Distributionally Robust Joint Chance-Constrained Optimal Power Flow using Relative Entropy

Eli Brock*, Haixiang Zhang*, Javad Lavaei, *Fellow, IEEE*, and Somayeh Sojoudi, *Senior Member, IEEE*

Abstract—Designing scalable and robust algorithms for the optimal power flow (OPF) problem is critical for the control of large-scale power systems under uncertainty. The chance-constrained AC OPF (CCOPF) problem provides a natural formulation of the trade-off between the operation cost and the constraint satisfaction rate. In this work, we propose a new data-driven algorithm for the CCOPF problem, which is based on the distributionally robust optimization (DRO). The proposed DRO approach achieves the *optimal efficiency* in the sense that (i) it finds the minimum-cost solution given the maximum rate of violating the constraints, and (ii) it uses an exact mixed-integer reformulation of chance constraints instead of inner approximations as used in existing literature. We apply the proposed algorithm to the semi-definite relaxation of the CCOPF problem and illustrate the advantage of our approach on IEEE benchmark power systems.

Index Terms—Distributionally robust optimization, optimal power flow, chance constraint.

I. INTRODUCTION

Developing resilient algorithms for the optimal power flow (OPF) problem is fundamental to efficient and reliable decision-making in large-scale energy systems. The OPF problem consists of minimizing some objective, including but not limited to generation costs, subject to the physics of the power network as well as additional constraints on power quality, safety, and reliability. Independent system operators solve OPF at several timescales, from hours to minutes ahead of the dispatch time, in order to manage the market and match supply to demand. Traditionally, the primary source of uncertainty in optimal power flow was stochastic loads. This uncertainty was handled through forecasts which were accurate enough that mismatches between supply and demand could be handled in real-time without a significant deviation from nominal network and market conditions. However, given the ongoing emergence of intermittent renewable generation, more sophisticated methods will be necessary to ensure that decisions can be made as efficiently as possible while being robust to large forecast errors.

The randomness in the constraints prohibits the application of optimization algorithms for deterministic problems and most stochastic optimization algorithms, which are often applicable to optimization problems that only contain randomness in the objective function. The robust optimization approach

*Equal Contribution.
Eli Brock and Somayeh Sojoudi are with the Department of EECS, UC Berkeley (e-mail: eli.brock@bekeley.edu, sojoudi@berkeley.edu). Haixiang Zhang is with the Department of Mathematics, UC Berkeley (e-mail: haixiang_zhang@bekeley.edu). Javad Lavaei is with the Department of IEOR, UC Berkeley (e-mail: lavaei@berkeley.edu). This work was supported in part by grants from NSF, Noyce Initiative, ONR, AFOSR, and ARO.

was proposed in [1] and [2] to find the worst-case solution, namely, the optimal decision that satisfies all constraints for all possible realizations of the randomness in the system. The robust optimization approach produces the most conservative solution and results in a high operational cost.

To improve the efficiency of the operation of power systems, it is often preferable to allow a small user-specified probability of violating the constraints in the OPF solution in exchange for a much better operational cost (small violations will later be handled via a real-time control mechanism). Chance-constrained optimal power flow (CCOPF) is a natural formulation for balancing the trade-off between efficiency and robustness [3]. In CCOPF, system operators attempt to find the minimum-cost solution which violates the constraints with a probability at most equal to a pre-defined parameter named the violation probability. Chance-constrained methods avoid the conservativeness associated with robust optimization, which insures an operating point that is feasible for all possible realizations of a system's forecast errors. Please refer to [3] and [4] for popular formulations of CCOPF.

A challenge for CCOPF is that the true underlying distribution of the random parameters is generally unknown and must be inferred from historical data. A conventional approach is the sample average approximation [5], which is easily applicable but may lead to a high-variance estimate of the true distribution. The scenario approach lower-bounds the number of samples required to achieve a given degree of confidence in the probability of satisfying the chance constraints [6] and is employed for CCOPF in [3] and [4]. However, the scenario approach is sample-intensive, may be overly conservative, and is often computationally complex. Additionally, more sample-efficient methods allow for samples over larger time horizons (i.e., a day instead of an hour) to be aggregated into a single realization of a random vector, which could reduce bias if forecast errors follow temporal patterns.

Distributionally robust optimization (DRO) alleviates the issue of unknown true distributions by enforcing the chance constraints for all distributions in an *ambiguity set* centered, in the sense of some characteristic metric of probability distributions, around the empirical distribution [7]. The idea is that, given enough samples, the true distribution is highly likely to fall inside the ambiguity set. A number of papers have applied distributionally robust optimization to OPF or related problems in energy systems. The authors of [8]–[11] employ *moment-based* ambiguity sets containing probability distributions with the first and second moments close to those of the empirical distribution. Li *et al.* [12] add a unimodality assumption to the moment-based sets to reduce the conservatism. Moment-

based ambiguity sets often yield exact tractable reformulations of the chance-constrained program, but they lose information about the true distribution revealed through other features of the data. *Metric-based* ambiguity sets, by contrast, are constructed using measures of distance between probability distributions, most often the Wasserstein metric, and are more expressive. The metric-based approach has the advantage that various statistical consistency and convergence guarantees can be established for DRO estimators [13], [14]. To reformulate the chance constraints as tractable constraints, inner approximations of Wasserstein metric-based ambiguity sets, such as hyper-cubes [15] and polytopes [16], have previously been studied. However, these inner approximations are overly conservative in practice and lead to pessimistic estimations.

All of the aforementioned DRO approaches are designed for *disjoint* chance constraints, in which each constraint individually must be satisfied with a given probability. The chance constraints in CCOPF are formulated disjointly for each two-sided constraint [11], [15], [16] or separately for each upper and lower bound [8], [9], [12]. *Joint* chance constraints, by contrast, require that a solution be feasible, that is, satisfies *all* constraints simultaneously, with a given probability. Given the same violation probability, joint chance constraints are clearly stronger than disjoint chance constraints. Joint chance constraints can be guaranteed by applying the Boole inequality to appropriately scaled disjoint chance constraints; see [17]. However, this approach is highly conservative and does not exploit the potential correlation between random variables in different constraints. Intuitively, when the randomness between constraints is highly correlated, joint chance constraints can be satisfied at a cost that is only slightly higher than that of the chance constraint of a single stochastic constraint. Yang *et al.* [18] build on the Boole inequality approach and achieve an inner approximation of a moment-based ambiguity set for the joint case.

The particularly interesting line of work [19]–[22] is inspired by [13], which provides a reformulation of Wasserstein metric-based DRO problems using conditional value-at-risk (CVaR). The two-part work [19]- [20] is the first to apply the CVaR reformulation to OPF by penalizing constraint violations in the objective function; however, this is not a chance-constrained approach and cannot guarantee the satisfaction of the constraints in any well-defined sense. Poolla *et al.* [21] approximate the joint chance constraints using the Boole inequality and reformulate them using CVaR. To achieve the reformulation, the authors use an inner approximation of the ambiguity set via a hyper-rectangle in the parameter space. Arab *et al.* [22] improve on [21] by using an ellipsoidal approximation, which reduces the conservativeness by exploiting the correlation between random variables. While the ellipse approximation improves on the hyper-rectangle approximation, the method in [22] remains overly conservative as a consequence of mismatch between the inner approximation and the ambiguity set; see Section IV for numerical illustrations. To address the above issues, we build upon our conference paper [23] tailored to a class of non-convex problems using DRO to study the CCOPF problem. Compared to [23], we develop strong theoretical results in the context of power systems for

TABLE I
COMPARISON OF RELEVANT CHANCE-CONSTRAINED OPF LITERATURE

	chance-constrained	joint	metric-based	exact reformulation
[8]	✓	✗	✗	✓
[16]	✓	✗	✓	✗
[11]	✓	✗	✗	✓
[15]	✓	✗	✓	✗
[9]	✓	✗	✗	✓
[12]	✓	✗	✗	✓
[18]	✓	✓	✗	✗
[19], [20]	✗	✗	✓	✓
[21]	✓	✓	✓	✗
[22]	✓	✓	✓	✗
this work	✓	✓	✓	✓

both the joint and disjoint cases, and we numerically illustrate the performances of our approach on benchmark IEEE power systems.

In this work, we expand upon the existing findings related to the DRO approach for CCOPF. Inspired by [14], we use a relative entropy-based ambiguity set in our DRO formulation and establish stronger theoretical guarantees than those in existing literature. Then, we apply the approximation of the chance-constraints from Roald *et al.* [3] and the semidefinite relaxation from Low *et al.* [24] to reformulate the problem as a mixed-integer program, which can be handled by existing optimization solvers. Moreover, we implement the algorithms on benchmark OPF problem instances, showcasing the advantages of our new formulation. We summarize our contributions in the following:

- Instead of the commonly used Wasserstein metric, our DRO formulation utilizes a relative entropy-based ambiguity set. We prove that the relative entropy-based formulation admits the *least conservative* DRO solution in the sense that the solution achieves the minimum possible generation cost under a certain asymptotic bound on out-of-sample performance.
- We provide the first *exact* reformulation of *joint* distributionally robust chance constraints over the ambiguity set. By comparison, existing works construct an approximation set of the ambiguity set and/or only consider disjoint chance constraints, which makes it challenging to control the trade-off between the efficiency and robustness of the solution. In our formulation, the balance can be effectively controlled by an input parameter. In addition, our reformulation always leads to a feasible problem, while existing approaches cannot guarantee the feasibility.
- We empirically compare the performance of our DRO approach with the state-of-the-art approach in [22] on the IEEE 14- and 118-bus test cases. We show that our approach is able to find more reliable and efficient solutions satisfying the joint chance constraints, while the approximation algorithm in [22] leads to overly conservative solutions.

Table I summarizes the relevant existing literature on DRO for power systems (most, but not all, of the listed papers focus on OPF) and illustrates our contributions. It is worth mentioning that all works in Table I except [22] use the common linearized DC approximation of the nonlinear power flow equations, though this approximation is not always coupled to the specific

handling of chance constraints. In comparison, we consider the full ACOPF problem in this work.

The remainder of the paper is organized as follows. In Section II, we first introduce the AC OPF problem and the corresponding joint chance constraint. Reformulations of the chance-constrained AC OPF problem, including the distributionally robust optimization approach, are derived in Section III. Finally, in Section IV, we implement the proposed algorithm to verify the theory and illustrate the superior empirical performances compared with existing algorithms. We conclude the paper in Section V. The proofs are provided in the appendix.

Notation: For every positive integer n, we define $[n] := \{1, \ldots, n\}$. The set of n-dimensional integer, real and complex vectors are denoted as $\mathbb{Z}^n$, $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. Similarly, we use $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ to denote the set of m-by-n real and complex matrices, respectively. The set of symmetric and positive semi-definite matrices of size n-by-n is denoted as $\mathbb{S}^n_+$. Let $\mathbf{1}_n$ and $\mathbf{0}_n$ be the vectors with all elements equal to 1 and 0, respectively. For every complex number x, the real and imaginary parts of x are denoted as $\Re(x)$ and $\Im(x)$, respectively. The same notation applies componentwise to complex vectors and matrices. Denote e_k as the k-th standard basis vector of $\mathbb{R}^n$. For any two matrices $X, Y \in \mathbb{R}^{m \times n}$, the innerproduct between them is defined as $\langle X, Y \rangle := \text{Tr}(X^T Y)$, where Tr stands for the trace. For each vector $\mathbf{v} \in \mathbb{R}^n$, we say $\mathbf{v} \le \mathbf{0}_n$ if $\mathbf{v}_k \le 0$ for all $k \in [n]$. Let $\| \cdot \|$ be the 2-norm of vectors. For a given vector $\mathbf{v} \in \mathbb{R}^n$, matrix $\text{diag}(\mathbf{v}) \in \mathbb{R}^{n \times n}$ is the diagonal matrix with diagonal entries from $\mathbf{v}$. We say $f(S) = o(S)$ if $\lim_{S \to \infty} f(S)/S = 0$.

II. AC OPF PROBLEM AND CHANCE CONSTRAINTS

In this section, we first introduce the notation for system variables and parameters in power flow equations. Then, we formulate the deterministic AC OPF problem as a quadratically constrained quadratic program (QCQP). Finally, we consider the stochastic OPF problem, where the system status is subject to random power injections. We formally define the joint and disjoint chance constraints and formulate the chance-constrained OPF problem.

A. QCQP Formulation of Deterministic AC OPF

Here, we present the deterministic AC OPF problem, namely, the AC OPF problem without unforecasted power injections, as a QCQP. Our formulation and most of our notation is based on [25]. Consider a power system with the set of buses $\mathcal{N} := [n]$. Define the system variables and parameters:

- $\mathbf{V} \in \mathbb{C}^n$: Vector of complex bus voltages.
- $\mathbf{I} \in \mathbb{C}^n$: Vector of complex nodal current injections.
- $\mathbf{Y} \in \mathbb{C}^{n \times n}$: Network admittance matrix, constructed such that $\mathbf{I} = \mathbf{Y}\mathbf{V}$.
- $P^D, Q^D \in \mathbb{R}^n$: Vectors of active and reactive loads, respectively. If bus k has no load, the k-th components are zero.
- $\overline{P}^G, \underline{P}^G \in \mathbb{R}^n$: Vectors of upper and lower active generation limits, respectively. If bus k has no generator, the k-th components are zero.

- $\overline{Q}^G, \underline{Q}^G \in \mathbb{R}^n$: Vectors of upper and lower reactive generation limits, respectively. If bus k has no generator or dispatchable reactive compensation device, the k-th components are zero.
- $\overline{V}, \underline{V} \in \mathbb{R}^n$: vectors of upper and lower voltage magnitude limits, respectively.
- c_{kd}: The d-th degree coefficient of the quadratic cost function for the k-th generator, where $d \in \{0, 1, 2\}$.

