ROBUST STOCHASTIC OPTIMIZATION WITH CONVEX RISK MEASURES: A DISCRETIZED SUBGRADIENT SCHEME

HAODONG YU
School of Statistics and Mathematics
Shanghai Lixin University of Accounting and Finance, China

JIE SUN∗
School of Mathematical Science, Chongqing Normal University, China
Faculty of Science and Engineering, Curtin University, Perth, Australia

(Communicated by Cedric Yiu)

Abstract. We study the distributionally robust stochastic optimization problem within a general framework of risk measures, in which the ambiguity set is described by a spectrum of practically used probability distribution constraints such as bounds on mean-deviation and entropic value-at-risk. We show that a subgradient of the objective function can be obtained by solving a finite-dimensional optimization problem, which facilitates subgradient-type algorithms for solving the robust stochastic optimization problem. We develop an algorithm for two-stage robust stochastic programming with conditional value at risk measure. A numerical example is presented to show the effectiveness of the proposed method.

1. Introduction. Let \((\Omega, \mathcal{F}, P)\) be a probability space, equipped with a sigma algebra \(\mathcal{F}\) and a probability measure \(P\). The classical stochastic optimization deals with

\[
\min_{x \in X} \mathbb{E}_P[F(x, \omega)],
\]

where \(F(x, \omega) : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}\) is a random function and \(\mathbb{E}_P[\cdot]\) is the expectation operator with respect to \(\omega\).

Assume that we can only obtain limited information about \(P\), in the distributionally robust model we consider the following robust stochastic model

\[
\min_{x \in X} \max_{P \in \mathcal{P}} \mathbb{E}_P[F(x, \omega)],
\]

where \(\mathcal{P}\) is a set of probability measures, called the ambiguity set, which is often defined by certain constraints on \(P\) such as the range of support and bounds on expectation, see [2, 16, 10] for examples. In particular, Wiesemann, Kuhn, and Sim [32] constructed a general framework which unifies several classes of moment constraints and showed that the corresponding problems are polynomial-time solvable.

2010 Mathematics Subject Classification. Primary: 90C15; 90C47.

Key words and phrases. Stochastic optimization, distributionally robust, convex risk measure, subgradient method, two-stage optimization problem.

This work is partially supported by Grants 11401384, 11671029, 71631008 and B16002 of National Natural Science Foundation of China and by Grant DP160102819 of Australian Research Council.

∗ Corresponding author.
The ambiguity set could be also defined through various distance (or divergence) measures. Hu et al. [14] and Calabrese [9] studied distributionally robust problems with the ambiguity set defined by the Kullback-Leibler (KL) divergence. Klabjan et al. [15] presented a robust inventory control model and introduced an ambiguity set of the demand distribution via the histogram and the χ^2-distance, etc.

Another important issue in distributionally robust optimization is the choice of objective function \( F(x, \omega) \). Mehrotra et al. [23] considered the distributionally robust least squares problems. In the robust two-stage stochastic programming, \( F(x, \omega) \) is defined as the optimal value of the second-stage problem

\[
\min_{y \in \mathcal{Y}(x, \omega)} g(x, y, \omega),
\]

where \( \mathcal{Y}(x, \omega) : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}^m \) is a multifunction, representing the second-stage feasible set. Shapiro et al. [28] studied the discrete distribution case and reformulated the problem to a standard stochastic optimization problem with the objective function being an expectation function. In [2, 16], Ang et al. and Li et al. respectively studied the distributionally robust two-stage problems in which the ambiguity sets are defined by the first and higher order moments of the random variables. In Sun et al. [31], the objective function is the sum of a quadratic cost of the first stage and a risk measure of the quadratic recourse cost of the second stage. Other related study on distributionally robust optimization and control can be found in [19, 17, 18].

In current literatures on distributionally robust optimization, most study focus on the case that the objective functions are expected value functions. From the statistical point of view, the mean measure describes the average risk of the system. However, in many cases, the expectation is not a reasonable risk measure. For instance, in portfolio optimization or risk management, the expectation measure is usually replaced by some mean-variance combination measures or the conditional value-at-risk (denoted CVaR). We note that most commonly used risk measures, including mean-variance and CVaR, are convex, it is therefore natural to consider a general framework of distributionally robust optimization model in which a convex risk measure is involved:

\[
\min_{x \in \mathcal{X}} \left\{ f(x) := \max_{P \in \mathcal{P}} \rho_P[F(x, \omega)] \right\},
\]

where for each fixed \( P \in \mathcal{P}, \rho_P : \mathcal{L}_p \rightarrow \mathbb{R} \) is a proper closed convex functional. See Lüthi and Doege [22] for more details of convex risk measure and its applications. Model (4) can also be viewed as a distributionally robust version of the so-called risk averse optimization:

\[
\min_{x \in \mathcal{X}} \rho_P[F(x, \omega)].
\]

Let \( \hat{\omega}(\omega), \hat{w}(\omega) \in \mathcal{L}, \omega \in \Omega \) be two random variables in the Banach space \( \mathcal{L}_p, 1 \leq p \leq +\infty \). By a risk measure we mean a functional \( \mathcal{L}_p \rightarrow \mathbb{R} \), where \( \mathbb{R} = \mathbb{R} \cup \{+\infty\} \). A risk measure \( \rho(\cdot) \) is said to be coherent if it satisfies the following four conditions:

C1 \( \rho(\cdot) \) is convex, i.e., if \( 0 \leq \lambda \leq 1 \), then \( \rho((1 - \lambda)\hat{z} + \lambda\hat{w}) \leq (1 - \lambda)\rho(\hat{z}) + \lambda\rho(\hat{w}) \);
C2 \( \rho(\cdot) \) is monotone, i.e., if \( \hat{z} \leq \hat{w} \) a.e., then \( \rho(\hat{z}) \leq \rho(\hat{w}) \);
C3 \( \rho(\cdot) \) is translation-equivalent, i.e., \( \rho(\hat{z} + \gamma) = \rho(\hat{z}) + \gamma \);
C4 \( \rho(\cdot) \) is positive homogeneous, i.e., if \( t > 0 \), then \( \rho(t\hat{z}) = t\rho(\hat{z}) \).

As shown in Proposition 4.14 of Föllmer and Schied [11] and Theorem 4(a) of Rockafellar [24], a coherent risk measure is necessarily representable by a support function and vice versa, the support function max_{P \in \mathcal{P}} \mathbb{E}_P[F(x, \omega)] that appears in (2) is a coherent risk measure of \( F(x, \omega) \). Theoretically speaking, this means if \( \rho(\cdot) \) is
is coherent, then (4) can be viewed as an equivalent form of (2) with the ambiguity set \( P \) has some particular form, see Ang et al. [3] for a detailed discussion and some examples. However, this approach cannot be used in some practical problems. The main difficulty is, in certain cases, coherency does not exist. For instance, in portfolio optimization or risk management, the mean-variance measures mentioned above is not coherent since it does not satisfies C4. Even if a risk measure is coherent, it is still not easy to obtain the exact form of a particular ambiguity set \( P \) which guarantees such equivalence.

In this paper, we propose a computational scheme of a general framework of distributionally robust optimization model (4), in which a convex non-coherent risk measure is involved. The computational scheme is based on the combination of the computation of the subgradient of \( f(\cdot) \) and the discretization of probability measures, with respect to the particular form of \( \rho(\cdot) \). We will show that, for most commonly used convex risk measures, problem (4) can be solved through this scheme.

In what follows, we establish the new computational scheme through two steps. In Section 2, we outline the building blocks of our discretized subgradient scheme, without respect to the particular form of \( \rho(\cdot) \). In Section 3, we specifically study three classes of risk measures, which in fact cover most commonly used risk measures, and show that the new computational scheme can be applied to these cases. Especially, a subgradient of the objective functions can be computed through solving a finite-dimensional optimization problem. As an application, or example, in Section 4, we present preliminary numerical results for a two-stage stochastic programming problem that shows the effectiveness of the proposed computational scheme.

2. Building blocks of the computational scheme. In this section, we present the main framework of the computational scheme. The main idea is, since \( \rho_P \) is convex, the function \( f(\cdot) \) in (4) is convex. This implies that, if we can compute an element of \( \partial f(\cdot) \), then problem (4) can be in principle solved through various subgradient-type algorithms, such as cutting plane method, ellipsoid method, etc. Hence in this section, we focus on the approach to compute a subgradient of \( f(x) \) in (4), while omit the discussion of the complete subgradient algorithms, since this is a well-developed method, those who are interested in the details of subgradient algorithms can see, e.g., [4] for reference.

In the first part of this section, we characterize the subdifferential of the function \( \rho_P[F(x, \omega)] \). In the second part, we study the inner maximum problem of (4) and show that for a general framework of \( \mathcal{P} \), the inner maximum problem can be reduced to a problem with discrete distributions.

