Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 21, No. 3, pp. 375–395, 2025

ON THE SUBDIFFERENTIAL OF SOME CRITICAL CONVEX RISK FUNCTIONS*

Haodong Yu, Aifan Ling† and Jie Sun

Abstract: Risk minimization models are commonly used in financial and economic decisions, but the solution approach is unmatured even for the convex risk functions. This paper aims at a subdifferential solution approach for three critical classes of convex risk functions: (1) risk functions of composite structure, (2) risk functions with centralized transformation, and (3) risk functions represented by optimal value functions. We analyze the subdifferential properties of the three critical classes of risk functions, which are closely related to the so-called risk quadrangle and covers many commonly used risk functions. We also discuss their applications on the subgradient method for the risk minimization problems with convex risk functions.

Key words: *convex risk function, subdifferential, composite risk measure, centralized transformation, optimal value function, risk quadrangle*

Mathematics Subject Classification: *90C15, 90C25*

1 Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Consider the stochastic optimization problem that minimizes a general convex risk function

$$\min_{x \in D_X} \rho[F(x, \omega)], \tag{1.1}$$

where $F(x, \omega) : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ is a random function, that is, for any fixed $x \in \mathbb{R}^n$, $F(x, \cdot)$ is $\mathcal{F}$-measurable. $\rho : \mathcal{Z} \to \bar{\mathbb{R}}$ is called a risk function, where $\mathcal{Z}$ is a space of random variables and $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. In practice, $F(x, \omega)$ usually represents the real "loss" of the decision maker, while ρ stands for a "measure" of the random loss, see [3, 26] for more discussions on this matter.

The simplest choice of ρ is the expectation operator $\mathbb{E}[\cdot]$, which is linear and reflects the neutral attitude of the decision maker towards the risk. The algorithms for (1.1) with the expectation operator have been widely studied, see e.g. [23, 12, 6] for a review.

Many applications (e.g., [26, 31, 16, 29]) require to deal with more general risk measures with convex functions, which reflects a risk averse attitude. In particular, this is related

*This work is partially supported by National Natural Science Foundation of China (NSFC), Grant No. 72071098, and the Fundamental Research Funds for the Central Universities, and Shanghai AI for Social Science Special Funding Project.

†Corresponding Author

 DOI: https:// doi.org/10.61208/pjo-2025-018

to the expected utility theory. If, for example, we set $\rho(Z) = -\mathbb{E}[V(Z)]$, where V is a regular concave utility function in economics, then problem (1.1) is equivalent with the next expectation utility maximization

$$\min_{x\in D_X} \mathbb{E}[-V(F(x,\omega))] = -\max_{x\in D_X} \mathbb{E}[V(F(x,\omega))]. \tag{1.2}$$

This shows the relationship between the risk optimization problem (1.1) and the expected utility approach.

The composite structure of the mapping $\rho(F(x,\omega))$ brings about certain complications in solving optimization problem (1.1). A popular approach is to transform the original problem to a min-max problem by using the dual representation of ρ. Specifically, let ρ be a convex, proper and lower semicontinuous risk function, then $\rho^{**} = \rho$ and it holds that

$$\rho(Z) = \sup_{Q\in\mathcal{Z}^*} \left[\mathbb{E}(ZQ) - \rho^*(Q)\right], Z \in \mathcal{Z} \tag{1.3}$$

where ρ^* is the conjugate function of ρ, and $\mathcal{Z}^*$ is the dual space of $\mathcal{Z}$. This means that problem (1.1) with a convex risk function can be transformed into a min-max problem as follws,

$$\min_{x\in D_X} \rho[F(x,\omega)] = \min_{x\in D_X} \sup_{Q\in\mathcal{Z}^*} \left[\mathbb{E}(F(x,\omega)Q(\omega)) - \rho^*(Q)\right], \quad \text{with } Z = F(x,\omega). \tag{1.4}$$

See [24] for the discussion on such methods for general convex risk functions.

However, when one wants to carry out such methods for practical problems, an immediately obstacle is that, it has to deal with the specific form of the conjugate $\rho^*(Q)$, together with the dual space $\mathcal{Z}^*$. Up till now, most literature actually focuses on the case where ρ is a coherent risk measure, which means a function satisfying the following four conditions (Artzner et al. [4]):

C1 (Convexity): For any $0 \le \lambda \le 1$, $\rho((1-\lambda)Z_1 + \lambda Z_2) \le (1-\lambda)\rho(Z_1) + \lambda\rho(Z_2)$;

C2 (Monotonicity): If $Z_1 \le Z_2$ a.e., then $\rho(Z_1) \le \rho(Z_2)$;

C3 (Translation Invariance): For $\gamma \in \mathbb{R}$, $\rho(Z+\gamma) = \rho(Z) + \gamma$;

C4 (Positive Homogeneity): If $t \ge 0$, then $\rho(tZ) = t\rho(Z)$.

The typical examples of the coherent risk measures include the expectation, $\text{EVaR}_{1-\alpha}$ (the entropic Value-at-Risk, with $\alpha \in (0,1)$), CVaR_α (conditional value at risk with $\alpha \in (0,1)$), see Example 5 and 6 in Section 5 for the details of the latter two examples.

A critical property of a coherent measure is that, the dual representation $\rho(\cdot)$ can be simplified as follows (see e.g. [3]),

$$\rho(Z) = \sup_{Q\in\mathcal{Q}} \mathbb{E}(ZQ), Z \in \mathcal{Z}, \tag{1.5}$$

where $\mathcal{Q}$ is called a risk envelop, which is a nonempty closed subset of $P := \{Q \in \mathcal{Z}^* : Q \succeq 0, \mathbb{E}(Q) = 1\}$. In this case, the 'sup' in equation (1.5) can be replaced by 'max'. If $\mathcal{Z} = \mathscr{L}^p(\Omega,\mathcal{F},P)$ with $p \in (1,+\infty)$, then $\mathcal{Q}$ is a closed set in its dual space $\mathscr{L}^q(\Omega,\mathcal{F},P)$ (q is determined by $1/q + 1/p = 1$).

The dual representation (1.5) of a coherent risk measure means that (1.4) can be simplified to the form below, which is also related to the so-called distributional robust optimization (DRO) problem (see e.g., [27, 15]),

$$\min_{x\in D_X} \rho[F(x,\omega)] = \min_{x\in D_X} \max_{Q\in\mathcal{Q}} \mathbb{E}\left[F(x,\omega)Q(\omega)\right], \quad \text{with } Z = F(x,\omega). \tag{1.6}$$

Following this approach, many methods have been proposed for problem (1.1) with coherent risk measures. See [17] for a review of this approach. The application of the coherent risk measures in portfolio optimization can be found in [2, 13].

From the practical perspective, however, the requirement on the risk function ρ to be coherent can be too strong in financial and economical applications. In particular, only the convexity condition C1 of ρ can be satisfied in most cases since it reflects a risk averse attitude. As to the conditions C2-C4, in many cases they are not satisfied simultaneously. The examples include the p-order lower semi-moment measure $\rho_a(Z) = \mathbb{E}(Z-\mathbb{E}(Z))_+^p$, where $(x)_+ = \max(x, 0)$, and the p-order lower partial moment measure $\rho_b(Z) = E[(Z - \tau)_+]^p$ (with τ be a fixed parameter). Clearly, ρ_a and ρ_b only have convexity, but do not satisfy the conditions C2-C4. Furthermore, it can be observed (see e.g. [9]) from the expected utility maximum problem (1.2) that, many utility functions U in (1.2) do not require the conditions C2-C4 .

In current literatures, there are some works that focus on the relaxation of the requirement of coherent risk measure and the extension of the min-max approach based on (1.4). For example, Follmer and Schied ([11]) proposed the so-called convex risk measure, which only requires ρ to satisfy C1-C3, they also explore the formulas for the conjugates of convex risk measures. (See also [10, 8] for reference.) On the whole, in the case that ρ fails to be coherent, the problem (1.1) will be much more complicated.

Even if ρ can be supposed to be a coherent measure, other problems may also cause difficulty for numerical methods. For instance, in many cases, the specific form of $\mathcal{Q}$ may cause the constraints to be too complex. Moreover, if $\mathcal{Q}$ is not polyhedral, then many algorithms will be unsuitable or inefficient.

In the current paper, we try to challenge the solution approach of problem (1.1) with a convex risk function ρ through an alternative way, i.e., the subgradient approach for problem (1.1). The focal point is the computation of subgradient of $\rho[F(x, \omega)]$. No matter whether ρ is coherent or not, if the composite mapping $\rho[F(x, \omega)]$ is a convex function with respect to x. Theoretically, we can use a subgradient-type of method to solve (1.1), as long as a subgradient of $\rho[F(x, \omega)]$ can be computed.

Recall that, by the theory of the convex optimization, the general form of the subdifferential of ρ is (see e.g., [24, 25])

$$\partial\rho(Z) = \arg\sup_{Q\in\mathcal{Z}^*} \left[\mathbb{E}(ZQ) - \rho^*(Q)\right]. \tag{1.7}$$

In the case when ρ is a coherent measure,

$$\partial\rho(Z) = \arg\sup_{Q\in\mathcal{Q}} \mathbb{E}(ZQ). \tag{1.8}$$

Theoretically, (1.7) means that (1.4) can be equivalently rewritten as follows:

$$\min_{x\in D_X} \rho[F(x, \omega)] = \min_{x\in D_X, Q^*\in\partial\rho(Z)} \mathbb{E}\left[F(x, \omega)Q^*(\omega) - \rho^*(Q^*)\right], \quad \text{with } Z(\omega) = F(x, \omega). \tag{1.9}$$

If we directly use (1.7) or (1.8) to obtain $\partial\rho(Z)$, we have to deal with the specific form of the conjugate $\rho^*(Q)$ and the risk envelop. As we discussed above, the structure of $\mathcal{Q}$ may not easy to obtain. On the other hand, the form (1.9) is very difficult for practical use. Thus, we will instead focus on the specific forms of the subdifferential of many practically used risk functions. Moreover, we notice that Rockafellar and Uryasev [20] proposed a new paradigm, called the risk quadrangle, which links a great variety of measures from different areas, especially the risk management and the statistics. Hence, our discussion is based on this new paradigm. We will discuss these three types of convex risk functions, which has close relationship with the risk quadrangle in detail and will provide the general form of the subdifferential of these risk functions. In summary, the contributions of this paper can be stated as follows.