Conventional OPF formulations consider fixed loads and dispatchable generators. To simplify the notation, the formulation used in this work incorporates renewable generators (without curtailment) into the load vectors as negative loads. Conversely, centrally dispatchable demand responses can be incorporated into the generator limits through appropriate adjustments.

To formulate the OPF problem as an optimization problem with real variables, we define the real vector

$$\mathbf{X} := \begin{bmatrix} \Re\{\mathbf{V}\} \\ \Im\{\mathbf{V}\} \end{bmatrix} \in \mathbb{R}^{2n}.$$

Additionally, for each $k \in \mathbb{N}$, we define the following matrices:

$$Y_k := e_k e_k^T Y,$$

$$\mathbf{Y}_k := \frac{1}{2} \begin{bmatrix} \Re\{Y_k + Y_k^T\} & \Im\{Y_k^T - Y_k\} \\ \Im\{Y_k - Y_k^T\} & \Re\{Y_k + Y_k^T\} \end{bmatrix},$$

$$\overline{\mathbf{Y}}_k := -\frac{1}{2} \begin{bmatrix} \Im\{Y_k + Y_k^T\} & \Re\{Y_k - Y_k^T\} \\ \Re\{Y_k^T - Y_k\} & \Im\{Y_k + Y_k^T\} \end{bmatrix},$$

$$\mathbf{M}_k := \begin{bmatrix} e_k e_k^T & 0 \\ 0 & e_k e_k^T \end{bmatrix}.$$

While the OPF problem can accommodate different choices of the objective function, we focus on a common total generation cost $f : \mathbb{R}^{2n \times 2n} \to \mathbb{R}$ as follows:

$$f(\mathbf{W}) := \sum_{k \in \mathcal{N}} \Big[c_{k0} + c_{k1} \left(\langle \mathbf{W}, \mathbf{Y}_k \rangle + P_k^D \right) + c_{k2} \left(\langle \mathbf{W}, \mathbf{Y}_k \rangle + P_k^D \right)^2 \Big].$$

We can now write the deterministic AC OPF problem as a real-valued QCQP in terms of $\mathbf{X}$:

$$\min_{\mathbf{X} \in \mathbb{R}^{2n}} \quad f(\mathbf{X}\mathbf{X}^T)$$

$$\text{s.t.} \quad \underline{P}_k^G - P_k^D \le \langle \mathbf{X}\mathbf{X}^T, \mathbf{Y}_k \rangle \le \overline{P}_k^G - P_k^D, \quad (1a)$$

$$\underline{Q}_k^G - Q_k^D \le \langle \mathbf{X}\mathbf{X}^T, \overline{\mathbf{Y}}_k \rangle \le \overline{Q}_k^G - Q_k^D, \quad (1b)$$

$$\underline{V}_k^2 \le \langle \mathbf{X}\mathbf{X}^T, \mathbf{M}_k \rangle \le \overline{V}_k^2, \quad (1c)$$

$$\forall k \in \mathcal{N},$$

where constraints (1a) and (1b) are the real and reactive power balance equations, respectively. These are *hard* constraints imposed by the laws of physics and cannot be violated. Moreover, for buses without generators, the lower and upper bounds in (1a) and (1b) are equal. Constraint (1c) limits the voltage magnitude at each bus. This is a *soft* constraint imposed by regulation or operator preference. It is physically possible to violate these constraints, and small violations may be tolerated if they are sparse and/or low in magnitude.

In our analysis, we neglect line flow limits to streamline the presentation and mathematical derivation. However, since

such constraints are in the form of (1a), our method can readily handle a more detailed formulation of AC OPF.

B. System Response to Unforecasted Power Injections

We now turn to rigorously formulating an approximate OPF problem based on the analysis in [3]. For the purposes of solving the OPF problem, the complex voltage vector serves as the decision variable as it fully specifies the operating point of the system; that is, given the complex voltage at each bus, the current and power injections can be computed. However, in practice, system operators cannot directly actuate voltage magnitudes and angles at all buses in the system. Instead, they control voltage magnitudes and active power injections at generator buses. Combined with the active and reactive demand from loads, the system naturally resolves to the complex voltage profile obtained as the OPF solution if the forecast is accurate.

We consider a random active power injection vector $\xi \in \mathbb{R}^n$, realized after the OPF decision is made. The random vector ξ represents the forecast error associated with either loads or intermittent renewable energy generators, such as wind turbines or solar panels. For simplicity, we will assume that the active power injection induces a proportional reactive power injection according to a constant power factor $\cos \phi$.

If the system operator leaves non-slack generator setpoints unchanged after the realization of the random variable, then the following variables are held constant:

1) Voltage magnitude and active power injection from generators at the set of generator buses $\mathcal{PV}$.
2) Active and reactive loads at the set of load buses $\mathcal{PQ}$.
3) Voltage magnitude and angle at the slack bus $P\theta$.

Under this response mechanism, the full aggregate active power imbalance induced by ξ is offset by the slack bus. Instead, we assume that an Automatic Generation Control (AGC) scheme is used to distribute the burden among the generators. The imbalance is divided among buses according to participation factors $\alpha \in \mathbb{R}^n$, where $\mathbf{1}_n^T \alpha = 1$ and $\alpha_k = 0$ for all $k \notin \mathcal{PV}$. In summary, the *known* change of the post-contingency system state is given by:

$$
\begin{aligned}
\Delta P_k &= \xi_k - \alpha_k \mathbf{1}_n^T \xi, \quad \forall k \in \mathcal{PQ} \cup \mathcal{PV}, \\
\Delta Q_k &= \gamma \xi_k, \quad \forall k \in \mathcal{PQ}, \\
\Delta |V_k| &= 0, \quad \forall k \in \mathcal{PV} \cup \{P\theta\}, \quad \Delta \theta_{P\theta} = 0,
\end{aligned} \tag{2}
$$

where we use symbol Δ to denote the change of corresponding system state variable, γ is equal to $\sqrt{\cos^{-2} \phi - 1}$ and θ_k is the voltage angle at bus k. After applying these changes, we can determine $2n$ of the post-contingency power flow variables and the other $2n$ variables are determined by solving the power flow equations.

Remark 1. *In [22], the participation factors in α account for the mismatch from forecast errors and the difference in resistive losses induced by the forecast errors. For simplicity, we do not adopt this approach as the difference in losses is small relative to the errors. In practice, the participation factor at the slack bus may be artificially lowered to offset the burden associated with resistive losses.*

Given voltage profile $\mathbf{X} \in \mathbb{R}^{2n}$ and forecast error $\xi \in \mathbb{R}^n$, denote the change in active power injections, reactive power injections and squared voltage magnitude, respectively, as

$$
\Delta P(\mathbf{X}, \xi), \quad \Delta Q(\mathbf{X}, \xi), \quad \Delta |V|^2(\mathbf{X}, \xi).
$$

Hence, the constraints in the OPF problem (1) become

$$
\begin{aligned}
\underline{P}_k^G - P_k^D &\leq \langle \mathbf{X}\mathbf{X}^T, \mathbf{Y}_k \rangle - \Delta P_k(\mathbf{X}, \xi) + \xi \leq \overline{P}_k - P_k^D, \\
\underline{Q}_k^G - Q_k^D &\leq \langle \mathbf{X}\mathbf{X}^T, \overline{\mathbf{Y}}_k \rangle - \Delta Q_k(\mathbf{X}, \xi) + \gamma \xi \leq \overline{Q}_k - Q_k^D, \\
\underline{V}_k^2 &\leq \langle \mathbf{X}\mathbf{X}^T, M_k \rangle + \Delta |V|_k^2(\mathbf{X}, \xi) \leq \overline{V}_k^2, \\
\forall k &\in \mathcal{N}.
\end{aligned} \tag{3}
$$

Notice that we include the forecast error explicitly in the active and reactive power balance equations. This is necessary because the bounds also change with the forecast error. For the notational simplicity, we write the constraints in (3) in the compact form:

$$
\mathcal{A}(\mathbf{X}\mathbf{X}^T) + \Delta(\mathbf{X}, \xi) \leq \mathbf{0}_{6n}, \tag{4}
$$

where $\mathcal{A} : \mathbb{R}^{2n \times 2n} \mapsto \mathbb{R}^{6n}$ is an affine operator, $\Delta : \mathbb{R}^{2n} \times \mathbb{R}^n \mapsto \mathbb{R}^{6n}$ characterizes the response to the random power injections, and the inequality is enforced componentwise.

C. Chance-constrained OPF

The randomness of ξ prohibits the application of most stochastic optimization algorithms to problems that involve constraint (4). As an alternative formulation, the chance constraints provide a practical way to enforce the stochastic constraint and quantify the satisfaction rate. Intuitively, the chance constraint requires that the stochastic constraint of (4) be satisfied with high probability. Let $\mathbb{P}_0(\cdot)$ be the probability with respect to the distribution of ξ and $\epsilon \in (0, 1]$ be the desired maximum violation probability. The *joint chance constraint* is defined as

$$
\mathbb{P}_0 \left[\mathcal{A}(\mathbf{X}\mathbf{X}^T) + \Delta(\mathbf{X}, \xi) \leq \mathbf{0}_{6n} \right] \geq 1 - \epsilon. \tag{5}
$$

Now, we can formulate the joint chance-constrained OPF (CCOPF) problem:

$$
\min_{\mathbf{X} \in \mathbb{R}^{2n}} f(\mathbf{X}\mathbf{X}^T) \quad \text{s.t.} \quad \text{chance constraint (5)}, \tag{6}
$$

The parameter ϵ effectively controls the trade-off between the reliability and efficiency of the solution to problem (6). With a smaller ϵ, the chance constraint becomes more restrictive and the operational cost becomes higher; and vice versa. Another similar chance constraint, named the *disjoint chance constraint*, can be written as

$$
\mathbb{P}_0 \left[\mathcal{A}_k(\mathbf{X}\mathbf{X}^T) + \Delta_k(\mathbf{X}, \xi) \leq 0 \right] \geq 1 - \epsilon, \quad \forall k \in [6n]. \tag{7}
$$

In the main manuscript, we focus on the joint chance constraint and leave the analysis of the disjoint case, as well as their generalizations, to the appendix.

III. REFORMULATIONS OF CCOPF

Although the CCOPF problem (6) is mathematically well-defined, the presence of uncertainty presents two challenges. First, since the change in the power system status is implicitly decided by the active power injection ξ through power flow equations, the function $\Delta(\cdot, \cdot)$ cannot be written in closed form. In Section III-A, we derive a linear approximation of $\Delta(\cdot, \cdot)$ and develop an efficient fixed-point iteration algorithm to find approximate solutions to the original non-linear formulation.

Second, the true distribution of ξ is unknown in most applications. Hence, it is not possible to enforce or verify the chance constraint (5). In Section III-B, we propose DRO-based reformulations of CCOPF, which only rely on historical samples of ξ. We show that the chance constraint is satisfied by the DRO solutions with high probability in terms of the sample complexity.

A. Linearization and Fixed-point Iteration Algorithm

To avoid the computation cost of solving $\Delta(\cdot, \cdot)$ via power flow equations, we construct linear approximations to the implicit function and design an iterative algorithm that converges to a reliable approximation solution in practice. First, we utilize the prior information that the forecast errors are relatively small in practice and approximate $\Delta(\mathbf{X}, \xi)$ with the first-order Taylor expansion around point $\xi = \mathbf{0}_n$. Namely, we have

$$\Delta(\mathbf{X}, \xi) \approx \Delta(\mathbf{X}, \mathbf{0}_n) + D_\Delta(\mathbf{X})\xi = D_\Delta(\mathbf{X})\xi,$$

where $D_\Delta(\mathbf{X}) \in \mathbb{R}^{6n \times n}$ is the Jacobian of $\Delta(\cdot, \cdot)$ with respect to the second input at point $(\mathbf{X}, \mathbf{0}_n)$. Given vector $\mathbf{X}$, the Jacobian can be computed in closed form; see the appendix for the derivation of $D_\Delta(\mathbf{X})$. Then, the approximate joint chance constraint is given by

$$\mathbb{P}_0\left[\mathcal{A}(\mathbf{X}\mathbf{X}^T) + D_\Delta(\mathbf{X})\xi \leq \mathbf{0}_{6n}\right] \geq 1 - \epsilon. \qquad (8)$$

The approximate disjoint chance constraint is defined in a similar way and we focus on the joint chance constraint in the remainder of this subsection. We note that this linearization approach is commonly used in CCOPF literature [3], [22].

Moreover, as proposed in [3], we further decouple the interaction between $\mathbf{X}$ and ξ through the fixed-point iteration. To be more specific, in the t-th iteration of the algorithm, the Jacobian $D_\Delta(\mathbf{X}_t)$ is fixed and we compute the new point $\mathbf{X}_{t+1}$ by solving problem (9). Note that we apply DRO-based algorithms in Section III-B to find solution $\mathbf{X}_{t+1}$ that satisfies the chance constraint with high probability. Then, the Jacobian at point $(\mathbf{X}_{t+1}, \mathbf{0}_n)$ is computed and used in the next iteration.

The pseudo-code of the heuristic algorithm is provided in Algorithm 1. Intuitively, if the initialization is close to the solution, the fixed-point iteration enjoys fast convergence. In most applications, the forecast errors are small relative to the forecasted power injections and thus, our approximation scheme is considerably accurate in the following sense:

1) The first-order approximation $\Delta(\mathbf{X}, \xi) \approx D_\Delta(\mathbf{X})\xi$ is acceptable under a wide range of operating conditions.

Algorithm 1 Fixed-point iteration for joint CCOPF problem.

