Assessing the Value of Including Unimodality Information in Distributionally Robust Optimization Applied to Optimal Power Flow

Bowen Li, Student Member, IEEE, Ruiwei Jiang, Member, IEEE, and Johanna L. Mathieu, Member, IEEE

Abstract—Uncertainties, such as renewable generation and load consumption, has been a major source of risk in power system planning. To safely incorporate these uncertainties, we need to ensure all physical constraints affected are satisfied at high probability level. To manage the uncertainties, different stochastic optimal power flow formulations have been proposed. Conventional approaches either provide over-conservative results or rely on accurate estimates on uncertainty distributions. Recently, an alternative distributionally robust optimal power flow formulations with only data-driven moment information are shown to provide better trade-off between objective and reliability. In this paper, we further consider a distributionally robust optimal power flow problems with both moment and unimodality information, based on the facts that most practical uncertainties are unimodal. We formulate the problem using chance constraint and provide various reformulations and approximations with efficient solving techniques. We evaluate the proposed approaches on modified IEEE 118/300-bus system with high penetrated renewable generation, and demonstrate the values of including the unimodality information and benefits against the conventional approaches.

Index Terms—Optimal power flow, chance constraint, distributionally robust optimization, \( \alpha \)-unimodality

I. INTRODUCTION

To resolve the uncertainty challenge and ensure the system operations are resilient to different uncertainties, previous works used chance constraint to formulate a stochastic optimal power flow (OPF) problem in which physical constraints are forced to be satisfied at high probability levels [1]–[7]. As conventional approaches to solve the chance constrained (CC) problem, we can either use randomized methodologies with empirical uncertainty scenarios [8], [9], various analytical reformulations assuming known distributions [10], [11], [12], and sample average approximation to approximate the probability level of constraint violation [13], [14]. For randomized methodologies such as scenario approach or probabilistically robust method, at high confidence level, they require large numbers of scenarios, high computational effort and provide overly conservative results against the pre-defined probability level. For analytical reformulations, less computational efforts are needed as the reformulation only uses key statistical information like moments but their results can suffer from low reliability as the assumption of the uncertainty distribution might be biased from the true one. For sample average approximation, it illustrates better performance with increasing number of samples but it also introduces heavy computation burdens as more binary variables and constraints are included at the same time. Hence, our objective of the work is to seek new approaches that are efficient to solve and generate high quality solutions (i.e., low objective costs and high reliability).

Recently, to robustify the results against the ambiguous distribution estimate, a distributionally robust (DR) optimization formulation is proposed, in which all the probabilistic constraints or expectation objectives are satisfied or evaluated considering the worst case probability distributions within a data-driven ambiguity set [15]–[19]. This new formulation is closely related to robust and stochastic optimization if the ambiguity set includes only support information or a singleton distribution. Hence, with proper statistical information involved in the ambiguity sets, DR optimization can achieve a better trade-off in terms of objective costs and reliability against current approaches and possibly presents reasonable computational tractability.

In previous DR studies on OPF problems, tractable reformulations are achieved if this ambiguity set is either moment based [17]–[23], statistical density-based [24]–[26] or includes structural information like symmetry [10], unimodality [10], [18], [27], [28], and log-concavity [29]. In [20], the authors consider an ambiguity set with first and second moments to solve a single period OPF problem. Their results present a good trade-off between objective performance and computational tractability. Similar with moment-based ambiguity set, [21] considers two-sided joint chance constraint for generation and transmission line limit constraints. In [10], [18], the authors use analytical reformulation techniques from a univariate random variable to consider structural properties like symmetry and unimodality (with no data-driven mode values). [24]–[26] use the ambiguity sets that quantify the difference between real distribution and data-based distribution with different metrics. Similar to [27], [28], in this paper, we incorporate a generalized multivariate unimodal distributions with fixed mode into known first and second moment in our ambiguity set. In practice, most uncertainties (i.e., wind generation) follow a bell-shape (unimodal) distribution with single peak and decaying tails. To further quantify the shape of the distribution, we use the concept called \( \alpha \)-unimodality with \( \alpha \) as the parameter.

As contributions of the paper, we first apply the results in [28], and propose multiple distributionally robust chance constrained (DRCC) optimization methodologies by incorporating...
the unimodality information into the moment-based ambiguity sets. To solve the problem, we give the exact reformulations and asymptotic sandwich approximations in second-order cone programs (SOCPs). For the exact reformulation, we provide an efficient solving algorithms to quickly identify the worst-case violated constraints through iterations. In addition, we define a new optimal parameter selection (OPS) problem to determine the optimal sandwich approximation of the DRCC problem. The OPS problem is equivalent to finding the optimal piecewise linear (PWL) outer approximation of a concave function. This also enables us a new off-line approximation against the current on-line approximation. As support, we give mathematical proofs for the optimality condition and existence guarantee and provide a heuristic algorithm to solve for the optimal parameters. Finally, we compare all the proposed formulations with analytical reformulation under Gaussian assumption, probabilistic robust method, and DR formulation without unimodality. Specifically, we evaluate each approach to gain insight of the underlying benefits and trade-offs in terms of objective costs, computational tractability, and empirical reliability. Meanwhile, we also assess the values of including unimodality information into DR approaches. Meanwhile, we demonstrate additional computational and performance benefits from our new OPS techniques and the off-line approximation.

The remainder of the paper is organized as follows. In Section [II], we introduce some fundamental concepts and generalize the results regarding DRCC formulations in [28]. In Section [III], we introduce the OPS problem for the sandwich approximations of DR chance constraints. In Section [IV], we test all the proposed approaches, algorithms, and newly developed techniques, compare with conventional methodologies and analyze the case study results. We specifically discuss the trade-offs and values of adding unimodality into moment-based DR optimization. Section [V] concludes the paper.

II. DISTRIBUTIONALLY ROBUST REFORMULATIONS

In this section, we first introduce the key fundamentals and our interested ambiguity sets in the DR reformulation. Note that the theoretical results in this section are from [28] with minor generalizations.

A. Fundamentals

In this paper, we assume the constraints under the uncertainty can be transformed into

\[ a(x)^\top \xi \leq b(x), \]  

(1)

where \( x \in \mathbb{R}^n \) represents the vector of decision variables and \( a(x) : \mathbb{R}^n \to \mathbb{R}^l \) and \( b(x) : \mathbb{R}^n \to \mathbb{R} \) represent two affine functions of \( x \). Uncertainty \( \xi \in \mathbb{R}^l \) represents a random vector defined on probability space \((\mathbb{R}^l, B^l, P_\xi)\) with Borel \( \sigma \)-algebra \( B^l \) and probability distribution \( P_\xi \). The relationships that \( a(x) \) and \( b(x) \) are affine in \( x \) are widely assumed in practical engineering problems (e.g., DC OPF).

To manage uncertain violations in (1), we can directly require that the constraint is satisfied with at least a probability threshold \( 1 - \epsilon \). This leads to the following chance constraint:

\[ \mathbb{P}_\xi (a(x)^\top \xi \leq b(x)) \geq 1 - \epsilon, \]  

(2)

where \( 1 - \epsilon \) normally takes a large value (e.g., 0.99 [30], [31]).