- We consider a risk minimization problem (1.1) with ρ being a convex risk function. In stead of solving the original problem through a min-max representation like (1.4), we provide a subdifferential approach to obtain the optimal solutions.
- We focus on the computation and theoretical properties of three classes of risk functions of which the subdifferential can be calculated directly. The discussed risk functions are all related to the risk quadrangle, which covers much more risk functions than the coherent or convex risk measure does.
- We discuss the approach to construct a subgradient-type algorithm for risk minimization problem and show the fundamental role of the computation of $\partial\rho(\cdot)$ in such method.

The remainder of this paper is organized as follows. In Section 2, we will recall the notion and the properites of the risk quadrangle, and then present three related types of risk functions. In the following three sections, we will discuss the properties and computation of these presented risk functions in details. Also included are some examples or applications of these discussed risk functions. Finally, in Section 6, we will show the approach to develop a standard subgradient algorithm for the problem (1.1) based on the subfifferential of ρ.

2 The Risk Quadrangle and Three Types of Risk Functions

Let us recall briefly the framework of the risk quadrangle theory by Rockafellar and Uryasev ([20]), which is based on their earlier work on the deviation measures (Rockafellar et al [21]). See also Rockafellar and Royset [18] for reference. This paradigm links a great variety of measures from different areas, classifies them into four classes, and establishes their connections.

Table 1: The risk quadrangle

	risk	$\mathcal{R}$	$\leftrightarrow$	$\mathcal{D}$	Deviation	
Optimization		$\uparrow\downarrow$	$\mathcal{S}$	$\uparrow\downarrow$		Estimation
	regret	$\mathcal{V}$	$\leftrightarrow$	$\mathcal{E}$	Error	

The four types of measures and their meanings are as follows:
$\mathcal{R}(Z)$ (risk measure): provides an overall measure of the "loss" or "damage" variable Z;
$\mathcal{D}(Z)$ (deviation measure): measures the "nonconstancy" in Z;
$\mathcal{E}(Z)$ (error measure): measures the "nonzeroness" in Z;
$\mathcal{V}(Z)$ (regret measure): measures the "regret" in facing the mix of outcomes of Z.
Moreover, the relationship among the four types of measures and the meaning of $\mathcal{S}$ are as follows.

The relationship illustrated above requires the regular properties of all these four types of measures. That is, all of the risk, deviation, regret, and error measures should be closed

Table 2: General Relationships ([20])

$\mathcal{R}(Z) = \mathbb{E}(Z) + \mathcal{D}(Z)$	$\mathcal{D}(Z) = \mathcal{R}(Z) - \mathbb{E}(Z)$
$\mathcal{V}(Z) = \mathbb{E}(Z) + \mathcal{E}(Z)$	$\mathcal{E}(Z) = \mathcal{V}(Z) - \mathbb{E}(Z)$
$\mathcal{R}(Z) = \inf_t \{t + \mathcal{V}(Z - t)\}$	$\mathcal{D}(Z) = \inf_t \mathcal{E}(Z - t)$
$\arg\min_t \{t + \mathcal{V}(Z - t)\} = \mathcal{S} = \arg\min_t \{\mathcal{E}(Z - t)\}$	

convex functions. Moreover, these measures should satisfy the following conditions respectively:

- (risk measure) $\mathcal{R}(C) = C$ for constant C, and $\mathcal{R}(Z) > \mathbb{E}(Z)$ for nonconstant Z;
- (deviation measure) $\mathcal{D}(Z) \geq 0$ for all Z, with $\mathcal{D}(0) = 0$ and $\mathcal{D}(Z) > 0$ for nonconstant Z;
- (error measure) $\mathcal{E}(Z) = 0$ but $\mathcal{E}(Z) > 0$ when $Z \not\equiv 0$;
- (regret measure) $\mathcal{V}(Z) = 0$ but $\mathcal{V}(Z) > \mathbb{E}(Z)$ when $X \not\equiv 0$.

Motivated by the risk quadrangle, in particular the above transition formulas illustrated in Table 2, in this paper we consider three special convex risk functions, including:

- a family of composite risk functions, which reflects the central place of the expectation measure in the overall risk quadrangle;
- risk functions with the centralized transformation, which is critical in the deviation measures;
- risk functions represented by optimal value functions, which covers the functions induced by the risk and regret measures.

The discussed families of risk functions contain the extensions and unifications of many commonly used risk measures. We will provide the general form of the subdifferential of these measures with the specific results of some important measures as their examples. Unless otherwise specified, in what follows, we consider $\mathcal{Z} := \mathscr{L}^2(\Omega, \mathcal{F}, P)$ to be the the space of random variable.

3 A Family of Composite Risk Functions

In the construction of the risk quadrangle, the expectation operator plays the central role. As is mentioned by the authors of [20] (see Page 19 of this paper), "the crucial role that $\mathbb{E}(X)$ has in the fundamental risk quadrangle is our guide".

In this section, we consider the family of risk functions as follows, which is the combination of a series of composite risk functions

$$\rho(Z) = \sum_{i=1}^{k} f_i(\mathbb{E}[g_i(Z)]). \tag{3.1}$$

To ensure the convexity of ρ, we give some assumptions on the functions $f_i(\cdot)$ and $g_i(\cdot)(i = 1, \cdots, k)$.

Assumption 3.1. Denote E be an open set in $\mathbb{R}$. Assume that, for all $i = 1, \cdots, k$,

(i) $f_i : E \to \mathbb{R}$ is a differentiable nondecreasing function;

(ii) $g_i : \mathbb{R} \to E$ is a convex nondecreasing function and the expectation $\mathbb{E}(g_i(Z))$ is well defined and proper;

(iii) $f_i(\mathbb{E}[g_i(Z)])$ is convex in Z.

Condition (i) and (ii) implies that the composite $f_i(\mathbb{E}[g_i(Z)])$ is nondecreasing. Together with Condition (iii), these further imply that the risk function defined by (3.1) is convex and monotone. In addition, by convex analysis, Condition (ii) implies that g_i is a continuous function.

Notice that in Assumption 3.1, the functions f_i itself is not required to be convex. This will broaden the scope of application of this framework. A typical example is the next measure, which is a combination of the expectation and the upper-semideviation of order p from the objective τ,

$$\rho(Z) = \mathbb{E}(Z) + c\left(\mathbb{E}(Z-\tau)_+^p\right)^{1/p}, \text{with } p \in [1, +\infty), \tag{3.2}$$

where $c > 0$. It can be observed that the second term of $\rho(\cdot)$ in (3.2) contains a non-convex outer function $f(z) = z^{1/p}$, while the entire risk measure is convex.

Proposition 3.1. Let $Z \in \mathcal{Z}$ and $g : \mathbb{R} \to \mathbb{R}$ be a finite-valued convex function and let $\partial g(Z) : \ \Omega \rightrightarrows \partial g(Z(\omega))$ be the set-valued function. Define the mapping $\tilde{g} : \mathcal{Z} \to \mathcal{Z}_1$ where $\tilde{g}(Z)(\omega) = g(Z(\omega))$. Suppose $\rho : \mathcal{Z}_1 \to \bar{\mathbb{R}}$ is a convex and proper mapping, which is differentiable and monotonically nondecreasing. Further denote the composite measure $\rho(\tilde{g}(Z))$ by $\tilde{\rho}(Z)$, then

$$\partial\tilde{\rho}(Z) = \nabla\rho(Y)\cdot S(\partial g(Z), \nabla\rho(Y)) \text{ with } Y = \tilde{g}(Z),$$

where

$$S(\partial g(Z), \nabla\rho(Y)) := \{\zeta \text{ is } \mathcal{F}\text{-measurable} : \nabla\rho(Y)\zeta \in \mathcal{Z}^*, \zeta(\omega) \in \partial g(Z(\omega))\text{a.e.}, \omega \in \Omega\}.$$

Proof. It is easy to verify that $\tilde{\rho}$ is convex and is directional differentiable. For any $Z, Z_0 \in \mathcal{Z}$, denote $h = Z - Z_0$, then the directionally derivative of $\tilde{\rho}$ at Z_0 along h is

$$\tilde{\rho}'(Z_0, h) = \langle\nabla\rho(Y_0), \tilde{g}'(Z_0, h)\rangle = \mathbb{E}[\nabla\rho(Y_0)\tilde{g}'(Z_0, h)],$$

where $Y_0 = g(Z_0)$. The first equality follows from the chain rule of the directionally derivative (see Proposition 2.47 in [7]). Choose any $\zeta \in S(\partial g(Z_0), \nabla\rho(Y_0))$, we have $\tilde{g}'(Z_0, h) \geq \zeta h$ a.e. and

$$\tilde{\rho}(Z) - \tilde{\rho}(Z_0) \geq \tilde{\rho}'(Z_0, h) \geq \mathbb{E}[\zeta\nabla\rho(Y_0)h] = \langle\zeta\nabla\rho(Y_0), h\rangle. \tag{3.3}$$

Hence, $\nabla\rho(Y_0)\zeta \in \partial\rho(Z_0)$. Consequently,

$$\nabla\rho(Y)\cdot S(\partial g(Z), \nabla\rho(Y)) \subseteq \partial\tilde{\rho}(Z).$$