1: **Input:** tolerance μ, maximum violation probability ϵ.
2: **Output:** robust solution $\mathbf{X}$.
3: **Initialization:**

$$\mathbf{X}_0 \leftarrow \arg\min_{\mathbf{X} \in \mathbb{R}^{2n}} f(\mathbf{X}\mathbf{X}^T) \quad \text{s.t. } \mathcal{A}(\mathbf{X}\mathbf{X}^T) \leq \mathbf{0}_{6n}.$$

4: **for** $t = 0, 1, \ldots$ **do**
5:     Update $\mathbf{X}_{t+1}$ to be the optimizer of

$$\min_{\mathbf{X} \in \mathbb{R}^{2n}} f(\mathbf{X}\mathbf{X}^T) \qquad (9)$$
$$\text{s.t. } \mathbb{P}_0\left[\mathcal{A}(\mathbf{X}\mathbf{X}^T) + D_\Delta(\mathbf{X}_t)\xi \leq \mathbf{0}_{6n}\right] \geq 1 - \epsilon.$$

              $\triangleright$ Solved by DRO-based algorithms.
6:     **If** $\|\mathbf{X}_{t+1} - \mathbf{X}_t\| \leq \mu$, **return** $\mathbf{X}_{t+1}$.
7: **end for**

---

2) The robust solution to the chance-constrained problem is expected to be not too far from to the deterministic solution (i.e., the solution with $\xi = 0$).

As a consequence, although there is no convergence guarantee, the fixed-point iteration exhibits efficient and robust convergence in practice; see the numerical experiments in Section IV and [3]. We also numerically illustrate the approximation quality for benchmark power systems instances in the appendix.

B. Distributionally Robust Optimization Approach

In this subsection, we develop exact reformulations of the approximate chance constraints in Section III-A, including but not limited to the joint chance constraint (8), based on DRO techniques. To preserve the generality of our results, we consider the general objective function and constraint function

$$g(\mathbf{X}) : \mathbb{R}^d \mapsto \mathbb{R}, \quad h(\mathbf{X}, \xi) : \mathbb{R}^d \times \mathbb{R}^n \mapsto \mathbb{R}^m,$$

where random vector $\xi \in \mathbb{R}^n$ obeys the distribution $\mathbb{P}_0$, and integers d and m are the size of input variable $\mathbf{X}$ and the number of constraints, respectively. In this subsection, we consider the optimization problem with stochastic constraints:

$$\min_{\mathbf{X} \in \mathbb{R}^d} g(\mathbf{X}) \quad \text{s.t. } h(\mathbf{X}, \xi) \leq \mathbf{0}_m. \qquad (10)$$

Note that our theory can be extended to the case when the randomness ξ also incurs in the objective function g or the feasible set is a convex subset of $\mathbb{R}^d$. We focus on the simpler problem (10) since our target is to solve the CCOPF problem (9). We make the following assumption:

Assumption 1. *The support of $\mathbb{P}_0$ belongs to a compact set $\Xi \subset \mathbb{R}^n$. Both functions $g(\cdot)$ and $h(\cdot, \cdot)$ are continuous. In addition, for every positive integer S and all realizations $\xi_1, \ldots, \xi_S \in \mathbb{R}^n$, problem*

$$\min_{\mathbf{X} \in \mathbb{R}^d} g(\mathbf{X}) \quad \text{s.t. } h(\mathbf{X}, \xi_j) \leq \mathbf{0}_m, \ \forall j \in [S]$$

is feasible and has a finite optimal value.

In our formulation of the CCOPF problem (9), Assumption 1 is satisfied unless, for instance, the reserve capacity of the conventional generators is insufficient to compensate for some

realizations of the forecast error. In this case, CCOPF would be infeasible.

In practice, the true distribution $\mathbb{P}_0$ is unknown and only limited historical samples may be available. Suppose that there are S independently and identically distributed samples, $\xi^1, \ldots, \xi^S$, generated from the distribution $\mathbb{P}_0$. We define the empirical distribution of ξ as

$$\hat{\mathbb{P}}_S := \frac{1}{S} \sum_{k \in [S]} \delta_{\xi^k},$$

where δ_ξ is the Dirac measure at ξ. The goal of the DRO approach is to find robust solutions that satisfy the chance constraint with high probability using empirical distribution $\hat{\mathbb{P}}_S$. Define the ambiguity set

$$\mathcal{D}_r(\mathbb{P}) := \{\mathbb{P}' \in \mathcal{P} \mid I(\mathbb{P}, \mathbb{P}') \leq r\}, \quad \forall \mathbb{P} \in \mathcal{P},$$

where $I(\cdot, \cdot)$ is the relative entropy [26], $r > 0$ is the radius and $\mathcal{P}$ is the family of Borel distributions with support in Ξ. The robustness of DRO solutions is guaranteed by the satisfaction of chance constraints under all distributions in the ambiguity set $\mathcal{D}_r(\hat{\mathbb{P}}_S)$. Other distributional metrics, such as the Wasserstein metric, are considered in CCOPF literature [21], [22]. In this work, however, we use the relative entropy due to the strong optimality guarantees it can provide; see Theorems 3-4 and [14], [26]. Intuitively, the large deviation theory guarantees that the relative entropy between the true data-generation distribution and the empirical distribution can be bounded by a value that depends on the sample size [26]. Hence, the true distribution is contained in the ambiguity set with high probability and the relative entropy-based ambiguity set is the "smallest" ambiguity set with such property [14].

For problem (10), the joint chance constraint is given by

$$\mathbb{P}_0[h(\mathbf{X}, \xi) \leq \mathbf{0}_m] \geq 1 - \epsilon. \tag{11}$$

Remark 2. *More generally, our results can be extended to the case when the joint constraints are defined by a convex cone*

$$\mathbb{P}_0[\omega^T h(\mathbf{X}, \xi) \leq 0, \quad \forall \omega \in \mathcal{W}] \geq 1 - \epsilon, \tag{12}$$

where $\mathcal{W}$ is the convex cone spanned by weight vectors[1] $\omega_1, \ldots, \omega_L$. Constraint (12) reduces to the cardinal case (11) when $L = m$ and $\omega_\ell = \mathbf{e}_\ell$ for all $\ell \in [m]$.

Define the α-quantile

$$q_\alpha(F, \mathbb{P}) := \sup\{q \mid \mathbb{P}[F(\xi) \leq q] \leq \alpha\}$$

for all $\alpha \in [0, 1]$, function $F(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}$ and distribution $\mathbb{P} \in \mathcal{P}$. Then, the chance constraint (11) can be equivalently written as

$$\begin{aligned} &\mathbb{P}_0[h_\ell(\mathbf{X}, \xi) \leq 0, \quad \forall \ell \in [m]] \geq 1 - \epsilon \\ \iff &\mathbb{P}_0[\bar{h}_\mathbf{X}(\xi) \leq 0] \geq 1 - \epsilon \\ \iff &q_{1-\epsilon}(\bar{h}_\mathbf{X}, \mathbb{P}_0) \leq 0, \end{aligned} \tag{13}$$

where we define

$$\bar{h}_\mathbf{X}(\xi) := \max_{\ell \in [m]} h_\ell(\mathbf{X}, \xi).$$

[1] A vector $\omega \in \mathbb{R}^m$ is called a weight vector if $\omega \geq \mathbf{0}_m$ and $\mathbf{1}_m^T \omega = 1$.

Adopting language from [14], we first introduce the distributionally robust predictor of the α quantile.

Definition 1 (Distributionally Robust Predictor). *For all $\epsilon \in [0, 1]$, $r > 0$, $\mathbf{X} \in \mathbb{R}^d$ and $\mathbb{P} \in \mathcal{P}$, the distributionally robust predictor is defined as*

$$\hat{q}_{1-\epsilon, r, \mathbb{P}}(\mathbf{X}) := \sup_{\mathbb{P}' \in \mathcal{D}_r(\mathbb{P})} q_{1-\epsilon}(\bar{h}_\mathbf{X}, \mathbb{P}').$$

For notational simplicity, when there is no confusion about $\hat{\mathbb{P}}_S$, we denote predictor $\hat{q}_{1-\epsilon, r, \hat{\mathbb{P}}_S}$ as $\hat{q}_{1-\epsilon, r, S}$.

Intuitively, the distributionally robust predictor is the *worst-case α-quantile* over all distributions in the relative entropy ball $\mathcal{D}_r(\mathbb{P})$. In the following lemma, we prove that the distributionally robust predictor is either a quantile of $\bar{h}_\mathbf{X}$ under the empirical distribution $\hat{\mathbb{P}}_S$ or the maximum value

$$h_\mathbf{X}^* := \max_{\xi \in \Xi} \bar{h}_\mathbf{X}(\xi).$$

Lemma 1. *For all $\epsilon \in [0, 1]$ and $r > 0$, there exists an integer $k(\epsilon, r, S) \in [S + 1]$ such that*

$$\hat{q}_{1-\epsilon, r, S}(\mathbf{X}) = \bar{h}_{k(\epsilon, r, S), \hat{\mathbb{P}}_S}(\mathbf{X}), \quad \forall \mathbf{X} \in \mathbb{R}^d,$$

where $\bar{h}_{k, \hat{\mathbb{P}}_S}(\mathbf{X})$ is the k-th smallest value in $\{\bar{h}_\mathbf{X}(\xi^j), j \in [S]\} \cup \{h_\mathbf{X}^*\}$. When there is no confusion, we denote for the notational simplicity

$$k := k(\epsilon, r, S) \quad and \quad \bar{h}_k(\mathbf{X}) := \bar{h}_{k(\epsilon, r, S), \hat{\mathbb{P}}_S}(\mathbf{X}).$$

In the case when $k = S + 1$, the evaluation of $\bar{h}_{S+1}(\mathbf{X})$ requires the knowledge of $h_\mathbf{X}^*$, which may be unknown in practice. Hence, we focus on the case when $k \in [S]$, which can be guaranteed by choosing suitable values of ϵ and r. Furthermore, the value of k can be computed by solving a convex optimization problem; see problem (28) in the appendix.

Then, we define the distributionally robust prescriptor.

Definition 2 (Distributionally Robust Prescriptor). *For all $\epsilon \in [0, 1]$ and $r > 0$, the distributionally robust prescriptor $\hat{\mathbf{X}}_{1-\epsilon, r, \mathbb{P}}$ is a quasi-continuous function of $\mathbb{P}$ that solves*

$$\min_{\mathbf{X} \in \mathbb{R}^d} g(\mathbf{X}) \quad \text{s.t.} \quad \hat{q}_{1-\epsilon, r, \mathbb{P}}(\mathbf{X}) \leq 0. \tag{14}$$

Similarly, when there is no confusion about $\hat{\mathbb{P}}_S$, we denote prescriptor $\hat{\mathbf{X}}_{1-\epsilon, r, \hat{\mathbb{P}}_S}$ as $\hat{\mathbf{X}}_{1-\epsilon, r, S}$.

By Lemma 1, the feasible set of problem (14) is a subset of $\{\mathbf{X} \in \mathbb{R}^d \mid \bar{h}_k(\mathbf{X}) \leq 0\}$. Thus, combining with Assumption 1, problem (14) has a finite optimal value and the distributionally robust prescriptor $\hat{\mathbf{X}}_{1-\epsilon, r, S}$ is well-defined.

Now, we provide a mixed-integer reformulation of (14) to compute the distributionally robust prescriptor. Choosing $C > 0$ to be a sufficiently large constant, we show that the distributionally robust prescriptor is a solution to

$$\min_{\mathbf{X} \in \mathbb{R}^d, \mathbf{b} \in \mathbb{Z}^S} g(\mathbf{X}) \tag{15}$$

$$\text{s.t.} \quad h_\ell(\mathbf{X}, \xi^j) \leq C\mathbf{b}_j, \quad \forall \ell \in [m], \ j \in [S],$$

$$\mathbf{1}_S^T \mathbf{b} \leq S - k, \quad \mathbf{b}_j \in \{0, 1\}, \quad \forall j \in [S].$$

Intuitively, the constraints in (15) enforce the joint chance constraint under the empirical distribution $\hat{\mathbb{P}}_S$. Namely, the

constraint $h(\mathbf{X}, \xi^j) \leq \mathbf{0}_m$ is satisfied by at least k samples. In the next theorems, we show that the chance constraint under the true distribution $\mathbb{P}_0$ can also be guaranteed by choosing k to be slightly larger than $(1-\epsilon)S$.

Theorem 2. *The solution to (15) is a distributionally robust prescriptor.*

For the CCOPF problem (9), the problem (15) is equivalent to a mixed-integer QCQP. In Section III-C, we apply the semi-definite relaxation to the QCQP and when the relaxation is exact, problem (9) is equivalent to a mixed-integer semi-definite program (MISDP), which can be handled by off-the-shelf convex optimization solvers.

Finally, we establish the theoretical properties of the distributionally robust prescriptor. First, we prove that the distributionally robust prescriptor satisfies joint chance constraint (13) with high probability in terms of the sample complexity S.

Theorem 3. *For all $\epsilon \in [0,1]$ and $r > 0$, it holds that*

$$\mathbb{P}_\infty \left[q_{1-\epsilon} \left(\bar{h}_{\hat{\mathbf{X}}_{1-\epsilon,r,S}}, \mathbb{P}_0 \right) \leq 0 \right] \geq 1 - \exp\left[-rS + o(S)\right], \tag{16}$$

where $\mathbb{P}_\infty$ is the probability measure of the sample path space of ξ under distribution $\mathbb{P}_0$. Furthermore, we have

$$\mathbb{P}_\infty \left[h_\ell \left(\hat{\mathbf{X}}_{1-\epsilon,r,S}, \xi \right) \leq 0, \quad \forall \ell \in [m] \right] \tag{17}$$
$$\geq 1 - \epsilon - \exp\left[-rS + o(S)\right].$$

In the regime when the support Ξ is a finite set, we can apply the strong large deviation principle [14] and derive the following finite-sample bound in the same way as Theorem 3:

$$\mathbb{P}_\infty \left[q_{1-\epsilon} \left(\bar{h}_{\hat{\mathbf{X}}_{1-\epsilon,r,S}}, \mathbb{P}_0 \right) \leq 0 \right] \geq 1 - (S+1)^d e^{-rS}. \tag{18}$$

Moreover, we show that the distributionally robust prescriptor achieves the minimum operational cost over all decisions that asymptotically satisfy the joint chance constraint (13).