B. Ambiguity Sets

Here, we introduce the two types of ambiguity sets we will consider in the DR reformulation. These sets are defined as combinations of moment and unimodality information. For unimodality, we consider the following generalized multivariate \( \alpha \) unimodality.

Definition II.1. (\( \alpha \)-Unimodality [32]) For any fixed \( \alpha > 0 \), a probability distribution \( P \) on \( \mathbb{R}^n \) is called \( \alpha \)-unimodal with mode \( \alpha \)-unimodality respectively; \( \mu \) and \( \Sigma \) denote the first and second moments of \( \xi \); and \( M(\xi) = m \) means that the true mode value of \( \xi \) is \( m \).

Next, before we introduce the DR reformulations, we will discuss the minor generalizations. In [28], the results are derived assuming the mode is at the origin. Without loss of generality, we can reformulate (1) as \( a(x)^\top (\xi - m) \leq b(x) - a(x)^\top m \) with \( \xi - m \) as our new random vector whose mode is at the origin and apply the results in accordingly. Meanwhile, we also require the following assumptions.

Assumption II.1. For \( U_\xi \), we assume that

\[ \left( \frac{\alpha + 2}{\alpha} \right) (\Sigma - \mu \mu^\top) > \frac{1}{\alpha^2} (\mu - m)(\mu - m)^\top. \]

Assumption II.2. For \( U_\xi \), we assume that \( a(x)^\top m \leq b(x) \).

Both assumptions are standard in the related literature [28], [33]–[35]. Assumption II.1 ensures that the corresponding \( D_\xi \neq 0 \). Assumption II.2 ensures that the constraint is satisfied at the mode. Furthermore, we assume \( \epsilon < 0.5 \) and \( \alpha \geq 1 \), since in practice the uncertainties will at least be univariate-unimodal.

C. DR reformulations with \( D_\xi \)

In this section, we consider DR chance constraint with \( D_\xi \):

\[ \inf_{P_\xi \in D_\xi} \mathbb{P}_\xi (a(x)^\top \xi \leq b(x)) \geq 1 - \epsilon, \]

(5)

and give its exact reformulation in SOCP.
Theorem II.1. (Theorem 2.2 in [36]) DR chance constraint (5) can be exactly reformulated as
\[
\sqrt{\left(\frac{1-\epsilon}{\epsilon}\right)} a(x)^\top (\Sigma - \mu \mu^\top) a(x) \leq b(x) - a(x)^\top \mu. \tag{6}
\]

D. DR Chance Constraint with \(\mathcal{U}_\xi\)

In this section, we consider DR chance constraint with \(\mathcal{U}_\xi\):
\[\inf_{p \in \mathcal{U}_\xi} \mathbb{P}_\xi (a(x)^\top \xi \leq b(x)) \geq 1 - \epsilon, \tag{7}\]
and we give the corresponding SOCP-based exact reformulation with its efficient solving algorithm and the asymptotic sandwich approximations by generalizing the results in [28].

1) Exact Reformulations:

Theorem II.2. (Theorem 1 in [28]) DR chance constraint (7) can be exactly reformulated as
\[
\sqrt{\frac{1-\epsilon - \tau^{-\alpha}}{\epsilon}} \|\Lambda a(x)\| \leq \tau \left(b(x) - a(x)^\top m\right)
- \left(\frac{\alpha + 1}{\alpha}\right) (\mu - m)^\top a(x), \quad \forall \tau \geq \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}, \tag{8}\]
where \(\Lambda := \left(\left(\frac{\alpha + 2}{\alpha}\right) (\Sigma - \mu \mu^\top)^\top - \frac{1}{\alpha} (\mu - m)(\mu - m)^\top\right)^{1/2}\).

Since parameter \(\tau\) has an infinite number of choices, the reformulation in Theorem II.2 also involves an infinite number of SOC constraints. To solve an optimization problem with (8), we give the following iterative algorithm Algorithm 1 based on the step of separation. Note that we assume the reformulated optimization problem in Step 1 can be solved directly.

Algorithm 1: Iterative solving algorithm

Initialization: \(i = 1, \tau_0 = \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}\).

Iteration \(i\):

Step 1: Solve the reformulated optimization problem with (8) using \(\tau_j\) for all \(j = 0, \ldots, i - 1\) and obtain optimal solution \(x^*_i\). All \(\tau_j\) values are collected from previous iterations;

Step 2 (Separation): Find worst case \(\tau^*\) that result in the largest violation of (8) under \(x^*_i\): IF \(\tau^*\) does not exist, STOP and RETURN \(x^*_i\) as optimal solution; ELSE GOTO Step 3;

Step 3: Set \(\tau_i = \tau^*\) and \(i = i + 1\), GOTO Step 1;

To efficiently perform Step 2 in Algorithm 1, we follow the proposition below.

Proposition II.1. (Proposition 3 in [28]) Define \(\mu^*_0 = \left(\frac{\alpha + 1}{\alpha}\right) (\mu - m)^\top a(x^*_i)\) and \(\Sigma^*_0 = \left(\frac{\alpha + 2}{\alpha}\right) a(x^*_i)^\top (\Sigma + mm^\top - m\mu^\top - \mu m^\top) a(x^*_i)\).

Then we have the following:

1) If \(a(x^*_i) = 0\), then constraints (8) are always satisfied;
2) If \(a(x^*_i) \neq 0\) and \(b(x^*_i) - a(x^*_i)^\top m = 0\), then \(x^*_i\) violates constraints (8) if and only if it violates them at \(\tau^* = \infty\).

3) If \(a(x^*_i) \neq 0\) and \(b(x^*_i) - a(x^*_i)^\top m > 0\), then \(x^*_i\) violates constraints (8) if and only if it violates them at \(\tau^* = \hat{\tau}\), where \(\hat{\tau}\) represents the minimizer of the one-dimensional problem
\[
\min_{\tau \geq \tau_0} \left(\left(\frac{1}{1-\epsilon - \tau^{-\alpha}}\right) \left(\left(\frac{\alpha^2}{2}\right)\Sigma^*_0 - \mu^*_0, \mu^*_0\right)^2 - \left(\frac{1-\epsilon - \tau^{-\alpha}}{\epsilon}\right) \left(\left(\frac{\alpha^2}{2}\right)\Sigma^*_0 - \mu^*_0\right)^2\right). \tag{9}\]
whose objective function is strongly convex. The minimizer can be efficiently found through golden section search in the interval \([\tau_0, \tau_u]\) where
\[
\tau_u = \frac{\mu^*_0}{b(x^*_i) - a(x^*_i)^\top m} + \frac{\alpha(1-\epsilon)\Sigma^*_0 - \mu^*_0}{2\epsilon(b(x^*_i) - a(x^*_i)^\top m)}. \tag{10}\]

In other words, if \(x^*_i\) can violate (8), the largest violation happens at \(\tau^*\) given in Proposition II.1.

