DISTRIBUTIONALLY ROBUST CHANCE CONSTRAINED PROBLEMS UNDER GENERAL MOMENTS INFORMATION

KE-WEI DING
School of Computer Science and Technology, Southwest Minzu University
Chengdu, Sichuan 610041, China
and
School of Mathematical Sciences
University of Electronic Science and Technology of China
Chengdu, Sichuan 611731, China

NAN-JING HUANG*
Department of Mathematics, Sichuan University
Chengdu, Sichuan 610064, China

YI-BIN XIAO
School of Mathematical Sciences
University of Electronic Science and Technology of China
Chengdu, Sichuan 611731, China

(Communicated by Bin Li)

Abstract. In this paper, we focus on distributionally robust chance constrained problems (DRCCPs) under general moments information sets. By convex analysis, we obtain an equivalent convex programming form for DRCCP under assumptions that the first and second order moments belong to corresponding convex and compact sets respectively. We give some examples of support functions about matrix sets to show the tractability of the equivalent convex programming and obtain the closed form solution for the worst case VaR optimization problem. Then, we present an equivalent convex programming form for DRCCP under assumptions that the first order moment set and the support subsets are convex and compact. We also give an equivalent form for distributionally robust nonlinear chance constrained problem under assumptions that the first order moment set and the support set are convex and compact. Moreover, we provide illustrative examples to show our results.

1. Introduction. Optimization theory and its related theory on variational inequalities, fixed point problems and equilibrium problems etc., which have been studied extensively by many scholars in recent years (see, for example, [9, 15, 16, 18, 20, 22, 23, 24, 25, 29, 32, 33]), are important research tools for many practical problems in engineering, mechanics, economics and other disciplines. Among these problems, distributionally robust chance constrained problem (DRCCP), which can
be viewed as a combination of chance constraint and distributionally robust optimization, has become a significant and effective approach to address practical optimization problems involving uncertainty and acquired a great development both in theory and applications [5, 13, 14, 27, 31, 35, 36, 37]. In this paradigm, the distribution of uncertain parameter is not precisely known and belongs to a given information set, and the optimal solution is required to be feasible for all the realizations of the uncertain distribution. DRCCP provides a computationally viable methodology for immunizing chance constrained programming against the distribution uncertainty.

Most distributionally robust chance constrained methods are developed with the purpose of achieving a computationally tractable model for corresponding DRCCPs, in which how to describe the distribution set is an important issue. Recently, Calafiore and El Ghaoui [5] presented that a distributionally robust individual chance constrained problem could be rewritten as a second-order cone problem under the exact second order moment information. Zymler et al. [36] obtained a semidefinite programming for a distributionally robust joint chance constraint under the exact second order moment information. Yang and Xu [34] proved that DRCCP for nonlinear uncertainties under the exact second order moment information was equivalent to a robust optimization problem. Yet, most previous results on the tractability of DRCCPs are restricted to the case that the distribution information is described by the exact second order moment information. As pointed out by Delage and Ye [6], it is not reliable when empirical moments have been treated as exact moments. In order to strengthen safety, they replaced the exact moments by the ambiguity moments information and discussed the distributionally robust optimization problem (DROP) under ellipsoidal moments information. Moreover, Natarajan et al. [17] considered a distributionally robust expected utility problem involving an interval moments uncertainty set. Wiesemann et al. [30] discussed DROPs by using a family of support subsets described by conic representable confidence sets and first order moment residing on an affine manifold to describe the distribution information set.

On the other hand, Ding et al. [7] discussed DRCCP under interval moments uncertainty set and obtained its equivalent convex programming. Zhang et al. [35] considered a distributionally robust chance constrained appointment scheduling problem under ellipsoidal moments uncertainty set and derived an approximate semidefinite programming model. Wang et al. [28] showed a distributionally robust chance constrained support vector machines problem could be transformed into a second order cone problem under the exact second order moment information, and discussed the problem under interval first order moment uncertainty, ellipsoid first order moment uncertainty and second order moment uncertainty controlled by $F$-norm, respectively.

We note that, in the research works mentioned above, DROPs and DRCCPs have been discussed under various uncertainty moments sets such as ellipsoidal moments sets, interval moments sets and so on. Clearly, the ellipsoidal moments sets and the interval moments sets are all convex and compact. Therefore, one natural question is: can we get the equivalent form for DRCCPs involving some convex and compact information sets?

The main purpose of this paper is to make a new attempt to study DRCCPs. The rest of this paper is organized as follows. In Section 2, we obtain the convex programming for DRCCP under assumptions that the first and second order
moments belong to the corresponding convex and compact sets, respectively. We also present some computationally tractable examples for the support functions of the matrix sets and the equivalent convex programming. Moreover, we apply one result to the worst case VaR optimization problem and obtain its closed form solution. To avoid the computation of support function about the matrix sets, we show the convex programming for DRCCP under assumptions that the information set of first order moment and the support subsets are convex and compact in Section 3. In Section 4, for distributionally robust nonlinear chance constrained problem under the set of first order moment and the support set are convex and compact, we obtain the equivalent form and show an example in which the equivalent form is convex. Finally, we summarize the results in Section 5.

2. DRCCP under the general first and second order moment information.

In this section, we consider the following DRCCP:

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad \inf_{P \in \mathcal{P}} \mathbb{P}(w_0(x) + w(x)^T \xi \leq 0) \geq 1 - \varepsilon, \quad x \in \chi,
\end{align*}
\]

where \( x \) is decision vector, \( \chi \subset \mathbb{R}^n \) is a convex and compact set, \( c \in \mathbb{R}^n \) is a deterministic cost vector, \( \xi \in \mathbb{R}^k \) is a random vector, and \( w_0(x), w(x) \) are two affine functions. Let \( \mathcal{P} \) denote the distribution information set of the involved probability distributions and include the true distribution \( \mathbb{P} \). By employing the model (1), Ghaoui et al. [8] discussed the worst case VaR problem and presented several convex reformulations, Zymler et al. [37] considered two types of the worst case VaR of nonlinear portfolio and developed tractable representations, respectively, and Wang et al. [28] showed a distributionally robust chance constrained support vector machines problem could be transformed into a second order cone problem.

Based on limited history data, it is hard to obtain the exact distribution information. In this section, we suppose the ambiguity information set has the following form

\[
\mathcal{P}_1 = \{ \mathbb{P} \in \mathcal{P} : \mu \in \mathcal{U}_\mu, \Sigma \in \mathcal{U}_\Sigma \},
\]

where \( \mathcal{P} \) denotes the set of all probability measures on the measurable space \((\mathbb{R}^k, \mathcal{B})\) with \( \mathcal{B} \) being the Borel \( \sigma \)-algebra on \( \mathbb{R}^k \). Denote by \( \mathcal{S}^k \) the space of symmetric matrices of dimension \( k \). Generally speaking, the exact first (respectively, second) order moment \( \mu \) (respectively, \( \Sigma \)) is not far away from the empirical first (respectively, second) order moment \( \mu_0 \) (respectively, \( \Sigma_0 \)). Thus, we can assume that

\[
\begin{align*}
\mathcal{U}_\mu = \{ \mu = \mu_0 + A\zeta \mid \zeta \in \mathcal{U}_1 \}, \quad \mathcal{U}_\Sigma = \{ \Sigma = \Sigma_0 + \Xi \mid \Xi \in \mathcal{U}_2 \},
\end{align*}
\]

where \( A, \Xi \in \mathcal{S}^k, \zeta \in \mathbb{R}^k \), and for \( i = 1, 2, \mathcal{U}_i \) is a given nonempty, convex and compact set with \( 0 \in \mathcal{U}_i \). The confidence of decision makers in \( \mu_0 \) and \( \Sigma_0 \) has been demonstrated by the magnitude of sets \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \), respectively.