Theorem 4. *Suppose that prescriptor $\tilde{\mathbf{X}}_{1-\epsilon,r,\mathbb{P}} \in \mathbb{R}^d$ is a quasi-continuous function of $\mathbb{P}$ and satisfies constraint (16). Then, we have*

$$\mathbb{P}_\infty \left[g\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S} \right) < g\left(\hat{\mathbf{X}}_{1-\epsilon,r,S} \right) \right] = 0,$$

where we denote $\tilde{\mathbf{X}}_{1-\epsilon,r,S} := \tilde{\mathbf{X}}_{1-\epsilon,r,\hat{\mathbb{P}}_S}$.

Compared with existing DRO formulations [21], [22], our formulation provides stronger guarantees in the following two senses. First, the DRO solution $\hat{\mathbf{X}}_{1-\epsilon,r,S}$ achieves the minimum possible generation cost over all robust solutions that satisfy the joint chance constraint (16). This optimality property arises from the choice of the relative entropy for the ambiguity set, and such property cannot be established by other distributional metrics, although the Wasserstein metric can provide similar high-probability bounds [13]. Second, the mixed-integer reformulation (15) is *exact*. In contrast, existing literature considered parameterized approximations to the ambiguity set, such as the hyper-rectangle [21] and the ellipsoid [22]. In practice, however, there is no guarantee that the ambiguity set is of the specified shape and thus, the

approximate DRO solution is usually overly conservative; see the comparison results in Section IV.

In practice, it is preferable for the user to first specify k and then compute the optimal ϵ and r to maximize the right-hand side of (17). Given $k \in [S]$ and $\epsilon \in [1 - k/S, 1]$, the maximal radius r such that $k(\epsilon, r, S) = k$ is given by

$$r = -\frac{k}{S} \log\left(\frac{S(1-\epsilon)}{k} \right) - \frac{S-k}{S} \log\left(\frac{S\epsilon}{S-k} \right),$$

where we define $0 \log 0 = 0$. Therefore, when the sample size S is sufficiently large, we ignore the $o(S)$ term on the right-hand side of (17) and solve the maximization problem

$$\epsilon^*_{k,S} := \underset{\epsilon \in [1-k/S,1]}{\arg\max} \; 1 - \epsilon - \frac{S^S}{k^k(S-k)^{S-k}}(1-\epsilon)^k \epsilon^{S-k}, \tag{19}$$

where we define $0^0 = 1$. The solution to the above problem maximizes the right-hand side of (17) and can be found by the bi-section algorithm.

C. Semi-definite Relaxation

In the last part of this section, we deal with the non-convexity of problem (9) induced by the quadratic parameterization $\mathbf{X}\mathbf{X}^T$. In the context of CCOPF problem, even with a fixed integer vector $\mathbf{b}$, problem (15) is still a non-convex QCQP and can be $\mathcal{NP}$-hard to solve in the worst case. To achieve the efficient and reliable operation of large-scale power systems, various techniques have been proposed to reduce the optimization complexity by utilizing the special structures of real-world power circuits. In literature, the semi-definite relaxation is widely applied to transform the non-convex QCQP to a semi-definite program (SDP); see [25] and [24] for semi-definite relaxations of OPF. More specifically, after making the change-of-variables $\mathbf{W} := \mathbf{X}\mathbf{X}^T$ and dropping the rank constraint $\mathrm{rank}(W) = 1$, we can apply the distrubutionally robust reformulation (15) to solve problem (9) by

$$\min_{\mathbf{W} \in \mathbb{S}^{2n}_+, \mathbf{b} \in \mathbb{Z}^S} \quad f(\mathbf{W}) \tag{20}$$
$$\text{s.t. } \mathcal{A}(\mathbf{W}) + D_\Delta(\mathbf{X}_t)\xi^j \leq C\mathbf{b}_j \cdot \mathbf{1}_{6n}, \quad \forall j \in [S],$$
$$\mathbf{1}_S^T \mathbf{b} \leq S - k, \quad \mathbf{b}_j \in \{0,1\}, \quad \forall j \in [S].$$

Note that the SDP part of (20) can be further written in the standard form using the Schur complement [25]. We denote problem (20) as the distributionally robust CCOPF (DRC-COPF) problem. With the rank constraint dropped, problem (20) is a MISDP, which can be solved efficiently by various solvers (e.g., YALMIP [27]).

The graphical structures of practical power networks, reflected in the algebraic properties of operator $\mathcal{A}$, guarantee the exactness of the semi-definite relaxation [24], [28], [29]. In this work, we make the assumption that the network admits an exact semi-definite relaxation, and our method can be readily extended to other formulations including DC OPF. Under the exact relaxation assumption, we are able to recover the rank-1 solution from the MISDP solution [25]. Therefore, in Step 5 of Algorithm 1, we first solve the relaxation (20) to find $\mathbf{W}_{t+1}$ and then generate the rank-1 solution $\mathbf{X}_{t+1}$ for the next iteration.

To further reduce the computational complexity of solving problem (20), we develop a heuristic algorithm that finds approximate solutions by solving a small number of SDPs. Intuitively, the proposed algorithm searches for the optimal integer vector $\mathbf{b}$ in a greedy way and avoids the mixed-integer part in problem (20). With a given k, the algorithm removes the "most restrictive sample" among the $k+1$ samples selected in the case of $k+1$. The heuristic algorithm is able to find nearly optimal approximate solutions for benchmark power systems and requires a much shorter running time; see Section IV and the appendix for more details.

IV. DEMONSTRATION ON IEEE TEST CASES

In this section, we apply our results to solve the joint CCOPF problem (6) on the IEEE 14- and 118-bus test cases. All system parameters are taken from the case data in MATPOWER [30], [31]. We make the following modifications to the system parameters:

1) As suggested in [25], a small resistance of 10^{-4} per-unit is added to each transformer to insure an exact semi-definite relaxation.
2) Wind generators are installed at n_W buses randomly chosen from load buses with nonzero active loads. The forecasted output of a generator is equal to a proportion η of the pre-installation load at that bus.

Note that wind generators could also have been installed at slack or generator buses. In our model, wind generator forecast errors account for the entire random power injection; that is, we assume loads are deterministic. Wind output forecast errors are taken from hour-ahead forecast errors from the National Renewable Energy Laboratory's Wind Integration National Dataset, which contains simulated forecast and output data for over 120,000 sites in the United States over seven years [32]. Specifically, each load bus chosen for a wind generator is assigned to a randomly selected site in Alameda County, California. The forecast errors are then scaled appropriately, assuming that the forecasted output is half of the installed capacity of the turbine. Turbines are selected from a single county and thus, their outputs are correlated; exploiting the correlations between random variables in different constraints is an important feature of joint chance-constrained methods.

For each network, we implement Algorithm 1 and use the DRO formulation (20) to solve (9). We use $\hat{S}$ training samples drawn independently from the full path of S samples. For comparison, we also solve (9) using the method from [22], which approximates a Wasserstein metric-based ambiguity set using a minimum-volume ellipsoid in the parameter space. To the best of our knowledge, this is the least conservative approximation of a metric-based ambiguity set for joint chance constraints in the literature (with ours being the first exact reformulation). The MISDP solver, the greedy algorithm and the method from [22] is called DRCCOPF-KL, DRCCOPF-G and DRCCOPF-E, respectively. For benchmarking, we also compare with the robust optimization (RO) approach, where all constraints satisfied for all S available samples, and the deterministic OPF approach, where each wind generator simply outputs its forecasted value. As the forecast errors
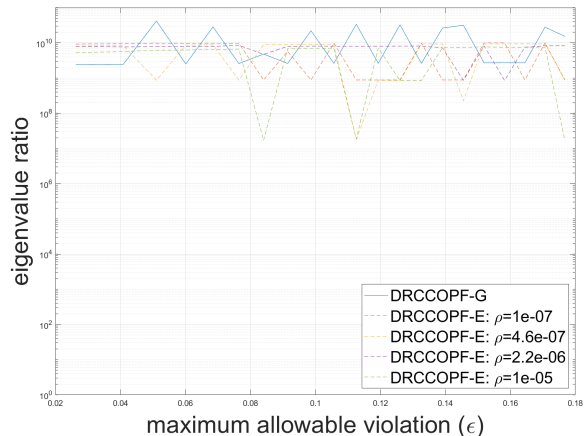

Fig. 1. Ratio between the second and third largest eigenvalues for IEEE 118-bus system.

are approximately zero-mean in practice, deterministic OPF essentially enforces the constraints in expectation. The RO approach gives the most conservative solution.

We simulate the IEEE 14- and 118-bus systems with parameters $\eta = 0.9$, $\cos\phi = 1$, and $n_W = 6$ for the 14-bus system and $n_W = 45$ for the 118-bus system. We generate $S = 2185$ samples and $\hat{S} = 200$ samples are used as training samples. Our hour-ahead forecast errors were taken from June 1 to August 31, 2007. To evaluate the performance under different robustness requirements, we performed a sweep over k from 180 to 200. Recall that for a given k and $\hat{S}$, the optimal ϵ can be computed from (19).

All simulations are performed in MATLAB 2023b. The MISDP problem (20) is solved using YALMIP [27] and all convex problems are solved using CVX [33]. We note that all simulation results in this section describe the performance of the voltage setpoint recovered from the solution to the MISDP (20) *after* applying Corollary 1 of [25].

A. Solution Recovery

First, we analyze the exactness of the semi-definite relaxation in Section III-C. Molzahn *et al.* [34] proposed the ratio between the second and third largest eigenvalues of the solution to the relaxed problem as a metric for exactness. If the eigenvalue ratio is high, the true power injections and voltage magnitudes at each bus will be close to those computed from the solution to the relaxed problem. We compute the eigenvalue ratio of solutions generated by DRCCOPF-G and DRCCOPF-E for the 118-bus system. As seen in Figure 1, both approaches produce solutions with eigenvalue ratios higher than, in the logarithmic sense, to the benchmark value of 10^7 from [34]. However, DRCCOPF-G often leads to slightly higher-quality solutions with ratios closer to 10^9 or $10^{1}0$. This is a promising feature of the method proposed here.

B. Efficiency and Robustness

Now, we compare the generation cost (efficiency) of different solutions under the same maximum allowable violation rate (robustness). For the DRCCOPF-G and DRCCOPF-KL approach, the violation rate can be effectively controlled by
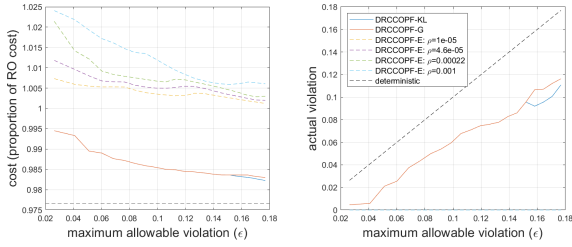
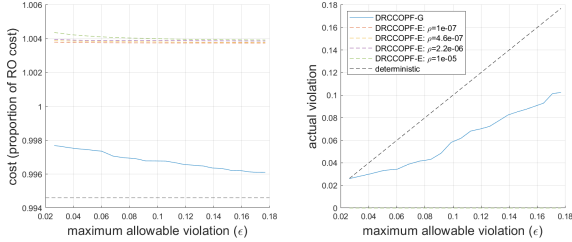Fig. 2. Performance comparison for 14-bus system.

Fig. 3. Performance comparison for 118-bus system.

Fig. 4. Distribution of selected generator outputs for DRCCOPF-G and DRCCOPF-E for the 14-bus system.

parameter k, while the DRCCOPF-E approach controls the robustness by parameters ϵ and ρ. We use the RO approach as the benchmark, which corresponds to the most conservative setting, and other costs are given as a proportion of the RO cost for reference. As another benchmark, we also illustrate the cost of deterministic OPF approach. All reasonable robust methods should always have higher costs than the deterministic approach, which, as a result, represents a lower bound on the achievable efficiency.

The results of the 14- and 118-bus systems are plotted in Figures 2 and 3, respectively. The left subplots compare the cost of the solution of DRCCOPF-G with that of DRCCOPF-E for several different Wasserstein radii ρ. The right subplots give the realized constraint violation rates for both methods on all of the S available samples. Note that the deterministic OPF method violates the constraints at a rate of more than 97%. This is a result of *joint* chance constraints, since violating even a single constraint constitutes a violation. Notice that DRCCOPF-E never violates the joint chance-constraint for any radius. For the 14-bus case, we also compare with the solutions of DRCCOPF-KL. We can observe that the solutions of DRCCOPF-KL and DRCCOPF-G exhibit very similar behaviors, which imply that the greedy algorithm finds nearly optimal solutions. Therefore, we focus on the DRCCOPF-G approach in the following discussion.

First, DRCCOPF-E is qualitatively more conservative than DRCCOPF-G in the sense that it fails to exploit the tolerance provided by non-zero values of ϵ to achieve lower objective values. In fact, for small ϵ and large ρ, DRCCOPF-E is even more inefficient than RO as a consequence of its approximation of the ambiguity set. By contract, distributionally robust methods are generally less conservative than RO. As shown in our results, the DRCCOPF-G approach closes a significant fraction of the gap between the robust and deterministic optimizations by approaching but never exceeding the prescribed maximum violation rate.

