Z - 1) - t (\mathbb{E} \varphi(Z) - \beta) \\ &= \inf_{\substack{\bar{\mu} \in \mathbb{R}, \\ t > 0}} t \left( \bar{\mu} + \beta + \sup_{Z \geq 0} \mathbb{E} \left( \left( \frac{X}{t} - \bar{\mu} \right) Z - \varphi(Z) \right) \right) \\ &= \inf_{\substack{\bar{\mu} \in \mathbb{R}, \\ t > 0}} t \left( \bar{\mu} + \beta + \sup_{Z \in L^1} \mathbb{E} \left( \left( \frac{X}{t} - \bar{\mu} \right) Z - \varphi(Z) \right) \right), \end{aligned}$$

where the last equality follows from the condition  $\varphi(z) = \infty$  for  $z < 0$ . Now observe that the inner function  $f(\omega, z) := \left( \frac{X(\omega)}{t} - \bar{\mu} \right) z - \varphi(z)$  is a normal convex integrand, as  $\varphi$  is lower semicontinuous and  $\text{int dom}(\varphi) = (0, \infty) \neq \emptyset$ . Moreover, as  $X \in L^\psi$ , it follows from (2.2) that

$$\sup_{Z \in L^1} \mathbb{E} \left( \left( \frac{X}{t} - \mu \right) Z - \varphi(Z) \right) \leq \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) < \infty$$

for some  $\mu \in \mathbb{R}$  and  $t > 0$ . Therefore, by inserting Theorem 5.1, we have that

$$\begin{aligned} \sup_{M_{\varphi,\beta}} \mathbb{E} X Z &\geq \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \mu + \beta + \sup_{Z \in L^1} \mathbb{E} \left( \frac{X}{t} - \mu \right) Z - \varphi(Z) \right) \\ &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \mu + \beta + \mathbb{E} \left( \sup_{z \in \mathbb{R}} \left( \frac{X}{t} - \mu \right) z - \varphi(z) \right) \right) \\ &= \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \mu + \beta + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right), \end{aligned}$$

which is the desired inequality. This completes the proof.  $\square$

**5.2. Consequences of the dual representation.** The  $\varphi$ -divergence risk measures derive its name from their relation to  $\varphi$  divergence. We provide this relation now explicitly and investigate the dual representation. We further relate the dual representation (5.1) to Haezendonck risk measures.

**Remark 5.3** (Alternative dual representation). Let  $M_{\varphi,\beta}$  as in (5.2) and  $Z \in M_{\varphi,\beta}$ . The random variable  $Z$  satisfies  $Z \geq 0$  and  $\mathbb{E} Z = 1$ . Therefore  $Q_Z$  defined as

$$Q_Z(B) := \mathbb{E}_P \mathbb{1}_B Z$$

is a probability measure.  $Q_Z$  is absolutely continuous with respect to  $P$  and Radon–Nikodym derivative  $\frac{dQ_Z}{dP} = Z$ . Hence we can reformulate the dual representation (5.1) as

$$(5.5) \quad \rho_{\varphi,\beta}(X) = \sup_{Q \ll P} \{ \mathbb{E}_Q X : D_\varphi(Q \parallel P) \leq \beta \},$$

where  $D_\varphi(Q \parallel P)$  is the  $\varphi$ -divergence defined in Remark 2.2.  $\rho_{\varphi,\beta}(X)$  can therefore be interpreted as the largest expected value  $\mathbb{E}_Q X$  over all probability measures  $Q$  within a  $\varphi$ -divergence ball around  $P$ . The divergence function  $\varphi$  characterizes the shape of the ball, while  $\beta$  determines the radius.

**Remark 5.4** (Relationship with Haezendonck risk measures). Suppose  $\varphi$  is a Young-function as in Definition 4.2. Then the dual representation in (5.1) rewrites as

$$\rho_{\varphi,\beta}(X) = \{ \mathbb{E} X Z : Z \geq 0, \mathbb{E} Z = 1, \|Z\|_{(\tilde{\varphi})} \leq 1 \}$$

where  $\tilde{\varphi}$  is the function  $\tilde{\varphi}(\cdot) = \frac{1}{\beta} \varphi(\cdot)$  and  $\|\cdot\|_{(\tilde{\varphi})}$  the corresponding Luxemburg norm (4.5). The dual norm of  $\|\cdot\|_{(\tilde{\varphi})}$  is the Orlicz norm  $\|\cdot\|_{\tilde{\psi}}$ , cf. (4.4), where  $\tilde{\psi}$  is the associated convex conjugate. Interchanging  $\|\cdot\|_{(\tilde{\varphi})}$  by  $\|\cdot\|_{\tilde{\psi}}$  we get

$$\rho(X) = \left\{ \mathbb{E} X Z : Z \geq 0, \mathbb{E} Z = 1, \|Z\|_{\tilde{\psi}} \leq 1 \right\},$$

which is the dual representation of the so-called *Haezendonck–Goovaerts* risk measure (see [9, Proposition 4]). It therefore turns out that the Haezendonck–Goovaerts risk measures are the natural dual counterparts of the  $\varphi$ -divergence risk measures, as the corresponding feasible sets are determined by norms which are dual to each other. For more information on Haezendonck–Goovaerts risk measures see [7], [9] and [18].

Employing the dual representation we derive a simple condition when the infimum in (3.1) is attained.

**Proposition 5.5** (Existence of minimizers). *Let  $X \in L^\psi$  and  $\bar{\alpha}$  be given by*

$$(5.6) \quad \bar{\alpha} := \max \left\{ \alpha \in [0, 1] : \varphi(0) \alpha + \varphi \left( \frac{1}{1-\alpha} \right) (1-\alpha) \leq \beta \right\}.$$

If

$$(5.7) \quad P(X = \text{ess sup}(X)) < 1 - \bar{\alpha}$$

holds true, then the infimum in the defining equation of the risk measure (3.1) is attained.

*Proof.* The assertion is shown in two parts. The first part demonstrates  $\rho_{\varphi, \beta}(X) < \text{ess sup}(X)$  while the second establishes that  $\rho_{\varphi, \beta}(X) = \text{ess sup}(X)$  holds if the infimum is not attained. The assertion then follows by contradiction.