Some examples of \( \mathcal{U}_i \) are as follows:

(i) Let \( \mathcal{U}_1 = \{ \zeta \mid \| \zeta \|_2 \leq \sqrt{\lambda_1} \} \) with \( A = \Sigma_0^{\frac{1}{2}} \) and \( \mathcal{U}_2 = \{ \Xi \mid \Xi \preceq (\lambda_2 - 1)\Sigma_0 \} \). They are exactly the ellipsoidal set

\[
(\mu - \mu_0)^T \Sigma_0^{-1}(\mu - \mu_0) \leq \lambda_1
\]

and LMI \( \Sigma \preceq \lambda_2 \Sigma_0 \) which have been mentioned in [6], respectively;
Then, the following two statements are equivalent:

\[ P \text{ distribution information set} \]

we are devoted to obtaining the tractable reformulation of DRCCP \((1)\) under the

strained problem discussed in [7] is a special case of DRCCP in this paper. Next,

Similarly, denote the support function of the set

\[ S \]

Theorem 2.2. DRCCP \((1)\) under the information set \(P_1\) is the following convex

\begin{equation}
\begin{aligned}
\min_{x,v,s} & c^T x \\
\text{s.t.} & w^0(x) + w^T(x)\mu_0 + \delta^* (Aw(x)|U_1) + \frac{1 - \varepsilon}{\varepsilon} s + v \leq 0, \\
& 4sv \geq w^T(x)\Sigma_0w(x) + \delta^* (w(x)w^T(x)|U_2), s \geq 0, x \in \chi.
\end{aligned}
\end{equation}

Proof. Firstly, we rewrite problem \((1)\) as the following equivalent form

\begin{equation}
\begin{aligned}
\min & c^T x \\
\text{s.t.} & \sup_{P \in P_1} \mathbb{P} \left( w^0(x) + w(x)^T \xi > 0 \right) \leq \varepsilon, \ x \in \chi.
\end{aligned}
\end{equation}

By constructing the information sets \(U = \{ (\mu, \Sigma) : \mu \in U_{\mu}, \Sigma \in U_\Sigma \} \) and

\[ P_{11} = \{ P \in \mathbb{P} : \mathbb{P} (\xi \in R^k) = 1, E[\xi] = \mu, E[(\xi - E[\xi])(\xi - E[\xi])^T] = \Sigma \}, \]

the chance constraint in \((4)\) can be reformulated as follows:

\begin{equation}
\begin{aligned}
\sup_{(\mu, \Sigma) \in U} \sup_{P \in P_{11}} \mathbb{P} \left( w^0(x) + w(x)^T \xi > 0 \right) \leq \varepsilon.
\end{aligned}
\end{equation}

Let \(1_S\) denote the indicator function of the set

\[ S = \{ \xi \in R^k : w_0(x) + w(x)^T \xi > 0 \}. \]

Then \(\sup_{P \in P_1} \mathbb{P} \left( w_0(x) + w(x)^T \xi > 0 \right)\) becomes \(\sup_{(\mu, \Sigma) \in U} \sup_{P \in P_{11}} \mathbb{E}_{P} [1_S] \).
Letting \( y = w(x)^T \xi \), similar to the discussion in the proof of Theorem 2.2 in [7], by employing the strong duality theorem of Iihii [10] and Lemma 2.1, we can reformulate the problem
\[
\sup_{(\mu, \Sigma) \in \mathcal{U}} \inf_{\lambda_0, \lambda_1, \lambda_2} \frac{1}{2} \lambda_0 + \lambda_1 (w(x)^T \mu) + \lambda_2 (w(x)^T \Sigma w(x) + (w(x)^T \mu)^2) \\
\text{s.t.} \quad \lambda_0 + \lambda_1 y + \lambda_2 y^2 \geq 0, \quad \forall y \in \mathbb{R}, \\
\lambda_0 + \lambda_1 y + \lambda_2 y^2 + \eta (-w_0(x) - y) \geq 1, \quad \eta \geq 0, \quad \forall y \in \mathbb{R},
\]
where \( \lambda_0, \lambda_1, \lambda_2 \in \mathbb{R} \) are the dual variables for the constraints. When \( \lambda_2 \leq 0 \), the feasible region is empty. Thus we get \( \lambda_2 > 0 \). Similar to the proof of Theorem 2.1 in [17], by taking \( y^* = \frac{-\lambda_1}{2\lambda_2} \) in the first constraint and \( y^* = \frac{-\lambda_1}{2\lambda_2} \) in the second constraint and using the following change of variables
\[
\lambda_0 = \eta (p + \frac{1}{4z} (t - w(x)^T \mu)^2), \quad \lambda_1 = \frac{\eta}{2z} (t - w(x)^T \mu), \quad \lambda_2 = \frac{\eta}{4z},
\]
the inner problem \( \sup_{p, t, z, \eta} \mathbb{E}_\mathcal{P} [1_s] \) can be reformulated as follows:
\[
\inf_{p, t, z, \eta} \eta p + \eta s \\
\text{s.t.} \quad s \geq \frac{t^2 + w(x)^T \Sigma w(x)}{4z}, \\
\eta p - \eta w_0(x) - 1 + \eta t - \eta w(x)^T \mu - \eta z \geq 0, \\
p \geq 0, \quad z \geq 0, \quad s \geq 0, \quad \eta \geq 0.
\]
Then, problem (6) can be rewritten as follows:
\[
\sup_{(\mu, \Sigma) \in \mathcal{U}} \inf_{p, t, z, \eta} \eta p + \eta s \\
\text{s.t.} \quad s \geq \frac{t^2 + w(x)^T \Sigma w(x)}{4z}, \\
\eta p - \eta w_0(x) - 1 + \eta t - \eta w(x)^T \mu - \eta z \geq 0, \\
p \geq 0, \quad z \geq 0, \quad s \geq 0, \quad \eta \geq 0.
\]
In what follows, we prove that problem (8) is equivalent to the following one:
\[
\inf_{p, t, z, \eta} \eta (p + s) \\
\text{s.t.} \quad s \geq \frac{t^2 + \max_{\Sigma \in \Delta_2} w(x)^T \Sigma w(x)}{4z}, \\
\eta p - \eta w_0(x) - 1 + \eta t - \eta \max_{\mu \in \mathcal{U}_\mu} w(x)^T \mu - \eta z \geq 0, \\
p \geq 0, \quad z \geq 0, \quad s \geq 0, \quad \eta \geq 0.
\]
For fixed \( x \), suppose that \((\mu^*, \Sigma^*, p^*, t^*, z^*, s^*, \eta^*)\) is the optimal solution of (8). We also assume that, for some \((\mu^1, \Sigma^1) \in \mathcal{U}, \) at least one of the constraints is violated for \((p^*, t^*, z^*, s^*, \eta^*)\), which means one of the following inequalities holds
\[
s^* < \frac{(t^*)^2 + w(x)^T \Sigma^1 w(x)}{4z^*},
\]
\[
\eta^* p^* < \eta^* w_0(x) + 1 - \eta^* t^* + \eta^* w(x)^T \mu^1 + \eta^* z^*.
\]
Suppose the first inequality holds. Then we can obtain
\[
 s = \frac{(t^*)^2 + w(x)^T \Sigma^1 w(x)}{4z^*}
\]
such that \((\mu^*, \Sigma^1, p^*, t^*, z^*, s, \eta^*)\) is feasible for (8). It follows that
\[
\eta^*(s + p^*) = \eta^*\left(\frac{(t^*)^2 + w(x)^T \Sigma^1 w(x)}{4z^*} + p^*\right)
\]
\[
> \eta^*(s^* + p^*)
\]
\[
\geq \eta^*\left(\frac{(t^*)^2 + w(x)^T \Sigma^* w(x)}{4z^*} + p^*\right),
\]
which is a contradiction with the fact that \((\mu^*, \Sigma^*, p^*, t^*, z^*, s^*, h^*)\) is the optimal solution of (8). Thus, we have
\[
s^* \geq \frac{(t^*)^2 + w(x)^T \Sigma w(x)}{4z^*}, \quad \forall \Sigma \in \mathcal{U}_\Sigma.
\]
Similarly, we can prove
\[
\eta^* p^* - \eta^* w_0(x) - 1 + \eta t^* - \eta w(x)^T \mu - \eta z^* \geq 0, \quad \forall \mu \in \mathcal{U}_\mu.
\]
In conclusion, we obtain that problem (8) is equivalent to problem (9). Then, problem (4) can be reformulated as
\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad \inf_{p, t, z, s, \eta} \eta(p + s) \leq \varepsilon, \\
& \quad s \geq \frac{t^2 + \max_{\Sigma \in \mathcal{U}_\Sigma} w(x)^T \Sigma w(x)}{4z}, \\
& \quad \eta p - \eta w_0(x) - 1 + \eta t - \eta \max_{\mu \in \mathcal{U}_\mu} w(x)^T \mu - \eta z \geq 0, \\
& \quad p \geq 0, \quad z \geq 0, \quad s \geq 0, \quad \eta \geq 0.
\end{align*}
\]
By the definition of support function, the above problem can be reformulated as
\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad \eta(p + s) \leq \varepsilon, \\
& \quad s \geq \frac{t^2 + \delta^* (w(x)w^T(x)|\mathcal{U}_\Sigma)}{4z}, \\
& \quad \eta p - \eta w_0(x) - 1 + \eta t - \eta \delta^* (w(x)|\mathcal{U}_\mu) - \eta z \geq 0, \\
& \quad p \geq 0, \quad z \geq 0, \quad s \geq 0, \quad \eta \geq 0.
\end{align*}
\]
Clearly, the third constraint of the above system does not hold when \(\eta = 0\), which implies that \(\eta > 0\). From the first and second inequalities of the above problem, we have
\[
\frac{1}{\eta} \geq \frac{p + s}{\varepsilon}, \quad z \geq \frac{t^2 + \delta^* (w(x)w^T(x)|\mathcal{U}_\Sigma)}{4s}.
\]
Then, (5) can be simplified as

\[-w^0(x) - \delta^*(w(x)|\mathcal{U}_4) \geq \frac{1}{\eta} - p - t + z\]

\[\geq \frac{1}{\varepsilon}(p + s) - p - t + \frac{t^2}{4s} + \frac{\delta^*(w(x)w^T(x)|\mathcal{U}_2)}{4s}\]

\[\geq \frac{1 - \varepsilon}{4s}s + \frac{\delta^*(w(x)w^T(x)|\mathcal{U}_2)}{4s}. \quad (10)\]