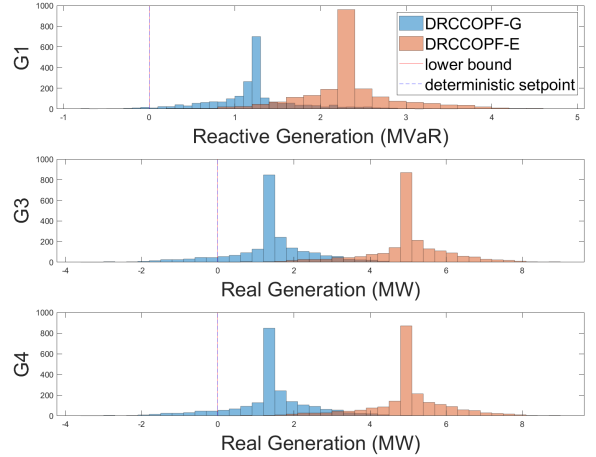Moreover, the running time of DRCCOPF-G is comparable with that of DRCCOPF-E. For the 118-bus system, DRCCOPF-G takes an average of 412 seconds to find the solution using a single 3.79-GHz CPU, while the DRCCOPF-E finishes in 235 seconds on average. However, the greedy search structure of DRCCOPF-G approach allows the application of parallel computing techniques, which will significantly improve the computational efficiency of the algorithm.

Furthermore, the DRCCOPF-G approach always generates a feasible solution as long as the deterministic problem is feasible for every sample (that is, as long as Assumption 1 is satisfied). This is because DRCCOPF-G basically requires satisfaction of the constraints for a subset of k samples. DRCCOPF-E and other existing approximate methods, by contrast, compute uncertainty margins indirectly using training samples, possibly rendering the problem infeasible. Indeed, we have observed that DRCCOPF-E is only feasible for a limited range of parameters and fail to find a solution for relatively large values of η (more renewable penetration) and small values of ϵ.

C. Demonstration on Selected Constraints

Finally, we select a few constraints to highlight the difference in performance. We focus on the generator outputs of three selected buses in the 14-bus system and plot the distribution of post-contingency generator outputs of DRCCOPF-G and DRCCOPF-E solutions. The results are shown in Figure 4, where the deterministic setpoint and lower bound are included for reference. Since the upper bound is not violated, it is not included in the histograms. From the results, we can see that DRCCOPF-G is significantly closer to the lower bound for real or reactive power for three selected generators, while DRCCOPF-E produces overly conservative power outputs.

V. CONCLUSION

In this work, we focus on the distributionally robust approach for the CCOPF problem. We propose a new DRO formulation based on the relative entropy, which achieves the optimal generation cost given the maximum violation rate. In addition, we provide an exact reformulation of the joint chance constraint, which guarantees the feasibility of the reformulated problem and leads to significantly better efficiency compared

with existing approaches based on inner approximation. Finally, numerical results on IEEE benchmark power systems are exhibited to show the superior performance of our approach compared to existing state-of-the-art approaches.

REFERENCES

[1] R. A. Jabr, S. Karaki, and J. A. Korbane, "Robust Multi-Period OPF With Storage and Renewables," IEEE Transactions on Power Systems, vol. 30, no. 5, pp. 2790–2799, Sep. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/6948280/

[2] R. Louca and E. Bitar, "Robust AC Optimal Power Flow," IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 1669–1681, May 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8392424/

[3] L. Roald and G. Andersson, "Chance-Constrained AC Optimal Power Flow: Reformulations and Efficient Algorithms," IEEE Transactions on Power Systems, vol. 33, no. 3, pp. 2906–2918, 2018.

[4] A. Venzke, L. Halilbasic, U. Markovic, G. Hug, and S. Chatzivasileiadis, "Convex Relaxations of Chance Constrained AC Optimal Power Flow," IEEE Transactions on Power Systems, vol. 33, no. 3, pp. 2829–2841, May 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8060613/

[5] 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, Aug. 2009. [Online]. Available: http://link.springer.com/10.1007/s10957-009-9523-6

[6] M. C. Campi, S. Garatti, and M. Prandini, "The scenario approach for systems and control design," Annual Reviews in Control, vol. 33, no. 2, pp. 149–157, Dec. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1367578809000479

[7] H. Rahimian and S. Mehrotra, "Distributionally Robust Optimization: A Review," 2019, publisher: arXiv Version Number: 1. [Online]. Available: https://arxiv.org/abs/1908.05659

[8] M. Lubin, Y. Dvorkin, and S. Backhaus, "A Robust Approach to Chance Constrained Optimal Power Flow With Renewable Generation," IEEE Transactions on Power Systems, vol. 31, no. 5, pp. 3840–3849, Sep. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7332992/

[9] Y. Zhang, S. Shen, and J. Mathieu, "Distributionally Robust Chance-Constrained Optimal Power Flow with Uncertain Renewables and Uncertain Reserves Provided by Loads," IEEE Transactions on Power Systems, pp. 1–1, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7478165/

[10] W. Wei, F. Liu, and S. Mei, "Distributionally Robust Co-Optimization of Energy and Reserve Dispatch," IEEE Transactions on Sustainable Energy, vol. 7, no. 1, pp. 289–300, Jan. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7323859/

[11] W. Xie and S. Ahmed, "Distributionally Robust Chance Constrained Optimal Power Flow with Renewables: A Conic Reformulation," IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 1860–1867, Mar. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7973099/

[12] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally Robust Chance-Constrained Optimal Power Flow Assuming Unimodal Distributions With Misspecified Modes," IEEE Transactions on Control of Network Systems, vol. 6, no. 3, pp. 1223–1234, Sep. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8771187/

[13] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," Mathematical Programming, vol. 171, no. 1-2, pp. 115–166, Sep. 2018. [Online]. Available: http://link.springer.com/10.1007/s10107-017-1172-1

[14] B. P. G. Van Parys, P. M. Esfahani, and D. Kuhn, "From Data to Decisions: Distributionally Robust Optimization Is Optimal," Management Science, vol. 67, no. 6, pp. 3387–3402, Jun. 2021. [Online]. Available: https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3678

[15] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally Robust Chance-Constrained Approximate AC-OPF With Wasserstein Metric," IEEE Transactions on Power Systems, vol. 33, no. 5, pp. 4924–4936, Sep. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8294298/

[16] A. Zhou, M. Yang, M. Wang, and Y. Zhang, "A Linear Programming Approximation of Distributionally Robust Chance-Constrained Dispatch With Wasserstein Distance," IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 3366–3377, Sep. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9026959/

[17] K. Baker and A. Bernstein, "Joint Chance Constraints in AC Optimal Power Flow: Improving Bounds Through Learning," IEEE Transactions on Smart Grid, vol. 10, no. 6, pp. 6376–6385, Nov. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8662704/

[18] L. Yang, Y. Xu, H. Sun, and W. Wu, "Tractable Convex Approximations for Distributionally Robust Joint Chance-Constrained Optimal Power Flow Under Uncertainty," IEEE Transactions on Power Systems, vol. 37, no. 3, pp. 1927–1941, May 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9548838/

[19] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers, "Data-Based Distributionally Robust Stochastic Optimal Power Flow—Part II: Case Studies," IEEE Transactions on Power Systems, vol. 34, no. 2, pp. 1493–1503, Mar. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8513855/

[20] ——, "Data-Based Distributionally Robust Stochastic Optimal Power Flow—Part I: Methodologies," IEEE Transactions on Power Systems, vol. 34, no. 2, pp. 1483–1492, Mar. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8513852/

[21] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein Distributionally Robust Look-Ahead Economic Dispatch," IEEE Transactions on Power Systems, vol. 36, no. 3, pp. 2010–2022, May 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9242289/

[22] A. Arab and J. E. Tate, "Distributionally Robust Optimal Power Flow via Ellipsoidal Approximation," IEEE Transactions on Power Systems, 2022, publisher: IEEE.

[23] E. Brock, H. Zhang, J. Mulvaney Kemp, J. Lavaei, and S. Sojoudi, "Distributionally Robust Optimization for Nonconvex QCQPs with Stochastic Constraints," in 2023 62th IEEE Conference on Decision and Control (CDC). IEEE, 2023, pp. 1–7.

[24] S. H. Low, "Convex Relaxation of Optimal Power Flow—Part I: Formulations and Equivalence," IEEE Transactions on Control of Network Systems, vol. 1, no. 1, pp. 15–27, Mar. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6756976/

[25] J. Lavaei and S. H. Low, "Zero Duality Gap in Optimal Power Flow Problem," IEEE Transactions on Power Systems, vol. 27, no. 1, pp. 92–107, Feb. 2012. [Online]. Available: http://ieeexplore.ieee.org/document/5971792/

[26] T. M. Cover, Elements of information theory. John Wiley & Sons, 1999.

[27] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," in 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508). Taipei, Taiwan: IEEE, 2004, pp. 284–289. [Online]. Available: http://ieeexplore.ieee.org/document/1393890/

[28] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," in 2012 IEEE Power and Energy Society General Meeting. IEEE, Jul. 2012.

[29] ——, "Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure," SIAM Journal on Optimization, vol. 24, no. 4, pp. 1746–1778, 2014, publisher: SIAM.

[30] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 12–19, Feb. 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5491276/

[31] R. D. Zimmerman and C. E. Murillo-Sánchez, "MATPOWER," Oct. 2020, language: en. [Online]. Available: https://zenodo.org/record/4074135

[32] B.-M. Hodge, "Final Report on the Creation of the Wind Integration National Dataset (WIND) Toolkit and API: October 1, 2013 - September 30, 2015," 2016.

[33] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," Mar. 2014. [Online]. Available: http://cvxr.com/cvx

[34] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco, "Implementation of a Large-Scale Optimal Power Flow Solver Based on Semidefinite Programming," IEEE Transactions on Power Systems, vol. 28, no. 4, pp. 3987–3998, Nov. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6510541/

[35] L. A. Wolsey and G. L. Nemhauser, Integer and combinatorial optimization. John Wiley & Sons, 1999, vol. 55.

[36] T. Neubrunn, "Quasi-continuity," Real Analysis Exchange, vol. 14, no. 2, pp. 259–306, 1988.

[37] Y. V. Prokhorov, "Convergence of Random Processes and Limit Theorems in Probability Theory," Theory of Probability & Its Applications, vol. 1, no. 2, pp. 157–214, Jan. 1956. [Online]. Available: http://epubs.siam.org/doi/10.1137/1101016

[38] P. M. Vaidya, "A new algorithm for minimizing convex functions over convex sets," Mathematical programming, vol. 73, no. 3, pp. 291–341, 1996, publisher: Springer.

APPENDIX

A. Heuristic Algorithm for MISDP

In this section, we describe a heuristic algorithm for the MISDP problem (20). Comparing with the internal algorithm of the YALMIP solver, the proposed algorithm achieves a much better computational complexity and is able to find solutions of the same quality, in terms of the objective function value and the constraint satisfaction rate.

Intuitively, the heuristic algorithm searches for the optimal integer vector $\mathbf{b}$ in a greedy way by gradually reducing the value of k. The algorithm starts with the most conservative case when $k = S$. In this case, the only feasible vector $\mathbf{b}$ is the zero vector $\mathbf{0}_S$ and problem (20) reduces to a SDP problem, which can be efficiently solved by a variety of optimization solvers. Then, for each integer $k_0 < S$, the algorithm searches for the optimal $\mathbf{b}$ for the case $k = k_0$ from that for the case $k = k_0 + 1$. More specifically, suppose that we have obtained an approximate solution $(\mathbf{W}^{k_0+1}, \mathbf{b}^{k_0+1})$ for the case when $k = k_0 + 1$. For each index $\ell \in [S]$ such that $\mathbf{b}_\ell^{k_0+1} = 0$, we construct the vector $\mathbf{b}^{k_0,\ell} \in \mathbb{R}^S$ by

$$\mathbf{b}_\ell^{k_0,\ell} = 1, \quad \mathbf{b}_j^{k_0,\ell} = \mathbf{b}_j^{k_0+1}, \quad \forall j \in [S]\backslash\{\ell\}, \quad (21)$$

and we solve the following SDP problem:

$$\min_{\mathbf{W} \in \mathbb{S}_+^{2n}} \quad f(\mathbf{W}) \quad (22)$$
$$\text{s.t.} \quad \mathcal{A}(\mathbf{W}) + D_\Delta \xi^j \le C \mathbf{b}_j^{k_0,\ell} \cdot \mathbf{1}_{6n}, \quad \forall j \in [S].$$

Let f_ℓ^* be the optimal objective function value of the above SDP problem. Then, the greedy algorithm chooses the most restrictive sample ξ^{ℓ^*} by

$$\ell^* := \arg\min_{\ell \in [S]} f_\ell^*, \quad \text{s.t.} \ \mathbf{b}_\ell^{k_0+1} = 0.$$

After removing the most restrictive sample, the approximate solution for the case $k = k_0$ is given by $(\mathbf{W}^{k_0}, \mathbf{b}^{k_0,\ell^*})$, where $\mathbf{W}^{k_0}$ is the solution to problem (22) with $\ell = \ell^*$. To further reduce the computational cost, we only need to consider indices $\ell \in [S]$ such that $\mathbf{b}_\ell^{k_0+1} = 0$ and problem (20) has active constraints with sample ξ^ℓ. Namely, the following componentwise inequality holds with equality for some components:

$$\mathcal{A}(\mathbf{W}^{k_0+1}) + D_\Delta \xi^\ell \le C \mathbf{b}_\ell^{k_0+1} \cdot \mathbf{1}_{6n}. \quad (23)$$

This is because otherwise if the above inequality holds strictly, the optimal objective value will be the same with sample ξ^ℓ (i.e., $\mathbf{b} = \mathbf{b}^{k_0+1}$) and without sample ξ^ℓ (i.e., $\mathbf{b} = \mathbf{b}^{k_0,\ell}$). Therefore, the value f_ℓ^* will not be the minimum among all choices of ℓ. The pesudo-code of the greedy is provided in Algorithm 2. We note that Algorithm 2 operates in the "top-down" style in the sense that it gradually decreases the value of k from S. Similarly, we can develop the "bottom-up" version of the greedy algorithm, which gradually increases the value of k from 0. In practice, the top-down algorithm is preferred since a large value of k is usually chosen to ensure a high constraint satisfaction rate. Therefore, the top-down algorithm requires fewer iterations to reach the targeted value of k.