It is reasonable to assume that for all \( x \in X \), there always exists \( P \in \mathcal{P} \) which maximizes the inner maximization problem of (4) as explained in Proposition 2.91 of Föllmer and Schied [11].

2.1. On the subgradients of \( f(x) \). In this part, we are concerned with the objective function of the outer minimization problem of (4) and begin with the next result.

**Lemma 2.1.** Denote \( \phi_P(x) := \rho_P[F(x, \omega)] \). Given \( x_0 \in X \), suppose \( P_0 \in \mathcal{P} \) is a probability measure which maximizes the inner problem of (4), that is, \( f(x_0) = \phi_{P_0}(x_0) = \rho_{P_0}[F(x_0, \omega)] \). If for any \( P \in \mathcal{P} \), \( \phi_P(\cdot) \) is convex, then the function \( f(\cdot) \) is convex as well and for any \( z \in \partial \phi_{P_0}(x_0) \), \( z \) is also a subgradient of \( f(\cdot) \) at \( x_0 \).
Corollary 1. The convexity is obvious. We focus on the subgradient part. For any $x_0$ given and a corresponding $P_0 \in \mathcal{P}$ such that $f(x_0) = \phi_{P_0}(x_0)$, choose any $z \in \partial \phi_{P_0}(x_0)$, it follows from the definition of $f(\cdot)$ and the subgradient of $\phi_{P_0}(\cdot)$ that, for any $x \in X$, we have
\[
  f(x) = \max_{P \in \mathcal{P}} \phi_P(x) \geq \phi_{P_0}(x) \\
  \geq \phi_{P_0}(x_0) + z^T (x - x_0) = f(x_0) + z^T (x - x_0).
\]
This shows that $z$ is a subgradient of $f(\cdot)$ at $x_0$. \hfill \Box

Next we characterize $\partial \phi_{P_0}(x_0)$ for given $P_0 \in \mathcal{P}$. For any given $x \in X$, $F(x, \cdot)$ is a random variable in $Z := \mathbb{L}_p(\Omega, \mathcal{F}, P)$, hence it is natural to define a mapping $\bar{F} : \mathbb{R}^n \to Z$, where $[\bar{F}(x)](\omega) = F(x, \omega)$. Furthermore, $\bar{F}$ is called to be convex if the random function $F(x, \omega)$ is convex with respect to $x$ for every $\omega \in \Omega$. We next characterize the subdifferential of $\phi(x)$, which, by Lemma 2.1, contains subgradients of $f(\cdot)$.

**Theorem 2.1.** Let $\bar{F} : \mathbb{R}^n \to Z$ be a convex mapping. Suppose that $\rho(\cdot)$ is a convex risk measure, finite valued and is continuous at $Z_0 := \bar{F}(x_0)$. If $\rho(\cdot)$ is a monotone measure, or the mapping $\bar{F}(\cdot)$ is affine, i.e., $[\bar{F}(x)](\omega) = A(\omega)^T x + b(\omega)$, with $\omega \in \Omega$, $A(\omega) \in \mathbb{R}^n$, $b(\omega) \in \mathbb{R}$, then the composite function $\phi(x) = \rho(\bar{F}(x))$ is convex at $x_0$ and
\[
  \partial \phi(x_0) = \text{cl} \left( \bigcup_{\zeta \in \partial \rho(Z_0)} \int_{\Omega} \partial_x F(x_0, \omega) \zeta(\omega) dP(\omega) \right).
\]

**Proof.** The case when $\rho(\cdot)$ is a monotone and convex risk measure is indeed Theorem 6.11 of [29]. In the case that $\bar{F}(\cdot)$ is an affine mapping, since $\rho(\cdot)$ is convex, then the composite function $\phi(\cdot)$ is also convex, indeed, for any $x_1, x_2 \in X$ and $t \in [0, 1]$,
\[
  \phi(tx_1 + (1 - t)x_2) = \rho(\bar{F}(tx_1 + (1 - t)x_2)) = \rho(t\bar{F}(x_1) + (1 - t)\bar{F}(x_2)) \\
  \leq t \rho(\bar{F}(x_1)) + (1 - t)\rho(\bar{F}(x_2)) = t \phi(x_1) + (1 - t)\phi(x_2).
\]
This shows the convexity of $\phi(\cdot)$.

Since $F(x, \omega) = A(\omega)^T x + b(\omega)$, obviously $F_{\omega}(x, h) = A(\omega)^T h$. It follows from Theorem 6.10 of [29] that
\[
  \phi'(x_0, h) = \sup_{\zeta \in \partial \rho(Z_0)} \eta_{\zeta}(h),
\]
where
\[
  \eta_{\zeta}(h) = \int_{\Omega} F_{\omega}(x, h) \zeta(\omega) dP(\omega) = \int_{\Omega} A(\omega)^T h \cdot \zeta(\omega) dP(\omega).
\]
Clearly, $\eta_{\zeta}(\cdot)$ is a linear function, and is therefore a convex, continuous and positively homogeneous function. Thus, $\eta_{\zeta}$ is the support function of $\partial \eta_{\zeta}(0)$. The remainder of this proof is the same as that of Theorem 6.11 of [29]. \hfill \Box

Lemma 2.1 and Theorem 2.1 imply the following result.

**Corollary 1.** Let $\bar{F} : \mathbb{R}^n \to Z$ be a convex mapping. For any $x_0 \in X$, denote $P_0 \in \mathcal{P}$ be a probability measure which maximizes the subproblem of (4), Suppose that $\rho_{P_0}(\cdot)$ is a convex risk measure, finite valued and is continuous at $Z_0 := \bar{F}(x_0)$.
If $\rho_{P_0}(\cdot)$ is a monotone measure, or the mapping $\tilde{F}(\cdot)$ is affine, then for any $\zeta \in \partial \rho(Z_0)$, we have
\[
\int_{\Omega} \partial_x F(x_0, \omega) \zeta(\omega) dP_0(\omega) \subseteq \partial f(x_0).
\] 
This result provides an approach to computing a subgradient of the objective function $f(\cdot)$.

2.2. The inner maximization problem. Now we turn to discuss the inner maximization of problem (4), i.e.,
\[
\max_{P \in \mathcal{P}} \rho_P[F(x, \omega)].
\] 
We focus on the form of ambiguity set $\mathcal{P}$, and leave the discussion on the specific expression of $\rho_P[\cdot]$ in Section 3. As is mentioned in Section 1, many types of ambiguity sets have been studied in the literature. Indeed, most of these ambiguity sets are constructed by using the moment information of the random parameters, of which some important models are as follows.

Example 1. Let
\[
\mathcal{P}_1 = \left\{ \int_{\Omega} dP = 1 : \mathbb{E}_P[\psi_i(\omega)] = b_i, i = 1, \cdots, p \quad \mathbb{E}_P[\psi_i(\omega)] \leq b_i, i = p + 1, \cdots, p + q \right\}. \tag{11}
\] 
This ambiguity set is considered in [29]. Note that if we set $\psi_i(\omega) = 1_{A_i}(\xi(\omega))$ for some $i \in \{1, \cdots, p + q\}$, and $A_i \in \mathbb{R}^n$, then $\mathbb{E}_P[\psi_i(\omega)] = P(\xi \in A_i)$. Thus, the expression (11) includes the probability constraints $P(\xi \in A_i) \in [\underline{p}_i, \bar{p}_i]$. Another important special case is the following ambiguity set studied in [2] and [12],
\[
\mathcal{P} = \left\{ \int_{\Omega} dP = 1 : \mathbb{E}_P(\xi^i) = \mu_0^i, \mathbb{E}_P(|\xi^i|^k) \leq \sigma_k^i, i = 1, \cdots, p; k = 2, \ldots, K \right\}. \tag{12}
\]

Example 2. Let
\[
\mathcal{P}_2 = \left\{ \int_{\Omega} dP = 1 : \mathbb{E}_P(\xi) = \mu_0, \mathbb{E}_P(\xi \xi') = \Sigma_0 + \mu_0 \mu_0^T \right\}. \tag{13}
\] 
This type of ambiguity set is studied by Bertsimas et al. [6], which supposes that the exact information about the mean and covariance matrix is known.

Example 3. Let
\[
\mathcal{P}_3 = \left\{ \int_{\Omega} dP = 1 : |\mathbb{E}_P(\xi^i - \mu_0^i)| \leq \sigma_i \gamma_i, i = 1, \cdots, n \quad \mathbb{E}_P(\xi \xi') \preceq \gamma_0 \Sigma_0 + \mu_0 \mu_0^T \right\}. \tag{14}
\] 
This ambiguity set is introduced in [20] and [21]. Ling et al. [20] used this set in their study on robust tracking error portfolio selection problem.

