A Two-Step Approach to Wasserstein Distributionally Robust Chance- and Security-Constrained Dispatch

Amin Maghami, Evrim Ursavas, and Ashish Cherukuri, Member, IEEE

Abstract—This paper considers a security constrained dispatch problem involving generation and line contingencies in the presence of the renewable generation. The uncertainty due to renewables is modeled using joint chance-constraint and the mismatch caused by contingencies and renewables are handled using reserves. We consider a distributionally robust approach to solve the chance-constrained program. We assume that samples of the uncertainty are available. Using them, we construct a set of distributions, termed ambiguity set, containing all distributions that are close to the empirical distribution under the Wasserstein metric. The chance constraint is imposed for all distributions in the ambiguity set to form the distributionally robust optimization problem. This problem is nonconvex and computationally heavy to solve exactly. We adopt a two-step approach to find an approximate solution. In the first step, we construct a polyhedral set in the space of uncertainty that contains enough mass under all distributions in the ambiguity set. This set is constructed by solving several two-dimensional distributionally robust problems. In the second step, we solve a linear robust optimization problem where the uncertain constraint is imposed for all uncertainty values lying in the polyhedral set. We demonstrate the scalability and robustness of our method using numerical experiments.

Index Terms—Optimization methods, power generation dispatch, uncertainty, power system security, renewable energy sources, economics, chance-constrained optimization, distributionally robust optimization.

I. INTRODUCTION

POWER systems are going through important changes, driven mainly by the increasing penetration of renewable sources. While using renewable energy is vital for mitigating climate change, they also bring significant challenges to the power grid. Renewable sources such as wind and solar are weather dependent resulting in uncertainty in production and difficulty in control. Furthermore, the intermittent nature of the production makes the grid more vulnerable to component failures rising the need for increased attention to secure the grid under such contingencies. Consequently, grid operators must clearly understand and accurately model the uncertainty and variability of renewable generation in order to achieve stable and optimal operation of the future power systems. In addition, a secure power system must be able to withstand all possible contingencies without violating physical limitations which may lead to cascading failures [1]. To avoid this, corrective N-1 security criterion is widely used within the dispatch or Optimal Power Flow (OPF) models, according to which the system is capable of finding a new operational solution after the occurrence of a single contingency (for example a line or a generator contingency), while no operational security limits (such as line flow limits) are violated. Despite the importance of N-1 criterion, the security constrained OPF problem has not been widely studied in the presence of renewable sources, except for few works [2], [3], [4], [5], [6], [7]. These works consider renewables to be uncertain, model the dispatch as stochastic optimization problems, and look for solutions where either information about the distribution is (partially) known or is inferred through samples. However, they do not consider the inherent robustness issues that stem from “uninfertility” of data about the uncertainty. That is, when the data is either scarce, corrupted, or sampled in a non-identical manner. This challenge is overcome by employing techniques of distributionally robust (DR) optimization, where decisions are made taking into account worst-case perturbations to the collected data. Motivated by this fact, in this paper, we propose a Wasserstein metric based DR approach for solving a dispatch problem that takes into account N-1 security constraints and the influence of uncertain renewable generation through chance-constraints. We then provide a two-step scalable method for approximating the solution of this DR problem.

A. Related Works

A comprehensive survey [5] on the stochastic security constrained dispatch problem has showed that considering forecast error is important and employing corrective control actions can reduce the excessive cost and increase the integration of renewable sources. Often, such security-constrained stochastic dispatch problems involve uncertainty in constraints which can be modeled in several different ways, from chance constraints [3], [8] to a robust formulation [9], [10]. We focus on the most popular choice among these, that of chance-constrained problems, where the uncertain constraint is required to hold with high probability. One way of solving these problems in a data-driven way is the scenario approach [11] where the underlying uncertain constraint defining the chance-constraint is required to hold for
all samples of the uncertainty. Such a method was studied in the context of security constrained dispatch in [2], [12], [13].

The scenario approach gets computationally burdensome when the dataset is large. To let go of this roadblock, the work [14] explores a two-step procedure for solving chance-constraints, where sample-based constraint is replaced with a robust one. This is similar to what we explore in the distributionally robust setting in the current work. An alternative way of handling chance constraints is to use analytical reformulations when the distribution is Gaussian, e.g., in the case of DC OPF [3], [7] and AC OPF [15], [16]. This is extended to mixture of Gaussians using linearization-based technique in [17]. However, no theoretical claims were made.