Conversely, choose any $\eta \in \partial\tilde{\rho}(Z_0)$. Notice that $\nabla\rho(Y_0) \geq 0$ a.e., then

$$\begin{aligned}\mathbb{E}[\nabla\rho(Y_0)\tilde{g}'(Z_0, h)] &= \mathbb{E}\left[\sup_{d(\omega)\in\partial g(Z_0(\omega))} \nabla\rho(Y_0)(\omega)d(\omega)h(\omega)\right] \\ &= \sup_{d\in S(\partial g(Z_0), \nabla\rho(Y_0))} \langle\nabla\rho(Y_0)d, h\rangle.\end{aligned}$$

The first equality follows from the properties of $\tilde{g}'(Z_0,h)$ and the measurability holds from Theorem 7.37 of in [22]. The second equality follows from the the interchangeability of minimization and integration theorem (see Theorem 14.60 in [22]). Hence,

$$\langle \eta, h\rangle \leq \tilde{\rho}'(Z_0,h) = \sup_{d\in S(\partial g(Z_0),\nabla\rho(Y_0))} \langle \nabla\rho(Y_0)d, h\rangle = \sup_{\xi\in A}\langle \xi, h\rangle,$$

where $A = \nabla\rho(Y_0)\cdot S(Z_0,\nabla\rho(Y_0))$, which is a closed convex set. Since h is an arbitrary direction, by the separate theorem of convex set, $\eta \in A$, it holds that $\partial\tilde{\rho}(Z) \subseteq \nabla\rho(Y)\cdot S(\partial g(Z),\nabla\rho(Y))$. □

Corollary 3.2. Suppose that $Z \in \mathcal{Z}$, and the function $g:\mathbb{R}\to\mathbb{R}$ satisfies Assumption 3.1(ii). Denote $\tilde{\rho}(Z) := \mathbb{E}(g(Z))$, then

$$\partial\tilde{\rho}(Z) = S(\partial g(Z)),\ \text{where } S(\partial g(Z)) := \{\zeta\in\mathcal{Z}^* : \zeta(\omega)\in\partial g(Z(\omega))\text{a.e.}, \omega\in\Omega\}.$$

Proof. In Proposition 3.1, let $\rho(Z) := \mathbb{E}(Z)$, then $\nabla\rho(Z) = 1$, and the asserted statement follows immediately. □

Based on Proposition 3.1 and Corollary 3.2, we can now discuss the subdifferential of the risk measure with the form of (3.1).

Proposition 3.3. Suppose that $Z \in \mathcal{Z}$, consider the risk measure $\rho(\cdot)$ defined by (3.1), with $f_i(\cdot)$ and $g_i(\cdot)$ satisfying Assumption 3.1, then:

$$\partial\rho(Z) = \mathcal{Q}_0 := \sum_{i=1}^{k} f_i'\left(\mathbb{E}[g_i(Z)]\right) S(\partial g_i(Z)). \tag{3.4}$$

Proof. First denote $\rho_i(Z) := \mathbb{E}(g_i(Z))$. By Corollary 3.2, we have that $\partial\rho_i(Z) = S(\partial g_i(Z))$. Given $Z_0 \in \mathcal{Z}$, for any $Z\in\mathcal{Z}$, denote $h = Z - Z_0$.

We first show that $\mathcal{Q}_0 \subseteq \partial\rho(Z_0)$. It follows from the convexity of ρ_i that $\rho_i(Z)-\rho_i(Z_0) \geq \rho_i'(Z_0,h)$, where $\rho_i'(Z_0,h)$ is the directional derivative of ρ_i at Z_0 along h. Also by convex analysis, we have that

$$\rho_i'(Z_0,h) = \sup_{\zeta\in\partial\rho_i(Z_0)} \langle\zeta, Z - Z_0\rangle. \tag{3.5}$$

Further denote the composite function $f_i\circ\rho_i(Z) := f_i(\rho_i(Z))$. Since f_i is differentiable, the chain rule holds, i.e.,

$$(f_i\circ\rho_i)'(Z_0,h) = f_i'(\rho_i(Z_0))\rho_i'(Z_0,h).$$

Since $f_i\circ\rho_i$ is convex, for any $\zeta_i\in\partial\rho_i(Z_0)$, we have

$$\begin{aligned} f_i(\rho_i(Z)) - f_i(\rho_i(Z_0)) &\geq (f_i\circ\rho_i)'(Z_0,h) \\ &= f_i'(\rho_i(Z_0))\rho_i'(Z_0,h) \\ &\geq f_i'(\rho_i(Z_0))\langle\zeta_i, Z - Z_0\rangle. \end{aligned}$$

The last inequality follows from the monontoncity of f_i and (3.5). This implies that $f_i'(\rho_i(Z_0))\zeta_i \in \partial(f_i\circ\rho_i)(Z_0)$.

Denote $\psi_i = f_i\circ\rho_i$. Thus, $\rho = \sum_{i=1}^k \psi_i$. It follows from Assumption 3.1 and the Moreau-Rockafellar Theorem (see e.g. Theorem 7.4 in [25]) that $\partial\rho = \sum_{i=1}^k \partial\psi_i$. Hence,

$$\sum_{i=1}^{k} f_i'(\rho_i(Z_0))\zeta_i \in \partial\rho(Z_0).$$

Now we prove $\partial\rho(Z) \subseteq \mathcal{Q}_0$. Since $\partial\rho = \sum_{i=1}^{k} \partial\psi_i$, it suffices to show that, if $\xi_i \in \partial(f_i \circ \rho_i)(Z_0)$, then $\xi_i \in f_i'(\rho_i(Z_0))\,\partial g_i(Z_0)$ a.e.

It follows from $\xi_i \in \partial(f_i \circ \rho_i)(Z_0)$ that

$$\langle \xi_i, Z - Z_0\rangle \leq (f_i \circ \rho_i)'(Z_0, h) = f_i'(\rho_i(Z_0))\rho_i'(Z_0, h).$$

If $f_i'(\rho_i(Z_0)) = 0$, then $\langle \xi_i, Z - Z_0\rangle \leq 0$. In this case, from the arbitrariness of $Z \in \mathcal{Z}$, we have that $\xi_i(\omega) = 0$ a.e., $\omega \in \Omega$. If, on the other hand, $f_i'(\rho_i(Z_0)) > 0$, then

$$\mathbb{E}\left[\frac{\xi_i}{f_i'(\rho_i(Z_0))}(Z - Z_0)\right] \leq \rho_i'(Z_0, h) \leq \rho_i(Z) - \rho_i(Z_0),$$

which implies $\xi_i / f_i'(\rho_i(Z_0)) \in \partial\rho_i(Z_0)$. Since $\partial\rho_i(Z_0) = S(\partial g_i(Z_0))$, then it holds that $\partial\rho(Z_0) \subseteq \mathcal{Q}_0$. □

If Assumption 3.1(i) is replaced by: i') $f_i : \mathbb{R} \to \mathbb{R}$ is a convex and monotonically nondecreasing function. Together with Condition (ii), this guarantees that Condition (iii) holds, i.e., $f_i(\mathbb{E}[g_i(Z)])$ are convex. In this case, the next result, which is similar to Proposition 3.3, can be established.

Proposition 3.4. Suppose that $Z \in \mathcal{Z}$, $\rho(\cdot)$ is defined by (3.1), with $f_i(\cdot)$ being convex and monotonically nondecreasing and $g_i(\cdot)$ satisfying Assumption 3.1 (ii). Then

$$\partial\rho(Z) \supseteq \mathcal{Q}_0 := \left\{\sum_{i=1}^{k} d_i S(\partial g_i(Z)) | d_i \in \partial f_i(\mathbb{E}[g_i(Z)])\right\}. \tag{3.6}$$

Proof. Since f_i is monotonically nondecreasing and convex, we have $d_i \geq 0$. Similar with in Proposition 3.3, denote $\rho_i(Z) := \mathbb{E}(g_i(Z))$. Choose any $\zeta_i \in S(\partial g_i(Z))$, by Proposition 3.3, $\zeta_i \in \partial\rho_i(Z)$, then

$$\begin{array}{rcl} f_i(\rho_i(Z)) - f_i(\rho_i(Z_0)) & \geq & d_i(\rho_i(Z) - \rho_i(Z_0)) \\ & \geq & d_i\langle \zeta_i, Z - Z_0\rangle. \end{array}$$

Consequently, $d_i\zeta_i \in \partial(f_i \circ \rho_i)$. □

As a special case, we consider the case when the probability space is discrete, denote $\{\omega_1, \cdots, \omega_S\}$ be the finite support of Ω, for fixed $Z \in \mathcal{Z}$, let $Z(\omega_s) = z_s$, $P(Z = z_s) = \alpha_s$, satisfying $\alpha_s \geq 0$ and $\Sigma_{s=1}^{S}\alpha_s = 1$. If Assumption 3.1 holds, then it follows from Proposition 3.3 that, a subgradient $q = (q_1, \cdots, q_S)^\top \in \partial\rho(Z)$ can be computed as follows

$$q_s = \sum_{i=1}^{k} f_i'\left(\sum_{s=1}^{S}\alpha_s g_i(z_s)\right) \cdot d_{is} \quad s = 1, \cdots, S, \tag{3.7}$$

where $d_{is} \in \partial g_i(z_s)$.

Finally, we discuss two examples of this type of risk functions.

Example 1. The first example is $\rho_1(Z) = \left(\mathbb{E}\left[e^Z\right]\right)^{1/p}$ where $p \in [1, +\infty)$. By defining $f(y) = y^{1/p}(y > 0)$ and $g(z) = e^z$, ρ_1 is an obvious example of (3.1) with $k = 1$. Let $Y = e^{Z/p}$, which is a convex transformation with respect to Z, then we have

$$\rho_1(Z) = \left(\mathbb{E}[Y^p]\right)^{1/p},$$

which is convex and monotonic for all $Y \geq 0$. Hence $\rho_1(Z)$ is a convex risk function. By using Proposition 3.3, it holds that, if $p > 1$

$$\partial\rho_1(Z) = \frac{1}{p}\frac{e^Z}{(\mathbb{E}[e^Z])^{1/q}},$$

where q satisfies $1/q + 1/p = 1$. If $p = 1$, then $\partial\rho_1(Z) = e^Z$.