Example 4. Let
\[
\mathcal{P}_4 = \left\{ \int_{\Omega} dP = 1 : \mathbb{E}_P[ (\xi - \mu_0)(\xi - \mu_0)^T ] \preceq \gamma_1 \Sigma_0 \quad \mathbb{E}_P \left[ \frac{\Sigma_0}{(\xi - \mu_0)^T \gamma_2} (\xi - \mu_0) \right] \succeq 0 \right\}. \tag{15}
\] 
This ambiguity set is proposed by Delage and Ye [10], they considered the confidential region estimation of the mean and covariance matrix and introduced the conic constraints.
All of the above sets can be unified and generalized by the following general framework.

\[
\mathcal{P} = \left\{ \int_{\Omega} dP = 1 : \begin{align*}
E_p[H(\xi)] &= 0, \\
E_p[G_i(\xi)] &\preceq 0, \quad \forall i \in I
\end{align*} \right\},
\]  

(16)

where \( H(\cdot) : \mathbb{R}^n \to \mathbb{R}^p \) and \( G_i(\cdot) : \mathbb{R}^n \to \mathbb{R}^{q_i} \) are continuously differentiable functions, \( I = \{1, \cdots, I\} \). Denote \( \tilde{H}(\omega) = H(\xi(\omega)), \ G_i(\omega) = G_i(\xi(\omega)), \omega \in \Omega \). Then (16) can also be expressed as follows:

\[
\mathcal{P} = \left\{ \int_{\Omega} dP = 1 : \begin{align*}
E_p[\tilde{H}(\omega)] &= 0, \\
E_p[\tilde{G}_i(\omega)] &\preceq 0, \quad \forall i \in I
\end{align*} \right\}.
\]

(17)

Notice that if \( q_i = 1 \) for some \( i \in I \), then the corresponding constraint will be \( E_p[G_i(\omega)] \leq 0 \). Thus, the ambiguity set (17) includes (11)-(15) as special cases. A computationally tractable case where \( \tilde{H}(\omega) \) is affine and \( \tilde{G}_i(\omega) \) is cone-representable has been considered by Bertsimas et al. [8].

It is generally challenging to find a continuous distribution \( P_0 \) such that \( \max_{P \in \mathcal{P}} \rho_P[F(x, \omega)] = \rho_{P_0}[F(x, \omega)] \). Luckily enough, the following conclusion indicates that one can concentrate on finding a discrete distribution \( P_0 \) for this purpose. Therefore the problem (10) can be converted to a finite-dimensional optimization problem.

**Theorem 2.2.** Let \( (\Omega, \mathcal{F}) \) be a measurable space, suppose that every finite subset of \( \Omega \) is \( \mathcal{F} \)-measurable and there exists a probability measure \( P_0 \) on \( (\Omega, \mathcal{F}) \) which satisfies (17). For any positive integer \( m \), denote \( \mathcal{P}_m = \mathcal{P} \cap \mathcal{B}_m \), with \( \mathcal{P} \) defined in (17) and \( \mathcal{B}_m \) being the set of probability measures on \( (\Omega, \mathcal{F}) \) which have a finite support of at most \( m \) points. Further denote \( q = \sum_{i=1}^{I} q_i^2 \) in (17), then the subset \( \mathcal{P}_{p+q+1} \) is not empty, i.e., there exists a probability measure \( \tilde{P} \) on \( (\Omega, \mathcal{F}) \) with a finite support of at most \( m \) points such that \( E_{\tilde{P}}[\tilde{H}(\omega)] = 0 \) and \( E_{\tilde{P}}[\tilde{G}_i(\omega)] \leq 0 \) \( (\forall i \in I) \).

**Proof.** Denote \( A_i = E_{P_0}[\tilde{G}_i(\omega)] \) \( (i \in I) \). It follows from (17) that \( A_i \leq 0 \). Denote \( \tilde{H}(\omega) = (h_1(\omega), \cdots, h_p(\omega))^T, \tilde{G}_i(\omega) = (g^j_{i,k}(\omega))_{j,k=1,\cdots,q_i} \). It follows from Richter-Rogosinski’s theorem (see e.g. [27]) that there exists a nonnegative measure \( \tilde{P} \) on \( (\Omega, \mathcal{F}) \) with a finite support of at most \( p + q + 1 \) points such that

\[
\begin{align*}
E_{\tilde{P}}[h_t(\omega)] &= E_{P_0}[|h_t(\omega)|] = 0 \quad t = 1, \cdots, p, \\
E_{\tilde{P}}[g^j_{i,k}(\omega)] &= E_{P_0}[|g^j_{i,k}(\omega)|] \quad j, k = 1, \cdots, q_i; \quad i = 1, \cdots, I, \\
\int_{\Omega} d\tilde{P} &= 1.
\end{align*}
\]

(18)

This implies that \( \tilde{P} \) is a probability measure, \( E_{\tilde{P}}[\tilde{H}(\omega)] = 0 \) and \( E_{\tilde{P}}[\tilde{G}_i(\omega)] = A_i \leq 0 \) \( (\forall i \in I) \). Hence \( \tilde{P} \in \mathcal{P}_m \). \( \square \)

Together with Corollary 1 and Theorem 2.2, we can build a discretized computational scheme for a subgradient of \( f(\cdot) \) at any \( x_0 \):

**Step 1.** Obtain a discrete distribution \( \tilde{P} \) which satisfies the inner problem (10);

**Step 2.** Choose any element of \( \partial_x F(x_0, \omega) \) and any \( \zeta \in \partial \rho(Z_0) \), where \( Z_0 = \tilde{F}(x_0) \), as defined in Corollary 1, then compute \( \int_{\Omega} \partial_x F(x_0, \omega) \zeta(\omega) dP(\omega) \), which must be an element of \( \partial f(x_0) \).
It remains in Section 3 to discuss the approach to obtain a discrete distribution \( \hat{P} \) in Step 1 and an element \( \zeta \in \partial P(Z_0) \) in Step 2, which involve the specific form of the risk measure \( \rho \). As to the previous task, we will show that, for several types of risk measures, the inner maximum problem (10) can be reduced to a finite-dimensional optimization problem.

3. The subgradients of three types of risk measures. As is mentioned above, the key points of implementing a subgradient algorithm for (4) is to compute \( \hat{P} \) mentioned in Theorem 2.2 and an element \( \zeta \in \partial P(Z_0) \), which depend on the specific form of the risk measure \( \rho \). In this section, we will discuss how to proceed these tasks for three types of measures, which covers many commonly used risk functions. We will show that, for all these risk measures, the subproblem (10) can be reduced to the case of discrete distribution, and a subgradient can be computed effectively.

3.1. Type 1. We begin with the risk measures of the following form:

\[
\rho_P(Z) = \sum_{i=1}^{k} f_i(\mathbb{E}_P[g_i(Z)]),
\]

where the functions \( f_i(\cdot), g_i(\cdot) (i = 1, \cdots, k) \) satisfy the following conditions:

**Assumption 1.**

(i) \( f_i(\cdot) \) are continuously differentiable and monotonically non-decreasing functions;
(ii) \( g_i(\cdot) \) are convex functions;
(iii) the composite measures \( f_i(\mathbb{E}_P[g_i(Z)]) \) are convex with respect to \( Z \).

The risk measure with the form of (19) is a general form containing many interesting special cases. We first consider the simplest case, i.e., \( k = 1 \), then

\[
\rho_P(Z) = f(\mathbb{E}_P[g(Z)]).
\]

An trivial case of (20) is when \( f(t) = t \), which means \( \rho_P(Z) = \mathbb{E}_P[g(Z)] \). Other important instances are as follows.

**Example 1.** (Moment of order \( p \) from a target) \( \rho(Z) = |\mathbb{E}[Z - \tau]|^p \), where \( \tau \in \mathbb{R} \) is a fixed value, \( f(t) = t^{1/p} \) and \( g(Z) = |Z - \tau|^p \).

**Example 2.** (Upper-semideviation of order \( p \) from a target) \( \rho(Z) = (\mathbb{E}(Z - \tau)_+)^{1/p} \), where \( \tau \in \mathbb{R} \) is a fixed value, \( f(t) = t^{1/p} \) and \( g(Z) = (Z - \tau)_+^{1/p} \).

Since for any \( Z \in \mathcal{Z}_p(\Omega,F,P) \), \( \|Z\|_p = (\mathbb{E}|Z|^p)^{1/p} \) is a norm of \( Z \), which is naturally convex, we are readily to get the convexity of the above two risk measure.

**Example 3.** \( \rho(Z) = \ln \mathbb{E}[e^Z] \), where \( f(t) = \ln(t) \) and \( g(Z) = e^Z \). This measure is the logarithmic version of the exponential utility function \( \mathbb{E}[e^Z] \), which is convex. The logarithmic operation keeps the convexity of the exponential utility function, and further enables the measure \( \rho(\cdot) \) to have translation equivalence, i.e., \( \ln \mathbb{E}[e^{Z+a}] = \ln \mathbb{E}[e^Z] + a \).