Since the greedy algorithm approximates the solution to MISDP (20) with a small number of SDP problems, the

Algorithm 2 Greedy algorithm for MISDP problem (20).

1: **Input:** samples $\xi^1, \ldots, \xi^S$, integer k_0.
2: **Output:** robust solution $\mathbf{X}$.
3: **Initialization:** let $\mathbf{b}^S \leftarrow \mathbf{0}_S$.
4: **for** $k = S - 1, S - 2, \ldots, k_0$ **do**
5:     Let $\mathcal{L} \subset [S]$ be the set of indices such that:
       1) $\mathbf{b}_\ell^{k+1} = 0$;
       2) inequality (23) does not hold strictly with sample ξ^ℓ.
6:     **for** $\ell \in \mathcal{L}$ **do**
7:         Construct vector $\mathbf{b}^{k_0, \ell} \in \mathbb{R}^S$ by (21).
8:         Apply Algorithm 1 to solve problem (22).
9:         Let f_ℓ^* be the optimal objective value.
10:     **end for**
11:     Let $\ell^* \leftarrow \arg\min_{\ell \in \mathcal{L}} f_\ell^*$ and $\mathbf{b}^k \leftarrow \mathbf{b}^{k, \ell^*}$.
12: **end for**
13: Apply Algorithm 1 to solve problem (21) with $\mathbf{b} = \mathbf{b}^{k_0}$.
14: Return the solution $\mathbf{X}$ to the above problem.

Fig. 5. Actual and approximate post-contingency generator outputs.

Fig. 6. Actual and approximate post-contingency squared voltage magnitudes.

running time of the greedy algorithm is much better than that of the off-the-shelf solvers, e.g., YALMIP. For example, using a single 3.79-GHz CPU, the greedy algorithm takes an average of 412 seconds to solve the 118-bus case for each $k \in \{180, \ldots, 200\}$, while the YALMIP solver takes more than two hours for a single iteration of Algortihm 1, which includes solving a single MISDP problem instance. In addition, we compare the solutions generated by the greedy algorithm and the YALMIP solver for the 14-bus system. The YALMIP solutions are shown to be optimal up to a small gap between the lower and upper bounds. The results are plotted in Figure 2. We can see that the greedy algorithm is able to find solutions of almost the same quality as the global optima. To be more concrete, the objective function values and constraint satisfaction rates of the two solutions are very close.

In summary, although there is no theoretical optimality guarantee, the heuristic greedy algorithm is able to find near-optimal solutions in an efficient and robust way for benchmark power systems.

B. Linearization Accuracy

To evaluate the accuracy of the first-order approximation presented in Section III-A for our test case, we compare the actual and first-order approximate post-contingency system responses for the 14-bus system. The responses are computed for all available samples. The operating point is obtained by DRCCOPF-KL with $k = \hat{S} = 100$. Actual system responses are computed by running the MATPOWER power flow solver. The generator outputs (on separate subplots) and squared voltage magnitudes (on a single plot, with a different color for each bus) are given in Figures 5 and 6, respectively. As shown by the figures, the approximation is quite accurate and appears appropriate for our problem instance. In Section IV, the approximate system response will be used to compute violation rates.
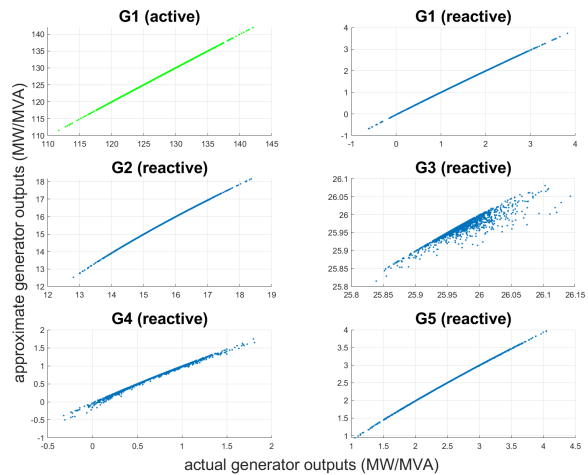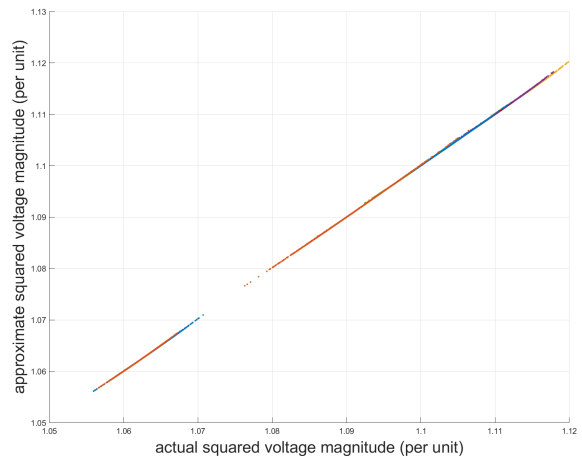
C. Derivation of Sensitivity Factor

In this section, we derive the sensitivity factor $D_\Delta(\mathbf{X})$, which is defined as the Jacobian of $\Delta(\mathbf{X}, \xi)$ with respect to ξ. More specifically, by the definition of $D_\Delta(\mathbf{X})$ and the constraints (3), we have

$$D_\Delta(\mathbf{X}) = \begin{bmatrix} \partial_\xi \Delta P(\mathbf{X}, \xi) - \mathbf{I}_n \\ -\partial_\xi \Delta P(\mathbf{X}, \xi) + \mathbf{I}_n \\ \partial_\xi \Delta Q(\mathbf{X}, \xi) - \gamma \mathbf{I}_n \\ -\partial_\xi \Delta Q(\mathbf{X}, \xi) + \gamma \mathbf{I}_n \\ \partial_\xi \Delta |V|^2(\mathbf{X}, \xi) \\ -\partial_\xi \Delta |V|^2(\mathbf{X}, \xi) \end{bmatrix} \in \mathbb{R}^{6n \times n}. \quad (24)$$

Therefore, the problem reduces to the calculation of the partial derivatives of ΔP, ΔQ and $\Delta |V|^2$ with respect to ξ.

We begin with the first-order approximation of the power

flow equations around an operating point $\mathbf{X}$:

$$\mathbf{J}\begin{bmatrix}\Delta\Theta \\ \Delta|V|\end{bmatrix} = \begin{bmatrix}\Delta P \\ \Delta Q\end{bmatrix}, \qquad (25)$$

where $\Theta \in \mathbb{R}^n$ is the vector of voltage angles and $\mathbf{J} \in \mathbb{R}^{2n \times 2n}$ is the Jacobian matrix, which can be computed by the implicit function theorem. For convenience, we provide the expression of $\mathbf{J}$:

$$\mathbf{J} = 2(\mathbf{I}_n \otimes \mathbf{X}^T)\begin{bmatrix}\mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \\ \overline{\mathbf{Y}}_1 \\ \vdots \\ \overline{\mathbf{Y}}_n\end{bmatrix}$$
$$\cdot \begin{bmatrix}-\operatorname{diag}(|V|)\operatorname{diag}(\sin\Theta) & \operatorname{diag}(\cos\Theta) \\ \operatorname{diag}(|V|)\operatorname{diag}(\cos\Theta) & \operatorname{diag}(\sin\Theta)\end{bmatrix},$$

where $\otimes$ denotes the Kronecker product and the magnitude, sine, and cosine operators are elementwise. Note that $\mathbf{J}$ depends on $\mathbf{X}$; we do not write this dependence to avoid clutter.

Then, applying the forecast error and AGC response (2), the equations (25) can be rewritten as:

$$\mathbf{J}\begin{bmatrix}\Delta\Theta \\ \Delta|V|\end{bmatrix} = \begin{bmatrix}\mathbf{I}_n - \alpha\mathbf{1}_n^T \\ \gamma\mathbf{I}_n\end{bmatrix}\xi + \begin{bmatrix}\delta P \\ \delta Q\end{bmatrix}, \qquad (26)$$

where δP and δQ are the additional changes in active and reactive power injections after non-slack generator setpoints are changed, respectively. Applying knowledge of different bus types (slack, generator, and load), many components in equations (26) are zero and we only need solve the sub-systems

$$\mathbf{J}_1\begin{bmatrix}\Delta\Theta_{\mathcal{PV}\cup\mathcal{PQ}} \\ \Delta|V||_{\mathcal{PQ}}\end{bmatrix} = \underbrace{\begin{bmatrix}(\mathbf{I}_n - \alpha\mathbf{1}_n^T)_{\mathcal{PV}\cup\mathcal{PQ}} \\ (\gamma\mathbf{I}_n)_{\mathcal{PQ}}\end{bmatrix}}_{:=\mathbf{G}_1}\xi,$$

$$\mathbf{J}_2\begin{bmatrix}\Delta\Theta_{\mathcal{PV}\cup\mathcal{PQ}} \\ \Delta|V||_{\mathcal{PQ}}\end{bmatrix} = \underbrace{\begin{bmatrix}(\mathbf{I}_n - \alpha\mathbf{1}_n^T)_{\{P\theta\}} \\ (\gamma\mathbf{I}_n)_{\{P\theta\}\cup\mathcal{PV}}\end{bmatrix}}_{:=\mathbf{G}_2}\xi + \begin{bmatrix}\delta P_{\{P\theta\}} \\ \delta Q_{\{P\theta\}\cup\mathcal{PV}}\end{bmatrix},$$

where $(\cdot)_{\mathcal{S}}$ denotes the rows indexed by members of $\mathcal{S}$. Solving the above sub-systems gives

$$\begin{bmatrix}\Delta\Theta_{\mathcal{PV}\cup\mathcal{PQ}} \\ \Delta|V||_{\mathcal{PQ}}\end{bmatrix} = \mathbf{J}_1^{-1}\mathbf{G}_1\xi,$$

$$\begin{bmatrix}\delta P_{\{P\theta\}} \\ \delta Q_{\{P\theta\}\cup\mathcal{PV}}\end{bmatrix} = (\mathbf{J}_2\mathbf{J}_1^{-1}\mathbf{G}_1 - \mathbf{G}_2)\xi.$$

Equivalently, we get the following rows of sensitivity factors

$$(\partial_\xi \delta P)_{\{P\theta\}} = (\mathbf{J}_2\mathbf{J}_1^{-1}\mathbf{G}_1 - \mathbf{G}_2)_1,$$

$$(\partial_\xi \delta Q)_{\{P\theta\}\cup\mathcal{PV}} = (\mathbf{J}_2\mathbf{J}_1^{-1}\mathbf{G}_1 - \mathbf{G}_2)_{2:1+|\{P\theta\}|+|\mathcal{PV}|},$$

$$(\partial_\xi \delta|V|)_{\mathcal{PQ}} = (\mathbf{J}_1^{-1}\mathbf{G}_1)_{|\mathcal{PV}|+|\mathcal{PQ}|+1:|\mathcal{PV}|+2|\mathcal{PQ}|},$$

where on the right-hand side, we use the MATLAB-style of row indexing. All other rows are zero.

Finally, combining the above results with (26), we can compute the sensitivity factors of the constraint functions on active power, reactive power, and squared voltage magnitude, respectively:

$$\partial_\xi \Delta P(\mathbf{X}, \xi) = \mathbf{I}_n - \alpha\mathbf{1}_n^T + \partial_\xi \delta P(\mathbf{X}, \xi),$$
$$\partial_\xi \Delta Q(\mathbf{X}, \xi) = \gamma\mathbf{I}_n + \partial_\xi \delta Q(\mathbf{X}, \xi),$$
$$\partial_\xi \Delta|V|^2(\mathbf{X}, \xi) = 2|V| \circ \partial_\xi \Delta|V|(\mathbf{X}, \xi),$$

where $\circ$ denotes the elementwise product. Substituting the partial derivatives into the expression (24), we get the sensitivity factor $D_\Delta(\mathbf{X})$.