Example 2. Another important example is $\rho_2(Z) := \ln\mathbb{E}(e^Z)$. This function is convex, monotone and translation-equivalent, i.e., it satisfies conditions C1-C3 but does not satisfy C4. This function can be further developed to a coherent risk measure, called EVaR_α ([1]), which will be discussed in Section 5.

It can be verified that Assumption 3.1 holds. Hence, by using Proposition 3.3, it readily holds that $\partial\rho(Z) = \frac{e^Z}{\mathbb{E}(e^Z)}$.

4 Risk Functions with Centralized Transformation

As is mentioned in Section 2, the deviation measure is to quantify the nonconstancy in a random variable X. In [21], a general deviation measure is defined as follows:

Definition 4.1. a function $\mathcal{D} : \mathscr{L}^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is called a general deviation measure if it satisfies the next four conditions:

D1: $\mathcal{D}(Z + C) = \mathcal{D}(Z)$ for all Z and constant C;
D2: $\mathcal{D}(0) = 0$ and $\mathcal{D}(\lambda Z) = \lambda\mathcal{D}(Z)$ for all Z and $\lambda > 0$;
D3: $\mathcal{D}(Z_1 + Z_2) \leq \mathcal{D}(Z_1) + \mathcal{D}(Z_2)$ for all Z_1 and Z_2;
D4: $\mathcal{D}(Z) \geq 0$ for all Z, with $\mathcal{D}(Z) > 0$ if Z is not a constant.

D2 is the positive homogeneity condition. D3 is called the subadditive condition. In case when D2 holds, D3 can be equivalently replaced by the convexity condition. Moreover, if the measure $\mathcal{D}$ is closed and convex, then D4, together with $\mathcal{D}(0) = 0$, implies that $\mathcal{D}$ is a regular deviation measure. On the other hand, conditions D1-D3 mean that the corresponding risk measure $\mathcal{R}$ generated by $\mathcal{R}(Z) = \mathbb{E}(Z)+\mathcal{D}(Z)$ is at least a convex risk measure (i.e., satisfies condition C1-C3).

It is easy to observe that, a risk function ρ satisfying D1 depends on $Z - \mathbb{E}(Z)$ (by setting $C = -\mathbb{E}(Z)$), which is usually called the centralization of the random variable Z. This means that, in practice, there exists lots of risk functions which can be formulated as $\rho(Z - \mathbb{E}(Z))$.

The composition $\rho(Z - \mathbb{E}(Z))$ can be viewed as an extension of the function family (3.1) discussed in the previous section. Since the inner operator includes both Z and $\mathbb{E}(Z)$, the problem now turns to be more complicated. Notice that $T_0(Z) := Z - \mathbb{E}(Z)$ is in fact a linear operator from $\mathcal{Z}$ to $\mathcal{Z}$, that is, for any $Z_1, Z_2 \in \mathcal{Z}$ and any scalars $a, b \in \mathbb{R}$, we have $T_0(aZ_1 + bZ_2) = a \cdot T_0(Z_1) + b \cdot T_0(Z_2)$.

Furthermore, for any linear operator T, an operator T^* is called its conjugate operator, if for any $Z \in \mathcal{Z}$ and $Z^* \in \mathcal{Z}^*$, it holds that $\langle Z^*, T(Z)\rangle = \langle T^*(Z^*), Z\rangle$. Notice that, For T_0 defined above, it can be verified that its conjugate operator is $T_0^*(Z^*) = Z^* - \mathbb{E}(Z^*)$, i.e., $\langle Z^*, Z - \mathbb{E}(Z)\rangle = \langle Z^* - \mathbb{E}(Z^*), Z\rangle$. Hence, the type of risk function $\rho(Z - \mathbb{E}(Z))$ can be covered by the risk function family $\rho(T(\cdot))$, in which T is a generic linear operator, with T^* as its conjugate.

In abstract, the subdifferential of a convex function with a linear operator has been discussed in locally convex spaces (see e.g. [30]). Here we only have to focus on the case when

$\mathcal{Z}$ is the space of random variables. We first cite the definition of Hadamard differentiability as follows.

Definition 4.2. ([7], Definition 2.45) Let X and Y be vector (linear) normed spaces and consider a mapping $g : X \to Y$, We say that g is directionally differentiable at x in the Hadamard sense if the directional derivative $g'(x,h)$ exists for all h and, moreover,

$$g'(x,h) = \lim_{t\downarrow 0, h'\to h} \frac{g(x+th') - g(x)}{t}.$$

If in addition $g'(x,h)$ is linear in h, it is said that g is Hadamard differentiable at x.

Proposition 4.3. Let ρ be a convex, proper, Hadamard directional differentiable risk measure, and $T : \mathcal{Z} \to \mathcal{Z}$ be a linear operator, with T^* be its conjugate operator. Denote $\tilde{\rho}(\cdot) := \rho(T(\cdot))$, then $\tilde{\rho}$ is convex and

$$\partial\tilde{\rho}(Z) = \mathcal{Q}_0 := cl\{T^*(\zeta) : \zeta \in \partial\rho(Y),\ \text{with } Y = T(Z)\}, \tag{4.1}$$

where cl is the closure operation.

Proof. The convexity of $\tilde{\rho}$ is obvious. To prove (4.1), for any fixed Z, first choose any $\xi \in \mathcal{Q}_0$. By the definition of $\mathcal{Q}_0$, there exists a series of $\zeta_n \in \partial\rho(Y)$, such that $T^*(\zeta_n) \to \xi$ (in the sense of $\|T^*(\zeta_n) - \xi\|_2 \to 0$). Denote $\xi_n := T^*(\zeta_n)$. For any $Z_1 \in \mathcal{Z}$, we have

$$\langle \xi, Z_1 - Z\rangle = \lim_{n\to\infty} \langle \xi_n, Z_1 - Z\rangle.$$

Furthermore, we have

$$\begin{aligned} \langle \xi_n, Z_1 - Z\rangle &= \langle T^*(\zeta_n), Z_1 - Z\rangle = \langle \zeta_n, T(Z_1 - Z)\rangle \\ &\le \rho(T(Z_1)) - \rho(T(Z)). \end{aligned}$$

The last inequality is by the definition of $\partial\rho(Y)$. Consequently, we have $\xi \in \partial\tilde{\rho}(Z)$, which means $\mathcal{Q}_0 \subseteq \partial\tilde{\rho}(Z)$.

Conversely, choose any $\xi_1 \in \tilde{\rho}(Z)$. Let $\tilde{\rho}'(Z,h)$ be the directional derivative of $\tilde{\rho}'$ at Z along the direction $h := Z_1 - Z$. Since $\rho(\cdot)$ is Hadamard directional differentiable, then the chain rule of directional derivative holds, i.e.,

$$\tilde{\rho}'(Z,h) = (\rho \circ T)'(Z,h) = \rho'(T(Z), T(h)).$$

It follows that

$$\begin{aligned} &\langle \xi_1, h\rangle \le \tilde{\rho}'(Z,h) \qquad (4.2)\\ =\ &\rho'(T(Z),T(h)) = \sup_{\zeta\in\partial\rho(Y)} \langle \zeta, T(h)\rangle \\ =\ &\sup_{\zeta\in\partial\rho(Y)} \langle T^*(\zeta), h\rangle = \sup_{\xi\in\mathcal{Q}_0} \langle \xi, h\rangle. \end{aligned}$$

Since the direction h is arbitrary, there must be that $\xi_1 \in \mathcal{Q}_0$. In fact, if $\xi_1 \notin \mathcal{Q}_0$, by the separate theorem of convex set, there must exist a direction h_0 such that

$$\langle \xi_1, h_0\rangle > \sup_{\xi\in\mathcal{Q}_0} \langle \xi, h_0\rangle,$$

which is contrary to (4.2). Hence, $\partial\tilde{\rho}(Z) \subseteq \mathcal{Q}_0$. The proof is completed. □

Remark: The Hadamard directional differentiability of ρ holds if $\mathcal{Z}$ is a Banach space and ρ is a continuous risk function (see e.g., Proposition 2.126 in [7]). Hence this condition is not difficult to satisfy.

Example 3. A commonly used deviation measure of this family is $\rho_3(Z) = \mathbb{E}[(Z-\mathbb{E}(Z))_+]^2$, called upper semi-variance, which is a special case of ρ_a (with $p = 2$) mentioned in Section 1. To use Proposition 4.3, we can define $\rho(Z) = \mathbb{E}[Z_+]^2$, and $T_0(Z) = Z - \mathbb{E}(Z)$, thus $\rho_3(Z) = \rho(T_0(Z))$. Denote $Y = T_0(Z)$. Since $\rho(\cdot)$ is convex and differentiable with $\nabla\rho(Y) = 2Y_+$, it holds that

$$\nabla\rho_3(Z) = T_0^*(\nabla\rho(Y)) = 2(Z - \mathbb{E}[Z])_+ - 2\mathbb{E}[(Z - \mathbb{E}[Z])_+].$$

Example 4. Based on ρ_1 in Example 1, we can construct a new deviation measure $\rho_4(Z) = \left(\mathbb{E}\left[e^{Z-\mathbb{E}Z}\right]\right)^{1/p}$ where $p \in [1, +\infty)$. By Proposition 4.3, it can be directly verified that, in case that $p > 1$, it holds that

$$\partial\rho_4(Z) = \frac{1}{pe^{\mathbb{E}(Z)/p}} \frac{e^Z - \mathbb{E}[e^Z]}{\left(\mathbb{E}\left[e^Z\right]\right)^{1/q}}.$$