To prove the first part let  $M_{\varphi, \beta}$  as in (5.2),  $\bar{\alpha}$  as in (5.6) and  $X \in L^\psi$  as in (5.7) be given. We choose  $Z \in M_{\varphi, \beta}$ ,  $\alpha \in (\bar{\alpha}, 1 - P(X = \text{ess sup}(X)))$  and  $U$  uniform distributed on  $[0, 1]$ . We further set  $\mu_\alpha^Z := \mathbb{E}(F_Z^{-1}(U) \mid 0 \leq U < \alpha)$  and  $\mu_{1-\alpha}^Z := \mathbb{E}(F_Z^{-1}(U) \mid \alpha \leq U \leq 1)$ . As  $F_Z^{-1}(U)$  and  $Z$  are identically distributed it follows that

$$\begin{aligned} 1 &= \mathbb{E}(F_Z^{-1}(U)) = \mu_\alpha^Z P(0 \leq U < \alpha) + \mu_{1-\alpha}^Z P(\alpha \leq U \leq 1) \\ &= \mu_\alpha^Z \alpha + \mu_{1-\alpha}^Z (1 - \alpha) \end{aligned}$$

and

$$\begin{aligned} \beta &\geq \mathbb{E}(\varphi(F_Z^{-1}(U))) \\ &= \mathbb{E}(\varphi(F_Z^{-1}(U)) \mid 0 \leq U < \alpha) \alpha + \mathbb{E}(\varphi(F_Z^{-1}(U)) \mid \alpha \leq U \leq 1) (1 - \alpha) \\ &\geq \varphi(\mu_\alpha^Z) \alpha + \varphi(\mu_{1-\alpha}^Z) (1 - \alpha) = \varphi(\mu_\alpha^Z) \alpha + \varphi\left(\frac{1 - \alpha \mu_\alpha^Z}{1 - \alpha}\right) (1 - \alpha) \end{aligned}$$

where we employed Jensen's inequality to obtain the second inequality. Additionally, by the definition of  $\bar{\alpha}$  in (5.6), we have that

$$\varphi(0) \alpha + \varphi\left(\frac{1}{1-\alpha}\right) (1-\alpha) > \varphi(0) \bar{\alpha} + \varphi\left(\frac{1}{1-\bar{\alpha}}\right) (1-\bar{\alpha}) = \beta.$$

From this and the continuity of  $\varphi$  we conclude that there exists a positive constant  $c$ , not depending on  $Z$ , such that  $\mu_\alpha^Z \geq c$  holds for every  $Z \in M_{\varphi, \beta}$ . Hence, by employing the covariance inequality in [43, Theorem 4], it follows that

$$\begin{aligned} \mathbb{E} X Z &\leq \int_0^1 F_X^{-1}(u) F_Z^{-1}(u) du \\ &= \int_0^\alpha F_X^{-1}(u) F_Z^{-1}(u) du + \int_\alpha^1 F_X^{-1}(u) F_Z^{-1}(u) du \\ &\leq F_X^{-1}(\alpha) \left( \int_0^\alpha F_Z^{-1}(u) du \right) + F_X^{-1}(1) \left( \int_\alpha^1 F_Z(u) du \right) \\ &\leq F_X^{-1}(\alpha) \alpha c + F_X^{-1}(1) (1 - \alpha c) \end{aligned}$$

and consequently

$$\begin{aligned}\rho_{\varphi,\beta}(X) &= \sup_{Z \in M_{\varphi,\beta}} \mathbb{E} X Z \leq F_X^{-1}(\alpha) \alpha c + F_X^{-1}(1) (1 - \alpha c) \\ &< F_X^{-1}(1) = \text{ess sup}(X),\end{aligned}$$

which demonstrates the first part.

For the second note that the infimum in (3.1) is *not* attained if and only if  $t$  inside of

$$\inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X}{t} - \mu \right) \right)$$

tends towards 0. Hence we have  $t^* = 0$  for the Lagrange multiplier  $t^*$  in (5.4). It thus follows that

$$\rho_{\varphi,\beta}(X) = \sup_{M_{\varphi,\beta}} \mathbb{E} X Z = \sup_{Z \geq 0} \mathbb{E} X Z - \mu^*(\mathbb{E} Z - 1) = \text{ess sup}(X).$$

This completes the proof.  $\square$

**5.3. Spectral representation.** The  $\varphi$ -divergence risk measure  $\rho_{\varphi,\beta}$  is coherent and law-invariant and thus has a Kusuoka representation (cf. [22]). We give the representation in terms of spectral risk measures, which is equivalent to the Kusuoka representation. We derive this representation from the dual (5.1) based on the general approach elaborated in [30].

**Proposition 5.6** (Spectral representation). *The spectral representation of a  $\varphi$ -divergence risk measure  $\rho_{\varphi,\beta}$  for  $X \in L^\psi$  is*

$$(5.8) \quad \rho_{\varphi,\beta}(X) = \sup_{\sigma} \int_0^1 \sigma(u) F_X^{-1}(u) du,$$

where the supremum is taken over all non-decreasing  $\sigma: [0, 1] \rightarrow [0, \infty]$  with  $\int_0^1 \sigma(u) du = 1$  and

$$\int_0^1 \varphi(\sigma(u)) du \leq \beta.$$

**Remark 5.7.** Every functional of the shape

$$\rho_{\sigma}(X) = \int_0^1 \sigma(u) F_X^{-1}(u) du,$$

where  $\sigma: [0, 1] \rightarrow [0, \infty]$  is non-decreasing with  $\int_0^1 \sigma(u) du = 1$ , is a coherent risk measure itself. It is called *distortion risk measure* in [26] or *spectral risk measure* in [1].

The spectral representation (5.8) is beneficial to derive bounds as

$$\rho_{\sigma}(X) \leq \rho_{\varphi,\beta}(X) \quad \text{for all } X \in L^\psi.$$

We provide an example next.