All the above mentioned methods assume that either the probability distribution is known or samples drawn from it in an independent manner are available. The former and the latter assumptions are very strong, as pointed out in [18] and [19], respectively. This fact motivates us to look at distributionally robust (DR) optimization, where decisions hedge against the lack of complete information about the uncertainty by considering a set of distributions. The DR problems take different forms, depending on the type of the set of distributions, termed ambiguity set, considered in the optimization problem. In [6], [20] moment-based ambiguity sets, that contain all distributions with a given first and second moment, are used. For such a setup, one still needs some data about the uncertainty to estimate the first and second moment that defines the ambiguity set. On the other hand, ambiguity sets constructed directly using data and a suitable distance metric in the space of distributions circumvent this modeling limitation. For these sets, the radius of the ambiguity set can be adjusted to control the level of conservativeness. In [21], Kullback-Leibler (KL) divergence distance is used for DR unit commitment problem. For uncertainties with finite support, the $\ell_1$- and $\ell_\infty$-norm based distance between distributions is employed in [22], [23], [24] for unit commitment problem and in [25] for security-constrained dispatch. Both KL and norm based metrics require the uncertainty to have finite support or only look for distributions supported on obtained historical data. The Wasserstein metric on the other hand can generalize to distributions supported on points that are not present in the historical data, thereby leading to better robustness performance. Recent works [19], [26], [27], [28], [29] demonstrate the effectiveness of this metric in various dispatch related problems. Nevertheless, with Wasserstein metric, as the number of available samples increases or the network becomes large, solving the DR chance-constrained problem exactly becomes computationally difficult. This is because the DR problem is equivalent to a bilinear [29] or a mixed-integer program [30], [31] with the number of constraints that scales with the product of the number of samples and size of the network. A convex approximation of DR chance constraints using CVaR constraints can bring down the computational burden [32]. However, the number of constraints of this LP still scales with the product of the number of samples, the network size, and the number of contingencies which is prohibitive. The main idea of our work is to explore methods that solve optimization problems having decision variables and constraints either independent of the size of the network or the number of samples. To this end, we build on the two-step approaches proposed in [33] and [19]. The idea is to first construct a set in the space of uncertainty such that the probability mass contained in this set for all distributions in the ambiguity set is high. Then, in the second step, a robust optimization problem where the uncertain constraint is enforced for all uncertain values in the constructed set is solved. Our work differs from [33] and [19] as we consider joint chance-contrains as opposed to individual ones in [33] and we take into account correlation among renewables when performing the first step, which is different from [19]. In the latter comparison, our method is less conservative as we are able to construct more “informative” sets in the first step while still retaining the tractability of the second step.

Our work is broadly also related to [34], [35], [36], [37], [38]. In [34], generation planning problem with distributionally robust CVaR constraints is considered where a box-type ambiguity set is used. The work [35] proposes a convex approximation of (DR) chance-constrained problems that outperform the CVaR-based convex approximation. They term this method as ALSO-X, which consists of solving a bilevel optimization problem using an iterative method. Recent works [36], [37] have looked at removing implausible distributions from the ambiguity set to bring down the conservativeness of DR methods. However, the computational burden of these methods is high for a large dataset. Finally, the first step of our method resembles the study by [38] that involves constructing polyhedral regions to describe the uncertainty.

B. Setup and Contributions

We begin with modeling the dispatch problem considering renewables and contingencies. We consider a security constrained dispatch problem with two stages of decision making, where the stages are separated by an event consisting of the realization of power provision by the renewable sources and contingency or failure in the network. For the first stage, the operator needs to determine for each conventional generator, the generation set point and the procurement of the reserve capacities. The latter include two different sets of reserves, ones that are activated to handle imbalances due to uncertainty and the others to counter the contingency. The second stage decisions are control policies that determine how the power imbalance is adjusted network-wide using reserves. Concurrently, the aim is to ensure that the line flow and reserve capacity limits are satisfied collectively with high probability. These considerations give rise to a chance-constrained optimization problem. We assume that the information about the uncertainty is known through a certain number of samples. We consider the ambiguity set to be all distributions with a compact support that are close to the empirical distribution in the Wasserstein metric and formulate the DR chance-constrained problem where the probabilistic constraint is required to hold for all distributions in the ambiguity set.