In case that $p = 1$, we have

$$\partial\rho_4(Z) = \frac{e^Z - \mathbb{E}[e^Z]}{e^{\mathbb{E}[Z]}}.$$

5 Risk Functions Represented by Optimal Value Functions

In the risk quadrangle, a regret measure $\mathcal{V}$ describes the displeasure in facing the mix of outcomes of X. In economics, a typical regret measure ν is defined as a negative utility (i.e., $\mathcal{V}(Z) = -u(-Z)$, where u is a utility function of $-Z$). As is shown in Table 2, the connection between a regret measure and its corresponding risk function can be established by

$$\mathcal{R}(Z) := \mathbb{E}(Z) + \inf_{t\in R} \mathcal{V}(Z - t). \tag{5.1}$$

In this section, we consider a broader form of risk functions as follows:

$$\rho(Z) := \inf_{t\in\mathcal{T}} \rho_0(Z, t), \tag{5.2}$$

here $\mathcal{T}$ is a convex set in $\mathbb{R}$ (in fact, this means $\mathcal{T}$ must be an interval). Suppose that ρ_0 is convex jointly in Z and t. Furthermore, for each $t \in \mathcal{T}$, $\rho_t(Z) := \rho_0(Z, t)$ is a convex and proper function. For each $Z \in \mathcal{Z}$, suppose that the optimal value of the right-side problem in (5.2) is finite and is attainable by some $t^* \in \mathcal{T}$. Hence, ρ is a convex and finite valued function and the "inf" in (5.2) can be replaced by "min". Further suppose ρ_t and ρ are both lower semi-continuous, then $\rho_t = \rho_t^{**}$, $\rho = \rho^{**}$, and both $\rho_t(\cdot)$ and $\rho(\cdot)$ are subdifferentiable.

Formula (5.2) covers (5.1) as an example by defining

$$\rho_0(Z, t) = t + \nu_0(Z - t) \tag{5.3}$$

in (5.2) and setting $\nu_0(Z) = \mathcal{V}(Z) + \mathbb{E}(Z)$. Another important example is as follows:

$$\rho(Z) := \inf_{t>0} t\tilde{\rho}(Z/t). \tag{5.4}$$

It can be verified that, no matter whatever $\mathcal{V}$ is, $\mathcal{R}$ in (5.1) always satisfies condition C3, i.e., the translation equivalence, and will keep C1,C2 and C4 if $\mathcal{V}$ satisfies these conditions.

Similarly, ρ defined in (5.4) always satisfies C4, i.e., the positive homogeneity, and will keep C1-C3 if $\tilde{\rho}$ satisfies these conditions. Thus these two families are often used to construct a new measure to improve the property of the original function $\mathcal{V}$ or $\tilde{\rho}$. Hence formula (5.2) is an important approach to construct new risk measures, in particular coherent or convex risk measures.

Now we start to explore the subdifferential of the risk function family (5.2). First notice that, in literature $\rho(Z)$ defined by (5.2) with $\mathcal{T} = \mathbb{R}$ is usually called a marginal function. Furthermore, a theoretical result on the subdifferential of a marginal function is as follows.

Proposition 5.1. ([30]) Let $\Phi : \mathbb{X} \times \mathbb{Y} \to \bar{\mathbb{R}}$ be a convex function and $h : \mathbb{Y} \to \bar{\mathbb{R}}$ be the marginal function associated to Φ (both $\mathbb{X}$ and $\mathbb{Y}$ are real linear spaces). Let $(\bar{x}, \bar{y}) \in \mathbb{X} \times \mathbb{Y}$ such that $\Phi(\bar{x}, \bar{y}) \in \mathbb{R}$, then

$$(0, y^*) \in \partial\Phi(\bar{x}, \bar{y}) \Leftrightarrow h(\bar{y}) = \Phi(\bar{x}, \bar{y}), \text{ and } y^* \in \partial h(\bar{y}).$$

In particular, the use of Proposition 5.1 involves the extension of ρ_0 to a new function $\hat{\rho}_0 : \hat{\mathcal{T}} \times \mathcal{Z} \to \bar{\mathbb{R}}$ with $\hat{\mathcal{T}} = \mathbb{R}$, of which its subdifferential coincides with ρ_0 on $\mathcal{T}$. It also involves the computation of $\partial\Phi$ with respect to the combination of Z and t. Hence, in many cases, the direct use of the above conclusion is not easy. In what follows, we further discuss the specific characteristic of $\partial\rho$.

Denote $\rho_t^*(\zeta)$ be the conjugate function of $\rho_t(Z)$. Since $\rho_t = \rho_t^{**}$ and $\rho = \rho^{**}$, then, $\rho(Z) = \inf_{t\in\mathcal{T}} \sup_{\zeta\in\mathcal{Z}^*} \{\langle\zeta, Z\rangle - \rho_t^*(\zeta)\}$.

On the other hand, the conjugate $\rho^*(\zeta)$ of $\rho(Z)$ can be expressed by

$$\begin{aligned} \rho^*(\zeta) &= \sup_{Z\in\mathcal{Z}} \left\{ \langle\zeta, Z\rangle - \inf_{t\in\mathcal{T}} \rho_0(Z, t) \right\} \\ &= \sup_{Z\in\mathcal{Z}} \sup_{t\in\mathcal{T}} \{\langle\zeta, Z\rangle - \rho_t(Z)\} \\ &= \sup_{t\in\mathcal{T}} \rho_t^*(\zeta). \end{aligned}$$

Consequently,

$$\rho(Z) = \sup_{\zeta\in\mathcal{Z}^*} \{\langle\zeta, Z\rangle - \rho^*(\zeta)\} = \sup_{\zeta\in\mathcal{Z}^*} \inf_{t\in\mathcal{T}} \{\langle\zeta, Z\rangle - \rho_t^*(\zeta)\}.$$

Hence, we have

$$\inf_{t\in\mathcal{T}} \sup_{\zeta\in\mathcal{Z}^*} \{\langle\zeta, Z\rangle - \rho_t^*(\zeta)\} = \sup_{\zeta\in\mathcal{Z}^*} \inf_{t\in\mathcal{T}} \{\langle\zeta, Z\rangle - \rho_t^*(\zeta)\}. \tag{5.5}$$

On the left side of (5.5), for each fixed $t \in \mathcal{T}$, a solution ζ of the inner maximum problem is also an element of $\partial\rho_t(Z)$. On the other hand, on the right side of (5.5), ζ is a solution of the outer problem if and only if $\zeta \in \partial\rho(Z)$. To take use of (5.5) to compute $\partial\rho(Z)$, we first need the next result to illustrate the relationship between $\partial\rho_t(Z)$ and $\partial\rho(Z)$.

Lemma 5.2. For given $Z \in \mathcal{Z}$, denote $\vartheta(Z) \subseteq \mathcal{T}$ be the solution set of the inner problem in (5.2). Suppose $\vartheta(Z) \neq \emptyset$, and denote $U(Z) := \bigcap_{t\in\vartheta(Z)} \partial\rho_t(Z)$, then

$$\partial\rho(Z) \subseteq U(Z).$$

Proof. For fixed $Z \in \mathcal{Z}$, choose any $d \in \partial\rho(Z)$ and $t_0 \in \vartheta(Z)$, for any $Z_1 \in \mathcal{Z}$, we have

$$\langle d, Z_1 - Z\rangle \leq \rho(Z_1) - \rho(Z) \leq \rho_0(Z_1, t_0) - \rho_0(Z, t_0).$$

Thus, $d \in \partial\rho_{t_0}(Z)$. It follows from the arbitrariness of t_0 that $\partial\rho(Z) \subseteq U(Z)$. □

Lemma 5.2 provides a framework to compute the $\partial\rho(Z)$.

Algorithm 1 (A Framework to compute $\partial\rho(Z)$)

Step 1. Compute all the solutions t^* which solves the inner problem in (5.2);

Step 2. Compute $U(Z)$, which consists of the common solutions ζ of the inner maximum problem in the left side of (5.5) for all t^* obtained by step 1;

Step 3. Choose $\zeta \in U(Z)$ such that there exists $t_*, t^* \in \mathcal{T}$, so that $t^* \in \vartheta(Z)$, (t_*, ζ) is the solutions of the right side problem in (5.5), and $\rho^*_{t_*}(\zeta) = \rho^*_{t^*}(\zeta)$, which implies the duality relationship holds.

The ζ obtained in Step 3 must be a subgradient of $\rho(Z)$. Furthermore, if $t^* = t_*$, then a saddle point is also obtained.

If the inner risk function $\rho_t(Z)$ is a coherent risk measure, or at least satisfies condition C1 and C4 (convexity and positive homogeneity), then $\rho^*_t(\zeta) = 0$ in the case when $\zeta \in \partial_t(0)$, and $\rho^*_t(\zeta) = +\infty$ in other cases. In such case, it can be verified that a subgradient ζ must be the solution of the problem

$$\begin{array}{cc} \sup\limits_{\zeta\in\mathcal{Z}^*} & \langle \zeta, Z\rangle \\ \text{s.t.} & \zeta \in \bigcap_{t\in\vartheta(Z)} \partial\rho_t(0). \end{array} \tag{5.6}$$

If, on the other hand, for some t, $\rho_t(Z)$ is not positive homogeneous, then the implementation of Algorithm 1 relies on the knowledge of $\rho^*_t(\zeta)$, which may be not easy to obtain its specific expression.

Lemma 5.2 implies that if $U(Z)$ contains only one element, then it must be the subgradient of $\rho(Z)$. This reminds us to explore that, in which cases, the converse statement holds, i.e., $U(Z) \subseteq \partial\rho(Z)$. If this relationship holds, then $\partial\rho(Z) = U(Z)$, and we only have to compute $U(Z)$ to obtain the subdifferential of $\rho(Z)$.