2) Asymptotic Sandwich Approximations: We notice that solving the exact reformulation can be cumbersome since we might have to deal with many separation problems and iterations. Hence, it is reasonable to have the sandwich approximations to bound the true objective value from both lower and upper directions. The approximations are asymptotic since they will converge to the true objective costs with more inputs.

The results are as follows.

Proposition II.2. Relaxed Approximation (Proposition 4 in [28]) For given integer \(K \geq 1\), and real number \(\tau_0 \leq n_1 < n_2 < \ldots < n_K \leq \infty\), (7) implies the SOC constraints
\[
\sqrt{\frac{1-\epsilon - n_k^{-\alpha}}{\epsilon}} \|\Lambda a(x)\| \leq n_k \left(b(x) - a(x)^\top m\right)
- \left(\frac{\alpha + 1}{\alpha}\right) (\mu - m)^\top a(x), \quad \forall k = 1, \ldots, K. \tag{11}\]

Proposition II.3. Conservative Approximation (Proposition 5 in [28]) For given integer \(K \geq 2\), and real number \(\tau_0 = n_1 < n_2 < \ldots < n_K = \infty\), we define a piece-wise linear function containing \((K - 1)\) pieces:
\[
g(\tau) = \min_{k=2,\ldots,K} \left\{ \sqrt{\frac{1}{\epsilon (1-\epsilon - n_k^{-\alpha})}} \left[\left(\frac{\alpha^{n_k-1}}{2}\right)\tau\right.ight.
+ 1 - \epsilon - \left(1 + \frac{\alpha}{2}\right) n_k^{-\alpha}\left.\right]\}. \tag{12}\]

Denote \(q_1 = \tau_0\) and \(q_2 < \ldots < q_{K-1}\) represent the \((K - 2)\) breakpoints of function \(g(\tau)\). Then, (7) is implied by the SOC constraints
\[
g(q_k)\|\Lambda a(x)\| \leq q_k \left(b(x) - a(x)^\top m\right)
- \left(\frac{\alpha + 1}{\alpha}\right) (\mu - m)^\top a(x), \quad \forall k = 1, \ldots, K - 1. \tag{13}\]

The converging performance of the sandwich approximation is directly affected by the selection of \(n_k\) for all \(k = 1, \ldots, K\). In [28], we proposed to use an on-line parameter selection scheme by using the worst case \(\tau^*\) values from the separation problem in Algorithm 1. However, these values are optimal for
the relaxed approximation (see Theorem II.2) but they do not have any direct relations to the conservative approximation. In later section, we will propose a new off-line version of OPS problem for the conservative approximation.

**Remark:** On the other hand, if the decision makers are more interested in the violation magnitude rather than violation probability, we can use other risk measures such as conditional value at risk (CVaR) which evaluates the conditional expectation of \(a(x)^\top \xi - b(x)\) on the right tail of its distribution. The details of solving DR CVaR constrained problem under \(D_\xi\) or \(U_\xi\) are discussed in [37] and [28] respectively.

### III. Optimal Parameter Selection

In this section, we define an OPS problem for the conservative approximation of the DR chance constraint. Based on [28], \(q_k\)'s in Proposition II.3 define the break points of a concave PWL function \(g(\tau)\) that outer approximates the nonlinear function

\[
f(\tau) = \sqrt{\frac{1 - \epsilon - \tau - \alpha}{\epsilon}} \quad \text{where} \quad \tau \in [\tau_0, \infty). \tag{14}
\]

Hence, equivalently we can relate the OPS problem to finding the optimal PWL outer approximation. Conventional approaches on finding optimal PWL approximation have been thoroughly discussed in [38]-[40]. They are not applicable to our problem because they do not consider outer approximation and assume the function has bounded domain. Based on their work, we make the following specific extensions for our OPS problem:

1) We give the optimality and existence guarantee for the optimal PWL outer approximation of a bounded concave function with unbounded domain. In addition, the idea can be easily applied to problem with bounded domain, convex functions, and inner approximations.

2) We give an heuristic searching algorithm to find the optimal approximation.

#### A. Optimality and Existence

First we define what is an "optimal" PWL approximation in our problem. Given a differentiable concave function \(v(\tau)\) on \(\tau \in [\tau, \infty)\) where \(\tau\) is finite (otherwise \(v(\tau)\) cannot be concave). Further we assume \(v(\tau)\) is bounded within domain to exclude the extreme cases such that the optimal solution does not exist. Denote an \(|S|\)-piece PWL outer approximation \(h(\tau) = \min_{s \in S} d_s \tau + f_s\) where \(S\) represents the index set of the pieces and \(d_s\) is non-increasing with increasing \(s\) (i.e., \(h(\tau)\) is concave). We further define \(h_s(\tau)\) represents the \(s\)-th piece in \(h(\tau)\). Suppose the domain of each piece in \(h(\tau)\) as \(\mathcal{H}_s\), then the error \(e\) of these two functions can be defined as the largest distance between these two functions

\[
e = \max_{s \in S} \max_{\tau \in \mathcal{H}_s} (d_s \tau + f_s - v(\tau)). \tag{15}
\]

Hence, the optimal solution describes the \(|S|\)-piece approximation that minimizes \(e\).

Next, we discuss the optimality condition such that: if there exists a PWL solution satisfying this condition then this solution is optimal. To start, we give the following lemma based on a simple setting.

**Proposition III.1.** (Single piece) If \(v(\tau)\) is a differentiable concave function on \([\tau, \tau]\), a linear function \(h(\tau) = d_0 \tau + f_0\) is the optimal linear outer approximation (i.e., \(|S| = 1\)) of \(v(\tau)\) if \(h(\tau)\) is tangent to \(v(\tau)\) and satisfy

\[
h(\tau) - v(\tau) = h(\tau) - v(\tau), \tag{16}
\]

\[
h(\tau) - v(\tau) = h(\tau) - v(\tau). \tag{17}
\]

The proof is neglected as the underlying idea is intuitive. If \(h(\tau)\) is tangent to \(v(\tau)\) at a different point from the current optimal selection, \(e\) will inevitably increase. Meanwhile, \(v(\tau)\) is bounded since it has a bounded domain and it needs to be differentiable and concave. From this simple example, we observe that the optimality condition requires a tangent relationship and compares the distance between \(h(\tau)\) and \(v(\tau)\) at the end points of the domain. Next, we show that for multi-piece approximation (i.e., \(|S| \geq 1\)), similar results can be attained.

**Theorem III.1.** (Multiple pieces) If \(v(\tau)\) is a bounded differentiable concave function on \([\tau, \tau]\), an \(|S|\)-piece PWL function \(h(\tau) = \min_{s \in S} d_s \tau + f_s\) is the optimal PWL outer approximation of \(v(\tau)\) if

1) The last piece of \(h(\tau)\) (i.e., when \(s = |S|\)) has to be tangent to \(v(\tau)\) at \(\tau = \infty\) with zero slope.

2) Each pieces of \(h(\tau)\) is tangent to \(v(\tau)\).