**Example 5.8** (AV@R bound). For some fixed  $\alpha \in (0, 1)$  we set  $\sigma_\alpha(\cdot) = \frac{1}{1-\alpha} \mathbb{1}_{[\alpha, 1]}(\cdot)$ . The associated distortion risk measure is

$$\rho_{\sigma_\alpha}(X) = \int_0^1 \sigma(u) F_X^{-1}(u) du = \frac{1}{1-\alpha} \int_\alpha^1 F_X^{-1}(u) du$$

which is called *Average Value-at-Risk* and denoted as  $\text{AV@R}_\alpha(X)$ . If

$$(5.9) \quad \int_0^1 \varphi(\sigma_\alpha(u)) du = \varphi(0)\alpha + \varphi\left(\frac{1}{1-\alpha}\right)(1-\alpha) \leq \beta$$

holds, then  $\sigma_\alpha$  is contained in the set of functions, over which the supremum on the left side of (5.8) is taken. We hence obtain

$$\text{AV@R}_\alpha(X) = \rho_{\sigma_\alpha}(X) \leq \rho_{\varphi, \beta}(X) \quad \text{for all } X \in L^\psi$$

for every  $\alpha$  such that (5.9) is satisfied. Therefore, by inserting definition of  $\bar{\alpha}$  in (5.6), we have that

$$\text{AV@R}_\alpha(X) \leq \rho_{\varphi, \beta}(X), \quad \alpha \leq \bar{\alpha}.$$

The latter inequality is of importance, as the Average Value-at-Risk is the most important risk measure in finance and in insurance. The inequality generalizes a corresponding inequality for the Entropic Value-at-Risk, cf. [2, Proposition 3.2].

## 6. CHARACTERIZATION OF THE DUAL AND APPLICATIONS

The Banach space  $L^\psi$  is, by Proposition 4.9, not reflexive, in general. By James's theorem, there are continuous linear functionals, which do not attain their supremum on the closed unit ball. This section characterizes functionals of the dual, which attain their supremum on the closed unit ball. We characterize the optimal dual random variables in (5.1) by an explicit relation to optimality of  $t$  and  $\mu$  in the defining equation (3.1). We further establish an explicit representation of the dual norm of  $\|\cdot\|_{\varphi, \beta}$ . We further specify conditions so that the optimal values in (3.1) can be derived based on a system of equations.

$\varphi$ -divergence risk measures are efficiently incorporated into portfolio optimization problems. We demonstrate this property in an explicit example.

**6.1. Characterizing equations.** To elaborate optimality inside of (3.1) and (5.1), we state some facts concerning the 'derivatives' of the convex function  $\varphi$  and its conjugate  $\psi$ . Even though they are not necessarily differentiable, they have subderivatives  $\varphi'$  and  $\psi'$  (see [11, Theorem 2.3.12], [32, Theorem 23.4]). These are functions, satisfying the equivalent relations

$$(6.1) \quad \psi'(x)(z-x) \leq \psi(z) - \psi(x) \quad \text{and} \quad \varphi'(y)(z-y) \leq \varphi(z) - \varphi(y),$$

and

$$(6.2) \quad x\psi'(x) = \varphi(\psi'(x)) + \psi(x) \quad \text{and} \quad y\varphi'(y) = \varphi(y) + \psi(\varphi'(y))$$

for all  $x, z \in \mathbb{R}, y \geq 0$ . The subderivatives  $\varphi'$  and  $\psi'$  are, in general, not unique. Nevertheless, they are uniquely determined, except for at most countably many

points. Any function satisfying (6.1) is non-decreasing and therefore measurable. Hence the system of equations

$$(6.3) \quad 1 = \mathbb{E} \psi' \left( \frac{X}{t} - \mu \right),$$

$$(6.4) \quad \beta = \mathbb{E} \varphi \left( \psi' \left( \frac{X}{t} - \mu \right) \right)$$

is well specified.

In what follows we demonstrate that solutions of the equations (6.3)–(6.4) characterize optimal solutions  $t^*$  and  $\mu^*$  in the defining equation (3.1). They specify the random variable  $Z^*$  in the dual space maximizing the functional  $\mathbb{E} X Z$  among all  $Z \in M_{\varphi, \beta}$ .

**Theorem 6.1.** *Let be  $X \in L^\psi$ ,  $M_{\varphi, \beta}$  as in (5.2) and  $\psi'$  satisfying (6.1). Suppose  $\mu^* \in \mathbb{R}$  and  $t^* > 0$  solve of the characterizing equations (6.3)–(6.4). Then they are the optimal values in (3.1). Furthermore, the random variable*

$$Z^* := \psi' \left( \frac{X}{t^*} - \mu^* \right)$$

*is optimal in (5.1), i.e.,*

$$\sup_{Z \in M_{\varphi, \beta}} \mathbb{E} X Z = \mathbb{E} X Z^* = t^* \left( \beta + \mu^* + \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right) \right)$$

*and  $Z^* \in M_{\varphi, \beta}$ .*

*Proof.* Let solutions  $t^* > 0$ ,  $\mu^* \in \mathbb{R}$  of (6.3) and (6.4) be given. The assertion  $Z^* \in M_{\varphi, \beta}$  is immediate by the equations (6.3), (6.4) and the fact that  $\varphi(x) = \infty$  holds for  $x < 0$ . Furthermore, by employing (6.2), we have that

$$\begin{aligned} & \mathbb{E} \left( \frac{X}{t^*} - \mu^* \right) Z^* - \varphi(Z^*) \\ &= \mathbb{E} \left( \frac{X}{t^*} - \mu^* \right) \psi' \left( \frac{X}{t^*} - \mu^* \right) - \varphi \left( \psi' \left( \frac{X}{t^*} - \mu^* \right) \right) = \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right). \end{aligned}$$

Hence by (6.3), (6.4) and Theorem 5.1 it follows that

$$\begin{aligned} \rho_{\varphi, \beta}(X) &= \sup_{Z \in M_{\varphi, \beta}} \mathbb{E} X Z \geq \mathbb{E} X Z^* \\ &= t^* \left( \frac{\mathbb{E} X Z^*}{t^*} - \mu^* (\mathbb{E} Z^* - 1) - (\mathbb{E} \varphi(Z^*) - \beta) \right) \\ &= t^* \left( \beta + \mu^* + \mathbb{E} \left( \frac{X}{t^*} - \mu^* \right) Z^* - \varphi(Z^*) \right) \\ &= t^* \left( \beta + \mu^* + \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right) \right) \geq \rho_{\varphi, \beta}(X). \end{aligned}$$

We therefore obtain  $\mathbb{E} X Z^* = \rho_{\varphi, \beta}(X)$  as well as

$$\rho_{\varphi, \beta}(X) = t^* \left( \beta + \mu^* + \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right) \right).$$



Thus  $\mu^*$ ,  $t^*$  and  $Z^*$  are optimal in (3.1) and (5.1), respectively. This is the assertion.  $\square$

**Remark 6.2.** Note that optimal values  $t^*$  and  $\mu^*$  in (3.1) may exist, although the characterizing system (6.3)–(6.4) *cannot* be solved. The existence of solutions depends on the specific choice of the subderivative  $\psi'$ .