The case when \( k > 1 \) can be used to combine different risk measures, for instance, if \( k = 2 \), we can define \( \rho(Z) = \mathbb{E}(Z) + [\mathbb{E}|Z - \tau|^p]^{1/p} \), where \( p \geq 2 \). In order to solve (4) with \( \rho \) having the form like (19), the first problem is to solve the subproblem (10). We have the next result.

**Proposition 1.** Suppose that \( P \) is a probability measure on a measurable space \((\Omega,F)\), and that every finite subset of \( \Omega \) is \( F \)-measurable. Let \( s = k+p+q+1 \), where \( k,p,q \) are defined in (19),(16) and Theorem 2.2 respectively. If the risk measure \( \rho_P \)
has the form of (19), then for any \( x \in X \), the inner maximum problem (10) is equivalent to the following problem

\[
\max_{\alpha_1, \cdots, \alpha_s \geq 0} \sum_{t=1}^{k} f_t \left( \sum_{j=1}^{s} \alpha_j g_t(F(x, \omega_j)) \right)
\]  
\( \text{s.t.} \quad \sum_{j=1}^{s} \alpha_j = 1, \)
\[ \sum_{j=1}^{s} \alpha_j \bar{H}(\omega_j) = 0, \]
\[ \sum_{j=1}^{s} \alpha_j G_i(\omega_j) \leq 0 \quad \forall i \in I. \]  

(21)

Proof. Since any feasible solution \( \alpha_1, \cdots, \alpha_s \geq 0 \) and \( \omega_1, \cdots, \omega_s \in \Omega \) of (21) is an element of \( P_s = P \cap B_s \), with \( P \) defined in (16) and \( B_m \) be the set of probability measures on \( (\Omega, \mathcal{F}) \) which has a finite support of at most \( s \) points. It is obvious that any optimal solution of (21) is also a feasible point of (10).

Conversely, if \( P_0 \) is an optimal solution of (10), it follows from Theorem 2.2 that there exists some \( P_s \in P_s \) such that \( \mathbb{E}_{P_0}[g(F(x, \omega))] = \mathbb{E}_{P_s}[g(F(x, \omega))] \) which implies that \( \rho_{P_0}[g(F(x, \omega))] = \rho_{P_s}[g(F(x, \omega))] \). This completes the proof.

The above result shows that we only have to consider the case when \( \Omega \) has a finite support. Without loss of generality, we can assume that \( (\Omega, \mathcal{F}) \subseteq (\mathbb{R}^n, \mathcal{B}) \), also suppose that \( \bar{H}(\cdot), \bar{G}_i(\cdot) \) are continuously differentiable, then problem (21) is a standard nonlinear programming with some conic constraints, which can be solved by various standard optimization techniques. Furthermore, notice that for any \( a, b \in \mathbb{R} \), since \( ab = \frac{1}{4}(a+b)^2-(a^2+b^2) \), which is in fact a D.C. function (differences of two convex functions, D.C.). Hence, under some mild assumptions, the subproblem (21) can be transformed to a D.C. programming and be solved by global optimization techniques to obtain its global solutions. Overall, as a subproblem, (21) can be solved effectively by current optimization methods.

Now we turn to another issue, i.e., the computation of an element of \( \partial \rho_{P_0}(Z_0) \) at \( Z_0 \) for fixed probability measure \( P_0 \). Note that, by Proposition 1, we only have to discuss the case that \( \Omega \) has a finite support.

**Proposition 2.** Let \( Z_0 \in \mathcal{L}_p(\Omega, \mathcal{F}, P) \), where \( p \in [1, +\infty) \) suppose that \( \{\omega_1, \cdots, \omega_s\} \) is a support of \( \Omega \), further suppose that \( P(Z_0(\omega_i) = \eta_i) = \alpha_i, \) where \( \alpha_i \geq 0, \sum_{i=1}^{s} \alpha_i = 1 \), consider the following risk measure \( \rho(\cdot) \)

\[
\rho(Z_0) = \sum_{t=1}^{k} f_t \left( \sum_{j=1}^{s} \alpha_j g_t(\eta_j) \right),
\]  

(22)

where \( f_t(\cdot) \) and \( g_t(\cdot) \) satisfy Assumption 1. Choose any \( Y_0 \in \mathcal{L}_q(\Omega, \mathcal{F}, P) \) (1/p + 1/q = 1) such that

\[
Y_0(\omega_i) \in \sum_{t=1}^{k} \nabla f_t \left( \sum_{j=1}^{s} \alpha_j g_t(\eta_j) \right) \partial g_t(\eta_i) \quad i = 1, \cdots, s,
\]  

(23)

then \( Y_0 \in \partial \rho(Z_0) \).

Proof. Choose any \( d_t \in \partial g_t(\eta_i), i = 1, \cdots, s, t = 1, \cdots k. \) Since \( \{\omega_1, \cdots, \omega_s\} \) is an support of \( \Omega \), for any \( Z \in \mathcal{L}_q(\Omega, \mathcal{F}, P) \), we have

\[
\rho(Z) - \rho(Z_0) = \sum_{t=1}^{k} f_t \left( \sum_{j=1}^{s} \alpha_j g_t(Z(\omega_j)) \right) - \sum_{t=1}^{k} f_t \left( \sum_{j=1}^{s} \alpha_j g_t(\eta_j) \right).
\]  

(24)
Denote $Z_j = Z(\omega_j)$. It follows that for any $x_0 \in \mathbb{R}^n$

$$\partial f_1(x_0) + \cdots + \partial f_k(x_0) \subseteq \partial \left( \sum_{i=1}^k f_i(x_0) \right).$$

(25)

Furthermore, since $f_i(\cdot)$ are continuously differentiable and monotonically nondecreasing functions, by Proposition 4.2.5 of [5], we have

$$\partial f_t \left( \sum_{j=1}^s \alpha_j g_t(\eta_j) \right) \supseteq \nabla f_t \left( \sum_{j=1}^s \alpha_j g_t(\eta_j) \right) \sum_{j=1}^s \alpha_j d_{jt}(Z_j - \eta_j).$$

(26)

Together with (25) and (26), we have

$$\rho(Z) - \rho(Z_0) \geq \sum_{t=1}^k \left[ \nabla f_t \left( \sum_{j=1}^s \alpha_j g_t(\eta_j) \right) \sum_{j=1}^s \alpha_j d_{jt}(Z_j - \eta_j) \right]$$

(27)

$$= \sum_{j=1}^s \left( \alpha_j (Z_j - \eta_j) \cdot \sum_{t=1}^k \left[ \nabla f_t \left( \sum_{j=1}^s \alpha_j g_t(\eta_j) \right) d_{jt} \right] \right)$$

$$= \sum_{j=1}^s \alpha_j Y_0(\omega_i)(Z_j - \eta_j)$$

From the definition of the subgradient the proposition holds.

Now we can propose a complete computational scheme for a subgradient of $f(\cdot)$ with $\rho$ has the form (19):

Algorithm 1. Computational Scheme for $\partial f(x_0)$

**Step 1.** Obtain a discrete distribution $\hat{P}$ as the solution of problem (21). Denote $\hat{P}(Z_0(\omega_i) = \eta_i) = \alpha_i$, where $\alpha_i \geq 0$, $\sum_{i=1}^s \alpha_i = 1$.

**Step 2.** Choose any element $d_j \in \partial_k F(x_0, \omega_j) \ (j = 1, \ldots, s)$ and any $Y_0$ satisfying (23), then compute $\sum_{j=1}^s \alpha_j Y_0(\omega_j)d_j$, which must be an element of $\partial f(x_0)$.

3.2. Type 2. In this subsection, we consider the following type of risk measure:

$$\rho_P(Z) = \inf_{t \in T} f \left( \mathbb{E}_P[g(Z, t)], t \right).$$

(28)

Here $t$ is a common parameter in $f(\cdot)$ and $g(\cdot)$, and $T$ is a subset of $\mathbb{R}$.

**Example 1.** The following risk measure is called conditional value-at-risk (CVaR$_\alpha$), where $f(x, t) = t + x/\alpha$ and $g(Z, t) = (Z - t)_+$. This measure is popularized by Rockafellar and Uryasev in [25] and [26], which attracts strong interests due to its coherency and other good properties.

$$\text{CVaR}_\alpha(Z) = \inf_{t \in \mathbb{R}} [t + \frac{1}{\alpha} \mathbb{E}(Z - t)_+].$$

(29)

**Example 2.** This measure is proposed in [1] and is called Entropic value-at-risk (EVaR$_\alpha$), where $f(x, t) = \frac{1}{t} \ln[x/\alpha]$ and $g(Z, t) = \mathbb{E}(e^{tZ})$. As is shown in [1], this measure is also a coherent measure.

$$\text{EVaR}_\alpha(Z) = \inf_{t \geq 0} \left\{ \frac{1}{t} \ln[\mathbb{E}(e^{tZ})/\alpha] \right\}. $$

(30)
As is mentioned in Section 1, coherency may not be satisfied by some important risk measures. These risk measures do not satisfy all the conditions C1–C4 in Section 1 required by a coherent risk measure. There are two important transformations to generate a new coherent risk measure if the original measure only violates condition C3 or C4. The first one is
\[
\bar{\rho}(Z) := \mathbb{E}(Z) + \inf_{t \in \mathbb{R}} \rho(Z - t).
\]
(31)
It is easy to verify that, whether \( \rho \) satisfies C3 or not, \( \bar{\rho} \) satisfies C3. If \( \rho \) satisfies conditions C1, C2 and C4, then \( \bar{\rho} \) also satisfies them.