3) At all \(|S|\) break points and end point (i.e., \(\tau\) of \(h(\tau)\)), the distances between \(h(\tau)\) and \(v(\tau)\) are equal.

The proof is given in Appendix A

Next, we show that, given the same \(v(\tau)\), there always exist a \(h(\tau)\) that satisfies all three statements in Theorem III.1.

**Theorem III.2.** (Existence) If \(v(\tau)\) is a bounded differentiable concave function on \([\tau, \tau]\), there always exist an \(|S|\)-piece PWL function \(h(\tau) = \min_{s \in S} d_s \tau + f_s\) that satisfies all three statements in Theorem III.1. The proof is given in Appendix B

#### B. Searching Algorithm

Here we provide a heuristic searching algorithm Algorithm 2 to solve for the optimal \(|S|\)-piece PWL approximation of \(f(\tau)\) on \([\tau_0, \infty)\) in (14). The algorithm is modified from the recursive descent algorithm in [39]. Before we give the algorithm, we define the following notations:

- \(I\): max number of iteration;
- \(\delta\): percentage tolerance as termination criteria;
- \(h^i(\tau)\): For iteration \(i\), the current first \(|S| - 1\) pieces of the \(|S|\)-piece PWL outer approximation of \(f(\tau)\). We exclude the last zero-slope piece since it is trivial. \(h^i_s(\tau)\) represents the \(s\)-th linear function in the PWL approximation.
- \(B^i \in \mathbb{R}^{|S|}\): For iteration \(i\), all the break points and end points in \(h^i(\tau)\). \(B^i_s\) represents the \(s\)-th entry;
- \(T^i \in \mathbb{R}^{|S| - 1}\): For iteration \(i\), all the tangent points between \(h(\tau)\) and \(f(\tau)\). \(T^i_s\) represents the \(s\)-th entry;
- \(E^i \in \mathbb{R}^{|S|}\): For iteration \(i\), distance between \(h(\tau)\) and \(f(\tau)\) at all \(B^i\)'s. Define the "true" error between last zero-slope piece and \(f(\tau)\) at \(B^i_{|S|}\) as \(e^i_T = (1 - \epsilon)/\epsilon - f(B^i_{|S|})\).
- \(d^i \in \mathbb{R}^{|S| - 1}\): For iteration \(i\), the modification on \(T^i\). \(d^i_s\) represents the \(s\)-th entry;
• $\Delta \in \mathbb{R}^2$: the step size for all iterations. $\Delta^i$ represents the $i$-th entry.

Algorithm 2: Heuristic searching algorithm

Initialization: $i = 1$, $\Delta = 1$, $\delta = 0.01$, $I = 50$, $B^1_{|S|} = 10$, calculate $T^1$ by evenly divide $[\tau_0, B^1_{|S|}]$ into $|S|$ pieces;

Iteration $i$:
Step 1: IF $i \leq I$, calculate the internal $B^i$ for $i = 2, ..., |S| - 1$ based on end points $\tau_0$ and $B^i_{|S|}$ as well as all $T^i$;

ELSE STOP and RETURN no convergence under current initialization.

Step 2: Calculate $E^i$ and $e_T^i$:
- IF $(1 + \text{tol}) \max(E^i) < e_T^i$ or $(1 + \text{tol}) E^i_{|S|} < e_T^i$, $B^i_{|S|} = 0.5(\tau^* + B^i_{|S|})$ where $h_{|S| - 1}(\tau^*) = \sqrt{(1 - \epsilon)/\epsilon}$
  GOTO Step 1;

- ELSE $e_T^i < \min(E^i)$ or $(1 + \text{tol}) e_T^i < E^i_{|S|}$, $B^i_{|S|} = 0.5(\tau^* + B^i_{|S|})$ where $h_{|S| - 1}(\tau^*) = \sqrt{(1 - \epsilon)/\epsilon}$
  GOTO Step 1;

- ELSE GOTO Step 3;

Step 3: IF $\max(E^i) \leq (1 + \text{tol}) \min(E^i)$, STOP and RETURN $h^i(\tau)$ combined with the zero-slope last piece as the optimal solution; ELSE GOTO Step 4;

Step 4: IF $i = 1$, GOTO Step 5; ELSEIF $\max(E^i) > \max(E^{i-1})$, $i = i - 1$, $\Delta^i = \Delta^i/2$, and GOTO Step 5; ELSE GOTO Step 5;

Step 5: For all $s = 1, ..., |S| - 1$,
$$d^i_s = \frac{\Delta^i (E^i_{s+1} - E^i_s)}{h_{s+1}-h_s} \frac{E^i_{s+1} - E^i_s}{B^i_{s+1} - B^i_s}.$$

Then, $T^{i+1} = T^i + d^i$ and $i = i + 1$. GOTO Step 1;

In the initialization step, tol, $\Delta$, $B^1_{|S|}$, and $I$ can use other reasonable values. In Step 2, we fix the last break point $B^i_{|S|}$ by comparing $E^i$ and $e_T^i$. Generally, if $E^i$ is larger, we should reduce $B^i_{|S|}$ and vice versa. In Step 5, the modification $d^i$ is based on the same derivations in [39]. When the algorithm terminates, the optimal break points $B^*$ can be used in the Proposition [1.3] that leads to an off-line conservative approximation.

IV. CASE STUDIES

A. Simulation Setup

We consider similar DC OPF problem with uncertain wind generation as in [28]. We assume that the system has $N_W$ wind power plants with wind forecast error $\tilde{w} \in \mathbb{R}^{N_W}$ (each element is represented as $\tilde{w}_l$), $N_G$ generators, and $N_B$ buses.

The wind forecast error will be compensated by generator reserves. Design variables include generation $P_G \in \mathbb{R}^{N_G}$, up and down reserve capacities $R^{up}\in \mathbb{R}^{N_G}$, $R^{dn}\in \mathbb{R}^{N_G}$, and a distribution vector $d_g \in \mathbb{R}^{N_G}$, which determines how much real-time reserve each generator provides to balance the overall wind forecast error. The full problem formulation is given below.

$$\min P^G_a \{ C_1 P_G + C^T_L P_G + C^T_R (R^{up}_G + R^{dn}_G) \}
 \text{s.t. } -P_t \leq A P_{mj} \leq P_t, \quad (18a)
 R_G = -d_G (\Sigma_{i=1}^{N_W} \tilde{w}_i), \quad (18b)
 P_{mj} = C_G (P_G + R_G) + C_W (P^W_G + \tilde{w}) - C_L P_L, \quad (18c)
 P_G \leq P_t + R_G \leq P^G, \quad (18d)
 -P^{dn}_G \leq R_G \leq P^{up}_G, \quad (18e)
 1_{1 \times N_{G}} d_G = 1, \quad (18f)
 1_{1 \times N_{B}} (C_G P_G + C^W P^W_G - C_L P_L) = 0, \quad (18g)
 P_G \geq 0_{N_{G} \times 1}, \quad d_G \geq 0_{N_{G} \times 1}, \quad (18h)
 R^{up}_G \geq 0_{N_{G} \times 1}, \quad R^{dn}_G \geq 0_{N_{G} \times 1}, \quad (18i)$$

where $[C_1] \in \mathbb{R}^{N_G \times N_{G}}, [C_2] \in \mathbb{R}^{N_G},$ and $[C_R] \in \mathbb{R}^{N_G}$ are cost parameters. Constraint (18a) bounds the power flow by the line limits $P_t$. The power flow is calculated from the power injections $P_{mj}$ in (18c) and the parameter matrix $A$ based on admittance and network connections. Constraint (18b) computes the real-time reserve value $R_G$ that is bounded by the reserve capacities $R^{up}_G$ and $R^{dn}_G$ in (18e). In (18c), $P^W_G$ is the wind forecast, $P_L$ is the load, and $C_G, C_W, C_L$ are matrices that map generators, wind power plants, and loads to buses; (18d) bounds generation within its limits; (18g) enforces power balance with and without wind forecast error; and (18h), (18i) ensure all decision variables are non-negative.