Nevertheless, further assumption on the random variable  $X$  and the function  $\psi$  can insure solutions of the system of equations. We present the corresponding result in Section 6.3 below.

**6.2. Dual norm.** This subsection addresses the dual norm

$$(6.5) \quad \|Z\|_{\varphi, \beta}^* := \sup_{\|X\|_{\varphi, \beta} \leq 1} \mathbb{E} X Z$$

of the  $\varphi$ -divergence norms given in (4.1). In what follows, we characterize (6.5) as an optimization problem in one variable, provided that  $\varphi$  satisfies the  $\Delta_2$ -condition (4.10).

Note that  $\varphi \in \Delta_2$  implies

$$\mathbb{E} \varphi(t|Z|) < \infty \text{ for some } t > 0 \iff \mathbb{E} \varphi(t|Z|) < \infty \text{ for all } t > 0$$

as well as  $(M^\psi)^* \cong L^\varphi$  and  $(L^\varphi)^* \cong L^\psi$  (see Proposition 4.9). Thus the expression in (6.5) is finite if and only if  $Z \in L^\varphi$ .

The following lemma states a specific transformation of a random variable  $Z \in L^\varphi$ , which we use later to characterize the dual norm.

**Lemma 6.3.** *Let  $\varphi \in \Delta_2$  and  $Z \in L^\varphi$ . There exists a continuous function  $c_Z: [\mathbb{E}|Z|, \infty) \rightarrow [0, 1]$  such that*

$$(6.6) \quad \mathbb{E} \max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} = 1$$

for all  $\lambda \in [\mathbb{E}|Z|, \infty)$ . If  $\mathbb{E} \varphi \left( \frac{|Z|}{\mathbb{E}|Z|} \right) > \beta$  in addition, then there is a number  $\lambda^* \in (\mathbb{E}|Z|, \infty)$  such that

$$\mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda^*), \frac{|Z|}{\lambda^*} \right\} \right) = \beta$$

is satisfied.

*Proof.* To establish the assertion we recall the *intermediate value theorem*, which states that the equation

$$f(x) = y$$

has a solution  $x^*$ , if  $f$  is continuous and there are  $x_1, x_2$  such that  $f(x_1) \leq y \leq f(x_2)$ .

Let  $Z \in L^\varphi$ . If  $Z$  is constant, the function  $c_Z(\lambda) := 1$  satisfies (6.6). We therefore assume that  $Z$  is non-constant and consider some fixed  $\lambda \in (\mathbb{E}|Z|, \infty)$ . Setting  $f(c) := \mathbb{E} \max \left\{ c, \frac{|Z|}{\lambda} \right\}$  we have that

$$|f(c_2) - f(c_1)| = \left| \mathbb{E} \max \left\{ c_2, \frac{|Z|}{\lambda} \right\} - \mathbb{E} \max \left\{ c_1, \frac{|Z|}{\lambda} \right\} \right| \leq |c_2 - c_1|$$

for all  $c_1, c_2 \in \left[ \operatorname{ess\,inf} \left( \frac{|Z|}{\lambda} \right), 1 \right]$ . Thus  $f$  is Lipschitz continuous and hence continuous. Further we have that

$$\begin{aligned} f \left( \operatorname{ess\,inf} \left( \frac{|Z|}{\lambda} \right) \right) &= \mathbb{E} \left( \frac{|Z|}{\lambda} \right) = \mathbb{E} \max \left\{ \operatorname{ess\,inf} \left( \frac{|Z|}{\lambda} \right), \frac{|Z|}{\lambda} \right\} \\ &< 1 \leq \mathbb{E} \max \left\{ 1, \frac{|Z|}{\lambda} \right\} = f(1) \end{aligned}$$

and thus, by employing the intermediate value theorem,  $f(c^*) = 1$  for some  $c^* \in \left( \operatorname{ess\,inf} \left( \frac{|Z|}{\lambda} \right), 1 \right]$ . Hence (6.6) has for a solution  $c^*(\lambda)$  for every  $\lambda \in (\mathbb{E}|Z|, \infty)$ , which is unique as  $f$  increases strictly on  $\left( \operatorname{ess\,inf} \left( \frac{|Z|}{\lambda} \right), 1 \right]$ . Therefore the function  $c_Z: [\mathbb{E}|Z|, \infty) \rightarrow [0, 1]$  given by

$$c_Z(\lambda) := \begin{cases} \operatorname{ess\,inf} \left( \frac{|Z|}{\mathbb{E}|Z|} \right) & \text{for } \lambda = \mathbb{E}|Z| \\ c^*(\lambda) & \text{for } \lambda \in (\mathbb{E}|Z|, \infty) \end{cases}$$

is well defined and satisfies (6.6) for every  $\lambda \in [\mathbb{E}|Z|, \infty)$ .

To demonstrate the continuity of  $c_Z$ , let  $\lambda_0 \in (\mathbb{E}|Z|, \infty)$  and  $\varepsilon > 0$ . Without loss of generality we may assume that  $\varepsilon$  is sufficiently small such that  $p = P \left( \frac{|Z|}{\lambda_0} \leq c_Z(\lambda_0) - \varepsilon \right) > 0$ . Choosing  $\delta \leq \lambda_0 \varepsilon p$  it follows that

$$\begin{aligned} \mathbb{E} \max \left\{ c_Z(\lambda_0) - \varepsilon, \frac{|Z|}{\lambda} \right\} &\leq \mathbb{E} \max \left\{ c_Z(\lambda_0) - \varepsilon, \frac{|Z|}{\lambda_0 - \delta} \right\} \\ &\leq \frac{\lambda_0}{\lambda_0 - \delta} \mathbb{E} \max \left\{ c_Z(\lambda_0) - \varepsilon, \frac{|Z|}{\lambda_0} \right\} \\ &\leq \frac{\lambda_0}{\lambda_0 - \delta} \left( \mathbb{E} \max \left\{ c_Z(\lambda_0), \frac{|Z|}{\lambda_0} \right\} - \varepsilon p \right) \\ &= \frac{\lambda_0}{\lambda_0 - \delta} (1 - \varepsilon p) \leq 1 \end{aligned}$$