Proposition 5.3. Suppose the next three conditions are satisfied: (i) ρ_0 is continuous jointly in Z and t on $\mathcal{Z} \times \mathcal{T}$; (ii) for any given $Z_0 \in \mathcal{Z}$, there exists a scalar β, a closed interval $[t_a, t_b] \subseteq \mathcal{T}$ $(-\infty < t_a < t_b < +\infty)$, and a neighbour of Z_0, such that for any $Z \in \mathcal{Z}$ within this neighbor, the level set $\text{lev}_\beta\rho_0(Z, \cdot) = \{t \in \mathcal{T} : \rho_0(Z, t) \le \beta\}$ is nonempty and is contained in $[t_a, t_b]$; (iii) given $t_0 \in \vartheta(Z)$ and a direction h, for any sequence $\{t_k\}$ converging to t_0, it holds that $\liminf_{k\to\infty} \rho'_{t_k}(Z, h) \ge \rho'_{t_0}(Z, h)$, then we have

$$\partial\rho(Z) = U(Z).$$

Proof. Given $Z_0 \in \mathcal{Z}$ and a direction h, for any stepsize $\alpha \in [0, 1]$, denote $Z_\alpha = Z_0 + \alpha h$, and t_α be any element of $\vartheta(Z_\alpha)$. In particular, if $\alpha = 0$, then let t_0 be any element of $\vartheta(Z_0)$. By condition (ii), for sufficiently small α, the level set $\text{lev}_\beta\rho_0(Z_\alpha, \cdot) \subseteq [a, b]$. Then choose any sequence $\{\alpha_k\} \to 0$, and correspondingly denote $t_k := t_{\alpha_k}$. Without loss of generality, we can suppose that $t_k := t_{\alpha_k}$ converges to a point $t^* \in \mathcal{T}$. On the other hand, by Proposition 4.4 in [7], conditions (i)-(iii) imply the continuity of the optimal solution of $\rho_0(Z, \cdot)$ on $\mathcal{T}$ with respect to t. Consequently, by $t_k \in \vartheta(Z_{\alpha_k})$, it readily holds that t^* must be an element of $\vartheta(Z_0)$.

Denote $Z_k = Z_0 + \alpha_k h$. By the definition of ρ, it holds that

$$\begin{aligned} \rho(Z_0 + \alpha_k h) - \rho(Z_0) &= \rho_0(Z_k, t_k) - \rho_0(Z_0, t^*) \\ &\ge \rho_0(Z_k, t_k) - \rho_0(Z_0, t_k) \\ &\ge \rho'_{t_k}(Z_0, a_k h) = \alpha_k \rho'_{t_k}(Z_0, h). \end{aligned}$$

Hence, the directional derivative of ρ satisfies

$$\begin{aligned}\rho'(Z_0,h) &= \lim_{k\to\infty}\frac{\rho(Z_0+\alpha_k h)-\rho(Z_0)}{a_k}\\ &\geq \liminf_{k\to\infty}\rho'_{t_k}(Z_0,h)\geq\rho'_{t_0}(Z_0,h).\end{aligned}$$

Consequently, for any $d\in U(Z_0)$, we have $\langle d,h\rangle\leq\rho'_{t_0}(Z_0,h)\leq\rho'(Z_0,h)$. Thus, $d\in\partial\rho(Z_0)$. Hence, $U(Z_0)\subseteq\partial\rho(Z_0)$. □

Corollary 5.4. Suppose that conditions (i) and (ii) in Proposition 5.3 holds, further suppose that (iii'): for given Z and any $t_0\in\vartheta(Z)$, it holds that $\liminf_{t\to t_0,t\in\mathcal{T}}\partial\rho_t(Z)\supseteq\partial\rho_{t_0}(Z)$, i.e., the point-to-set mapping $t\to\partial\rho_t(Z)$ is lower semi-continuous at t_0, then $\partial\rho(Z)=\bigcap_{t\in\vartheta(Z)}\partial\rho_t(Z)$.

Proof. Choose any $d\in U(Z)$, for any $t_0\in\vartheta(Z)$, we have $d\in\partial\rho_{t_0}(Z)$. Since $\liminf_{t\to t_0,t\in\mathcal{T}}\partial\rho_t(Z)\supseteq\partial\rho_{t_0}(Z)$, let $\{t_k\}\subseteq\mathcal{T}$ be any sequence converges to t_0, there exists a sequence $\tilde{d}_k\in\partial\rho_{t_k}(Z)$ such that $\tilde{d}_k\to d$, which implies that $\lim_{k\to\infty}\langle\tilde{d}_k,h\rangle=\langle d,h\rangle$ for any direction h.

Taking into a subsequence if necessary, we suppose that the sequence $\rho'_{t_k}(Z,h)$ converges to a limit value, then

$$\liminf_{k\to\infty}\rho'_{t_k}(Z,h)=\liminf_{k\to\infty}\sup_{d_k\in\partial\rho_{t_k}(Z)}\langle d_k,h\rangle\geq\liminf_{k\to\infty}\langle\tilde{d}_k,h\rangle=\langle d,h\rangle.$$

Hence condition (iii) in In Proposition 5.3 holds, this implies $\partial\rho(Z)=U(Z)$. □

Remark: Recall that, to derive (5.5), we have assumed that $\rho(\cdot)$ is lower semi-continuous. In fact, the conditions (i)-(iii) in Proposition 5.3 guarantee the continuity of $\rho(\cdot)$. This is due to Proposition 4.4 in [7], which is about the continuity of the optimal value function with respect to the parameters.

Example 5. In this example, we consider the next measure, called $\mathrm{EVaR}_{1-\alpha}$ (Entropic Value-at-Risk), which was first studied in [1],

$$\rho_{1-\alpha}(Z):=\inf_{t>0}t\ln\left(\frac{1}{\alpha}\mathbb{E}\left[e^{Z/t}\right]\right),\quad\alpha\in(0,1].\tag{5.7}$$

In this example, we consider $\mathcal{Z}$ to be $\mathscr{L}^{M+}(\Omega,\mathcal{F},P)$, which means that for any $Z\in\mathcal{Z}$, the moment generating function $\mathbb{E}\left[e^{sZ}\right]<+\infty$ for all $s\geq 0$.

As is shown in [1], $\mathrm{EVaR}_{1-\alpha}$ is a coherent measure. This measure can be derived by using (5.4) to homogenize the measure $\rho(Z)=\ln\mathbb{E}(e^Z)-\ln\alpha$, which we have discussed in Example 2.

For $\mathrm{EVaR}_{1-\alpha}$, we have $\rho_0(Z,t)=t\ln\left(\frac{1}{\alpha}\mathbb{E}\left[e^{Z/t}\right]\right)$, and $\mathcal{T}=(0,+\infty)$ in (5.2). Notice that $\lim_{t\downarrow 0}t\ln\left(\mathbb{E}\left[e^{Z/t}\right]\right)=\text{ess-sup}(Z)$. Consequently, if $\text{ess-sup}(Z)<+\infty$, then $\mathcal{T}$ can be extended to $\bar{\mathcal{T}}=[0,+\infty)$.

Furthermore, the case when $\alpha=1$ is trivial. Indeed, by setting $s=1/t$, we have

$$\lim_{t\to\infty}\rho_0(Z,t)=\lim_{s\to 0^+}\frac{1}{s}\ln\mathbb{E}[e^{sZ}]=\lim_{s\to 0^+}\frac{\mathbb{E}[Ze^{sZ}]}{\mathbb{E}[e^{sZ}]}=\mathbb{E}(Z).$$

The last equality follows from the Lebesgue dominated convergence theorem (which implies $\lim_{s\to 0^+}\mathbb{E}[Ze^{sZ}]=\mathbb{E}[Z]$ and $\lim_{s\to 0^+}\mathbb{E}[e^{sZ}]=1$). Furthermore, since $\rho_0(Z,\cdot)$ is convex, it must be monotonic non-increasing with respect to t. Hence, in the case that $\alpha = 1$, we have $\rho(Z) = \mathbb{E}(Z)$, which implies $\nabla\rho(Z) = 1$.

In what follows, we suppose $\alpha \in (0, 1)$. For simplicity we only consider the case when the distribution of Z is continuous, i.e., the cumulative probability function $F(z) := P(Z \le z)$ is continuous.

For fixed $Z_0 \in \mathcal{Z}$, denote t^* be the optimal solution for minimizing $\rho_0(Z_0, \cdot)$ on $\mathcal{T}$. Let z_1 :=ess-sup(Z_0) and choose any z_2 such that $P(Z_0 \le z_2) := p > 1 - \alpha$. Obviously, if $z_1 = +\infty$, then $t^* > 0$. If, on the other hand, we suppose that $z_1 < +\infty$, then

$$\rho_0(Z_0, t) \le \varphi(t) := t \left[\ln\left(pe^{z_2/t} + (1-p)e^{z_1/t}\right) - \ln\alpha\right].$$

It can be verified that $\lim_{t\to 0^+} \varphi(t) = z_1$, and for sufficiently small $t > 0$, we have $\varphi'(t) < 0$. This means there exists some $t > 0$ such that $\varphi(t) <$ess-sup(Z_0), which implies that $t^* > 0$. Consequently, either ess-sup$(Z_0) = +\infty$ or not, we have $t^* > 0$.

Now we turn to show that condition (i), (ii) and (iii') in Corollary 5.4 hold. Firstly, condition (i) follows from the expression of $\rho_{1-\alpha}$ immediately.

Denote $f(t) := t\ln(\frac{1}{\alpha}\mathbb{E}[e^{Z_0/t}])$, $t > 0$. By using Theorem 7.44 in [25], it can be verified that $f(t)$ is differentiable, and

$$f'(t) = \ln\mathbb{E}[e^{Z_0/t}] - \ln\alpha - \frac{1}{t}\cdot\frac{\mathbb{E}[Ze^{Z_0/t}]}{\mathbb{E}[e^{Z_0/t}]}.$$

Since $t^* > 0$ is the optimal solution, by $f'(t^*) = 0$ we have that:

$$t^*\ln\mathbb{E}[e^{Z_0/t^*}] - t^*\ln\alpha = \frac{\mathbb{E}[Z_0e^{Z_0/t^*}]}{\mathbb{E}[e^{Z_0/t^*}]}. \tag{5.8}$$

For this t^*, as is discussed in Example 2, the subdifferential of $\rho_{t^*}(Z)$ can be computed as follows,

$$\partial\rho_{t^*}(Z) = \frac{e^{Z/t^*}}{\mathbb{E}[e^{Z/t^*}]}. \tag{5.9}$$

Hence, condition (iii') in Corollary 5.4 holds.