We test our approaches on modified IEEE 118-bus and 300-bus systems with network and cost parameters from [41]. We set $C_R = 10C_2$. We add wind power plants to all generation buses to create high uncertainty dimension. Meanwhile, for high wind penetration, we add in total 400/2000 MW wind forecast generation for 118/300-bus system respectively. Specifically, we scale the wind forecast data so that the forecast generation on each generation buses is proportional to its generation limit.

For uncertainty data, we also consider two sources with distinct properties:

- $a)$ Data Source 1 (DS1): We use the same wind forecast and error scenarios as in [6]. The data are generated based on Markov-Chain Monte Carlo mechanism [42] based on real wind power forecast and actual data from Germany. The data has good forecast with small error-forecast ratio. For each wind bus, we randomly selected the wind errors from the same wind data pool without spatial correlation.

- $b)$ Data Source 2 (DS2): The RE-Europe data set [43] contains hourly wind forecasts and realizations based on European energy system and has strong spatio-temporal correlations. However, the raw data set contains poor forecast values (i.e., extreme error-forecast ratio). Hence, we apply a scaling and filter to scale down the forecast error while maintaining its distribution shape and correlations. The details are demonstrated in Appendix [C].

We solve all optimization problems using CVX with the Mosek solver [44, 45]. We set $\epsilon = 5\%$ and $\alpha = 1$. The latter is valid because, in general, wind forecast error vector is also marginally unimodal (details in Appendix [D]). For data input, we use 5000/8000 randomly generated data points for 118/300-bus system to construct $D_\xi$ and $U_\xi$. More data is needed for 300-bus system since the uncertainty dimension is larger. In addition, we use 15/20 bins to determine the loca-
tions of mode \( m \) for DS1/DS2 by identifying the highest bin in the histogram. Further, we randomly select same number of data points to run out-of-sample tests to evaluate reliability. We define the reliability as the percentage of wind forecast errors for which all chance constraints are satisfied. To guarantee the credibility of the result, we perform 3 parallel tests by randomly reselect the data input.

As benchmark, we consider the following conventional approaches:

- **Probabilistically robust method** [9]: The method enforces the constraints affected by uncertainties to be robust against a scenario-based set. This set is constructed priorly using sufficient number of randomly selected uncertainty realizations.
- **Analytical reformulation with Gaussian assumption** [4], [5], [7]: Here, we assume the uncertainty follows a multivariate normal distribution given the data-driven moment information. Then all the chance constraints can be exactly reformulated as SOCPs.

**B. Results**

1) **Convergence of Algorithm 2:** Here, we demonstrate the convergence performance of Algorithm 2 under different \(|S|\) values. In Fig. 1 we observe that when \(|S|\) increases, the optimal approximation error \( e \) will decrease and the total number of iteration grows almost linearly. In addition, the algorithm also demonstrates a fast convergence rate at different \(|S|\) values although it is heuristic.

2) **Costs, Computational time, and Reliability:** Here, we compare probabilistically robust method (PR), analytical reformulation under Gaussian assumption (GA), DR approach with only moment information (DR-M), and DR approach with both moment and unimodality information (DR-U) on objective costs, reliability, and the overall computational requirement. The results are summarized in Table I. To quantify the comparisons, we define a percentage difference (Diff) on objective cost and reliability against our benchmarks PR and GA as they normally bounds the other approaches. Specifically, we define Diff of GA to be 0 and Diff of PR to be 100%. Diffs of other approaches are calculated as \((X-GA)/(PR-GA)\) where \(X\) can be result of DR-M or DR-U. Next, we define the improvement (Improv) of a certain approach to be the ratio of its Diff on reliability to its Diff on cost. Both PR and GA have Improvs of 1 (i.e., we assume 0/0 = 1 for GA). If the Improv of a certain approach is large, we conclude that this approach has better trade-off between cost and reliability than PR and GA benchmarks. Specifically, it means the approach achieves high reliability with low cost increment.

From Table I we see that, at any system dimensions and data sets, PR provides overly conservative results with highest objective costs and 100% reliability far from our pre-defined probability level of 95%. GA provides the least conservative results with lowest objective costs and lowest reliability that always violates the 95% requirement. For DR-U and DR-M, their costs and reliability are between PR and GA and satisfy the 95% requirement. Specifically, DR-U provides higher objective costs and higher reliability since it only considers moment information in the ambiguity sets. If we compare the Diff and Improv of DR-U and DR-M, we conclude DR-U provides the best trade-off in terms of cost and reliability and hence has the best solution quality. In terms of computation time, GA, PR, and DR-M finishes in single run and DR-U requires an iterative solving algorithm and hence takes the longest computational time. As system dimension increases, we observe the resulted computational burdens but similar relationships between approaches still remain. Comparing DS1 and DS2, solutions from DS1 are more stable with less deviations over the parallel tests. Meanwhile, cost Diffs and simulation time from DS1 are much smaller than the ones from DS2 but reliability Diffs and Improvs are much larger.

3) **Convergence of Algorithm 1:** Here, we analyze the computational performance of Algorithm 1 as well as its approximate solutions within the iterations. Table II summarizes the time percentage of separation (i.e., Step 2 in Algorithm 1), and the total number of iterations. The rest of the time percentage is used to solve for optimal solutions as in Step 1 of Algorithm 1. We observe that generally DS2 requires larger number of iterations than DS1 and 300-bus system requires heavier computational effort in the separation step than 118-bus system.

We further use the 4 test cases with maximum number of iterations in each group of Bus/ Data Set combinations to analyze the time decomposition in each iteration as shown in Fig. 2. We see that, in all the cases, the time of the separation step almost remains constant. With larger number of iterations, the overall simulation time slightly increases as more constraints are added in Step 1 of Algorithm 1. Additionally, cases with 300-bus system or DS2 requires longer simulation time in each iteration.