For investigating the convexity of problem (4), we introduce a new variable \(v\) and so (10) can be simplified as follows:

\[
\begin{align*}
&\begin{cases}
w^0(x) + w^T(x)\mu_0 + \delta^*(Aw(x)|\mathcal{U}_1) + \frac{1 - \varepsilon}{4s}s + v \leq 0, \\
4v \geq w^T(x)\Sigma_0w(x) + \delta^*(w(x)w^T(x)|\mathcal{U}_2), s \geq 0.
\end{cases}
\end{align*}
\]

It is well known that \(x^T\Sigma x\) is convex in \(x\) when \(\Sigma > 0\). This fact shows that, for any \(\theta \in [0, 1]\) and \(x_1, x_2 \in \mathbb{R}^n\), we have

\[
\delta^*(((\theta x_1 + (1 - \theta)x_2)(\theta x_1 + (1 - \theta)x_2)^T|\mathcal{U}_2) = \max_{\Sigma \in \mathcal{U}_2} \{\theta x_1^T\Sigma x_1 + (1 - \theta)x_2^T\Sigma x_2\}
\]

\[\leq \theta \max_{\Sigma \in \mathcal{U}_2} \{x_1^T\Sigma x_1\} + (1 - \theta)\max_{\Sigma \in \mathcal{U}_2} \{x_2^T\Sigma x_2\}
\]

and so \(\delta^*(xx^T|\mathcal{U}_2)\) is convex in \(x\). Since \(w(x)\) is affine with respect to \(x\), we know that \(\delta^*(w(x)w^T(x)|\mathcal{U}_2)\) is convex in \(x\). It is easy to see that

\[
\delta^*(w(x)w^T(x)|\mathcal{U}_2) = w^T(x)\Sigma_0w(x) + \delta^*(w(x)w^T(x)|\mathcal{U}_2)
\]

and

\[
\delta^*(w(x)|\mathcal{U}_4) = w^T(x)\mu_0 + \delta^*(Aw(x)|\mathcal{U}_1).
\]

From the above discussion, we get the conclusion. \(\square\)

**Remark 2.1.** We note that, if \(\mathcal{U}_1 = \{0\}\) and \(\mathcal{U}_2 = \{0\}\), i.e., \(\mu_0\) and \(\Sigma_0\) are the exact first and second order moments, then \(\delta^*(Aw(x)|\mathcal{U}_1) = 0\) and \(\delta^*(w(x)w^T(x)|\mathcal{U}_2) = 0\). In this case, it follows from (10) that the constraints in problem (3) reduces to the following form

\[
w^0(x) + w^T(x)\mu_0 + \sqrt{\frac{1 - \varepsilon}{\varepsilon}}\sqrt{w^T(x)\Sigma_0w(x)} \leq 0,
\]

which is the same as (27) in Theorem 2 of [28] by setting \(w^0(x) = 1 - \xi_i - y_ib\) and \(w(x) = -y_iw\).

**2.1. Examples.** Theorem 2.2 shows that DRCCP (1) under the information set \(\mathcal{P}_1\) can be reformulated as a convex programming. It follows from (3) that the key issue for the computationally tractability of the convex programming is the computation of support functions about the vector set and matrix set. Concerning with the computationally tractable support functions about the vector sets, we refer the reader to [3]. For matrix \(\Xi\), if we define \(\mathcal{U}_2 \triangleq \mathcal{U}(vec(\Xi))\), then we can deal with the support function of the matrix set by employing the method of [3]. Nevertheless,
it is necessary to consider the support function about the matrix set directly, since the structure of the matrix is destroyed if we transform the matrix into the vector form. Therefore, we next show some examples for the support functions about the matrix sets, in which the term \( \delta^*(w(x)w^T(x)\|\mathcal{U}_2) \) can be reformulated as some computationally tractable forms.

**Example 2.1.** Let \( \mathcal{U}_2 = \{ \Xi \in S^k | \Theta_j \preceq \Xi \preceq \Xi_j, j = 1, 2, \ldots, J \} \), where \( \Theta_j \succeq 0 \) and \( \Xi_j > 0 \) are given matrices. By Schur complement, the support function of \( \mathcal{U}_2 \) is given as follows:

\[
\delta^*(w(x)w^T(x)\|\mathcal{U}_2) = \min_{Y_j, Z_j} \sum_{j=1}^J (Z_j - Y_j) \cdot \Theta_j
\]

s.t. \( \begin{pmatrix} \sum_{j=1}^J (Z_j - Y_j) & w(x) \\ w(x)^T & 1 \end{pmatrix} \succeq 0, Y_j \succeq 0, Z_j \succeq 0. \)

**Example 2.2.** Let \( \mathcal{U}_2 = \{ \Xi \in S^k | tr(\Xi D\Xi) \leq \tau \} \), where \( \tau > 0 \) and \( D \succ 0 \). By using the inequality

\[
tr(AB) \leq \sqrt{tr(A^TDA)tr(B^TD^{-1}B)},
\]

the support function of \( \mathcal{U}_2 \) can be represented as follows:

\[
\delta^*(w(x)w^T(x)\|\mathcal{U}_2) = \sqrt{\tau} \sqrt{tr(w(x)D^{-1}w(x)w^T(x))}.
\]

If \( D \) is the unit matrix, then \( \mathcal{U}_2 = \{ \Xi \in S^k | \|\Xi\|_F \leq \sqrt{\tau} \} \) and the corresponding support function reduces to \( \sqrt{\tau} \|w(x)\|_2^2 \).

**Example 2.3.** Let \( \mathcal{U}_2 = \{ \Xi \in S^k | \|\Xi\|_{\sigma_p} \leq \tau \} \) with \( \tau > 0 \), where the Schatten norm \( \|\cdot\|_{\sigma_p} \) is defined in [4] by

\[
\|A\|_{\sigma_p} \triangleq \begin{cases} \left( \sum_{i=1}^k \sigma_i^p (A) \right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \sigma_{\max} (A), & p = \infty. \end{cases}
\]

Here \( A \in S^k \) and \( \sigma_i (A) \) is the absolute value of the \( i \)-th largest eigenvalue of \( A \) which has real eigenvalues. Let \( A, B \in S^k \) and \( p, q \in [1, +\infty] \) satisfy \( \frac{1}{p} + \frac{1}{q} = 1 \). Then it is easy to see that \( tr(A^T B) \leq \|A\|_{\sigma_p} \|B\|_{\sigma_q} \). In particular,

\[
tr(A^T B) \leq \sigma_{\max} (A) \|B\|_{\sigma_1}.
\]

Thus, the support function can be given as follows:

\[
\delta^*(w(x)w^T(x)\|\mathcal{U}_2) = \tau \|w(x)w^T(x)\|_{\sigma_q} = \tau (\|w^T(x)w(x)\|^q + 0^q + \cdots + 0^q)^{\frac{1}{q}} = \tau w^T(x)w(x).
\]

**Example 2.4.** Let \( \mathcal{U}_2 = \{ \Xi \in S^k | tr(C_j\Xi) \leq b_j, \Xi \succeq 0, j = 1, 2, \ldots, J \} \), where \( C_j \in S^k \) and \( b_j \in \mathbb{R} \) are given for \( j = 1, 2, \ldots, J \). For the support function of \( \mathcal{U}_2 \), we have

\[
\delta^*(w(x)w^T(x)\|\mathcal{U}_2) = \min_{\lambda_j} \sum_{j=1}^J (\lambda_j b_j)
\]

s.t. \( \begin{pmatrix} \sum_{j=1}^J (\lambda_j C_j) & w(x) \\ w(x)^T & 1 \end{pmatrix} \succeq 0, \lambda_j \succeq 0, j = 1, 2, \ldots, J. \)
Example 2.5. Let \( U_2 = \{ \Sigma \in \mathbb{S}^k | \varphi(\Sigma) \leq \tau \} \) with
\[
\varphi(\Sigma) = \begin{cases} 
\sum_{j=1}^k \beta_j \text{tr}(D_j(\Sigma - \Xi_0)), & \text{if } \Sigma \succeq \Xi_0, \\
0, & \text{if else},
\end{cases}
\]
where \( \beta_j \in \mathbb{R}, D_j \in \mathbb{S}^k \), for \( j = 1, 2, \cdots, J \), and \( \Xi_0 \succeq 0 \) are given. Then the support function of \( U_2 \) can be given as follows:
\[
\delta^* (w(x)w^T(x)|U_2) = \min_{\lambda} \lambda \tau + \lambda \sum_{j=1}^J \beta_j \text{tr}(D_j \Xi_0)
\]
\[
\text{s.t. } \left( \sum_{j=1}^J \frac{\beta_j D_j}{\lambda} w(x) \right) \geq 0, \lambda \geq 0.
\]

In the above examples, we drive some computationally tractable support functions about the matrix sets. Next, we show two computationally tractable examples for problem (3).