and similarly

$$\begin{aligned} \mathbb{E} \max \left\{ c_Z(\lambda_0) + \varepsilon, \frac{|Z|}{\lambda} \right\} &\geq \mathbb{E} \max \left\{ c_Z(\lambda_0) + \varepsilon, \frac{|Z|}{\lambda_0 + \delta} \right\} \\ &\geq \frac{\lambda_0}{\lambda_0 + \delta} \mathbb{E} \max \left\{ c_Z(\lambda_0) + \varepsilon, \frac{|Z|}{\lambda_0} \right\} \\ &\geq \frac{\lambda_0}{\lambda_0 + \delta} \left( \mathbb{E} \max \left\{ c_Z(\lambda_0), \frac{|Z|}{\lambda_0} \right\} + \varepsilon p \right) \geq 1 \end{aligned}$$

for all  $\lambda \in (\lambda_0 - \delta, \lambda_0 + \delta)$ . We thus get that  $c_Z(\lambda) \in [c_Z(\lambda_0) - \varepsilon, c_Z(\lambda_0) + \varepsilon]$  for all  $\lambda \in (\lambda_0 - \delta, \lambda_0 + \delta)$ , by the intermediate value theorem. This establishes the continuity of  $c_Z$  on  $(\mathbb{E}|Z|, \infty)$ . The (right side) continuity in  $\lambda = \mathbb{E}|Z|$  follows from the fact that

$$\mathbb{E} \max \left\{ \operatorname{ess\,inf} \left( \frac{|Z|}{\mathbb{E}|Z|} \right) + \varepsilon, \frac{|Z|}{\mathbb{E}|Z|} \right\} > 1$$

holds for every  $\varepsilon > 0$ . This demonstrates the first part of the assertion.

For the second we assume  $\mathbb{E} \varphi \left( \frac{|Z|}{\mathbb{E}|Z|} \right) > \beta$  and define the function  $g(\lambda) := \mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} \right)$ . By

$$\mathbb{E} \max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} = 1 \text{ for all } \lambda \in (\mathbb{E}|Z|, \infty),$$

we observe that  $\max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} \rightarrow 1$  almost surely, for  $\lambda \rightarrow \infty$ . It is hence sufficient to show that  $g$  is continuous, as then the assertion follows from

$$0 = \varphi(1) = \lim_{\lambda \rightarrow \infty} g(\lambda) < \beta < \mathbb{E} \varphi \left( \frac{|Z|}{\mathbb{E}|Z|} \right) = g(\mathbb{E}|Z|)$$

and the intermediate value theorem. Let  $(\lambda_n)_{n \in \mathbb{N}} \subset (\mathbb{E}|Z|, \infty)$  such that  $\lambda_n \rightarrow \lambda_0 \in [\mathbb{E}|Z|, \infty)$ . Choosing a number  $M \geq 1$  such that  $\varphi$  is non-decreasing and non-negative for all  $x \geq M$ , we have the estimation

$$\begin{aligned} \left| \varphi \left( \max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} \right) \right| &\leq \sup_{x \in [0, M]} |\varphi(x)| + \varphi \left( \max \left\{ M, \frac{|Z|}{\lambda} \right\} \right) \\ (6.7) \qquad \qquad \qquad &\leq \sup_{x \in [0, M]} |\varphi(x)| + \varphi \left( \max \left\{ M, \frac{|Z|}{\mathbb{E}|Z|} \right\} \right) \end{aligned}$$

for all  $\lambda \in [\mathbb{E}|Z|, \infty)$ . As (6.7) is integrable we can interchange limit and expectation by Lebesgue's Dominated convergence theorem, and thus get

$$\begin{aligned} g(\lambda_0) &= \mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda_0), \frac{|Z|}{\lambda_0} \right\} \right) = \mathbb{E} \left( \lim_{n \rightarrow \infty} \varphi \left( \max \left\{ c_Z(\lambda_n), \frac{|Z|}{\lambda_n} \right\} \right) \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda_n), \frac{|Z|}{\lambda_n} \right\} \right) = \lim_{n \rightarrow \infty} g(\lambda_n) \end{aligned}$$

by Proposition 2.3 (i) and continuity of  $c_Z$ . This demonstrates continuity of  $g$  and consequently the assertion.  $\square$

The dual norm allows the following explicit expression, which reduces the problem to an optimization exercise in a single variable.

**Theorem 6.4.** *For  $\varphi \in \Delta_2$  and  $Z \in L^\varphi$  it holds that*

$$\|Z\|_{\varphi, \beta}^* = \inf \left\{ \lambda \geq \mathbb{E}|Z| : \mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda), \frac{|Z|}{\lambda} \right\} \right) \leq \beta \right\},$$

where  $c_Z$  is the function in Lemma 6.3.

*Proof.* Let be  $Z \in L^\varphi$  and  $M_{\varphi, \beta}$  as in (5.2). If  $\mathbb{E} \varphi \left( \frac{|Z|}{\mathbb{E}|Z|} \right) \leq \beta$  holds, we have that  $\frac{|Z|}{\mathbb{E}|Z|} \in M_{\varphi, \beta}$  and therefore

$$\mathbb{E} X Z \leq \mathbb{E}|Z| \frac{\mathbb{E}|X||Z|}{\mathbb{E}|Z|} \leq \mathbb{E}|Z| \sup_{Y \in M_{\varphi, \beta}} \mathbb{E}|X|Y = \mathbb{E}|Z| \|X\|_{\varphi, \beta}$$

by Theorem 5.1. Hence it holds  $\|Z\|_{\varphi, \beta}^* \leq \mathbb{E}|Z|$ . Conversely, by (3.3), we get that  $\|\text{sign}(Z)\|_{\varphi, \beta} \leq 1$  and thus  $\|Z\|_{\varphi, \beta}^* \geq \mathbb{E}|Z|$ , as  $\mathbb{E} Z \text{sign}(Z) = \mathbb{E}|Z| \geq \mathbb{E}|Z| \|\text{sign}(Z)\|_{\varphi, \beta}$ . We therefore obtain  $\|Z\|_{\varphi, \beta}^* = \mathbb{E}|Z|$ .