The second transformation is
\[
\hat{\rho}(Z) := \inf_{t \geq 0} t\rho(Z/t).
\]
(32)
Similar to the previous transformation, it is not difficult to verify that, \( \hat{\rho} \) always satisfies C4. If \( \rho \) satisfies conditions C1–C3, then \( \hat{\rho} \) also satisfies them.

Notice that, if \( \rho \) is of the form (20), then both \( \bar{\rho} \) and \( \hat{\rho} \) are of the form of (28). In this sense, (28) is more general than (20), which covers many kinds of risk measures.

When the risk measure \( \rho \) has the form (28), problem (4) has the following form.
\[
\min_{x \in X} \max_{P \in \mathcal{P}} \inf_{t \in T} f \left( \mathbb{E}_P[g(F(x, \omega))] , t \right).
\]
(33)
To solve (33), we need more assumptions.

**Assumption 2.** i) \( f(y, t) \) is continuously differentiable and is concave in \( y \); ii) for \( P \in \mathcal{P} \) given, the composite measure \( f \left( \mathbb{E}_P[g(Z, \omega)] , t \right) \) is convex jointly in \( Z \) and \( t \).

It is easy to verify that both the CVaR\( _\alpha \) and EVaR\( _\alpha \) satisfy the above assumptions. We first discuss the inner problem (10), i.e., the inner max-inf problem of Assumption 2(i), we have that
\[
\phi(Z) := \mathbb{E}(Z) + \inf_{t \in \mathbb{R}} \rho(Z - t).
\]
(34)
It follows from the well-known minimax equality theorem (see [30] for a reference) that strong duality holds.

**Lemma 3.1.** ([30]) Suppose that \( \varphi_Z(P, t) \) is convex in \( t \) and concave in \( P \), the ambiguity set \( \mathcal{P} \) is compact, and \( \varphi_Z(\cdot, \cdot) \) is continuous jointly in \( P \) and \( t \), then the following strong duality holds:
\[
\sup_{P \in \mathcal{P}} \inf_{t \in T} \varphi_Z(P, t) = \inf_{t \in T} \sup_{P \in \mathcal{P}} \varphi_Z(P, t).
\]
(34)
It follows that the inner problem (10) with \( \rho \) being of the form of (28) is equivalent to
\[
\inf_{t \in T} \max_{P \in \mathcal{P}} f \left( \mathbb{E}_P[g(Z_0), t] \right).
\]
(35)
with \( Z_0 = F(x_0, \cdot) \).

**Theorem 3.1.** Suppose that there exists \( t_0 \in T \), \( P_0 \in \mathcal{P} \) which optimizes (35), and that for the corresponding measurable space \( (\Omega, \mathcal{F}) \), every finite subset of \( \Omega \) is \( \mathcal{F} \)-measurable. Further denote \( s = p + q + 2 \), where \( p, q \) are defined in (16) and Theorem 2.2 respectively, then problem (35) is equivalent to the following problem:
\[
\min_{t \in T} \max_{\alpha_1, \ldots, \alpha_s, \omega_1, \ldots, \omega_s \in \Omega} f \left( \sum_{j=1}^s \alpha_j g(F(x, \omega_j), t) \right)
\]
(36)
s.t. \( \sum_{j=1}^s \alpha_j = 1 \), \( \sum_{j=1}^s \alpha_j H(\omega_j) = 0 \), \( \sum_{j=1}^s \alpha_j G_i(\omega_j) \leq 0 \) \( \forall i \in I \).
Proof. Assume that \( t_0 \in T \), \( P_0 \in \mathcal{P} \) is a solution of (35), by Theorem 2.2, there exists \( P_* \in \mathcal{P}_m \) where \( m = p + q + 2 \), such that \( \mathbb{E}_{P_0}[g(Z_0, t_0)] = \mathbb{E}_{P_*}[g(Z_0, t_0)] \). It follows that for \( t_0 \in T \) given, the inner max problem of (35) is equivalent to finding \( \alpha_1, \ldots, \alpha_s \geq 0 \) and \( \omega_1, \ldots, \omega_s \in \Omega \) satisfying the inner maximum problem of (36).

Denote \( \vartheta(t) \) to be the optimal value of the following sub maximum problem of (36),

\[
\max_{\omega_1, \ldots, \omega_s \in \Omega} f \left( \sum_{j=1}^{s} \alpha_j g(F(x, \omega_j), t), t \right)
\]

s.t.
\[
\sum_{j=1}^{s} \alpha_j = 1, \\
\sum_{j=1}^{s} \alpha_j H_\omega(\omega_j) = 0, \\
\sum_{j=1}^{s} \alpha_j \tilde{G}_\omega(\omega_j) \leq 0 \quad \forall \omega \in \mathcal{I}.
\]

By Assumption 2(ii), it holds that for given \( Z_0 = F(x_0, \cdot) \), \( f (\mathbb{E}_P[g(Z_0, t)], t) \) is convex in \( t \). Thus, for any \( \alpha_1^0, \ldots, \alpha_s^0 \geq 0 \) and \( \omega_1^0, \ldots, \omega_s^0 \in \Omega \) given, the function \( f (\sum_{j=1}^{s} \alpha_j^0 g(F(x, \omega_j^0), t), t) \) is subdifferentiable with respect to \( t \). These results show that we can use the following search method to solve the subproblem (35).

**Algorithm 2. A bisection algorithm**

- **Step 1**: Choose \( t_0 = (t_L + t_R)/2 \). Solve the subproblem (37) at \( t_0 \) to get \( \alpha_1^0, \ldots, \alpha_s^0 \) and \( \omega_1^0, \ldots, \omega_s^0 \);

- **Step 2**: Compute \( z \in \partial f \left( \sum_{j=1}^{s} \alpha_j^0 g(F(x, \omega_j^0), t_0), t_0 \right) \);

- **Step 3**: Stopping criterion: if \( z = 0 \) or \( t_R - t_L < \epsilon \);

- **Step 4**: Update rule: If \( z > 0 \), let \( t_R = t_0 \); if \( z < 0 \), let \( t_L = t_0 \); then return to step 1;

**Proposition 3.** Suppose that Assumption 2 holds, and that every finite subset of \( \Omega \) is \( \mathcal{F} \)-measurable, then at each iteration of Algorithm 3.1, the interval will always contain the optimal solution of \( \min_{t \in T} \vartheta(t) \).

Proof. For any \( t_0 \in T \) fixed, denote \( \alpha_1^0, \ldots, \alpha_s^0 \geq 0 \) and \( \omega_1^0, \ldots, \omega_s^0 \in \Omega \) be the solution of problem (37). Since \( \vartheta(t) = \max_{P \in \mathcal{P}} f(\mathbb{E}_P[g(Z_0, t)], t) \), same as the proof of Lemma 2.1, it holds that \( \vartheta(t) \) is a convex function, and for any \( z \in \partial f \left( \sum_{j=1}^{s} \alpha_j^0 g(F(x, \omega_j^0), t_0), t_0 \right) \), \( z \) is also a subgradient of \( \vartheta(\cdot) \) at \( t_0 \).

If \( z = 0 \), we immediately get that \( 0 \not\in \partial \vartheta(t_0) \), which means that \( t_0 \) is the minimal point of \( \vartheta(\cdot) \). Otherwise, the algorithm will retain the interval such that \( z(t - t_0) \leq 0 \). We claim that this area will contain \( t_* \). Indeed, if \( z(t - t_0) > 0 \), it follows from the convexity of \( \vartheta(\cdot) \) that

\[
\vartheta(t) \geq \vartheta(t_0) + z(t - t_0) > \vartheta(t_0),
\]

which means \( t \) is impossible to be a minimal point.

As a result, the length \( l_n = t_R - t_L \) of the interval containing \( t_* \) vanishes at the rate \( l_n = (1/2)^n l_0 \), hence the algorithm will converge to a solution \( t_* \).