Example 2.6. Let \( U_1 = \{ \zeta \in \mathbb{R}^k | \|\zeta\|_{p_1} \leq \rho_1 \} \) and \( U_2 = \{ \Xi \in \mathbb{S}^k | vec(\Xi)\|_{p_2} \leq \rho_2 \} \) with constants \( \rho_1, \rho_2 > 0 \) and \( p_1, p_2 \geq 1 \). Then, the support functions of sets \( U_1 \) and \( U_2 \) can be given as follows:
\[
\delta^* (Aw(x)|U_1) = \rho_1 \|Aw(x)\|_{q_1}, \delta^* (w(x)w^T(x)|U_2) = \rho_2 \|w(x)\|_{q_2}^2
\]
with \( \frac{1}{p_1} + \frac{1}{q_1} = 1 \) and \( \frac{1}{p_2} + \frac{1}{q_2} = 1 \). In this case, the problem (3) becomes
\[
\min_{x,v,s} c^T x
\]
\[
\text{s.t. } w^0(x) + w^T(x)\mu_0 + \rho_1 \|Aw(x)\|_{q_1} + \frac{1}{\varepsilon} s + v \leq 0,
\]
\[
4sv \geq w^T(x)\Sigma_0 w(x) + u, \sqrt{u} \geq \sqrt{p_2} \|w(x)\|_{q_2}, s > 0, x \in \chi.
\]

Example 2.7. Let \( U_1 = \{ \zeta \in \mathbb{R}^k | \|\zeta\|_2 \leq \rho_1, \|\zeta\|_1 \leq \rho_2 \} \) and
\[
U_2 = \{ \Xi \in \mathbb{S}^k | \|\Xi\|_{\nu} \leq \theta, tr(C\Xi) \leq \tau \}
\]
with \( \rho_1, \rho_2, \theta > 0 \) and \( C \in \mathbb{S}^k \). Then, the support functions of sets \( U_1 \) and \( U_2 \) can be obtained as follows:
\[
\delta^* (Aw(x)|U_1) = \min_{u_1, u_2} \left\{ \rho_1 \|u_1\|_2 + \rho_2 \|u_2\|_{\infty} | u_1 + u_2 = Aw(x) \right\},
\]
\[
\delta^* (w(x)w^T(x)|U_2) = \min_{\lambda \geq 0} \left\{ \tau \lambda + \theta \sqrt{\text{tr}((w(x)w^T(x)-\lambda C)(w(x)w^T(x)-\lambda C))} \right\}.
\]

In this case, the problem (3) becomes
\[
\min_{x,v,s} c^T x
\]
\[
\text{s.t. } 4sv \geq w^T(x)\Sigma_0 w(x) + \tau \lambda + \theta \sqrt{\text{tr}((w(x)w^T(x)-\lambda C)(w(x)w^T(x)-\lambda C))},
\]
\[
w^0(x) + w^T(x)\mu_0 + \rho_1 \|u_1\|_2 + \rho_2 \|u_2\|_{\infty} + \frac{1}{\varepsilon} s + v \leq 0,
\]
\[
u_1 + u_2 = Aw(x), \lambda \geq 0, s > 0, x \in \chi.
2.2. An application to the risk management. Recently, the Value-at-risk (VaR) has become one of the most popular risk measures in risk management. Usually, the VaR can be defined as follows:

\[ V(x) = \min \{ \gamma \mid P(-x^T \xi \leq \gamma) \geq 1 - \varepsilon \}, \quad (11) \]

where \( x \) denotes the vector of asset weights, and \( \xi \) denotes the random vector of relative asset returns. When the distribution of returns is normal distribution \( N(\mu, \Sigma) \), the VaR can be expressed as

\[ V(x) = \kappa(\varepsilon) \sqrt{x^T \Sigma x - x^T \mu}, \quad (12) \]

where \( \kappa(\varepsilon) = -\Phi^{-1}(\varepsilon) \), and \( \Phi^{-1}(\cdot) \) is the inverse cumulative standard normal distribution. In practice, the exact distribution is not easy to acquire based on limited history data. In 2003, Ghaoui et al. [8] presented the exact formulation for (11) by setting \( \kappa(\varepsilon) = \sqrt{\varepsilon^{-1}(1 - \varepsilon)} \) under the first and second order moments information. But the optimal portfolio strategy may become infeasible when we treat the empirical moments as the exact moments. In this subsection, we investigate the VaR (11) under the information set \( \mathcal{P}_1 \). Let \( w_0(x) = -\gamma \) and \( w(x) = -x \). From Theorem 2.2, we have

\[ \gamma \geq (1 - \varepsilon)z + \frac{1}{4\varepsilon} \left( x^T \Sigma_0 x + \delta^* (xx^T | \mathcal{U}_2) \right) + \delta^* (-Ax|\mathcal{U}_1) - \mu_0^T x \]

\[ \geq \sqrt{\frac{1 - \varepsilon}{\varepsilon}} \sqrt{x^T \Sigma_0 x + \delta^* (xx^T | \mathcal{U}_2) + \delta^* (-Ax|\mathcal{U}_1) - \mu_0^T x}. \]

Thus, we obtain

\[ V(x) = \sqrt{\frac{1 - \varepsilon}{\varepsilon}} \sqrt{x^T \Sigma_0 x + \delta^* (xx^T | \mathcal{U}_2) + \delta^* (-Ax|\mathcal{U}_1) - \mu_0^T x}. \quad (13) \]

It is worth to mentioning that (13) will degenerate to (12) with \( \kappa(\varepsilon) = \sqrt{\varepsilon^{-1}(1 - \varepsilon)} \) when we set \( \mathcal{U}_1 = 0 \) and \( \mathcal{U}_2 = 0 \) (no uncertainty about \( \mu \) and \( \Sigma \)).

Next we give two examples to show how to use (13) to compute the VaR.

**Example 2.8.** Let \( \mathcal{U}_1 = \{ \zeta \in \mathbb{R}^k \mid ||\zeta||_{p_1} \leq \tau_1 \} \) and \( \mathcal{U}_2 = \{ \Xi \in \mathbb{S}^k \mid ||\Xi||_{\sigma_p_2} \leq \tau_2 \} \), where \( \tau_1, \tau_2 > 0 \) and \( p_1, p_2 \geq 1 \). By \( x^T y \leq ||x||_{p_1} ||y||_{q_1} \) and \( tr(X^T Y) \leq ||X||_{\sigma_p_2} ||Y||_{\sigma_q_2} \), one has

\[ \delta^*(-Ax|\mathcal{U}_1) = \tau_1 ||Ax||_{q_1}, \quad \delta^*(xx^T | \mathcal{U}_2) = \tau_2 x^T x, \]

where \( \frac{1}{p_1} + \frac{1}{q_1} = 1 \) and \( \frac{1}{p_2} + \frac{1}{q_2} = 1 \). Then, the VaR can be given as follows:

\[ V(x) = \sqrt{\frac{1 - \varepsilon}{\varepsilon}} \sqrt{x^T (\Sigma_0 + \tau_2 E)x + \tau_1 ||Ax||_{q_1} - \mu_0^T x}, \]

where \( E \) is the unit matrix.