D. Proof of Lemma 1

Proof of Lemma 1. We first show that in the definition of predictor $\hat{q}_{1-\epsilon, r, S}$, the supremum can be restricted to the set of distributions in $\mathcal{D}_r(\hat{\mathbb{P}}_S)$ that are absolutely continuous with respect to $\hat{\mathbb{P}}_S$ except on the set

$$\Xi^*(\mathbf{X}) := \left\{\xi \mid \bar{h}_{\mathbf{X}}(\xi) = h_{\mathbf{X}}^*\right\}.$$

The proof is the same as that of Lemma 2 of [14] except the bound on the expectation, i.e., the second last inequality in the proof. To deal with this issue, we only need to prove that for all $\mathbf{X} \in \mathbb{R}^d$, $p \in [0, 1]$, $\xi^* \in \Xi^*(\mathbf{X})$, and $\mathbb{P}_c, \mathbb{P}_\perp \in \mathcal{P}$ such that $\mathbb{P}_c \ll \hat{\mathbb{P}}_S$ and $\mathbb{P}_\perp \perp \mathbb{P}_c$[2], it holds that

$$q_{1-\epsilon}\left(\bar{h}_{\mathbf{X}}, \mathbb{P}'\right) \geq q_{1-\epsilon}\left(\bar{h}_{\mathbf{X}}, \mathbb{P}''\right), \qquad (27)$$

where

$$\mathbb{P}' := p \cdot \mathbb{P}_c + (1-p) \cdot \delta_{\xi^*}, \quad \mathbb{P}'' := p \cdot \mathbb{P}_c + (1-p) \cdot \mathbb{P}_\perp.$$

Let $F'(h)$ and $F''(h)$ be the cumulative distribution function of $\bar{h}_{\mathbf{X}}(\xi)$ under distribution $\mathbb{P}'$ and $\mathbb{P}''$, respectively. By the definition of quantile, to prove inequality (27), it is sufficient to show that

$$F'(h) \geq F''(h), \quad \forall h \in \mathbb{R},$$

which is equivalent to

$$\mathbb{E}_{\xi \sim \mathbb{P}'}\left[\mathbf{1}(\bar{h}_{\mathbf{X}}(\xi) \leq h)\right] \geq \mathbb{E}_{\xi \sim \mathbb{P}''}\left[\mathbf{1}(\bar{h}_{\mathbf{X}}(\xi) \leq h)\right], \quad \forall h \in \mathbb{R},$$

where $\mathbf{1}(\gamma(\nu, \xi) \leq \gamma)$ is the indicator function. This can be proved in the same way as the proof in [14]. As a result, there exists a distribution that attains $\hat{q}_{1-\epsilon, r, \hat{\mathbb{P}}_S}(\mathbf{X})$ and has support in $\{\xi^j, j \in [S]\} \cup \Xi^*(\mathbf{X})$, which implies the existence of an integer $k \in [S+1]$ such that

$$\hat{q}_{1-\epsilon, r, \hat{\mathbb{P}}_S}(\mathbf{X}) = \bar{h}_{k, \hat{\mathbb{P}}_S}(\mathbf{X}).$$

Next, we prove that integer k does not depend on $\mathbf{X}$ and $\hat{\mathbb{P}}_S$. Let $\mathbb{P}_{1-\epsilon, r, S}$ be the aforementioned worst-case distribution that attains $\hat{q}_{1-\epsilon, r, S}(\mathbf{X})$. Assume without loss of generality that

$$\bar{h}_{\mathbf{X}}(\xi^1) \leq \cdots \leq \bar{h}_{\mathbf{X}}(\xi^S).$$

Define vector $\mathbf{p} \in \mathbb{R}^{S+1}$ as

$$\mathbf{p}_j := \mathbb{P}_{1-\epsilon, r, S}(\xi^j), \quad \forall j \in [S], \quad \mathbf{p}_{S+1} := \mathbb{P}_{1-\epsilon, r, S}[\Xi^*(\mathbf{X})].$$

---

[2]For distributions $\mathbb{P}, \mathbb{P}' \in \mathcal{P}$, we use $\mathbb{P} \ll \mathbb{P}'$ and $\mathbb{P} \perp \mathbb{P}'$ to denote the case when $\mathbb{P}$ is absolutely continuous and singular with respect to $\mathbb{P}'$, respectively.

Then, by problem (33) in [14], the integer k is the solution to

$$\max_{k\in[S],\mathbf{p}\in\mathbb{R}^{S+1}} k \tag{28}$$

$$\text{s.t. } \sum_{j\in[k]}\mathbf{p}_j \leq 1-\epsilon, \ \mathbf{1}_{S+1}^T\mathbf{p}=1, \ \mathbf{p}\geq\mathbf{0}_{S+1},$$

$$-\frac{1}{S}\sum_{j\in[S]}\log(S\mathbf{p}_j) \leq r,$$

which is independent of $\mathbf{X}$ and $\hat{\mathbb{P}}_S$. Intuitively, k is the largest integer such that the probability $\mathbb{P}_{1-\epsilon,r,S}$ on the smallest k samples is at most $1-\epsilon$ and the relative entropy constraint is not violated. □

E. Proof of Theorem 2

Proof of Theorem 2. The formulation (15) is based on the big-M method [35]. If the variable $\mathbf{b}_j = 1$, since the constant C is sufficiently large, there is no constraint on $h_\ell(\mathbf{X},\xi^j)$. Otherwise if the variable $\mathbf{b}_j = 0$, the first constraint becomes

$$h_\ell(\mathbf{X},\xi^j) \leq 0, \quad \forall\ell\in[m],$$

which is equivalent to the condition $\bar{h}_\mathbf{X}(\xi^j) \leq 0$. With a given $\mathbf{X}\in\mathbb{R}^d$, the constraint $\mathbf{1}_S^T\mathbf{b} \leq S-k$ requires that the above condition holds for at least k samples. To achieve the minimum over $\mathbf{X}$, the condition $\mathbf{b}_j = 0$ should hold for the k indices that correspond to the k smallest values in $\{\bar{h}_\mathbf{X}(\xi^j), j\in[S]\}$. In other words, the constraints in (15) are equivalent to

$$\bar{h}_k(\mathbf{X}) \leq 0.$$

Combining with Lemma 1, we get the desired result. □

F. Proof of Theorem 3

Proof of Theorem 3. By the definition of the prescriptor $\hat{\mathbf{X}}_{1-\epsilon,r,S}$, we have

$$\hat{q}_{1-\epsilon,r,S}\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}\right) \leq 0.$$

By a similar technique to the proof of Lemma 1, the results of Theorem 10 of [14] also holds for the predictor $\hat{q}_{1-\epsilon,r,S}$ and we have

$$\limsup_{S\to\infty}\frac{1}{S}\log\mathbb{P}_\infty\left[\hat{q}_{1-\epsilon,r,S}\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}\right)\right.$$
$$\left. < q_{1-\epsilon}\left(\bar{h}_{\hat{\mathbf{X}}_{1-\epsilon,r,S}},\mathbb{P}_0\right)\right] \leq -r.$$

Combining the above two inequalities, we get

$$\mathbb{P}_\infty\left[q_{1-\epsilon}\left(\bar{h}_{\hat{\mathbf{X}}_{1-\epsilon,r,S}},\mathbb{P}_0\right)\leq 0\right] \geq 1-\exp\left[-rS+o(S)\right].$$

By the definition of the quantile and applying the union bound, it follows that

$$\mathbb{P}_\infty\left[h_\ell\left(\hat{\mathbf{X}}_{1-\epsilon,r,S},\xi\right)\leq 0, \quad \forall\ell\in[m]\right]$$
$$\geq 1-\epsilon-\exp\left[-rS+o(S)\right].$$

which is the desired result of this theorem. □

G. Proof of Theorem 4

Proof of Theorem 4. We first construct a set where the distributionally robust predictor $\hat{q}_{1-\epsilon,r,S}$ takes positive value. Assume conversely that

$$p_S := \mathbb{P}_\infty\left[g\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S}\right) < g\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}\right)\right] > 0.$$

Since the prescriptor $\hat{\mathbf{X}}_{1-\epsilon,r,S}$ attains the minimal objective value under the constraint $\bar{h}_k(\mathbf{X})\leq 0$, we have

$$\mathbb{P}_\infty\left[\bar{h}_k\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S}\right)>0\right]\geq p_S.$$

Since $\mathbb{P}_\infty(\mathbf{b}>z)$ is a right-continuous function of $z\in\mathbb{R}$ for every random variable $\mathbf{b}$, there exists a sufficiently small constant $\tau>0$ such that

$$\mathbb{P}_\infty\left[\bar{h}_k\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S}\right)>\tau\right]\geq p_S/2>0.$$

Consider the set

$$\mathcal{X}_S := \left\{\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S},\hat{\mathbb{P}}_S\right) \mid \bar{h}_k\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S}\right)>\tau\right\}\subset\mathbb{R}^d\times\mathcal{P}.$$

Since $\tilde{\mathbf{X}}_{1-\epsilon,r,S}$ is a quasi-continuous function of the empirical distribution $\hat{\mathbb{P}}_S$, the set $\mathcal{X}_S$ is a non-empty quasi-open set [36, Prop. 1.2.4] under the product topology of the Euclidean topology on $\mathbb{R}^d$ and the weak topology on $\mathcal{P}$. Therefore, the interior of $\mathcal{X}_S$, denoted as $\mathcal{X}_S^\circ$, is non-empty.

Now, we construct a data-driven predictor $\tilde{q}_{1-\epsilon,r,\mathbb{P}}$ that is continuous and does not dominate the distributionally robust predictor $\hat{q}_{1-\epsilon,r,\mathbb{P}}$. For every point $(\mathbf{X},\mathbb{P})\in\mathcal{X}_S$, we define

$$\mathrm{d}(\mathbf{X},\mathbb{P}) := \min\left\{\mathrm{dist}\left[(\mathbf{X},\mathbb{P}),\mathcal{X}_S^c\right],\tau\right\},$$

where $\mathcal{X}_S^c := (\mathbb{R}^d\times\mathcal{P})\backslash\mathcal{X}_S$ is the complementary set of $\mathcal{X}_S$ and the distance function is induced by the Euclidean 2-norm on $\mathbb{R}^d$ and the Prokhorov metric [37] on $\mathcal{P}$. Since the distance function is continuous, the function $\mathrm{d}(\cdot,\cdot)$ is also continuous and takes positive values on $\mathcal{X}_S^\circ$. We define

$$\tilde{q}_{1-\epsilon,r,\mathbb{P}}(\mathbf{X}) := \hat{q}_{1-\epsilon,r,\mathbb{P}}(\mathbf{X})-\mathrm{d}(\mathbf{X},\mathbb{P}), \quad \forall(\mathbf{X},\mathbb{P})\in\mathbb{R}^d\times\mathcal{P}.$$

It follows from the definition of d and $\mathcal{X}_S$ that

$$0\leq\tilde{q}_{1-\epsilon,r,\mathbb{P}}(\mathbf{X})\leq\hat{q}_{1-\epsilon,r,\mathbb{P}}(\mathbf{X}), \quad \forall(\mathbf{X},\mathbb{P})\in\mathcal{X}_S, \tag{29}$$

where the second inequality holds strictly on $\mathcal{X}_S^\circ$. Note that the predictor $\tilde{q}_{1-\epsilon,r,S} := \tilde{q}_{1-\epsilon,r,\hat{\mathbb{P}}_S}$ is a data-driven predictor since it only relies on the empirical distribution $\hat{\mathbb{P}}_S$.

Finally, we show that $\tilde{q}_{1-\epsilon,r,\mathbb{P}}$ is feasible for problem (5) in [14], namely,

$$\limsup_{S\to\infty}\frac{1}{S}\log\mathbb{P}_\infty\left[\tilde{q}_{1-\epsilon,r,S}(\mathbf{X})<q_{1-\epsilon}\left(\bar{h}_\mathbf{X},\mathbb{P}_0\right)\right]\leq -r. \tag{30}$$

Since condition (30) is satisfied by $\hat{q}_{1-\epsilon,r,S}$ and

$$\tilde{q}_{1-\epsilon,r,S}(\mathbf{X}) = \hat{q}_{1-\epsilon,r,S}(\mathbf{X}), \quad \forall\mathbf{X}\in\mathbb{R}^d \text{ s.t. } \mathbf{X}\neq\tilde{\mathbf{X}}_{1-\epsilon,r,S},$$

we only need to show

$$\limsup_{S\to\infty}\frac{1}{S}\log\mathbb{P}_\infty\left[\tilde{q}_{1-\epsilon,r,S}\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}\right)\right. \tag{31}$$
$$\left. < q_{1-\epsilon}\left(\bar{h}_{\tilde{\mathbf{X}}_{1-\epsilon,r,S}},\mathbb{P}_0\right)\right]\leq -r.$$

Since prescriptor $\tilde{\mathbf{X}}_{1-\epsilon,r,S}$ satisfies condition (16), it holds that

$$\limsup_{S\to\infty} \frac{1}{S} \log \mathbb{P}_\infty \left[q_{1-\epsilon} \left(\bar{h}_{\tilde{\mathbf{X}}_{1-\epsilon,r,S}}, \mathbb{P}_0 \right) < 0 \right] \leq -r.$$

Combining with the property (29), we get the desired result (31).

In summary, we have constructed a predictor $\tilde{q}_{1-\epsilon,r,S} \left(\hat{\mathbf{X}}_{1-\epsilon,r,\mathbb{P}} \right)$ that is continuous and feasible for problem (5) in [14], but it does not dominate the distributionally robust predictor $\hat{q}_{1-\epsilon,r,S} \left(\hat{\mathbf{X}}_{1-\epsilon,r,\mathbb{P}} \right)$. However, this is contradictory with Theorem 10 in [14], which claims that the distributionally robust predictor is the strong solution to problem (5). $\qquad\square$

H. Disjoint Chance Constraint

In this section, we extend the theory in Section III-B to disjoint chance constraints. Using the same notation as Section III-B, the disjoint chance constraint (7) can be written as

$$\mathbb{P}_0 \left[h_\ell(\mathbf{X}, \xi) \leq 0 \right] \geq 1 - \epsilon, \quad \forall \ell \in [m]. \tag{32}$$

With the same violation probability ϵ, the disjoint chance constraint is less restrictive than the joint counterpart (11). On the other hand, if we choose ϵ to be ϵ/m in (32), Boole's inequality leads to

$$\mathbb{P}_0 \left[h(\mathbf{X}, \xi) > \mathbf{0}_m \right] \leq \sum_{\ell\in[m]} \mathbb{P}_0 \left[h_\ell(\mathbf{X}, \xi) > 0 \right] \leq \epsilon,$$

which implies that the joint chance constraint holds with violation probability ϵ. More generally, with the disjoint chance constraint, we are able to bound the probability that at least s constraints are violated for all $s \in [m]$. More specifically, define the indicator function

$$\mathbf{1}_\ell(\mathbf{X}, \xi) := \begin{cases} 1 & \text{if } h_\ell(\mathbf{X}, \xi) > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \forall \ell \in [m].$$

Then, it holds that

$$\mathbb{E}_0 \left[\mathbf{1}_\ell(\mathbf{X}, \xi) \right] = \mathbb{P}_0 \left[h_\ell(\mathbf{X}, \xi) > 0 \right] \leq \epsilon,$$

where $\mathbb{E}_0$ is the expectation under the true distribution $\mathbb{P}_0$. Using Markov's inequality, we get

$$\mathbb{P}_0 \left[h_\ell(\mathbf{X}, \xi) > 0 \text{ for at least } s \text{ indices } \ell \right]$$
$$= \mathbb{P}_0 \left[\sum_{\ell\in[m]} \mathbf{1}_\ell(\mathbf{X}, \xi) \geq s \right] \leq \frac{\mathbb{E}_0 \left[\sum_{\ell\in[m]} \mathbf{1}_\ell(\mathbf{X}, \xi) \right]}{s}$$
$$\leq \frac{m\epsilon}{s}.$$

In the following, we consider two different generalizations of chance constraint (32), which we denote as the *finite* case and the *infinite* case.