Denote $z^* = \rho_0(Z_0, t^*)$, i.e., z^* is the optimal value. As is analyzed previously, $z^* <$ess-sup(Z_0). Since ρ_0 is continuous, there exist neighbours $N(Z_0, \delta_1) = \{Z \in \mathcal{Z} : \|Z - Z_0\|_p \le \delta_1\}$ and $N(t^*, \delta_2) = \{t \in \mathcal{T} : |t - t^*| \le \delta_2\}$, and scalars $0 < p_1 < \alpha < p_2$, $z_3 > z^* > z_4$, such that for all $(Z, t) \in N(Z_0, \delta_1) \times N(t^*, \delta_2)$, it holds that $p_1 \le P(Z \ge z_3) < \alpha$, $P(Z \ge z_4) \ge p_2 > \alpha$, and $\rho_0(Z, t) \le \frac{2}{3}z^* + \frac{1}{3}z_3$. Consequently, we have

$$t\ln\left(\frac{1}{\alpha}\mathbb{E}\left[e^{Z/t}\right]\right) \ge t\ln\left[\frac{p_1}{\alpha}e^{z_3/t}\right] = z_3 + t\ln(p_1/\alpha), \tag{5.10}$$

and

$$t\ln\left(\frac{1}{\alpha}\mathbb{E}\left[e^{Z/t}\right]\right) \ge z_4 + t\ln(p_2/\alpha). \tag{5.11}$$

Now we choose $\beta = \frac{1}{3}z^* + \frac{2}{3}z_3$. For any $Z \in N(Z_0, \delta_1)$, the level set $\text{lev}_\beta\rho_0(Z, \cdot)$ is nonempty. Furthermore, notice that $z_4 < \beta < z_3$, and $p_1 < \alpha < p_2$, hence, $\rho_0(Z, t) \le \beta$. Together with (5.10) and (5.11), this implies that

$$t_a := \frac{z^* - z_3}{3\ln(p_1/\alpha)} \le t \le t_b := \frac{\beta - z_4}{\ln(p_2/\alpha)}.$$

i.e., $\text{lev}_\beta \rho_0(Z,\cdot) \subseteq [t_a, t_b] \subseteq \mathcal{T}$. Hence, condition (ii) in Corollary 5.4 holds.

Consequently, by Corollary 5.4, we have $\partial\rho(Z) = \bigcap_{t\in\vartheta(Z)} \partial\rho_t(Z)$. Furthermore, similar with the proof of Lemma 3.1 in [1], it can be verified that $\bar{\rho}_0(Z_0,\cdot)$ is strictly convex. Hence, t^* is the unique solution. Together with (5.9), we have

$$\partial\rho_{1-\alpha}(Z) = \frac{e^{Z/t^*}}{\mathbb{E}[e^{Z/t^*}]}.$$

Example 6. As is mentioned previously, (5.1) or its equivalent form (5.3), is an important approach to generate various risk measures from some basic regret measures. An important instance is the optimized certainty equivalence (OCE), proposed by Ben-Tal et al. [5]. The OCE can be defined by setting ν in (5.3) as follows (see [26])

$$\nu(Z) := \gamma_1 \mathbb{E}\left[Z_+\right] - \gamma_2 \left[Z_-\right] \quad \text{with } 0 \le \gamma_2 < 1 < \gamma_1. \tag{5.12}$$

The OCE further incorporates the CVaR_α (conditional value at risk with the risk level $\alpha \in (0,1)$) as a special case. The $\text{CVaR}_\alpha(Z)$ (defined by $\text{CVaR}_\alpha(Z) = \inf_{t\in\mathbb{R}}[t + \frac{1}{\alpha}\mathbb{E}(Z-t)_+]$, see e.g. [19]), can be derived from the OCE by setting $\alpha = 1/\gamma_1$ and choosing $\gamma_2 = 0$ in (5.12).

Now we use Algorithm 1 to discuss the subdifferential of the general form (5.3), together with the results on the OCE and CVaR_α as examples.

Step 1: For fixed $Z \in \mathcal{Z}$, denote $h_Z(t) := \nu(z-t)$, which is viewed as a function with respect to t. Then $\vartheta(Z) = \{t \in \mathcal{T} | 0 \in 1 + \partial h_Z(t)\}$. In many applications, the measure ν has the form $\nu(Z) = \mathbb{E}\left[g(Z)\right]$. Thus, $h_Z(t) = \mathbb{E}\left[g(z-t)\right]$. It follows from Theorem 7.47 in [25] that, if the function $g : \mathbb{R} \to \mathbb{R}$ is convex, lower semicontinuous, and $\nu(Z)$ is finite for all $Z \in \mathcal{Z}$, then

$$\partial h_Z(t) = -\int_\Omega \partial g(Z(\omega) - t) dP(\omega).$$

As an instance, for OCE, $g(Z) = \gamma_1 Z_+ - \gamma_2 Z_-$. It can be verified that, the solution set $\vartheta(Z)$ is a closed interval, denoted $[t_*, t^*]$, which consists of the value t such that $P(Z < t) \le 1-\alpha \le P(Z \le t)$ where $\alpha = (1-\gamma_2)/(\gamma_1 - \gamma_2)$. In other words, it contains the $(1-\alpha)$-quantile of the cdf $F_Z(z) := P(Z \le z)$ (including the end points).

Step 2: Since $U(Z) := \bigcap_{t\in\vartheta(Z)} \partial\rho_t(Z)$, and $\rho_t(Z) = t + \nu(Z-t)$. For fixed $t \in \mathbb{R}$,

$$\partial\rho_t(Z) = -\partial\nu(Y), \quad \text{with } Y = Z - t.$$

If $\nu(Z) = \mathbb{E}\left[g(Z)\right]$, then $U(Z)$ consists of $\zeta \in Z^*$ such that

$$\zeta(\omega) \in -\bigcap_{t\in\vartheta(Z)} \partial g(Y(\omega)) \quad \text{a.e. } \omega \in \Omega, \quad \text{with } Y = Z - t.$$

For OCE, for simplicity, suppose that $t_* < t^*$, then it can be verified that $U(Z)$ consists of $\zeta \in Z^*$ such that

$$\zeta(\omega) = \begin{cases} \gamma_1 & \text{if } Z(\omega) \ge t^*, \\ \gamma_2 & \text{if } Z(\omega) \le t_*, \\ [\gamma_2, \gamma_1] & \text{if } Z(\omega) \in (t_*, t^*). \end{cases} \tag{5.13}$$

Step 3: First compute the conjugate of $\rho_t(Z)$,

$$\begin{aligned} \rho_t^*(\zeta) &= \sup_{Z\in\mathcal{Z}} \{\langle Z, \zeta\rangle - t - \nu(Z-t)\} \\ &= \sup_{Y\in\mathcal{Z}} \{\langle Y + t, \zeta\rangle - t - \nu(Y)\} \\ &= \nu^*(\zeta) + t(E[\zeta] - 1). \end{aligned}$$

Since ζ satisfies both sides of (5.5) and the whole equality holds, then $\mathbb{E}(\zeta) = 1$, otherwise the rightside of (5.5) will be equal to $-\infty$. Conversely, for any $\zeta \in U(Z)$ such that $\mathbb{E}(\zeta) = 1$, in (5.5) we have

$$\langle \zeta, Z \rangle - \rho_t^*(\zeta) = \langle \zeta, Z \rangle - \nu^*(\zeta).$$

This implies that the equality in (5.5) holds. Consequently,

$$\partial \rho_5(Z) = \{\zeta \in Z^* : \mathbb{E}(\zeta) = 1, \zeta \in U(Z)\}. \tag{5.14}$$

If $\nu(Z) = \mathbb{E}\left[g(Z)\right]$, then (5.14) takes the from

$$\partial \rho_5(Z) = \left\{ \zeta \in Z^* : \mathbb{E}(\zeta) = 1, \zeta(\omega) \in - \bigcap_{t \in \vartheta(Z)} \partial g(Y(\omega)), Y = Z - t \right\}.$$

For the OCE, suppose that $t_* < t^*$, then its subdifferential is:

$$\partial OCE(Z) = \left\{ \zeta \in Z^* : \mathbb{E}(\zeta) = 1, \begin{array}{ll} \zeta(\omega) = \gamma_1 & \text{if } Z(\omega) \geq t^*, \\ \zeta(\omega) = \gamma_2 & \text{if } Z(\omega) \leq t_*, \\ \zeta(\omega) \in [\gamma_2, \gamma_1] & \text{if } Z(\omega) \in (t_*, t^*). \end{array} \right. \tag{5.15}$$

Finally, by setting $\alpha = 1/\gamma_1$ and $\gamma_2 = 0$ in (5.15), we directly obtain the subdifferential of CVaR_α as follows

$$\partial CVaR_\alpha(Z) = \left\{ \zeta \in Z^* : \mathbb{E}(\zeta) = 1, \begin{array}{ll} \zeta(\omega) = 1/\alpha & \text{if } Z(\omega) \geq t^*, \\ \zeta(\omega) = 0 & \text{if } Z(\omega) \leq t_*, \\ \zeta(\omega) \in [0, 1/\alpha] & \text{if } Z(\omega) \in (t_*, t^*). \end{array} \right.$$

6 On the Subgradient Approach

As an application of the previous results, in this section, we discuss the way to develop a subdifferetial approach for the risk minimization problem (1.1) based on the knowledge of the subdifferetial of $\rho(\cdot)$.