**Example 2.9.** Let \( \mathcal{U}_1 = \{ \zeta \in \mathbb{R}^k \mid C\zeta \leq c \} \) and \( \mathcal{U}_2 = \{ \Xi \in \mathbb{S}^k \mid \Sigma_1 \preceq \Xi \preceq \Sigma_2 \} \), where \( \Sigma_2 \succeq 0 \) and \( C \in \mathbb{R}^{k \times k} \) with \( c \in \mathbb{R}^k \). Then, the support function can be rewritten as follows:

\[ \delta^*(-Ax|\mathcal{U}_1) = \min_{u \geq 0} \{ u^T C^T u + Ax = 0 \}, \]

\[ \delta^*(xx^T | \mathcal{U}_2) = x^T \Sigma_2 x. \]

Thus, we have

\[ V(x) = \min_{u \geq 0} \left\{ \sqrt{\frac{1 - \varepsilon}{\varepsilon}} \sqrt{x^T (\Sigma_0 + \Sigma_2) x + C^T u - \mu_0^T x} \mid C^T u + Ax = 0 \right\}. \]
3. DRCCP under the general first order moment and support subsets information. In section 2, we obtain the equivalent convex programming for DRCCP under the information set \(P_1\). We notice that the term \(\delta^*(w(x)^Tw(x)|\mathcal{U}_2)\) in the equivalent convex programming only has a few tractable forms which have limited the practical applications of this model. In 2014, Wiesemann et al. [30] discussed distributionally robust programming under the conic represented distribution sets with support and first order moment information. In the WKS-type of ambiguity set, the authors used the interval probability of support subsets instead of variance to describe the deviation. Many stochastic optimization problems have been studied under the WKS-type of ambiguity set [11, 12]. In this section, we consider the following information set

\[
P_2 = \left\{ \mathbb{P} \in \mathcal{P} : \mathbb{E}_\mathbb{P}[\xi] = \mu \in \mathcal{U}_\mu, \, \mathbb{P}(\xi \in \mathcal{X}_j) \in [p_j, \bar{p}_j], \, \forall \, j \in \mathcal{J} \right\},
\]

where \(\mathcal{X}_0, \mathcal{X}_j\) are convex and compact sets and \(\mathcal{U}_\mu\) is defined in (2), \(\mathcal{J} = \{1, 2, \ldots, J\}\). Suppose that \(p_j, \bar{p}_j \in [0, 1]\) and \(p_j \leq \bar{p}_j\), for all \(j \in \mathcal{J}\). We remark that the information set \(P_2\) where it is assumed that the set of first order moment and the support subsets are convex and compact is a generalization of the conic represented information set considered by Wiesemann et al. [30]. We also suppose that \(P_2\) satisfies the following conditions:

(C1) There exists a distribution \(\mathbb{P} \in P_2\) such that \(\mathbb{P}(\xi \in \mathcal{X}_j) \in (p_j, \bar{p}_j)\), whenever \(p_j < \bar{p}_j\) for \(j \in \mathcal{J}\).

(C2) \(\mathcal{X}_j\) is bounded and almost contains all realizations of random variable such that \(\mathbb{P}(\xi \in \mathcal{X}_j) = 1\).

(C3) For all \(j, j' \in \mathcal{J}\) and \(j \neq j'\), either \(\mathcal{X}_j \not\subset \mathcal{X}_{j'}\), \(\mathcal{X}_{j'} \not\subset \mathcal{X}_j\) or \(\mathcal{X}_j \cap \mathcal{X}_{j'} = \emptyset\).

Condition (C1) ensures that we can apply the strong dual theorem in [21] to reformulate DRCCP (1). Condition (C2) states that \(\mathcal{X}_j\) contains the support information of random variable \(\xi\). Condition (C3) implies that a strict partial order on the confidence sets \(\mathcal{X}_j\) regarding the \(\triangleleft\)-relation and requires that the incomparable sets must be disjoint. We denote

\[
\mathfrak{A}(j) = \{ j' \in \mathcal{J} : \mathcal{X}_j \not\subset \mathcal{X}_{j'} \}. \tag{14}
\]

Next, we devote to getting the equivalent form for DRCCP (1) under the information set \(P_2\) which satisfies conditions (C1)-(C3).

**Theorem 3.1.** DRCCP (1) under the information set \(P_2\) can be rewritten as the following convex programming

\[
\min \ c^T x
\]

s.t. \(\delta^*(A_\beta|\mathcal{U}_1) + \beta^T \mu_0 + \sum_{j \in \mathcal{J}} (\lambda_j \bar{p}_j - \gamma_j p_j) - \varepsilon t \leq 0\),

\[
\delta^*( - \beta|\mathcal{X}_j) - \sum_{j' \in \mathfrak{A}(j)} (\lambda_{j'} - \gamma_{j'}) \leq 0, \, j \in \mathcal{J},
\]

\[
t + w^0(x) - \sum_{j' \in \mathfrak{A}(j)} (\lambda_{j'} - \gamma_{j'}) + \delta^*(w(x) - \beta|\mathcal{X}_j) \leq 0, \, j \in \mathcal{J},
\]

\[
t > 0, \, \lambda_j \geq 0, \, \gamma_j \geq 0, \, \forall j \in \mathcal{J},
\]

where \(x \in \mathbb{R}^n, t \in \mathbb{R}\) and \(\beta, \lambda_j, \gamma_j \in \mathbb{R}^k\) are decision variables.
Proof. By constructing the information set 
\[ P_{21} = \{ P \in \mathcal{P} : \mathbb{E}_P[\xi] = \mu, P(\xi \in X_j) \in [\underline{p}_j, \bar{p}_j], \forall j \in J \}, \]
the chance constraint in (1) can be reformulated as follows
\[ \sup_{\mu \in \mathcal{U}} \sup_{P \in P_{21}} \mathbb{E}_P[1_S] \leq \varepsilon, \tag{16} \]
where \( 1_S \) represents the indicator function of the set
\[ S = \{ \xi \in \mathbb{R}^k : w_0(x) + w(x)^T \xi > 0 \}. \]
Then, the inner subordinate problem \( \sup_{P \in P_{21}} \mathbb{E}_P[1_S] \) can be unfolded as
\[ \inf_{\beta, \gamma} \beta^T \mu + \sum_{j \in J} (\lambda_j \bar{p}_j - \gamma_j p_j) \tag{17} \]
s.t. \( \beta^T \xi + \sum_{j \in J} (\lambda_j - \gamma_j) 1_{(\xi \in X_j)} \geq 1, \forall \xi \in X_j, \forall j \in J, \lambda_j \geq 0, \gamma_j \geq 0, \forall j \in J, \]
where \( \beta \in \mathbb{R}^k, \gamma, \lambda_j \in \mathbb{R} \) are the dual variables for the constraints. Obviously, the strong dual theorem holds by condition (C1). By Lemma 2.1, the problem (17) is equivalent to
\[ \inf_{\beta, \gamma, \lambda} \beta^T \mu + \sum_{j \in J} (\lambda_j \bar{p}_j - \gamma_j p_j) \tag{18} \]
s.t. \( \beta^T \xi + \sum_{j \in J} (\lambda_j - \gamma_j) 1_{(\xi \in X_j)} - \eta (w_0(x) + w(x)^T \xi) \geq 1, \forall \xi \in X_j, \forall j \in J, \lambda_j \geq 0, \gamma_j \geq 0, \eta \geq 0, \forall j \in J. \]
By condition (C3), we divide \( X_j \) into nonempty and disjoint sets
\[ X_j = X_j \setminus \bigcup_{j' \in \mathcal{D}(j)} X_{j'}, \quad j = 1, 2, \cdots, J, \]
where \( \mathcal{D}(j) \) denotes the index set of strict subsets of \( X_j \). Then, the left hand term of (16) can be rewritten as
\[ \sup_{\mu \in \mathcal{U}, \beta, \gamma, \lambda, \eta} \beta^T \mu + \sum_{j \in J} (\lambda_j \bar{p}_j - \gamma_j p_j) \]
s.t. \( \beta^T \xi + \sum_{j' \in \mathcal{S}(j)} (\lambda_{j'} - \gamma_{j'}) \geq 0, \forall \xi \in \bar{X}_j, j \in J, \lambda_j \geq 0, \gamma_j \geq 0, \eta \geq 0. \]
Similar to the proof of Theorem 2.1, the above problem can be reformulated as

\[
\inf_{\beta, \gamma_j, \lambda_j, \eta} \delta^*(\beta|\mathcal{U}_\mu) + \sum_{j \in \mathcal{J}} (\lambda_j \bar{p}_j - \gamma_j p_j) 
\]

subject to

\[
\beta^T \xi + \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) \geq 0, \quad \forall \xi \in \bar{X}_j, \ j \in \mathcal{J},
\]

\[
\beta^T \xi + \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) - \eta(w^0(x) + w(x)^T \xi) \geq 1, \quad \forall \xi \in \bar{X}_j, \ j \in \mathcal{J},
\]

\[
\lambda_j \geq 0, \ \gamma_j \geq 0, \ \forall j \in \mathcal{J}, \ \eta \geq 0.
\]

Next we claim that \( \eta > 0 \). Suppose \( \eta = 0 \), then the constraints which contain \( \eta \) become

\[
\beta^T \xi + \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) \geq 1, \quad \forall \xi \in \bar{X}_j, \ j \in \mathcal{J}.
\]

Taking integral operation on both sides of the above constraints, we have

\[
\beta^T \int_{\xi \in \bar{X}_j} \xi \zeta(d\xi) + \sum_{j \in \mathcal{J}} (\lambda_j - \gamma_j) \int_{\xi \in X_j} \zeta(d\xi) \geq \int_{\xi \in X_j} \zeta(d\xi), \ j \in \mathcal{J},
\]

where \( \zeta(\cdot) \) is the exact nonnegative distribution measure. Then, we have

\[
\beta^T \mu + \sum_{j \in \mathcal{J}} (\lambda_j - \gamma_j) p_j \geq 1.
\]