Now assume  $\mathbb{E} \varphi \left( \frac{|Z|}{\mathbb{E} |Z|} \right) > \beta$ . Employing Lemma 6.3 we get a number  $\lambda^* \in (\mathbb{E} |Z|, \infty)$  such that

$$(6.8) \quad \mathbb{E} \max \left\{ c_Z(\lambda^*), \frac{|Z|}{\lambda^*} \right\} = 1 \quad \text{and} \quad \mathbb{E} \varphi \left( \max \left\{ c_Z(\lambda^*), \frac{|Z|}{\lambda^*} \right\} \right) = \beta$$

holds. Setting  $Z^* := \max \left\{ c_Z(\lambda^*), \frac{|Z|}{\lambda^*} \right\}$ , and observing  $Z^* \in M_{\varphi, \beta}$  as well as  $\frac{|Z|}{\lambda^*} \leq Z^*$ , it follows from Theorem 5.1 that

$$\frac{\mathbb{E} X Z}{\lambda^*} \leq \frac{\mathbb{E} |X| |Z|}{\lambda^*} \leq \mathbb{E} |X| Z^* \leq \|X\|_{\varphi, \beta}$$

for every  $X \in L^\psi$ . We therefore conclude  $\|Z\|_{\varphi, \beta}^* \leq \lambda^*$ .

To establish the converse inequality, we consider the random variable  $X^* := \max \left\{ 0, \varphi' \left( \frac{|Z|}{\lambda^*} \right) - \varphi'(c_Z(\lambda^*)) \right\}$ , where  $\varphi'$  corresponds to the function in (6.1). Invoking (6.1) and (6.2), we obtain that

$$\begin{aligned} \mathbb{E} \psi \left( \varphi' \left( \frac{|Z|}{\lambda^*} \right) \right) &= \mathbb{E} \frac{|Z|}{\lambda^*} \varphi' \left( \frac{|Z|}{\lambda^*} \right) - \varphi \left( \frac{|Z|}{\lambda^*} \right) \\ &\leq \mathbb{E} \varphi \left( \frac{2|Z|}{\lambda^*} \right) - 2 \varphi \left( \frac{|Z|}{\lambda^*} \right) < \infty \end{aligned}$$

as  $Z \in L^\varphi$  and  $\varphi \in \Delta_2$ . Thus  $\varphi' \left( \frac{|Z|}{\lambda^*} \right) \in L^\psi$  and consequently  $X^* \in L^\psi$ . Further, as  $\varphi'$  is non-decreasing, we observe that

$$\begin{aligned} X^* + \varphi'(c_Z(\lambda^*)) &= \max \left\{ \varphi'(c_Z(\lambda^*)), \varphi' \left( \frac{|Z|}{\lambda^*} \right) \right\} \\ &= \varphi' \left( \max \left\{ c_Z(\lambda^*), \frac{|Z|}{\lambda^*} \right\} \right) = \varphi'(Z^*) \end{aligned}$$

and hence

$$\begin{aligned} \mathbb{E}(X^* + \varphi'(c_Z(\lambda^*))) Z^* &= \mathbb{E} \varphi'(Z^*) Z^* = \mathbb{E} \psi(\varphi'(Z^*)) + \varphi(Z^*) \\ &= \mathbb{E} \psi((X^* + \varphi'(c_Z(\lambda^*))) + \varphi(Z^*) \end{aligned}$$

by (6.2). Employing this as well as (5.3) and (6.8), we obtain

$$\begin{aligned} \rho_{\varphi, \beta}(X^*) &\geq \mathbb{E} X^* Z^* = -\varphi'(c_Z(\lambda^*)) + \beta + \mathbb{E}((X^* + \varphi'(c_Z(\lambda^*))) Z^* - \varphi(Z^*)) \\ &= -\varphi'(c_Z(\lambda^*)) + \beta + \mathbb{E} \psi(X^* + \varphi'(c_Z(\lambda^*))) \geq \rho_{\varphi, \beta}(X^*) \end{aligned}$$

and therefore  $\rho_{\varphi, \beta}(X^*) = \mathbb{E} X^* Z^*$ . Observing that  $Z^*$  equals  $\frac{|Z|}{\lambda^*}$  on the set where  $X^*$  differs from 0, we finally get that

$$\begin{aligned} \frac{\mathbb{E} \operatorname{sign}(Z) X^* Z}{\lambda^*} &= \frac{\mathbb{E} X^* |Z|}{\lambda^*} = \mathbb{E} X^* Z^* \\ &= \rho_{\varphi, \beta}(X^*) = \|X^*\|_{\varphi, \beta} = \|\operatorname{sign}(Z) X^*\|_{\varphi, \beta} \end{aligned}$$

as  $X^*$  is non-negative. This establishes  $\|Z\|_{\varphi, \beta}^* \geq \lambda^*$  and thus the theorem.  $\square$

**6.3. Existence of solutions of the characterizing equations.** For completeness we provide conditions to guarantee that the system (6.3)–(6.4) is solvable. The solutions  $t^*$  and  $\mu^*$  identify the optimal solution in the initial problem (3.1). This is of importance in numerical evaluations of  $\rho_{\varphi,\beta}(X)$ .

**Theorem 6.5.** *Let be  $X \in M^\psi$ ,  $X \geq 0$  and  $\varphi \in \Delta_2$ . Further suppose there are optimal values  $t^* > 0$  and  $\mu^* \in \mathbb{R}$  inside of (3.1) (i.e.,  $P(X = \text{ess sup}(X)) < 1 - \bar{\alpha}$  by Proposition 5.5). If  $\psi$  is differentiable, then  $t^*$  and  $\mu^*$  solve the equations (6.3) and (6.4) for the normal derivative  $\psi'$ . If  $X$  is continuously distributed, then  $t^*$  and  $\mu^*$  solve the equations (6.3) and (6.4) for any subderivative  $\psi'$  satisfying (6.1).*