We first define the finite case of disjoint chance constraint. Given L weight vectors $\omega_1, \ldots, \omega_L \in \mathbb{R}^m$, the disjoint chance constraint is defined as

$$\mathbb{P}_0 \left[\omega_\ell^T h(\mathbf{X}, \xi) \leq 0 \right] \geq 1 - \epsilon, \quad \forall \ell \in [L]. \tag{33}$$

The cardinal case (32) is a special case with $L = m$ and $\omega_\ell = \mathbf{e}_\ell$ for all $\ell \in [m]$. Basically, the reformulation of the finite case can be derived in a similar way as that of the

joint chance constraint. Therefore, we omit the proofs for the disjoint chance constraint and use the same notation as Section III-B. Choosing $C > 0$ to be a sufficiently large constant, the distributionally robust prescriptor $\hat{\mathbf{X}}_{1-\epsilon,r,S}$ is a solution to

$$\min_{\mathbf{X}\in\mathbb{R}^d, \mathbf{B}\in\mathbb{Z}^{S\times L}} g(\mathbf{X}) \tag{34}$$
$$\text{s.t. } \omega_\ell^T h(\mathbf{X}, \xi^j) \leq C\mathbf{B}_{j,\ell},$$
$$\mathbf{1}_S^T \mathbf{B}_{:,\ell} \leq S - k, \quad \mathbf{B}_{j,\ell} \in \{0,1\},$$
$$\forall \ell \in [L], \ j \in [S],$$

where the integer $k \in [\ell]$ is defined in Lemma 1 as a function of ϵ, r and S. Problem (34) is the disjoint counterpart of problem (15) and can be formulated as a MISDP for the CCOPF problem if the semi-definite relaxation is exact. Similarly, we can prove that the distributionally robust prescriptor achieves the optimal cost among solutions that satisfy disjoint chance constraint (33) with high probability.

Theorem 5. *For all $\epsilon \in [0, 1]$ and $r > 0$, it holds that*

$$\mathbb{P}_\infty \left[q_{1-\epsilon} \left(\omega_\ell^T h(\hat{\mathbf{X}}_{1-\epsilon,r,S}, \cdot), \mathbb{P}_0 \right) \leq 0 \right] \tag{35}$$
$$\geq 1 - \exp\left[-rS + o(S) \right], \quad \forall \ell \in [L],$$

which leads to

$$\mathbb{P}_\infty \left[\omega_\ell^T h \left(\hat{\mathbf{X}}_{1-\epsilon,r,S}, \xi \right) \leq 0 \right]$$
$$\geq 1 - \epsilon - \exp\left[-rS + o(S) \right], \quad \forall \ell \in [L].$$

Furthermore, suppose that prescriptor $\tilde{\mathbf{X}}_{1-\epsilon,r,\mathbb{P}} \in \mathbb{R}^d$ is a quasi-continuous function of $\mathbb{P}$ and satisfies constraint (35). Then, we have

$$\mathbb{P}_\infty \left[g\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S} \right) < g\left(\hat{\mathbf{X}}_{1-\epsilon,r,S} \right) \right] = 0,$$

where we denote $\tilde{\mathbf{X}}_{1-\epsilon,r,S} := \tilde{\mathbf{X}}_{1-\epsilon,r,\hat{\mathbb{P}}_S}$.

Next, we extend the disjoint chance constraint to a more general case. Instead of a finite number of weight vectors, the infinite case is defined by a set of weight vectors $\mathcal{W}$, which can contain an infinite number of elements. The infinite case of disjoint chance constraint is then formulated as

$$\mathbb{P}_0 \left[\omega^T h(\mathbf{X}, \xi) \leq 0 \right] \geq 1 - \epsilon, \quad \forall \omega \in \mathcal{W}. \tag{36}$$

Hence, the finite case can be viewed as a special example of the infinite case, where the set $\mathcal{W}$ only contains a finite number of weight vectors. As an example of the infinite case, the set $\mathcal{W}$ can be the set of all weight vectors:

$$\mathcal{W} = \left\{ \omega \in \mathbb{R}^m \mid \mathbf{1}_m^T \omega = 1, \omega \geq \mathbf{0}_m \right\}.$$

In this case, the constraint (36) enforces that all convex combinations of stochastic constraints are satisfied with high probability. More generally, in certain applications, the constraints can be divided into several groups. We can choose the set $\mathcal{W}$ to be the union of weight vectors of a subset of indices:

$$\mathcal{W} = \bigcup_{k\in[L]} \left\{ \omega \in \mathbb{R}^m \mid \mathbf{1}_m^T \omega = 1, \omega \geq \mathbf{0}_m, \omega_\ell = 0, \forall \ell \notin \mathcal{I}_k \right\},$$

where $\mathcal{I}_k \subset [m]$ are disjoint subsets. Similar to the finite case, the chance constraint (36) can be reformulated as a

MISDP. However, the MISDP contains an infinite number of constraints and thus, is considerably more challenging to solve. More specifically, for each $\omega \in \mathcal{W}$, the constraint requires that there exists a vector $\mathbf{b}^\omega \in \mathbb{Z}^S$ such that

$$\omega^T h(\mathbf{X}, \xi^j) \leq C\mathbf{b}_j^\omega, \quad \mathbf{1}_S^T \mathbf{b}^\omega \leq S - k, \quad (37)$$
$$\mathbf{b}_j^\omega \in \{0, 1\}, \quad \forall j \in [S].$$

To deal with this challenge, we develop an iterative algorithm to approximate the constraint (37). In the t-th iteration, we use a finite set of weight vectors $\mathcal{W}_t$ to approximate the set $\mathcal{W}$. The algorithm proceeds in two stages:

1) With a fixed set $\mathcal{W}_t$, the algorithm generates an approximate distributionally robust prescriptor $\hat{\mathbf{X}}_t$ by solving problem (34) with weight vectors in $\mathcal{W}_t$;
2) With a fixed solution $\hat{\mathbf{X}}_t$, the algorithm finds the weight vector ω_t that violates constraint (37) by the largest margin. If there does not exist such weight vectors, we know that the constraint (37) is satisfied and the algorithm is terminated. Otherwise, we add vector ω_t to set $\mathcal{W}_t$.

The pseudo-code of the aforementioned algorithm is provided in Algorithm 3. For the CCOPF problem, if the set $\mathcal{W}$ is a polyhedral, problem (38) becomes a MISDP and problem (39) becomes a MIP. In this case, the algorithm runs efficiently in practice and exhibits good empirical performances; see more details in Section III of [23]. If the set $\mathcal{W}$ has certain special structure, the initial set $\mathcal{W}_1$ can be chosen based on the prior information about $\mathcal{W}$. For example, if $\mathcal{W}$ is a polyhedral, we can initialize $\mathcal{W}_1$ to contain all extreme points of the polyhedral. In the general case when the set $\mathcal{W}$ is not a polyhedral or even non-convex, problem (39) can be more challenging to solve.

Problem (39) is also based on the big-M method. If the variable $\mathbf{b}_j = 1$, since the constant C is sufficiently large, there is no constraint on s. Otherwise if the variable $\mathbf{b}_j = 0$, the constraint requires that

$$\omega^T h\left(\hat{\mathbf{X}}_t, \xi^j\right) \geq s.$$

This means that s should be the minimal value of the left-hand side over all indices j such that $\mathbf{b}_j = 0$. With a given $\omega \in \mathbb{R}^m$, to maximize the value of s, variable $\mathbf{b}_j$ is equal to one for indices with the k largest values of the left-hand side. Then, the optimal value of s should be the k-th largest value of the left-hand side over all samples. If we further minimize over the weight vector ω, the condition (37) holds if and only if the optimal value s_t is non-positive. In addition, if $s_t > 0$, the corresponding vector ω_t provides a weight vector such that condition (37) is violated by the largest margin.

Since problem (38) usually involves cone constraints, such as the semi-definite constraint in the CCOPF case, Algorithm 3 does not fit into the framework of classical cutting-plane methods, e.g., [38]. Therefore, the convergence of Algorithm 3 cannot be directly derived from those of existing cutting-plane methods and we leave the theoretical analysis to future works.

Algorithm 3 Algorithim for the infinite case of disjoint chance constraints.

1: **Input:** Set of weight vectors $\mathcal{W}$, empirical distribution $\hat{\mathbb{P}}_S$, number of iterations t_{max}, parameters ϵ, r.
2: **Output:** Approximate prescriptor $\hat{\mathbf{X}}_{1-\epsilon,r,S}$.
3: Compute k by solving (28).
4: Initialize $\mathcal{W}_1 \leftarrow \emptyset$.
      $\triangleright$ Alternatively, initialize with a finite subset of $\mathcal{W}$.
5: **for** $t = 1, 2, \ldots, t_{max}$ **do**
6:    Let $\hat{\mathbf{X}}_t$ be a solution to:

$$\min_{\mathbf{X} \in \mathbb{R}^d, \mathbf{B} \in \mathbb{Z}^{S \times L_t}} g(\mathbf{X}) \quad (38)$$
$$\text{s.t. } \omega_\ell^T h(\mathbf{X}, \xi^j) \leq C\mathbf{B}_{j,\ell},$$
$$\mathbf{1}_S^T \mathbf{B}_{:,\ell} \leq S - k, \quad \mathbf{B}_{j,\ell} \in \{0, 1\},$$
$$\forall \ell \in [L_t], j \in [S],$$

   where we define $L_t = |\mathcal{W}_t|$ and $\mathcal{W}_t = \{\omega_1, \ldots, \omega_{L_t}\}$.
7:    Let $(s_t, \omega_t, \mathbf{b}_t)$ be a solution to:

$$\max_{s \in \mathbb{R}, \omega \in \mathcal{W}, \mathbf{b} \in \mathbb{Z}^S} s, \quad (39)$$
$$\text{s.t. } \omega^T h\left(\hat{\mathbf{X}}_t, \xi^j\right) \geq s + C\mathbf{b}_j,$$
$$\mathbf{1}_S^T \mathbf{b} \leq S - k, \quad \mathbf{b}_j \in \{0, 1\}, \quad \forall j \in [S].$$

8:    **if** solution $s_t \leq 0$ **then** $\triangleright$ condition (37) is satisfied.
9:       **break**
10:    **end if**
11:    Update $\mathcal{W}_{t+1} \leftarrow \mathcal{W}_t \cup \{\omega_t\}$.
12: **end for**
13: Return the last iterate of $\hat{\mathbf{X}}_t$ as $\hat{\mathbf{X}}_{1-\epsilon,r,S}$.

In this work, we assume that the minimum-cost solution $\hat{\mathbf{X}}_{1-\epsilon,r,S}$ can be found, namely, it is a solution to the following optimization problem:

$$\min_{\mathbf{X} \in \mathbb{R}^d, \mathbf{b}^w \in \mathbb{R}^S} g(\mathbf{X})$$
$$\text{s.t. } \quad \text{constraint (37) is satisfied for all } \omega \in \mathcal{W}.$$

The next theorem claims that the solution $\hat{\mathbf{X}}_{1-\epsilon,r,S}$ satisfies a similar optimality condition as the finite case.

Theorem 6. *For all $\epsilon \in [0, 1]$ and $r > 0$, it holds that*

$$\mathbb{P}_\infty\left[q_{1-\epsilon}\left(\omega^T h(\hat{\mathbf{X}}_{1-\epsilon,r,S}, \cdot), \mathbb{P}_0\right) \leq 0\right] \quad (40)$$
$$\geq 1 - \exp\left[-rS + o(S)\right], \quad \forall \omega \in \mathcal{W},$$

which leads to

$$\mathbb{P}_\infty\left[\omega^T h\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}, \xi\right) \leq 0\right]$$
$$\geq 1 - \epsilon - \exp\left[-rS + o(S)\right], \quad \forall \omega \in \mathcal{W}.$$

Furthermore, suppose that prescriptor $\tilde{\mathbf{X}}_{1-\epsilon,r,\mathbb{P}} \in \mathbb{R}^d$ is a quasi-continuous function of $\mathbb{P}$ and satisfies constraint (40). Then, we have

$$\mathbb{P}_\infty\left[g\left(\tilde{\mathbf{X}}_{1-\epsilon,r,S}\right) < g\left(\hat{\mathbf{X}}_{1-\epsilon,r,S}\right)\right] = 0,$$

where we denote $\tilde{\mathbf{X}}_{1-\epsilon,r,S} := \tilde{\mathbf{X}}_{1-\epsilon,r,\hat{\mathbb{P}}_S}$.

We omit the proof due to its similarity to the proof of Theorems 3 and 4.