From \( p_j \leq \bar{p}_j \leq \bar{p}_j \), one has

\[
\delta^*(\beta|\mathcal{U}_\mu) + \sum_{j \in \mathcal{J}} (\lambda_j \bar{p}_j - \gamma_j p_j) \geq \beta^T \mu + \sum_{j \in \mathcal{J}} (\lambda_j - \gamma_j) p_j \geq 1,
\]

which is a contradiction with

\[
\delta^*(\beta|\mathcal{U}_\mu) + \sum_{j \in \mathcal{J}} (\lambda_j \bar{p}_j - \gamma_j p_j) \leq \varepsilon.
\]

Thus, we have \( \eta > 0 \). Divided by \( \eta \) in all constraints of problem (19) and substituted \( (\frac{\beta}{\eta}, \frac{\lambda_{j'}}{\eta}, \frac{\gamma_j}{\eta}, \frac{1}{\eta}) \) by \( (\beta, \lambda_{j'}, \gamma_{j'}, t) \), problem (19) becomes

\[
\inf_{t} \frac{1}{t} \left( \delta^*(\beta|\mathcal{U}_\mu) + \sum_{j \in \mathcal{J}} (\lambda_j \bar{p}_j - \gamma_j p_j) \right)
\]

subject to

\[
\beta^T \xi + \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) \geq 0, \quad \forall \xi \in \bar{X}_j, \ j \in \mathcal{J},
\]

\[
(\beta - w(x))^T \xi + \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) - w^0(x) \geq t, \quad \forall \xi \in \bar{X}_j, j \in \mathcal{J},
\]

\[
t > 0, \ \lambda_j \geq 0, \ \gamma_j \geq 0, \ \forall j \in \mathcal{J}.
\]

The robust counterparts of constraints (21) are

\[
\max_{\xi \in \bar{X}_j} \left\{ -\beta^T \xi - \sum_{j' \in \mathcal{A}(j)} (\lambda_{j'} - \gamma_{j'}) \right\} \leq 0, \ \forall j \in \mathcal{J}.
\]
Since the left side of the above inequality is linear in $\xi$, we obtain that the optimal solution must be taken on the boundary of $\tilde{X}_j$. According to the same boundary of $X_j$ and $\tilde{X}_j$, the above problem can be reformulated as

$$\max_{\xi \in X_j} \left\{-\beta^T \xi - \sum_{j' \in \Phi(j)} (\lambda_{j'} - \gamma_{j'}) \right\} \leq 0, \quad \forall j \in J.$$ 

By similar discussion for constraint (22) and the definition of support function, problem (20) can be reformulated as follows

$$\inf_{t} \frac{1}{2} \left( \delta^*(\beta|\mathcal{U}_0) + \sum_{j \in J} (\lambda_j p_j - \gamma_j p_j) \right)$$

s.t. $\delta^*(-\beta|X_j) - \sum_{j' \in \Phi(j)} (\lambda_{j'} - \gamma_{j'}) \leq 0, \quad j \in J$,

$$t + w^0(x) - \sum_{j' \in \Phi(j)} (\lambda_{j'} - \gamma_{j'}) + \delta^*(w(x) - \beta|X_j) \leq 0, \quad j \in J,$$

$$t > 0, \quad \lambda_j \geq 0, \quad \gamma_j \geq 0, \quad \forall j \in J.$$

Thus, constraint (16) can be rewritten as

$$\begin{align*}
\delta^*(A\beta|\mathcal{U}_0) + \beta^T \mu_0 + \sum_{j \in J} (\lambda_j p_j - \gamma_j p_j) - \varepsilon t & \leq 0, \\
\delta^*(-\beta|X_j) - \sum_{j' \in \Phi(j)} (\lambda_{j'} - \gamma_{j'}) & \leq 0, \quad j \in J, \\
t + w^0(x) - \sum_{j' \in \Phi(j)} (\lambda_{j'} - \gamma_{j'}) + \delta^*(w(x) - \beta|X_j) & \leq 0, \quad j \in J, \\
t & > 0, \quad \lambda_j \geq 0, \quad \gamma_j \geq 0, \quad \forall j \in J.
\end{align*}$$

This completes the proof.

Theorem 3.1 shows that DRCCP (1) under the information set $\mathcal{P}_2$ which satisfies conditions (C1)-(C3) can be rewritten as an equivalent convex programming. The equivalent tractable form of DRCCP (1) is difficult to acquire when $\mathcal{P}_2$ does not satisfy condition (C3). By choosing a partition for $X_j$ and each element of this partition is required to be satisfied condition (C3), instead we can obtain a convex approximation for DRCCP (1). Compared to the equivalent form (3) of DRCCP under the information set $\mathcal{P}_1$, we only need to compute the support function about vector sets in problem (15). We remark that the method for solving the distributionally robust programming under the conic represented distribution sets proposed in [30] is not suitable for solving problem (15).

Now we give the following examples to illustrate the efficiency of our results.

**Example 3.1.** Let $X_j = \{ \xi \in \mathbb{R}^k | \| \xi - \mu_0\|_2 \leq \tau_j \}$ and

$$\mathcal{U}_l = \{ \xi \in \mathbb{R}^k | h_l(\xi) \leq 0, l = 1, 2, \ldots, L \},$$

where $\tau_3 > \tau_2 > \tau_1 > 0$, $h_l(\xi)$ is convex and $\bigcap_{l=1}^L ri(dom h_l) \neq \emptyset$. It follows that

$$\delta^*(y|X_j) = y^T \mu_0 + \tau_j \| y \|_2 (j = 1, 2, 3),$$

$$\delta^*(A\beta|\mathcal{U}_l) = \min_{\delta \geq 0} \left\{ \sum_{l=1}^L \theta_l h_l^T \left( \frac{v_l}{\theta_l} \right) \left| \sum_{l=1}^L v_l = A\beta \right. \right\}.$$
Then, problem (15) becomes

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t. } & \quad \sum_{i=1}^{L} \theta_i h_i^T \left( \frac{v_i}{\theta_i} \right) + \beta^T \mu_0 + \sum_{j=1}^{3} (\lambda_j p_j - \gamma_j p_j) - \varepsilon t \leq 0,
\end{align*}
\]

\[
-\beta^T \mu_0 + \tau_j \| \beta \|_2 - \sum_{j'=j}^{3} (\lambda_{j'} - \gamma_{j'}) \leq 0, \quad j = 1, 2, 3,
\]

\[
t + w^0(x) + (w(x) - \beta)^T \mu_0 + \tau_j \|(w(x) - \beta)\|_2 - \sum_{j'=j}^{3} (\lambda_{j'} - \gamma_{j'}) \leq 0, \quad j = 1, 2, 3,
\]

\[
\sum_{i=1}^{L} v_i = A \beta, \quad \theta \geq 0, \quad t > 0, \quad \lambda_j \geq 0, \quad \gamma_j \geq 0, \quad \forall j = 1, 2, 3.
\]

**Example 3.2.** Let \( X_j = \{ \xi \in \mathbb{R}^k \| \xi - \mu_j \|_p \leq \tau_j \}, j = 1, 2, \quad X_3 = \text{conv}\{X_1, X_2\}, \) where \( \tau_2 > 0, \tau_1 > 0, p \geq 1 \) and \( X_1 \cap X_2 = \emptyset \). Let

\[
U_l = \left\{ \zeta \in \mathbb{R}^k \mid \sum_{i=1}^{k} \zeta_i \log \left( \frac{\zeta_i}{\zeta_0^l} \right) \leq \rho \right\},
\]

where \( \zeta_0^l (l = 1, \ldots, k) \) are given nominal values and \( \rho > 0 \). Then, we have

\[
\delta^*_l (y|X_j) = y^T \mu_j + \tau_j \| y \|_q, \quad j = 1, 2,
\]

\[
\delta^*(y|X_3) = \max_{j=1,2} \{ y^T \mu_j + \tau_j \| y \|_q \},
\]

\[
\delta^*(A \beta|U_1) = \min_{u \geq 0} \left\{ \rho u + \sum_{l=1}^{k} \theta_k \zeta_0^l u \exp \left( \frac{a_l^T \beta}{u} - 1 \right) \right\},
\]

where \( a_l \) is the \( j \)-row of matrix \( A \) and \( \frac{1}{p} + \frac{1}{q} = 1 \). Thus, the problem (15) reduces to the following problem:

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t. } & \quad \rho u + \sum_{l=1}^{k} \theta_k \zeta_0^l u \exp \left( \frac{a_l^T \beta}{u} - 1 \right) + \beta^T \mu_0 + \sum_{j=1}^{3} (\lambda_j p_j - \gamma_j p_j) - \varepsilon t \leq 0,
\end{align*}
\]

\[
-\beta^T \mu_j + \tau_j \| \beta \|_q - \sum_{j'=j}^{3} (\lambda_{j'} - \gamma_{j'}) \leq 0, \quad j = 1, 2,
\]

\[
t + w^0(x) + (w(x) - \beta)^T \mu_j + \tau_j \|(w(x) - \beta)\|_q - \sum_{j'=j}^{3} (\lambda_{j'} - \gamma_{j'}) \leq 0, \quad j = 1, 2,
\]

\[
t > 0, \quad u > 0, \quad \lambda_j \geq 0, \quad \gamma_j \geq 0, \quad \forall j = 1, 2, 3.
\]