We have shown that the exact solution of DR-M normally requires long simulation time. Next, we check if the solutions from inner iterations can be good approximates of the final optimal solution. As an example, we use 36-iteration test case in 118/DS2 and calculate its reliability and optimality gap as shown in Fig. 3. We observe that solutions from inner iterations (i.e., lower bound approximation) are not suitable to approximate the final solution. The reason is that even at low optimality gap (< 1%), it may have disastrous reliability (< 70%). In addition, we also observe that higher objective cost does not always guarantee higher reliability in out-of-sample tests.

4) **Conservative Approximations and OPS:** In this section, we provide several other options to approximate the optimal solutions by using the conservative approximation together.
| Bus/Data Set | GA  | PR  | DR-M | DR-U |
|--------------|-----|-----|------|------|
| 118/DS1      |     |     |      |      |
| min          | 3509| 3310| 3466 | 3340 |
| avg          | 3310| 3310| 3468 | 3344 |
| max          | 3349| 3356| 3470 | 3344 |
| 300/DS1      |     |     |      |      |
| min          | 15125| 14409| 15880| 15724 |
| avg          | 15161| 15200| 16956| 15777 |
| max          | 15200| 15200| 17052| 15880 |

Table I: Objective costs, reliability (%), and computational time (second)

| Time Percentage of Separation (%) and Iteration Number |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Bus/Data Set                    | 118/DS1         | 118/DS2         | 300/DS1         | 300/DS2         |
| Percentage Iteration            | min avg max     | min avg max     | min avg max     | min avg max     |
| 118/DS1                         | 86.2 86.8 87.4  | 5 5 5           | 85.4 85.6 85.9  | 26 32 36        |
| 118/DS2                         | 59.1 59.1 59.2  | 6 7 7           | 94.1 94.3 94.5  | 14 23 33        |

Fig. 2. Total simulation time and time of separation in each iteration.

Fig. 3. Optimality Gap and reliability of inner-iteration solutions of Algorithm 1 (red dash marks 1% threshold optimality gap).

Fig. 4. Overall performance of the conservative approximations on 118-bus system with DS1 (red dash marks 1% threshold optimality gap).

with OPS. We propose 5 different options to carry out conservative approximation: Given $K$ in Proposition II.3.

- UB [28]: an on-line approach that uses the worst case $\tau^*$ on the violated constraints from the separation problem in Algorithm 1 for all iterations before $K - 1$.
- OPS0: an on-line approach that uses the OPS solutions $|S| = K - 1$ on the violated constraints from the separation problem in Algorithm 1.
- OPS1: an off-line approach that uses the OPS solutions $|S| = K - 1$ on all the chance constraints.
- OPS2: an aggregated version of OPS0 that use all the OPS solutions with $1 \leq |S| \leq K - 1$.
- OPS3: an aggregated version of OPS1 that use all the OPS solutions with $1 \leq |S| \leq K - 1$.

Note that if $|S| = 1$, solution of Algorithm 2 trivially coincides with Proposition II.3 with $K = 2$. Next, we use the same 4 test cases with maximum number of iterations in each group of bus/data set combinations to analyze their overall performance. For UB, we check solutions resulted from all iterations in the separation problem. For OPS0/1/2/3 options, we limit $|S| \leq 5$ for 118-bus system and $|S| \leq 4$ for 300-bus system.

Figure 4 shows the comparison of conservative approximations on 118-bus system with DS1. For objective cost, UB fails to converge into 1% optimality gap as $K$ increases to the maximum number of iterations. OPS0 and OPS1 demonstrate better convergence rate and optimality gap but fails to
guarantee a cost reduction as $|S|$ increases. The reason is that the
OPS solution for smaller $|S|$ values is not a subset of the OPS solution for larger $|S|$. On the other hand, we also see that
OPS2 and OPS3 show similar convergence rate as well as non-
increasing costs as $|S|$ increases since they aggregate the OPS
solutions. In addition, OPS1 and OPS3 successfully converge
1% optimality gap when $|S| \geq 2$. In terms of simulation
time, we observe that all the approximation solutions except
the last solutions of UB/OPS0/OPS1 take less time than
exactly solving the DR-U (483.4s). Specifically, the off-line
OPS options: OPS1 and OPS3 show overall faster simulation
time than the other on-line options. Actually, the first few
solutions in OPS1 and OPS3 even take similar amount of time
as the single-run approaches (i.e., GA, PR, and DR-M). On-
line options UB/OPS0/OPS2 show linear relationship between
simulation time and $K - 1$ or $|S|$. However, the simulation
time in OPS1 and OPS3 grows faster as $|S|$ increases. For
reliability, all the approximation solutions satisfy the 95% level
and follow the tendency of the optimality gap. Next, if we
shift to 300-bus system as shown in Fig. 7 we observe similar
properties as in Fig. 6 and all on-line and off-line options give
solutions with better performance much faster than solving
DR-U (15720.6s). We also see changes on the simulation time
difference between the on-line and off-line options similar to
Fig. 5 due to the increase on problem dimension. Further, we
also see an oscillation on the reliability of UB which is similar
to Fig. 5 which shows that reliability is not always driven by
the cost.

In addition to the results above, we also observe that
conservative approximations on results from DS1 generally
has smaller optimality gap than DS2. The simulation time
advantages of the off-line options start to decrease when
system dimension increases. In general, we can see that,
conservative approximations are suitable to approximate the
optimal solution since they can have smaller optimality gaps
and higher reliability. Especially With OPS, we are able
to use the off-line approximations or improve the on-line
approximations to achieve better convergence rate and high-
quality approximate solutions with less simulation time.

V. CONCLUSIONS

In this paper, we proposed multiple DRCC methodologies
by incorporating the unimodality information into the moment-
based ambiguity sets. Specifically, we gave the exact reformu-
lization and its efficient solving algorithm based on step of
separation as well as an asymptotic sandwich approximation.
Both reformulation and approximation are in SOC constraints.
Furthermore, we defined a new optimal parameter selection...
problem that determines the optimal setup of the conserva-
tive approximations and provided mathematical support and
solving algorithm. To evaluate the proposed approaches and
ideas, we compared with probabilistically robust method, ana-
lytical reformulation under Gaussian assumption, and DRCC
approach with only moment information. As simulation setup,
we performed case studies on modified IEEE 118/300-bus
system using two wind uncertainty data sets. We observed that
including unimodality information in moment-based DRCC
approach greatly improves the trade-off between objective
costs and reliability compared with other approaches but
also require longer simulation time due to the separation
algorithm takes more iterations than single-run approaches.
Detail analysis also shows that in separation algorithm, the
computation effort almost maintains at the same level through
iterations. Additionally, if we want to find optimal approximate
efficiently, conservative approximations are more reliable than
relaxed approximation. Furthermore, with optimal parameter
selection, we are able to develop multiple on-line and off-line
options that outperforms the existing conservative approxima-
tion approach by providing solutions with low optimal gap,
satisfied reliability level, and much less simulation time.

As future work, we are interested in comparing distribu-
tionally robust approaches using different ambiguity sets.
Meanwhile, we will study their performances on optimization
problems other than OPF and try to further improve the
performance of current distributionally robust approaches.