We now turn to discuss the expression of \( \zeta \in \partial \rho(Z_0) \). Since \( \rho \) is a proper lower semicontinuous risk measure, by the discussion in Section 6.3 in [29], a general result
is as follows:

$$\partial \rho(Z) = \arg\max_{\zeta \in \text{dom}(\rho^*)} \{ \langle \zeta, Z \rangle - \rho^*(\zeta) \},$$

(38)

where $\rho^*$ is the conjugate of $\rho$ with $\text{dom}(\rho^*)$ being its domain and $\langle \zeta, Z \rangle := \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega)$. Furthermore, if $\rho$ is positively homogeneous, then $\text{dom}(\rho^*) = \partial \rho(0)$, and

$$\partial \rho(Z) = \arg\max_{\zeta \in \text{dom}(\rho^*)} \{ \langle \zeta, Z \rangle \}.$$

(39)

Hence, for the two examples in this subsection, we only have to focus on the solution method for (38) or (39).

**Case 1:** In Example 1, $\rho(\cdot)$ is the CVaR$_{\alpha}$. It follows from the analysis in [29] that

$$\text{dom}(\rho^*) = \{ \zeta \in L_{\infty}(\Omega, F, P) : \zeta(\omega) \in [0, \alpha^{-1}], \text{a.e.}, \omega \in \Omega, \mathbb{E}[\zeta] = 1 \}. \quad (40)$$

Denote $p := P(Z > \text{VaR}_{\alpha}(Z))$, $p_0 := P(Z = \text{VaR}_{\alpha}(Z))$, where $\text{VaR}_{\alpha}(Z)$ is the Value-at-Risk of $Z$, i.e., $\text{VaR}_{\alpha}(Z) = \inf \{ z : P(Z \leq z) \geq 1 - \alpha \}$. Since we obtain a finite support discrete probability measure from (36), it follows that $p < \alpha$ and $p_0 > 0$. Together with (39) and (40), we have that the random variable $Y$ defined as follows is an element of $\partial \text{CVaR}_{\alpha}(Z)$.

$$Y(\omega) = \begin{cases} \alpha^{-1} & \text{if } Z(\omega) > \text{VaR}_{\alpha}(Z), \\ 1 - \frac{p}{\alpha} \frac{1}{p_0} & \text{if } Z(\omega) = \text{VaR}_{\alpha}(Z), \\ 0 & \text{if } Z(\omega) < \text{VaR}_{\alpha}(Z). \end{cases} \quad (41)$$

In Section 4, we will apply the above conclusions to present an algorithm for two-stage robust stochastic programming with CVaR measure.

**Case 2:** In Example 2, $\rho$ is the Entropic Value-at-Risk. Denote $\hat{\rho}(Z) = \ln[\mathbb{E}(e^Z)/\alpha]$. It can be verified that $\hat{\rho}(\cdot)$ is convex, and that $\rho(Z) = \inf_{t > 0}\{ t \hat{\rho}(Z/t) \}$. It follows that

$$\text{dom}(\rho^*) = \{ \zeta \in \mathbb{Z}^* : \langle \zeta, Z \rangle \leq \hat{\rho}(Z), \forall Z \in \mathbb{Z}^* \}.$$ \quad (42)

That is,

$$\text{dom}(\rho^*) = \{ \zeta \in \mathbb{Z}^* : \mathbb{E}[\zeta Z] \leq \ln[\mathbb{E}(e^Z)/\alpha], \forall Z \in \mathbb{Z}^* \}. \quad (43)$$

Let $t \in T$, $\alpha^0_1, \cdots, \alpha^0_s \geq 0$ and $\omega^0_i, \cdots, \omega^0_s \in \Omega$ be the solution of (36), since the Entropic VaR$_{\alpha}$ is a coherent risk measure, by (39), the random variable $\zeta \in \partial \rho(Z)$ can be obtained by solving the problem below.

$$\max_{\zeta} \sum_{i=1}^{s} \alpha^0_i \zeta(\omega^0_i) Z(\omega^0_i) \quad (44)$$

s.t. \quad \sum_{i=1}^{s} \alpha^0_i \zeta(\omega^0_i) \eta_i \leq \ln \left[ \sum_{i=1}^{s} \alpha^0_i e^{\eta_i} \right] - \ln(\alpha) \quad \forall \eta \in \mathbb{R}^s.$$

The next result says that problem (44) can be solved by using an ellipsoid method, hence we can compute an element of $\partial \rho(Z_0)$ effectively.

**Proposition 4.** Suppose that the optimal solution set of problem (44) is non-empty, then (44) can be solved to any precision $\epsilon$ in time polynomial in $\log(1/\epsilon)$ and in the size of problem.
Proof. First notice that (44) is a convex problem with linear objective function. Furthermore, for any given \( \zeta(\omega_i^0), i = 1, \cdots, s \), the feasibility of the constraint in (44) can be verified by solving the following convex problem (Notice that the convexity of this problem follows from the convexity of \( \hat{f} \)).

\[
\min \ln \left[ \sum_{i=1}^s \alpha_j^0 e^{\eta_i} \right] - \sum_{i=1}^s \alpha_j^0 \zeta(\omega_i^0)\eta_i. \tag{45}
\]

If \( \eta \) is an infeasible point of (44), a hyperplane that separates \( \eta \) from the feasible set can be generated in time polynomial in the dimension of \( \eta \). Hence, by Lemma 2 in [10], can be solved to any precision \( \epsilon \) in time polynomial by using the ellipsoid method proposed in [13]. \( \square \)

Similar to Subsection 3.1, we can propose a complete computational scheme for a subgradient of \( f(\cdot) \) with \( \rho \) be a EVaR measure.

Algorithm 3. Computational Scheme for \( \partial f(x_0) \)

**Step 1.** Obtain a discrete distribution \( \hat{P} \) by solving problem (36) through Algorithm 2. Denote \( \hat{P}(Z_0(\omega_i) = \eta_j) = \alpha_i \), where \( \alpha_i \geq 0, \Sigma_i=1 \alpha_i = 1; \)

**Step 2.** Choose any element of \( d_j \in \partial_x F(x_0, \omega_j) \) \((j = 1, \cdots, s)\) and compute an element \( \zeta \in \partial \rho(Z_0) \) by obtaining a solution of (44), then compute \( \sum_{j=1}^s \alpha_j \zeta(\omega_j)d_j \), which must be an element of \( \partial f(x_0) \).

3.3. Type 3. In this subsection, we consider the following two risk measures.

**Example 1.** (Mean-deviation of order \( p \)) Let \( c \geq 0, p \in [1, +\infty) \), and for any \( Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P) \) let

\[
\rho(Z) = E(Z) + c (E[Z - E(Z)]^p)^{1/p} . \tag{46}
\]

An important special case of this measure is \( \rho(Z) = E(Z) + c \sqrt{Var(Z)} \), with \( p = 2 \).

**Example 2.** (Mean-semideviation of order \( p \)) Let \( c \geq 0, p \in [1, +\infty) \), and for any \( Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P) \) let

\[
\rho(Z) = E(Z) + c (E(Z - E(Z))^p)^{1/p} . \tag{47}
\]

This measure is an modification of the previous example, which is commonly used in portfolio optimization.

Similar to the previous two subsections, we first discuss the specific form of inner problem (10). Notice that these two measures can be unified as follows:

\[
\rho(Z) = E(Z) + c (E[h(Z - E(Z))]^p)^{1/p}, \tag{48}
\]

with \( h(\cdot): \mathbb{R} \to \mathbb{R} \). In example 1, \( h(z) = |z| \), while in example 2, \( h(z) = z_+ \). Thus, we now consider the following problem:

\[
\max_{P \in \mathcal{P}} E_P(Z) + c (E_P[h(Z - E_P(Z))]^p)^{1/p} . \tag{49}
\]

**Theorem 3.2.** Suppose that for the measurable space \( (\Omega, \mathcal{F}) \), every finite subset of \( \Omega \) is \( \mathcal{F} \)-measurable. Further denote \( s = p + q + 3 \), where \( p, q \) are defined in (16) and Theorem 2.2 respectively, then for any \( Z \) fixed, the problem (49) is equivalent to the
following problem:
\[
\max_{\alpha_1, \ldots, \alpha_s \geq 0, \mu \in \mathbb{R}, \omega_1, \ldots, \omega_s \in \Omega} \mu + c \left( \sum_{j=1}^{s} \alpha_j h(Z_j - \mu)^p \right)^{1/p}
\]  
\[
\text{s.t.} \quad \sum_{j=1}^{s} \alpha_j = 1,
\sum_{j=1}^{s} \alpha_j Z_j = \mu,
\sum_{j=1}^{s} \alpha_j \bar{G}_i(\omega_j) = 0,
\sum_{j=1}^{s} \alpha_j \bar{H}_i(\omega_j) \leq 0 \quad \forall i \in I.
\]  
\]

\[\text{Proof.}\] Problem (49) is obviously equivalent to the following problem:
\[
\max_{P \in \mathcal{P}, \mu \in \mathbb{R}} \mu + c \left( \mathbb{E}_P[h(Z - \mu)^p] \right)^{1/p}
\]  
\[
\text{s.t.} \quad \mathbb{E}_P(Z) = \mu.
\]  
By the expression of (16), the optimal solution of (51) is a feasible solution of (50). Conversely, if \(P_0 \in \mathcal{P}, \mu_0 \in \mathbb{R}\) is an optimal solution of (50), by Theorem 2.2, there exists \(\tilde{P}_0 \in \mathcal{P} = \mathcal{P} \cap \mathcal{B}_s, s = p + q + 3\) (see Theorem 2.2 for the definition of \(\mathcal{B}_s\)) such that
\[
\mathbb{E}_{\tilde{P}_0}[h(Z - \mu_0)] = \mathbb{E}_{P_0}[h(Z - \mu_0)],
\]  
and
\[
\mathbb{E}_{\tilde{P}_0}(Z) = \mu_0.
\]  
Consequently, \((\tilde{P}_0, \mu_0)\) is still an optimal solution of (50), and is also a feasible solution of (51), and the optimal values of these two problems are equal. Hence, problem (50) is equivalent to (51).