4. Distributionally robust nonlinear chance constrained problem under the general first order moment and support information. In the former discussions, the constraint function in DRCCP (1) is bilinear in \( x \) and \( \xi \). In many practical problems the constraint function may have more complicate and flexible
structures. In this section, we consider the following distributionally robust nonlinear chance constrained problem

\[
\min \ c^T x
\]
\[
\text{s.t. } \inf_{P \in \mathcal{P}_3} \mathbb{P}(g(\xi, x) \leq 0) \geq 1 - \varepsilon, \ x \in \chi,
\]
where \(g(\cdot, x)\) is concave for all \(x \in \mathbb{R}^n\). We notice that the nonlinear constraint function will result in unnecessarily conservation about DRCCP (1). Here we consider the following distribution set under support and first moment information

\[
\mathcal{P}_3 = \{ P \in \mathcal{P} : \mathbb{P}(\xi \in \mathcal{U}_\xi) = 1, \ \mu \in \mathcal{U}_\mu \},
\]
where \(\mathcal{U}_\xi\) is a convex and compact set and \(\mathcal{U}_\mu\) is defined in (2).

Let \(f : \mathbb{R}^k \to (-\infty, +\infty]\). We denote the convex conjugate of \(f\) by

\[
f^*(y) = \sup_{x \in \text{dom}(f)} \{ x^T y - f(x) \}, \ \forall y \in \mathbb{R}^k
\]
and the concave conjugate of \(f\) by

\[
f_*(y) = \inf_{x \in \text{dom}(-f)} \{ x^T y - f(x) \}, \ \forall y \in \mathbb{R}^k,
\]
where \(\text{Dom}(f)\) is the effective domain of \(f\).

We need the following lemma to show our main result.

**Lemma 4.1.** [1] Let \(f, -g : \mathbb{R}^k \to (-\infty, +\infty]\) be two proper, convex and lower-semicontinuous functionals. If there exists an element \(x_0 \in \text{Dom}(f) \cap \text{Dom}(g)\) such that either \(f\) or \(g\) is continuous at \(x_0\), then the following equality holds:

\[
\min \{ f(x) - g(x) : x \in \mathbb{R}^k \} = \max \{ g_*(y) - f^*(y) : y \in \mathbb{R}^k \}.
\]

Now we give our main result in this section.

**Theorem 4.2.** Distributionally robust nonlinear chance constrained problem (23) under the information set \(\mathcal{P}_3\) can be reformulated as

\[
\min \ c^T x
\]
\[
\text{s.t. } s_0 + \delta^*(A s_1|\mathcal{U}_1) + s_1^T \mu_0 - \varepsilon t \leq 0,
\]
\[
\delta^*(w|\mathcal{U}_\xi) - g_*(w + s_1, x) \leq s_0 - t,
\]
\[
\delta^*(-s_1|\mathcal{U}_\xi) \leq s_0, \ t > 0,
\]
where \(x \in \mathbb{R}^n\), \(t, s_0 \in \mathbb{R}\) and \(w, s_1 \in \mathbb{R}^k\) are decision variables.

**Proof.** By constructing the following information set

\[
\mathcal{P}_{31} = \{ P \in \mathcal{P} : E_P[\xi] = \mu, \ P(\xi \in \mathcal{U}_\xi) = 1 \},
\]
the chance constraint in (23) becomes

\[
\sup_{\mu \in \mathcal{U}_\mu} \sup_{P \in \mathcal{P}_{31}} E_P[1_S] \leq \varepsilon,
\]
where \(1_S\) is the indicator function of set \(S = \{ \xi \in \mathbb{R}^k : g(\xi, x) > 0 \}\). Then, the dual of the subordinate problem \(\sup_{P \in \mathcal{P}_{31}} E_P[1_S]\) can be unfolded as follows:

\[
\inf_{s_0, s_1} s_0 + s_1^T \mu
\]
\[
\text{s.t. } s_0 \geq -s_1^T \xi, \ \forall \xi \in \mathcal{U}_\xi,
\]
\[
s_0 \geq 1 - s_1^T \xi, \ \forall g(\xi, x) > 0, \ \forall \xi \in \mathcal{U}_\xi,
\]
where \( s_0 \in \mathbb{R}, s_1 \in \mathbb{R}^k \) are the dual variables. By Lemma 2.1, the above problem becomes

\[
\inf_{\eta,s_0,s_1} \quad s_0 + s_1^T \mu \\
\text{s.t.} \quad s_0 \geq -s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \\
\text{s.t.} \quad s_0 \geq -s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \\
\text{s.t.} \quad s_0 - \eta g(\xi, x) \geq 1 - s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \eta \geq 0.
\]

Similarly, the left hand term of (25) can be rewritten as

\[
\inf_{\eta,s_0,s_1} \quad s_0 + \delta^*(s_1 | \mathcal{U}_\mu) \\
\text{s.t.} \quad s_0 \geq -s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \\
\text{s.t.} \quad s_0 - \eta g(\xi, x) \geq 1 - s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \eta \geq 0.
\]

Suppose that \( \eta = 0 \). Then the second inequality reduces to \( s_0 \geq 1 - s_1^T \xi \). By taking expectation operation on both sides, we get \( s_0 + s_1^T \mu \geq 1 \), which is a contradiction with \( s_0 + \delta^*(s_1 | \mathcal{U}_\mu) \leq \varepsilon \). Thus, we have \( \eta > 0 \). Divided by \( \eta \) in all constraints of the problem (26) and substituted \((\frac{s_0}{\eta}, \frac{s_1}{\eta}, t)\) by \((s_0, s_1, 1)\), the problem (26) becomes

\[
\inf_{t,s_0,s_1} \quad \frac{1}{\eta}(s_0 + \delta^*(s_1 | \mathcal{U}_\mu)) \\
\text{s.t.} \quad s_0 \geq -s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, \\
\text{s.t.} \quad s_0 - g(\xi, x) \geq t - s_1^T \xi, \forall \xi \in \mathcal{U}_\xi, t > 0.
\]

By Lemma 4.1, the robust counterpart of the last constraint is

\[
s_0 - t \geq \max_{\xi \in \mathcal{U}_\xi} \left\{ g(\xi, x) - s_1^T \xi \right\} \\
= \max_{\xi \in \mathbb{R}^k} \left\{ g(\xi, x) - s_1^T \xi - \delta(\xi | \mathcal{U}_\xi) \right\} \\
= \min_w \left\{ \delta^*(w | \mathcal{U}_\xi) - [g(\xi, x) - s_1^T \xi] \right\} \\
= \min_w \left\{ \delta^*(w | \mathcal{U}_\xi) - g_s(w + s_1, x) \right\}.
\]

Similarly, the robust counterpart of (25) is

\[
s_0 + \delta^*(A s_1 | \mathcal{U}_1) + s_1^T \mu_0 - \varepsilon t \leq 0, \\
\delta^*(w | \mathcal{U}_\xi) - g_s(w + s_1, x) \leq s_0 - t, \\
\delta^*(-s_1 | \mathcal{U}_\xi) \leq s_0, \quad t > 0,
\]

where \( t, s_0, s_1, w \) are decision variables. This completes the proof.

Theorem 4.2 shows that DRCCP (23) under the information set \( \mathcal{P}_3 \) can be reformulated as the problem (24). The key issue for the computationally tractability of the equivalent form (24) is the computation of support functions about the vector set and the conjugate function. Concerning with the computationally tractable conjugate functions, we refer the reader to [3].

Now we give an example to illustrate DRCCP (23).

**Example 4.1.** Let \( \mathcal{U}_1 = \{ \zeta \in \mathbb{R}^k ||\zeta||_2 \leq \rho_1, ||\zeta||_1 \leq \rho_2 \}, \mathcal{U}_\xi = \{ \zeta \in \mathbb{R}^k ||\zeta||_1 \leq \rho_3 \}, \) and

\[
g(\xi, x) = \sum_{i=1}^n f_i(\xi)x_i,
\]
where \( f_i(\cdot) \) is a concave function and \( x_i > 0 \), \( \rho_1 > 0 \) and \( \rho_2 > \rho_3 > 0 \). Then it is easy to see that

\[
g_*(w + s_1, x) = \sup_v \left\{ \sum_{i=1}^{n} x_i(f_i)_* \left( \frac{v_i}{x_i} \right) \left| \sum_{i=1}^{n} v_i = w + s_1 \right. \right\}.
\]

Thus, problem (23) becomes

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad s_0 + \rho_1 \|u_1\|_2 + \rho_2 \|u_2\|_\infty + s_1^T \mu_0 - \varepsilon t \leq 0, \\
& \quad \rho_3 \|w\|_\infty - \sum_{i=1}^{n} x_i(f_i)_* \left( \frac{v_i}{x_i} \right) \leq s_0 - t, \\
& \quad \rho_3 \|s_1\|_\infty \leq s_0, \ u_1 + u_2 = As_1, \sum_{i=1}^{n} v_i = w + s_1, \ t > 0.
\end{align*}
\]

The above problem is convex in variables \( x, v, w, s_0, s_1, t \).