First, the random function $F(x, \omega)$ now can be viewed as a mapping (also denoted by F) from $\mathbb{R}^n$ to $\mathcal{Z}$, i.e., it maps $x \in \mathbb{R}^n$ to a random variable $[F(x)](\cdot) \in \mathcal{Z}$. If elements of the subdifferential $\partial\rho(\cdot)$ can be obtained, then we can further consider the subgradient of the composition function $\rho(F(x))$, which is the objective function of the problem (1.1). We cite the following conclusion, which is an extension of Theorem 6.11 in [25].

Theorem 6.1. ([28]) Let $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that $\rho(\cdot)$ is a convex risk function, finite valued and is continuous at $Z_0 := F(x_0)$. If $\rho(\cdot)$ is monotone, or the mapping $F(\cdot)$ is affine, i.e., $[F(x)](\omega) = A(\omega)^T x + b(\omega)$, with $\omega \in \Omega$, $A(\omega) \in \mathbb{R}^n$, $b(\omega) \in \mathbb{R}$, then the composite function $\phi(x) = \rho(F(x))$ is convex at x_0 and

$$\partial \phi(x_0) = \text{cl}\left(\bigcup_{\zeta \in \partial \rho(Z_0)} \int_\Omega \partial_x F(x_0, \omega) \zeta(\omega) dP(\omega) \right). \tag{6.1}$$

The above theorem shows that an element of $\partial\phi(x_0)$ can be computed by (6.1) through the following procedure.

Procedure for subgradient of ϕ:

Step 1: For given x_0, choose any $\zeta \in \partial\rho(Z_0)$ where $Z_0 = F(x_0)$;
Step 2: Compute $d = \int_\Omega \partial_x F(x_0, \omega)\zeta(\omega)dP(\omega)$. Then $d \in \partial\phi(x_0)$.

It can be found that Step 1 is the core of the above procedure, and is related to the expression and properties of the risk function ρ. Hence it is the main issue of this paper and has been carefully discussed in Section 3-5.

Step 2 is associated with the inner mapping F and the probability measure $P(\omega)$, and is separated from the risk function ρ. If the support of Ω is finite, denoted by $\{\omega_1, \cdots, \omega_K\}$, which means the random variables defined on $(\Omega, \mathcal{F}, P)$ all have discrete distributions, then Step 2 can be carried out by computing

$$d = \sum_{k=1}^{K} \partial_x F(x_0, \omega_k)\zeta(\omega_k)P(\omega_k).$$

In the case when the distribution is not discrete, Step 2 can be approximated through a sample average approximation (SAA) method. Without loss of generality, suppose that $\zeta(\omega_i) \neq \zeta(\omega_j)$ for any $\omega_i \neq \omega_j$ in Ω. We can generate independent identically distributed random samples $\zeta^1, \cdots, \zeta^N$ from the distribution of ζ, with N be the sample size, and choose

$$d \in \frac{1}{N}\sum_{j=1}^{N} \zeta^j \partial_x F(x_0, \zeta^j)$$

as an estimation of the subgradient of ϕ at x_0.

Whenever the elements of $\partial\phi(\cdot)$ is computable, the risk minimization problem (1.1) with the form

$$\min_{x \in D_X} \phi(x) \tag{6.2}$$

is solvable for many subgradient type algorithms. A typical example is the subgradient projection algorithm, which generates the iteration as follows:

$$x^{k+1} = \mathrm{Proj}_{D_X}(x^k - \alpha_k d^k) \tag{6.3}$$

where d^k is an subgradient of ϕ at x^k, α_k is a positive stepsize, and Proj_{D_X} is the projection operation onto the set D_X. The convergence analysis of such method is standard.

Finally, recall that, by theorem 6.1, the usage of the subgradient-type algorithms requires ρ to satisfy at most conditions R1 and R2, which is weaker than the coherence required by most dual-representation methods. In particular, if $F(x, \omega)$ is affine with respect to x, then ρ is only required to have convexity. Another case is, if ρ is a convex, differentiable risk function, F is differentiable mapping (with respect to x), the composition ϕ is a convex function, then ϕ is differentiable and

$$\nabla\phi(x_0) = \int_\Omega \nabla_x F(x_0, \omega)\zeta(\omega)dP(\omega) \tag{6.4}$$

which can also be viewed as a special case of (6.1).

Another important subgradient-type method is the Benders decomposition algorithm, which are widely used in stochastic programs. In such algorithm, the knowledge of the subgradient is also crucial for computing the cutting planes of the objective functions. We omit the detailed discussion here. See e.g., [23, 14] for details of such method.

7 Conclusion

Risk minimization problems are widely used in financial and economic decisions. We consider the risk minimization problem with a convex risk function. Being different from the dual representation approach, a subdifferential approach is analyzed for three classes of convex risk functions. The discussion is embeded into the the risk quadrangle theory and covers a large number of practically used risk functions. The results in this paper are basic building blocks for designing subgradient-type algorithms for risk minimization problems and they are helpful in better understanding on the properties of convex risk functions as well.

References

[1] A. Ahmadi-Javid, Entropic value-at-risk: A new coherent risk measure, *J. Optim. Theory Appl.* 155 (2012) 1105–1123.

[2] T.D. Aktürk and Ç. Ararat, Portfolio optimization with two coherent risk measures, *J. Global Optim.* 78 (2020) 597–626.

[3] M. Ang, J. Sun and Q. Yao, On the dual representation of coherent risk measures, *Ann. Oper. Res.* 262 (2018) 29–46.

[4] P. Artzner, F. Delbaen, J.M. Eber and D. Heath, Coherent measures of risk, *Math. Finance* 9 (1999) 203–228.

[5] A. Ben-Tal and M. Teboulle, An old-new concept of convex risk measures: The optimized certainty equivalent, *Math. Finance* 17 (2007) 449–476.

[6] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 2011.

[7] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[8] R.I. Bot, N. Lorenz and G. Wanka, Dual representations for convex risk measures via conjugate duality, *J. Optim. Theory Appl.* 144 (2010) 185–203.

[9] P.C. Fishburn, Utility theory, *Manag. Sci.* 14 (1968) 335–378.

[10] H. Föllmer and A. Schied, Convex measures of risk and trading constraints, *Finance Stoch.* 6 (2002) 429–447.

[11] M. Fritelli and E.R. Gianin, Putting order in risk measures, *J. Banking Finan.* 26 (2002) 1473–1486.

[12] P. Kall and J. Mayer, *Stochastic Linear Programming: Models, Theory, and Computation*, Springer-Velag, New York, 2011.

[13] V.S. Kirilyuk, Polyhedral coherent risk measure and distributionally robust portfolio optimization, *Cybern. Systems Anal.* 59 (2023) 90–100.

[14] C. Li and I.E. Grossmann, A generalized Benders decomposition-based branch and cut algorithm for two-stage stochastic programs with nonconvex constraints and mixed-binary first and second stage variables, *J. Global Optim.* 75 (2019) 247–272.

[15] Y. Li and W. Xing, Globalized distributionally robust optimization based on samples, *J. Global Optim.* 88 (2024) 871–900.

[16] A. Ling, J. Sun, N. Xiu and X. Yang, Robust two-stage stochastic linear optimization with risk aversion, *European J. Oper. Res.* 256 (2017) 215–229.

[17] R.T. Rockafellar, Coherent approaches to risk in optimization under uncertainty, in: *Tutorials in Operations Research*, INFORMS, Maryland, 2007, pp. 38–61.

[18] R.T. Rockafellar and J.O. Royset, Superquantiles and their applications to risk, random variables, and regression. in: *Tutorials in Operations Research*, INFORMS, Maryland, 2013, pp. 151–167.

[19] R.T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *J. Risk.* 2 (2000) 21–42.

[20] R.T. Rockafellar and S. Uryasev, The fundamental risk quadrangle in risk management, optimization and statistical estimation, *Surv. Oper. Res. Manag. Sci.* 18 (2013) 33–53.

[21] R.T. Rockafellar, S. Uryasev and M. Zabarankin, Generalized deviations in risk analysis, *Finance Stoch.* 10 (2006) 51–74.

[22] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 2009.

[23] Ruszczyński, A. Decomposition methods, in: *Stochastic Programming. Vol. 10 of Handbooks in operations research and management science*, Shapiro and Ruszczynski(eds.), Elsevier, Amsterdam, 2003, pp.141–211.

[24] A. Ruszczyński and A. Shapiro, Optimization of convex risk functions, *Math. Oper. Res.* 31 (2006) 433–452.

[25] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[26] J. Sun, X. Yang, Q. Yao and M. Zhang, Risk minimization, regret minimization and the progressive hedging algorithm, *Math. Program.* 181 (2020) 509–530.

[27] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.* 62 (2014) 1203–1466.

[28] H. Yu and J. Sun, Robust stochastic optimization with convex risk measures: A discretized subgradient scheme, *J. Ind. Manag. Optim.* 17 (2021) 81–99.

[29] H. Yu, Scenario decomposable subgradient projection method for two-stage stochastic programming with convex risk measures, *J. Appl. Math. Comput.* 69 (2023) 2389–2419.

[30] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World scientific, Singapore, 2002.

[31] M. Zhang, L. Hou, J. Sun and A. Yan, A model of multistage risk-averse stochastic optimization and its solution by scenario-based decomposition algorithms, *Asia-Pac. J. Oper. Res.* 37 (2020) 1–21.

Manuscript received 13 February 2024
revised 6 May 2024
accepted for publication 23 July 2024

Haodong Yu
School of Statistics and Mathematics
Shanghai Lixin University of Accounting and Finance
Shanghai 201209, China
E-mail address: nianchuixiao@msn.com

Aifan Ling
School of Economics and Finance
Shanghai International Studies University
Shanghai 201620, China
E-mail address: aiffling@163.com

Jie Sun
School of EECMS, Curtin University
Perth 6845, Australia

School of Business, National University of Singapore
Singapore 119245, Singapore
E-mail address: Jie.Sun@curtin.edu.au