*Proof.* Let non-negative  $X \in M^\psi$  and minimizers  $t^* > 0$  and  $\mu^* \in \mathbb{R}$  inside of in (3.1) be given. By the non-negativity of  $X$  we have that  $\rho_{\varphi,\beta}(X) = \|X\|_{\varphi,\beta}$ . Therefore it exists a random variable  $Z \in (M^\psi)^* = L^\varphi$  such that

$$\|Z\|_{\varphi,\beta}^* = 1 \quad \text{and} \quad \mathbb{E} X Z = \|X\|_{\varphi,\beta} = \rho_{\varphi,\beta}(X)$$

by the Hahn-Banach theorem ([23, p. 112 Corollary 2]). As we have shown in the proof of Theorem 6.4, there is  $Z^* \in L^\varphi$  with  $Z^* \in M_{\varphi,\beta}$  and  $|Z| \leq Z^*$ . Therefore, as  $X \geq 0$ , we have that  $\mathbb{E} X Z \leq \mathbb{E} X Z^*$ . Conversely, it holds that  $\mathbb{E} X Z^* \leq \|X\|_{\varphi,\beta} = \mathbb{E} X Z$ , as  $Z^*$  is feasible inside of  $M_{\varphi,\beta}$ , from which we conclude  $\mathbb{E} X Z = \mathbb{E} X Z^*$ . Applying the Fenchel–Young inequality (2.2) we obtain

$$\begin{aligned} \mathbb{E} X Z^* &\leq \mathbb{E} X Z^* + t^* (\beta - \mathbb{E} \varphi(Z^*)) + t^* \mu^* (1 - \mathbb{E} Z^*) \\ &= t^* \left( \beta + \mu^* + \mathbb{E} \left( \frac{|X|}{t^*} - \mu^* \right) Z^* - \varphi(Z^*) \right) \\ &\leq t^* \left( \beta + \mu^* + \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right) \right) = \|X\|_{\varphi,\beta}. \end{aligned}$$

By  $\mathbb{E} X Z^* = \|X\|_{\varphi,\beta}$  it follows that neither of the upper inequalities is strict and hence  $\mathbb{E} \varphi(Z^*) = \beta$  as well as

$$(6.9) \quad \mathbb{E} \left( \frac{X}{t^*} - \mu^* \right) Z^* - \varphi(Z^*) = \mathbb{E} \psi \left( \frac{X}{t^*} - \mu^* \right).$$

If  $\psi$  is differentiable, the only function establishing equality inside of Fenchel–Young inequality (2.2) is the derivative  $\psi'$  (see (6.2)). In any other case it holds strict inequality. Hence by (6.9), we have that  $Z^* = \psi' \left( \frac{X}{t^*} - \mu^* \right)$  almost surely and therefore

$$1 = \mathbb{E} Z^* = \mathbb{E} \psi' \left( \frac{X}{t^*} - \mu^* \right) \quad \text{and} \quad \beta = \mathbb{E} \varphi(Z^*) = \mathbb{E} \varphi \left( \psi' \left( \frac{X}{t^*} - \mu^* \right) \right).$$

Thus  $t^*$  and  $\mu^*$  solve the equations (6.3), (6.4).

Now assume  $X$  is continuously distributed. Then the random variables  $\psi' \left( \frac{X}{t^*} - \mu^* \right)$  coincide almost surely, for every subderivative  $\psi'$  of  $\psi$ . This follows from the fact that the subderivatives  $\psi'$  of  $\psi$  are uniquely determined, apart from at most countably many points. Furthermore, by the same argument as above, we have that  $Z^* = \psi' \left( \frac{X}{t^*} - \mu^* \right)$  almost surely and thus the assertion.  $\square$

**6.4. Application in finance.** In what follows we highlight the benefits of  $\varphi$ -divergence risk measures for a problem in optimizing a portfolio (cf. also Rockafellar et al. [37]). To this end set

$$W := \left\{ w = (w_1, \dots, w_n) \in \mathbb{R}^n : w_i \geq 0 \text{ and } \sum_{i=1}^n w_i = 1 \right\}$$

and consider random variables  $X_1, \dots, X_n \in L^\psi$ .  $X_i$  is the loss of the  $i$ -th asset and  $W$  constitutes all possible portfolio allocations. By denoting  $X_w := w_1 X_1 + \dots + w_n X_n$  the associated optimization problem is

$$\min_{w \in W} \rho_{\varphi, \beta}(X_w) = \min_{w \in W} \inf_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X_w}{t} - \mu \right) \right),$$

which determines the portfolio allocation with minimal risk based on the risk measure  $\rho_{\varphi, \beta}$ . One may restate this expression as

$$\begin{aligned} \min_{w \in W} \rho_{\varphi, \beta}(X_w) &= \min_{w \in W} \min_{\substack{\mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X_w}{t} - \mu \right) \right) \\ (6.10) \qquad &= \min_{\substack{w \in W, \\ \mu \in \mathbb{R}, \\ t > 0}} t \left( \beta + \mu + \mathbb{E} \psi \left( \frac{X_w}{t} - \mu \right) \right). \end{aligned}$$

The striking benefit in (6.10) is that it is sufficient to execute a single minimization problem with only two additional variables instead of two nested minimization problems when employing (5.5). This reduces the complexity of the problem significantly. Similar results are available for Haezendonck–Goovaerts risk measures in [8] as for Average Value-at-Risk in [40].

## 7. SUMMARY

Coherent risk measures are of fundamental importance in mathematical finance. They constitute convex functionals on appropriate Banach spaces for which the entire and rich theory of convex analysis and convex duality applies.

This paper addresses a specific risk functional based on  $\varphi$ -divergence. The  $\varphi$ -divergence is a non-symmetric distance, it is used to quantify aberrations from a given probability measure.  $\varphi$ -divergence generalizes Kullback–Leibler divergence, which is nowadays exhaustively used in data science.

We characterize the corresponding Banach space in detail and elaborate the dual norm. The space is an Orlicz space and, in general, not reflexive.

The specific form of the  $\varphi$ -divergence risk measure allows a rich variety of equivalent expressions. They can be employed mutually to exploit the specific properties in given applications. We also exemplify the properties for a typical problem in mathematical finance.

## REFERENCES

- [1] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, Journal of Banking & Finance **26** (2002), 1505–1518.