5. Conclusions. In this paper, we mainly focus on the equivalent determinate forms for DRCCPs under several general convex moments information sets. We demonstrate some examples for the support functions about some matrix sets and the equivalent convex programming. We also show the closed form solution for worst case VaR optimization problem. Compared with the results of [28], we directly obtain the equivalent convex programming form for DRCCP and establish a uniform framework for DRCCP under general second order moments information. How to obtain more computationally tractability examples for support function about matrix set is still an open problem. The equivalent determinate form for distributionally robust nonlinear chance constrained problem under information sets \( \mathcal{P}_1 \) and \( \mathcal{P}_2 \) also need further discussion.

As pointed out by [2], the larger the uncertainty set \( \mathcal{P} \) is, the worse the optimal solution’s performance level will be under \( \mathcal{P} \) for distributionally robust optimization problems. Clearly, DRCCPs will be more conservative when the magnitude of uncertainty set \( \mathcal{P} \) becomes bigger. Therefore, how to reduce the conservatism of DRCCP would really deserve to be taken into consideration.

We note that, in some real applications, the sets \( \mathcal{U}_u \) and \( \mathcal{U}_\Sigma \) might be unknown and need to be estimated. Sun and Xu [26] presented a comprehensive convergence analysis of distributionally robust optimization problem where the ambiguity sets were described by a sequential increasing distribution information. Thus, it is important and necessary to consider the perturbation analysis of DRCCPs, which requires our further investigation.

Acknowledgements. The authors would like to thank the editors and the reviewers for their helpful comments and suggestions, which have improved the presentation of the paper. This work was supported by the National Natural Science Foundation of China (11471230, 11671282, 11771067), the Applied Basic Project of Sichuan Province (2019YJ0204) and the Fundamental Research Funds for the Central Universities, Southwest Minzu University (2018NQN28).
REFERENCES

[1] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces*, Springer, New York, 2012.
[2] A. Ben-Tal, D. Bertsimas and D. Brown, A soft robust model for optimization under ambiguity, *Operations Research*, 58 (2010), 1220–1234.
[3] A. Ben-Tal, D. Hertog and J. Vial, Deriving robust counterparts of nonlinear uncertain inequalities, *Mathematical Programming*, 149 (2015), 265–299.
[4] D. S. Bernstein, *Matrix Mathematics*, Princeton University Press, New Jersey, 2009.
[5] G. Calafiore and L. El Ghaoui, On distributionally robust chance-constrained linear programs with applications, *Journal of Optimization Theory and Applications*, 130 (2006), 1–22.
[6] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Operations Research*, 58 (2010), 595–612.
[7] K. W. Ding, M. H. Wang and N. J. Huang, Distributionally robust chance constrained problem under interval distribution information, *Optimization Letters*, 13 (2018), 1315–1328.
[8] L. Ghaoui, M. Oks and F. Oustry, Worst-case value-at-risk and robust portfolio optimization: A conic programming approach, *Operations Research*, 51 (2003), 543–556.
[9] R. Hu, Y.-B. Xiao, N.-J. Huang and X. Wang, Equivalence results of well-posedness for split variational-hemivariational inequalities, *J. Nonlinear Convex Anal.*, 20 (2019), 447–459.
[10] K. Isii, On sharpness of Tchebychev-type inequalities, *Annals of the Institute of Statistical Mathematics*, 14 (1962), 185–197.
[11] B. Li, J. Sun, H. Xu and M. Zhang, A class of two-stage distributionally robust stochastic games, *Journal of Industrial and Management Optimization*, 15 (2019), 387–400.
[12] 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, *Applied Mathematical Modelling*, 58 (2018), 86–97.
[13] 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 Transactions on Wireless Communications*, 16 (2017), 464–474.
[14] B. Li, Y. Rong, J. Sun and K. L. Teo, A distributionally robust minimum variance beamformer design, *IEEE Signal Processing Letters*, 25 (2018), 105–109.
[15] J. Lu, Y.-B. Xiao and N.-J. Huang, A Stackelberg quasi-equilibrium problem via quasi-variational inequalities, *Carpathian Journal of Mathematics*, 34 (2018), 355–362.
[16] W. Li, Y.-B. Xiao, N.-J. Huang and Y. J. Cho, A class of differential inverse quasi-variational inequalities in finite dimensional spaces, *Journal of Nonlinear Sciences and Applications*, 10 (2017), 4532–4543.
[17] K. Natarajan, M. Sim and J. Uichanco, Tractable robust expected utility and risk models for portfolio optimization, *Mathematical Finance*, 20 (2010), 695–731.
[18] A. Petrusel, G. Petrusel, Y.-B. Xiao and J.-C. Yao, Fixed point theorems for generalized contractions with applications to coupled fixed point theory, *Journal of Nonlinear and Convex Analysis*, 19 (2018), 71–87.
[19] I. Polik and T. Terlaky, A survey of the S-lemma, *SIAM Review*, 49 (2007), 371–481.
[20] Q.-Y. Shu, R. Hu and Y.-B. Xiao, Metric characterizations for well-posedness of split hemi-variational inequalities, *J. Inequal. Appl.*, (2018), 17 pp.
[21] A. Shapiro and A. Kleywegt, Minimax analysis of stochastic problems, *Optimization Methods & Software*, 17 (2002), 523–542.
[22] M. Sofonea, Y.-B. Xiao and M. Couderc, Optimization problems for elastic contact models with unilateral constraints, *Z. Angew. Math. Phys.*, 70 (2019), 17 pp.
[23] M. Sofonea, Y.-B. Xiao and M. Couderc, Optimization problems for a viscoelastic frictional contact problem with unilateral constraints, *Nonlinear Anal. Real World Appl.*, 50 (2019), 86–103.
[24] M. Sofonea, A. Matei and Y.-B. Xiao, Optimal control for a class of mixed variational problems, Submitted.
[25] M. Sofonea and Y.-B. Xiao, Boundary optimal control of a nonsmooth frictionless contact problem, *Comput. Math. Appl.*, 78 (2019), 152-165.
[26] H. Sun and H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Mathematics of Operations Research*, 41 (2016), 377–401.
[27] X. Tong, H. Sun, X. Luo and Q. Zheng, Distributionally robust chance constrained optimization for economic dispatch in renewable energy integrated systems, *Journal of Global Optimization*, 70 (2018), 131–158.

[28] X. Wang, N. Fan and P. Pardalos, Robust chance-constrained support vector machines with second-order moment information, *Annals of Operations Research*, 263 (2018), 45–68.

[29] Y.-M. Wang, Y.-B. Xiao, X. Wang and Y. J. Cho, Equivalence of well-posedness between systems of hemivariational inequalities and inclusion problems, *J. Nonlinear Sci. Appl.*, 9 (2016), 1178–1192.

[30] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Operations Research*, 62 (2014), 1358–1376.

[31] W. Xie and S. Ahmed, On deterministic reformulations of distributionally robust joint chance constrained optimization problems, *SIAM Journal on Optimization*, 28 (2018), 1151–1182.

[32] Y.-B. Xiao and M. Sofonea, On the optimal control of variational-hemivariational inequalities, *Journal of Mathematical Analysis and Applications*, 475 (2019), 364–384.

[33] Y.-B. Xiao and M. Sofonea, Generalized penalty method for elliptic variational-hemivariational inequalities, *Applied Mathematics and Optimization*, (2019).

[34] W. Yang and H. Xu, Distributionally robust chance constraints for non-linear uncertainties, *Mathematical Programming*, 155 (2016), 231–265.

[35] Y. Zhang, S. Shen and S. Erdogan, Distributionally robust appointment scheduling with moment-based ambiguity set, *Operations Research Letters*, 45 (2017), 139–144.

[36] S. Zymler, D. Kuhn and B. Rustem, Distributionally robust joint chance constraints with second-order moment information, *Mathematical Programming*, 137 (2013), 167–198.

[37] S. Zymler, D. Kuhn and B. Rustem, Worst-case value at risk of nonlinear portfolios, *Management Science*, 59 (2009), 172–188.

Received December 2018; revised March 2019.

*E-mail address: dingkeweimath@163.com*

*E-mail address: nanjinghuang@hotmail.com*

*E-mail address: xiaoyb9999@hotmail.com*