APPENDIX A

PROOF OF THEOREM III.1

Statement 1: since \( v(\tau) \) is bounded and concave with an
unbounded domain, we are certain that \( v(\tau) \) is non-decreasing
and its slope converges to zero as \( \tau \to \infty \). In addition, since
the last piece of \( h(\tau) \) is a linear function. It must have zero
slope otherwise it will either go to \( \infty \) or cut \( v(\tau) \) as \( \tau \to \infty \).
Meanwhile, it should be tangent to \( v(\tau) \) (i.e., asymptote to
\( v(\tau) \)) otherwise you can shift it down and obtain smaller \( e \).

Statement 2: since \( v(\tau) \) is non-decreasing, we know \( e_s \geq 0
\) for all \( s \in S \) otherwise the last piece will cut \( v(\tau) \). Hence,
if any piece of \( h(\tau) \) is not tangent to \( v(\tau) \), we can start by
shifting any non-touching pieces in \( h(\tau) \) down until it touches
the \( v(\tau) \) at a single point. Then, all the pieces in \( h(\tau) \) will
either be tangent to \( v(\tau) \) or cross \((\tau, v(\tau))\) with a larger slope
than \( v'(\tau) \). If it is the later case, we can rotate the piece until
it is tangent at \( \tau \). Both of the movements will guarantee
to reduce \( e \).

Statement 3: we prove with contradiction. Assume we have
a \( h_1(\tau) \) satisfying all three statements with equal distances \( e_1
\) (i.e., also the largest distance) at all break points, we want to
show there does not exist a \( h_2(\tau) \) satisfying Statements 1 and
2 but has the largest distance \( e_2 < e_1 \). Compare \( h_2(\tau) \) and
\( v(\tau) \), the largest distance only happens at all the break points
or \( \tau \). Hence, this means all the pieces excluding the last one
of \( h_2(\tau) \) have to be tangent at a smaller value than \( h_1(\tau) \) but
this will always make the distance on the LBP larger than \( e_1 \) since
the last piece has to satisfy the Statement 2. In other words,
you can not reduce all the distance at the break points strictly
smaller than \( e_1 \). This contradicts with our assumption \( e_2 < e_1
\) and hence \( h_1(\tau) \) gives us the optimal largest distance.

APPENDIX B

PROOF OF THEOREM III.2

Statements 1 and 2 in Theorem III.1 are easy to satisfy.
Hence, we only need to prove that there exists a \( h(\tau) \) that
satisfies the Statements 1 and 2 also satisfying the Statement
3. We prove this by mathematical induction.

First, we confirm this is true if we only use single-piece
linear approximation. For all \( \mathcal{L} \subseteq [\tau, \infty) \), Statement 3 can be
achieved with the single-piece approximation. Specifically, if
\( \mathcal{L} = [l_1, l_2] \), then Proposition III.1 applies. If \( \mathcal{L} = [l_1, \infty) \),
we can’t use the single-piece approximation but need a slack
piece tangent to \( v(\tau) \) at \( \tau = \infty \). With the slack piece, we can
achieve Statement 3 similar to Proposition III.1.

Next, we assume that for all \( \mathcal{L} \subseteq [\tau, \infty) \), we can realize
Statement 3 with the \( C \)-piece PWL approximation. Similarly,
if \( \mathcal{L} = [l_1, \infty) \), we need a slack piece tangent to \( v(\tau) \) at \( \tau = \infty \).
We need to prove the same property holds with the \( C + 1
\)-piece approximation.

a) Case 1 \( \mathcal{L} = [l_1, l_2] \): If we increase the piece number
from \( C \) to \( C + 1 \), we will have an additional break point
in the approximation. Denote the LBP before \( l_2 \) as \( B_{C} \),
we perform the optimal \( C \)-piece approximation and single-
piece approximation (i.e., satisfying Statement 3) on \([l_1, B_{C}]\)
and \([B_{C}, l_2]\) respectively with errors \( e_{C} \) and \( e_{S} \). When \( B_{C}
\) increases from \( l_1 \) and \( l_2 \), \( e_{C} \) increases from zero to its possible
maximum value and \( e_{S} \) decreases from its possible maximum
value to zero. Hence, we claim that there exists a \( B_{C} \in [l_1, l_2]\)
such that \( e_{C} = e_{S} > 0 \) and the combined approximation has
\( C + 1 \) pieces.

b) Case 2 \( \mathcal{L} = [l_1, \infty) \): We start with the optimal \( C
\)-piece approximation (i.e., satisfying Statement 3) under our
assumption with error \( e_{C} \). Denote the LBP as \( B_{L} \). On \([l_1, B_{L}]\),
the current approximation is also optimal. Next, we add a
new piece by perform an optimal single-piece approximation
on \([B_{L}, \infty)\) with error \( e_{L} \). Similar to Case 1, there exists a
\( B_{L} \in [l_1, l_2] \) such that \( e_{C} = e_{L} > 0 \) and the combined
approximation has \( C + 1 \) pieces.

Hence, with any finite piece number, we can always con-
struct an optimal PWL approximation satisfying all three
statements.

APPENDIX C

STATISTICAL ANALYSIS AND FILTER ON DS2

First, we demonstrate the poor forecast quality in DS2 by
showing the distribution of error-forecast ratio for an instance
node. Intuitively, this error is lower bounded by \(-100\% \). In
Fig. 8 we see that the original data contains extreme outliers
as well as large forecast error within the 25-75 percent quantile
box.

To obtain usable data for simulations while maintaining the
distribution, we first scale all the wind errors down by
60% and filter out extreme error points that is larger than 100%
of the corresponding forecast values. The resulted histogram
and the box plot are shown in Fig. 9 and demonstrate a more
reasonable range and distribution.

APPENDIX D

VALIDATION OF UNIMODALITY

Here, we validate the unimodality of the wind forecast error
from DS1 and DS2 using 5000 data samples. In Figs. 10 and
we present the histograms of univariate and bivariate wind forecast errors with 15/20 bins for DS1/DS2.

In general, both figures empirically justify our assumption that the probability distribution of wind forecast errors is unimodal and satisfies Assumption II.1. Furthermore, we also observe that the distribution of DS2 is more skewed than DS1.

REFERENCES

[1] H. Zhang and P. Li, “Chance constrained programming for optimal power flow under uncertainty,” *IEEE Trans Power Systems*, vol. 26, no. 4, pp. 2417–2424, 2011.

[2] R. A. Jabr, “Adjustable robust OPF with renewable energy sources,” *IEEE Trans Power Systems*, vol. 28, no. 4, pp. 4742–4751, 2013.

[3] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, “A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration,” *IEEE Trans Power Systems*, vol. 28, no. 4, 2013.

[4] D. Bienstock, M. Chertkov, and S. Harnett, “Chance-constrained optimal power flow: risk-aware network control under uncertainty,” *SIAM Review*, vol. 56, no. 3, pp. 461–495, 2014.

[5] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, “Analytical reformulation of security constrained optimal power flow with probabilistic constraints,” in *IEEE PowerTech Conference*, Grenoble, France, 2013.