Consequently, we can solve the inner maximum problem by solving (50). Now we turn to discuss the subgradient of \(\rho\). Notice that the previous two risk measures satisfy condition C1, C3, and C4 of a coherent measure, but generally do not satisfy the monotonicity condition C2. Hence, to apply Corollary 1 to compute \(\partial f(x)\), we assume that \(F(\cdot, \omega)\) is an affine function, i.e., \(F(x, \omega) = A(\omega)^T x + b(\omega)\). Based on such assumption, an element of \(\partial \rho(Z)\) can be computed directly by using the discussion in Example 6.19 and Example 6.20 of [29] respectively and we omitted these expression here.

4. Two-stage stochastic linear program: An application.

4.1. Implementation issues. In this section, as an application, we discuss the distributionally robust two-stage stochastic programming problem. As is mentioned in Section 1, this problem can be viewed as the case when \(F(x, \omega)\) is the optimal value of a two-stage programming. We consider the following distributionally robust two-stage problem:

\[
\min_{x \in \mathcal{X}} \max_{P \in \mathcal{P}} c^T x + \rho_P[F(x, \xi)]
\]  
\[
(52)
\]
where
\[
F(x, \xi) = \min\{q^T y : Ax + Dy = h, y \geq 0\},
\]  
\[
(53)
\]
and \(\mathcal{P}\) is the ambiguity set. Here \(\xi = (q, A, h)\) are parameters with uncertainty. Or, it can be written as \(\xi(\omega) = (q(\omega), A(\omega), h(\omega)), \omega \in \Omega\). The matrix \(D\) is fixed, in this case the problem (52) is said to have fixed recourse. We study the case when
\( \rho \) is the CVaR\(_\alpha\), i.e., \( \rho(Z) = \inf_{t \in \mathbb{R}} \{ t + \frac{1}{\alpha} \mathbb{E}(Z - t)_+ \} \), and the ambiguity set \( \mathcal{P} \) is defined by (11).

Suppose that for all \( x \in X \) and \( \xi, \{ y \geq 0 : Ax + Dy = h \} \) is nonempty and \( F(x, \xi) > -\infty \), it follows from the duality theory of linear programming that (53) is equivalent to

\[
F(x, \xi) = \max_z \{ (h - Ax)^T z : D^T z \leq q \}. \tag{54}
\]

To solve (52), the first step is to solve the subproblem (10). As is shown in Theorem 3.1, the subproblem (10) is equivalent to (36). Together with (53) and the expression of CVaR\(_\alpha\), we further have the following equivalent form of the inner maximum problem.

\[
\begin{align*}
\min_{t \in \mathcal{T}} & \quad \max_{\omega_1, \ldots, \omega_s \in \Omega} \; c^T x + t + \frac{1}{\alpha} \sum_{j=1}^s \alpha_j [(h(\omega_j) - A(\omega_j)x)^T z_j - t]_+ \tag{55} \\
\text{s.t.} & \quad \sum_{j=1}^s \alpha_j = 1, \quad \alpha_1, \ldots, \alpha_s \geq 0, \\
& \quad \sum_{j=1}^s \alpha_j \psi_i(\omega_j) = b_i, \; i = 1, \ldots, p, \\
& \quad \sum_{j=1}^s \alpha_j \psi_i(\omega_j) \leq b_i, \; i = p + 1, \ldots, p + q, \\
& \quad D^T z_j \leq q(\omega_j), \; j = 1, \ldots, s.
\end{align*}
\]

Denote \( Z_j := (h(\omega_j) - A(\omega_j)x)^T z_j \) and

\[
R(t) := t + \frac{1}{\alpha} \sum_{j=1}^s \alpha_j (Z_j - t)_+.
\tag{56}
\]

Without loss of generality, we assume that \( Z_1 < \cdots < Z_s \). It is not difficult to compute the subdifferential of \( R(\cdot) \) at \( t_0 \):

\[
\partial R(t_0) = \begin{cases} 
1 - 1/\alpha & \text{if } t < Z_1, \\
1 - \frac{1}{\alpha} \sum_{j=l+1}^s \alpha_j & \text{if } Z_l < t < Z_{l+1}, \\
1 - \frac{1}{\alpha} \sum_{j=1}^l \alpha_j - \frac{1}{\alpha} \sum_{j=l+1}^s \alpha_j & \text{if } t = Z_l, \\
1 - \frac{\alpha_s}{\alpha} & \text{if } t = Z_s, \\
1 & \text{if } t > Z_s.
\end{cases}
\tag{57}
\]

Consequently, we can use Algorithm 2 to solve (55). Notice that the objective function in this case is \( R(\cdot) \), and at each iteration, we choose \( z \in \partial R(t) \) by using expression (57).

The next step is to solve the outer minimum problem of (52) by using a subgradient algorithm, in which the main issue is the convexity of \( F(x, \xi) \) and its subdifferential. We cite the following result, which, together with the coherence property of CVaR\(_\alpha\), shows that the assumption in Corollary 1 holds for problem (52).
Proposition 5. ([29]) For any $\xi$ given, the function $F(x, \xi)$ is convex. Moreover, if $F(x_0, \xi)$ is finite for given $x_0 \in X$ and $\xi$, then $F(\cdot, \xi)$ is subdifferentiable at $x_0$, and
\[
\partial F(x_0, \xi) = -A^T V(x_0, \xi)
\]
where
\[
V(x_0, \xi) := \arg \max_{D^T \pi \leq q} \pi^T (h - Ax).
\]

We thereby present a complete computational scheme for $\partial f(x_0)$ for two-stage robust stochastic programming with CVaR measure.

Algorithm 4. Computational Scheme for $\partial f(x_0)$

Step 1. Obtain a solution of problem (55) by using Algorithm 2, in which the subgradient vector $z$ is computed by (57). A discrete distribution $\tilde{P}$ is constructed by denoting $\tilde{P}(Z_0(\omega_i) = \eta_i) = \alpha_i$, where $\alpha_i \geq 0$, $\sum_{i=1}^s \alpha_i = 1$;

Step 2. Compute an element of $d_j \in \partial_x F(x_0, \omega_j)$ for $j = 1, \ldots, s$ by (58) and (59);

Step 3. Compute $Y_0$ by (41), which is an element of $\partial \rho(Z_0)$, then compute $\sum_{j=1}^s \alpha_j Y_0(\omega_j) d_j$, which must be an element of $\partial f(x_0)$.

4.2. An example. We consider the following example, which is first introduced in Bertsimas and Freund [7]. Here we use the following slightly modified version.

Example 1. A company manager is considering the amount of two types of steel to purchase (at $25/lb for Steel A and $18/lb for Steel B) for producing wrenches and pliers in next month. The manufacturing process involves molding the tools on a molding machine and then assembling the tools on an assembly machine. Here are the technical data.

| Table 1. parameters of the test problem |
|----------------------------------------|
| $w$ (Wrench) | $p$ (Plier) |
|----------------|--------------|
| $x$: Steel A(lbs.) | 1.5 | 1 |
| $y$: Steel B(lbs.) | 1 | 2 |
| Molding Machine (hours) | 1 | 1 |
| Assembly Machine (hours) | .3 | .5 |
| Contribution to Earnings ($/1000 units) | 130 | 100 |

There are uncertainties that will influence his decision. 1. The total available molding hours of next month (denoted $h_1$) could be either 21,000 or 25,000 at 50% possibility for each case. 2. The total available assembly hours of next month (denoted $h_2$) could be 8000 or 10,000, with 50/50 chance. We also assume that the random variable $h_1$ and $h_2$ are independent with each other, which implies that the combined distribution of $(h_1, h_2)$ is as follows:

| Table 2. combined distribution of $h$ |
|-------------------------------|
| $h_2$ | $h_1$ |
| 21000 | 25000 |
| 8000 | 0.25 | 0.25 |
| 10000 | 0.25 | 0.25 |
The manager would like to plan, in addition to the amount of the two types of steel to purchase, for the production of wrenches and pliers of next month so as to maximize the expected net revenue of this company.