- [2] A. Ahmadi-Javid, *Entropic Value-at-Risk: A new coherent risk measure*, Journal of Optimization Theory and Applications **155** (2012), 1105–1123.
- [3] A. Ahmadi-Javid, *Addendum to: Entropic Value-at-Risk: A new coherent risk measure*, Journal of Optimization Theory and Applications **155** (2012), 1124–1128.
- [4] A. Ahmadi-Javid and A. Pichler, *An analytical study of norms and Banach spaces induced by the entropic value-at-risk*, Mathematics and Financial Economics **11** (2017), 527–550.
- [5] P. Artzner, F. Delbaen and D. Heath, *Thinking coherently*, Risk **10** (1997), 68–71.
- [6] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, *Coherent Measures of Risk*, Mathematical Finance **9** (1999), 203–228.
- [7] F. Bellini and E. Rosazza Gianin, *On Haezendonck risk measures*, Journal of Banking & Finance **32** (2008), 986–994.
- [8] F. Bellini and E. Rosazza Gianin, *Optimal portfolios with Haezendonck risk measures*, Statistics & Decisions, 26, 01 2008. doi:10.1524/std.2008.0915.24
- [9] F. Bellini and E. Rosazza Gianin, *Haezendonck-Goovaerts risk measures and Orlicz quantiles*, Insurance: Mathematics and Economics **51** (2012), 107–114.
- [10] F. Bellini, B. Klar, A. Müller and E. Rosazza Gianin, *Generalized quantiles as risk measures*, Insurance: Mathematics and Economics **54** (2014), 41–48.
- [11] R. I. Boş, S.-M. Grad and G. Wanka, *Duality in Vector Optimization*, Springer, 2009.
- [12] T. Breuer and I. Csiszár, *Measuring distribution model risk*, Mathematical Finance, 2013. doi:10.1111/ma.12050. 2, 3
- [13] T. Breuer and I. Csiszár, *Systematic stress tests with entropic plausibility constraints*, Journal of Banking & Finance **37** (2013), 1552–1559.
- [14] P. Cheridito and T. Li, *Dual characterization of properties of risk measures on Orlicz hearts*, Mathematics and Financial Economics **2** (2008), 29–55.
- [15] P. Cheridito and T. Li, *Risk measures on Orlicz hearts*, Mathematical Finance **19** (2009), 189–214.
- [16] F. Delbaen, *Remark on the paper “Entropic Value-at-Risk: A new coherent risk measure” by Amir Ahmadi-Javid*, in: Risk and Stochastics, P. Barrieu (ed), World Scientific, 2015.
- [17] F. Delbaen and K. Owari, *Convex functions on dual Orlicz spaces*, Positivity **23** (2009), 1051–1064.
- [18] M. Goovaerts, D. Linders, K. V. Weert and F. Tank, *On the interplay between distortion, mean value and the Haezendonck-Goovaerts risk measures*, Insurance: Mathematics and Economics **51** (2012), 10–18.
- [19] T. Kalmes and A. Pichler, *On Banach spaces of vector-valued random variables and their duals motivated by risk measures*, Banach Journal of Mathematical Analysis **12** (2018), 773–807.
- [20] M. A. Krasnosel’skii and Y. B. Rutickii, *Convex Functions and Orlicz Spaces*, Noordhoff Groningen, 1961.

- [21] P. Krokmal, M. Zabarankin and S. Uryasev, *Modeling and optimization of risk*, Surveys in Operations Research and Management Science **16** (2011), 49–66.
- [22] S. Kusuoka, *On law invariant coherent risk measures*, in: Advances in mathematical economics, volume 3, chapter 4, Springer, 2001, pp. 83–95.
- [23] D. G. Luenberger, *Optimization by Vector Space Methods*, Decision and control. Wiley, New York, NY, 1969.
- [24] 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.
- [25] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM Journal on Optimization **13** (2002), 60–78.
- [26] G. Ch. Pflug, *On distortion functionals*, Statistics and Risk Modeling (formerly: Statistics and Decisions) **24** (2006), 45–60.
- [27] A. Pichler, *The natural Banach space for version independent risk measures*, Insurance: Mathematics and Economics **53** (2013), 405–415.
- [28] A. Pichler, *A quantitative comparison of risk measures*, Annals of Operations Research **254** (2017), 251–275.
- [29] A. Pichler and R. Schlotter, *Entropy based risk measures*, European Journal of Operational Research, 2018. doi:10.1016/j.ejor.2019.01.016.
- [30] A. Pichler and A. Shapiro, *Minimal representations of insurance prices*, Insurance: Mathematics and Economics **62** (2015), 184–193.
- [31] L. Pick, A. Kufner, O. John, and S. Fučík, *Function Spaces*, De Gruyter Series in Nonlinear Analysis and Applications 14. Walter de Gruyter & Co., Berlin, 2013.
- [32] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [33] R. T. Rockafellar, *Integral functionals, normal integrands and measurable selections*, in: Nonlinear Operators and the Calculus of Variations, Springer, 1976, pp. 157–207.
- [34] R. T. Rockafellar and J. O. Royset, *Random variables, monotone relations, and convex analysis*, Mathematical Programming **148** (2014), 297–331.
- [35] R. T. Rockafellar and J. O. Royset, *Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity*, SIAM Journal on Optimization **25** (2015), 1179–1208.
- [36] R. T. Rockafellar and J. O. Royset, *Superquantile/ CVaR risk measures: second-order theory*, Annals of Operations Research **262** (2016), 3–28.
- [37] R. T. Rockafellar, J. O. Royset and S. I. Miranda, *Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk*, European Journal of Operational Research **234** (2014), 140–154.
- [38] R. T. Rockafellar and S. Uryasev, *The fundamental risk quadrangle in risk management, optimization and statistical estimation*, Surveys in Operations Research and Management Science **18** (2013), 33–53.
- [39] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk **2** (2000), 21–41.



- [40] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, Journal of Banking and Finance **26** (2002), 1443–1471.
- [41] R. Rockafellar, S. Uryasev and M. Zabarankin, *Generalized deviations in risk analysis*, Finance and Stochastics **10** (2006), 51–74.
- [42] R. Rockafellar, S. Uryasev, and M. Zabarankin, *Risk tuning with generalized linear regression*, Math. Oper. Res. **33** (2008), 712–729.
- [43] S. Wang and J. Dhaene, *Comonotonicity, correlation order and premium principles*, Insurance: Mathematics and Economics **22** (1998), 235–242.

*Manuscript received March 14 2020*

*revised May 10 2020*

P. DOMMEL

University of Technology, Chemnitz, Germany

*E-mail address:* paul.dommel@s2013.tu-chemnitz.de

A. PICHLER

University of Technology, Chemnitz, Germany

*E-mail address:* alois.pichler@math.tu-chemnitz.de