[6] M. Vrakopoulou, B. Li, and J. Mathieu, “Chance constrained reserve scheduling using uncertain controllable loads Part I: Formulation and scenario-based analysis,” *IEEE Trans Smart Grid* (in press), 2017.

[7] B. Li, M. Vrakopoulou, and J. Mathieu, “Chance constrained reserve scheduling using uncertain controllable loads Part II: Analytical reformulation,” *IEEE Trans Smart Grid* (in press), 2017.

[8] M. Campi, G. Calafiore, and M. Prandini, “The scenario approach for systems and control design,” *Annual Reviews in Control*, vol. 33, no. 2, pp. 149–157, 2009.

[9] K. Margellos, P. Goulart, and J. Lygeros, “On the road between robust optimization and the scenario approach for chance constrained optimization problems,” *IEEE Trans Automatic Control*, vol. 59, no. 8, pp. 2258–2263, 2014.

[10] L. Roald, F. Oldewurtel, B. V. Parys, and G. Andersson, “Security constrained optimal power flow with distributionally robust chance constraints,” arXiv preprint arXiv:1508.06061, 2015.

[11] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, “Sample average approximation method for chance constrained programming: Theory and applications,” *Journal of Optimization Theory and Applications*, vol. 142, no. 2, pp. 399–416, 2009.

[12] S. Ahmed and A. Shapiro, *Solving chance-constrained stochastic programs via sampling and integer programming*. INFORMS, 2008.

[13] L. E. Ghaoui, M. Oks, and F. Oustry, “Worst-case value-at-risk and robust optimization and the scenario approach for chance constrained portfolio optimization,” *IEEE Trans Power Systems*, vol. 26, no. 3, pp. 667–674, 2011.

[14] E. Delage and Y. Ye, “Distributionally robust optimization under moment uncertainty with application to data-driven problems,” *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[15] B. Stellato, *Data-driven chance constrained optimization*. Master thesis, ETH Zurich, 2014.

[16] R. Jiang and Y. Guan, “Data-driven chance constrained stochastic program,” *Mathematical Programming*, vol. 158, no. 1, pp. 291–327, 2016.

[17] M. Lubin, Y. Dvorkin, and S. Backhaus, “A robust approach to chance constrained optimal power flow with renewable generation,” *IEEE Trans Power Systems*, vol. 31, no. 5, pp. 3840–3849, 2016.

[18] T. Summers, J. Warrington, M. Morari, and J. Lygeros, “Stochastic optimal power flow based on conditional value at risk and distributional robustness,” *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 116–125, 2015.

[19] R. Miett and Y. Dvorkin, “Data-driven distributionally robust optimal power flow for distribution systems,” *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 363–368, 2018.

[20] Y. Zhang, S. Shen, and J. L. Mathieu, “Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads,” *IEEE Trans Power Systems*, vol. 32, no. 2, pp. 1378–1388, 2017.

[21] W. Xie and S. Ahmed, “Distributionally robust chance-constrained optimal power flow with renewables: A conic reformulation,” *IEEE Trans Power Systems*, vol. 33, no. 2, pp. 1860–1867, 2018.

[22] X. Lu, K. W. Chan, S. Xia, B. Zhou, and X. Luo, “Security-constrained multi-period economic dispatch with renewable energy utilizing distributionally robust optimization,” *IEEE Trans Sustainable Energy*, 2018.

[23] 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*, vol. 70, no. 1, pp. 131–158, 2018.
[24] Y. Guo, K. Baker, E. Dall’Anese, Z. Hu, and T. Summers, “Stochastic optimal power flow based on data-driven distributionally robust optimization,” in IEEE Annual American Control Conference, Milwaukee, WI, 2018.

[25] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, “Distributionally robust chance-constrained approximate ac-opf with wasserstein metric,” IEEE Trans Power Systems, vol. 33, no. 5, pp. 4924–4936, 2018.

[26] C. Wang, R. Gao, F. Qiu, J. Wang, and L. Xin, “Risk-based distributionally robust optimal power flow with dynamic line rating,” IEEE Trans Power Systems, pp. 1–1, 2018.

[27] B. Li, R. Jiang, and J. L. Mathieu, “Distributionally robust risk-constrained optimal power flow using moment and unimodality information,” in IEEE Conference on Decision and Control, Las Vegas, NV, 2016.

[28] B. Li, R. Jiang, and J. Mathieu, “Ambiguous risk constraints with moment and unimodality information,” Mathematical Programming (Accepted), 2017.

[29] B. Li, J. L. Mathieu, and R. Jiang, “Distributionally robust chance constrained optimal power flow assuming log-concave distributions,” in Power Systems Computation Conference, Dublin, Ireland, 2018.

[30] A. Charnes, W. Cooper, and G. Symonds, “Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil,” Management Science, vol. 4, no. 3, pp. 235–263, 1958.

[31] B. Miller and H. Wagner, “Chance constrained programming with joint constraints,” Operations Research, vol. 13, no. 6, pp. 930–945, 1965.

[32] S. W. Dharmadhikari and K. Joag-Dev, Unimodality, convexity, and applications. Academic Press, 1988.

[33] G. Hanasusanto, Decision Making under Uncertainty: Robust and Data-Driven Approaches. PhD thesis, Imperial College London, 2015.

[34] B. V. Parys, P. Goulart, and D. Kuhn, “Generalized Gauss inequalities via semidefinite programming,” Mathematical Programming, vol. 156, no. 1, pp. 271–302, 2016.

[35] B. V. Parys, P. Goulart, and M. Morari, “Distributionally robust expectation inequalities for structured distributions,” Mathematical Programming (in press), 2017.

[36] M. Wagner, “Stochastic 0–1 linear programming under limited distributional information,” Operations Research Letters, vol. 36, no. 2, pp. 150–156, 2008.

[37] S. Zymler, D. D. Kuhn, and B. Rustem, “Distributionally robust joint chance constraints with second-order moment information,” Mathematical Programming, vol. 137, no. 1, pp. 167–198, 2013.

[38] A. Imamoto and B. Tang, “A recursive descent algorithm for finding the optimal minimax piecewise linear approximation of convex functions,” in World Congress on Engineering and Computer Science, San Francisco, CA, 2008.

[39] J. Vandewalle, “On the calculation of the piecewise linear approximation to a discrete function,” IEEE Trans on Computers, vol. 24, pp. 843–846, 1975.

[40] C. Coffrin, D. Gordon, and P. Scott, “Nesta, the nieta energy system test case archive,” arXiv preprint arXiv:1411.0359v5, 2016.

[41] G. Papaefthymiou and B. Klockl, “MCMC for wind power simulation,” IEEE Trans Energy Conversion, vol. 23, no. 1, pp. 234–240, 2008.

[42] T. V. Jensen and P. Pinson, “Re-europe, a large-scale dataset for modeling a highly renewable european electricity system,” Scientific data, vol. 4, pp. 170–175, 2017.

[43] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” [http://cvxr.com/cvx], 2014.