Indeed, if we choose CVaR as the risk measure, the original example is to solve the following problem

\[
\min_{x,y \geq 0} 25x + 18y + \text{CVaR}_\alpha[F(x, y, h)]
\]

(60)

where \( F(x, y, h) \) is the optimal value of the next linear program, with \( h \) being a random vector,

\[
F(x, y, h) = \max_{w, p \geq 0} -130w - 100p
\]

s.t.

\[
0.3w + 0.5p \leq h_1,
\]

\[
1.5w + p - x \leq 0,
\]

\[
w + 2p - y \leq 0.
\]

Together with (29), problem (60) is equivalent to the next problem:

\[
\min_{x, y, t, w_i, p_i} \quad \min_{x, y, t, w_i, p_i} 25x + 18y + t + \frac{1}{\alpha} \sum_{i=1}^{4} 0.25[-130w_i - 100p_i - t]_+
\]

(62)

where \( h_{1i} \) and \( h_{2i} \) are defined as follows:

|   | 1     | 2     | 3     | 4     |
|---|-------|-------|-------|-------|
| \( h_{1i} \) | 21000 | 21000 | 25000 | 25000 |
| \( h_{2i} \) | 8000  | 10000 | 8000  | 10000 |

By setting \( \alpha = 0.1 \), the solution of Problem (60) is \( x = 15002, y = 9969 \), minimal cost \( = -484819 \), and the production plans under various scenarios are as follows.

|   | 1     | 2     | 3     | 4     |
|---|-------|-------|-------|-------|
| \( w_i \) | 7988  | 9969  | 8000  | 9969  |
| \( p_i \) | 77    | 10    | 27    | 10    |

4.3. Distributionally robustness of example 1. It is not difficult to compute the means and variances of \( h_1 \) and \( h_2 \)

\[
\mathbb{E}(h_1) = 23000, \mathbb{E}(h_2) = 9000, \mathbb{E}(h_1^2) = 533 \times 10^6, \mathbb{E}(h_2^2) = 82 \times 10^6,
\]

\[
21000 \leq h_1 \leq 25000, 8000 \leq h_2 \leq 10000.
\]

(63)
Now, we formulate the distributionally robust problem as follows:

$$\min_{x \geq 0} \max_{P \in \mathcal{P}} 25x + 18y + \text{CVaR}_\alpha[F(x, y, h)]$$

(64)

where $F(x, y, h)$ is defined by (61) and $\alpha = 0.1$. We define the following ambiguity set

$$\mathcal{P} = \left\{ \int_\Omega dP = 1 : \mathbb{E}_P[h] = \mu, \mathbb{E}_P[h^2] \leq \sigma, \Omega = \{h : l_1 \leq h \leq l_2\} \right\}.$$  

(65)

We define $h^2 := (h^2_1, h^2_2)^T$, and

$$\mu = \begin{bmatrix} 23000 \\ 9000 \end{bmatrix}, \sigma = \begin{bmatrix} 533 \\ 82 \end{bmatrix} \times 10^6, l_1 = \begin{bmatrix} 21000 \\ 8000 \end{bmatrix}, l_2 = \begin{bmatrix} 25000 \\ 10000 \end{bmatrix}.$$  

(66)

The numerical result is listed as follows. $x = 2068$, $y = 2189$, minimal cost $= -128824$, the worst-case distribution of $(h_1, h_2)$ is as follows:

|   | Pro   | 0.5   | 0.0134 | 0.0539 | 0.0006 | 0.4321 |
|---|-------|-------|--------|--------|--------|--------|
| $h_1$ | 22355 | 22216 | 21008  | 22239  | 24021  |
| $h_2$ | 8000  | 10000 | 10000  | 10000  | 10000  |

Table 5. worst-case distribution of $h$

Compared with the results of the original models, the solution of the distributional robust model is more conservative. However, we can still expect considerable returns at the worst case, which shows the effectiveness of the robust distributional method.

5. **Conclusion.** We proposed a subgradient-type of computational scheme for solving stochastic optimization problems with various convex risk measures. We designed a discretization approach, which converts the computation of a subgradient of the objective function into a finite-dimensional optimization problem. We in particular studied three types of risk measures, which cover many commonly used models. It is shown that the algorithm can be effectively implemented for these measures. Finally, we presented numerical results for solving a two-stage stochastic linear program with CVaR being the risk measure.

**REFERENCES**

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

[2] J. Ang, F. Meng and J. Sun, Two-stage stochastic linear programs with incomplete information on uncertainty, *European J. Oper. Res.*, 233 (2014), 16–22.

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

[4] D. P. Bertsekas, *Convex optimization algorithms*, Athena Scientific, Belmont, MA, 2015.

[5] D. P. Bertsekas, A. Nedi and A. E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Belmont, MA, 2003.

[6] D. Bertsimas, X. V. Doan and K. Natarajan, Models for minimax stochastic linear optimization problems with risk aversion, *Math. Oper. Res.*, 35 (2010), 580–602.

[7] D. Bertsimas and R. Freund, *Data, Models, and Decisions: The Fundamentals of Management Science*, South-Western College Publishing, Cincinnati, OH, 2000.

[8] D. Bertsimas, M. Sim and M. Zhang, Adaptive distributionally robust optimization, *Manag. Sci.*, (2018).

[9] G. C. Calalore, Ambiguous risk measures and optimal robust portfolios, *SIAM J. Optim.*, 18 (2007), 853–877.
[10] E. Delage and Y. Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Oper. Res.*, **58** (2010), 595–612.

[11] H. Föllmer and A. Schied, *Stochastic finance*, Walter de Gruyter & Co., Berlin, 2002.

[12] S. Gao, L. Kong and J. Sun, Robust two-stage stochastic linear programs with moment constraints, *Optimization*, **63** (2014), 829–837.

[13] M. Grötschel, L. Lovász, and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica*, **1** (1981), 169–197.

[14] Z. Hu and J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, Available from: [http://www.optimization-online.org/DB_HTML/2012/11/3677.html](http://www.optimization-online.org/DB_HTML/2012/11/3677.html).

[15] D. Klabjan, D. Simchi-Levi and M. Song, Robust stochastic lot-sizing by means of histograms, *Prod. Oper. Manag.*, **22** (2013), 691–710.

[16] B. Li, X. Qian, J. Sun, K. L. Teo and C. Yu, A model of distributionally robust two-stage stochastic convex programming with linear recourse, *Appl. Math. Model.*, **58** (2018), 86–97.

[17] B. Li, Y. Rong, J. Sun and K. L. Teo, A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems, *IEEE Trans. Signal Process.*, **16** (2017), 464–474.

[18] B. Li, Y. Rong, J. Sun and K. L. Teo, A distributionally robust minimum variance beamformer design, *IEEE Signal Process. Lett.*, **25** (2018), 105–109.

[19] B. Li, J. Sun, H. L. Xu and M. Zhang, A class of two-stage distributionally robust games, *J. Ind. Manag. Optim.*, **15** (2019), 387–400.

[20] A. Ling, J. Sun and X. Yang, Robust tracking error portfolio selection with worst-case downside risk measures, *J. Econom. Dynam. Control*, **39** (2014), 178–207.

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

[22] H-J. Lüthi and J. Doege, Convex risk measures for portfolio optimization and concepts of flexibility, *Math. Program.*, **104** (2005), 541–559.

[23] S. Mehrotra and H. Zhang, Models and algorithms for distributionally robust least squares problems, *Math. Program.*, **146** (2014), 123–141.

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

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

[26] R. T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, *J. Banking & Finance*, **26** (2002), 1443–1471.

[27] W. W. Rogosinski, Moments of non-negative mass, *Proc. Roy. Soc. London Ser. A*, **245** (1958), 1–27.

[28] A. Shapiro and S. Ahmed, On a class of minimax stochastic programs, *SIAM J. Optim.*, **14** (2004), 1237–1249.

[29] A. Shapiro, D. Dentcheva and A. Ruszczyski, *Lectures on Stochastic Programming: Modeling and Theory*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, PA, 2009.

[30] M. Sion, On general minimax theorems, *Pacific J. Math.*, **8** (1958), 171–176.

[31] J. Sun, L. Liao and B. Rodrigues, Quadratic two-stage stochastic optimization with coherent measures of risk, *Math. Program.*, **168** (2018), 599–613.

[32] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.*, **62** (2014), 1358–1376.

Received March 2018; revised July 2018.

E-mail address: nianchuiqiao@msn.com
E-mail address: jie.sun@curtin.edu.au