Our first contribution is the two-step procedure to approximate the solution of the DR problem. In the first step, we assume that the renewables are partitioned into sets where two
renewables are correlated only if they belong to a partition. This encapsulates the scenario where renewables in a certain geographical location will have similar pattern of energy generation. For each partition, we construct a polyhedral set that contains enough probability mass under all distributions in the ambiguity set. We construct these sets by solving several two-dimensional DR chance-constrained problems. In the second step, we do robust optimization with respect to the Cartesian product of all these polyhedral sets. The optimal solution of the robust problem is guaranteed to be feasible for the DR chance-constrained problem. Our second contribution involves demonstration of our two-step method with a numerical example of IEEE RTS 24-bus system. We show how our method is scalable in terms of computational ease as the number of samples grows while providing a tunable parameter to trade off the cost and violation frequency of the obtained decision.

The rest of the paper is organized as follows. Section II models the optimization problem capturing the dispatch decision. Section III considers the DR version of the problem, presents the two-step approach, derives the computational tractability of the method, and analyzes the guarantee of the obtained solution. Finally, Section IV presents the numerical example.

II. PROBLEM STATEMENT

In this section, we define the N-1 security constrained dispatch problem in the presence of uncertainty caused due to renewable sources. Throughout the paper we use the linearized DC power flow model assuming that all lines have zero resistance. As a consequence, we neglect any power losses. The aim is to optimize the cost of energy generation and reserve procurement while satisfying constraints that encode various capacity limits, reserve activation for handling discrepancies caused by contingencies and renewables, and probabilistic flow constraints due to line capacity limits. Let $\mathcal{G}, \mathcal{W}, \mathcal{D},$ and $\mathcal{C}$ be the set of conventional generators, renewable sources, loads, and power lines, respectively. Each load $d \in \mathcal{D}$ is associated with a demand $P_d \geq 0$ that is assumed to be known and fixed. Let $\mathcal{C} := \{0, 1, 2, \ldots, N_C\}$ be the set enumerating the possible line and generator contingencies. Each element $c \in \mathcal{C}$ with $c \geq 1$ represents a scenario where one component, either a line or a conventional generator has failed. The element $0 \in \mathcal{C}$ corresponds to the case where the system is intact. We assume that the network remains connected under any line contingency. For notational ease, we divide the set of contingencies $\mathcal{C}$ into ones where a generator fails $C_g \subset \mathcal{C}$ and ones where a line fails $C_l \subset \mathcal{C}$. Note that $\mathcal{C} = C_g \cup C_l \cup \{0\}$ and $C_g \cap C_l = \emptyset$. For $c \in C_g$, the generator under contingency is denoted by $g_c \in \mathcal{G}$. We assume that the problem is solved for dispatch of generation and reserves at a particular time $t$ during the day. We consider two sets of decision variables, ones that are made before and after encountering contingency and uncertainty, respectively. We assume that the problem is solved approximately 15 minutes before observing the actual realization of the renewable generation and the possible contingency. The first set includes the planned power generation for the intact system, denoted $P_g^p, g \in \mathcal{G}$, and two types of reserves that handle generation mismatch due to uncertainty and generator contingency, respectively. For the renewable uncertainty, we denote $r^+_{g} \geq 0, g \in \mathcal{G}$ as the up and down reserve capacities. The values $r^+_{g}$ and $r^-_{g}$ denote the upper and lower bound on the change in generation from the planned value $P_g^p$ that generator $g$ can execute using reserves. Moreover, $r^\text{con}_g \geq 0, g \in \mathcal{G}$ is the reserve capacity of $g$ to handle the change in power due to generator contingencies. We decide on the planned generation $P_g^p$ in such a way that it allows a feasible power flow under any line contingency. Thus, these events do not require reserve capacity. We assume that we have access to the forecast power production $P_w^f$ by each renewable generator $w \in \mathcal{W}$ at the first stage, that is, before the actual uncertainty and contingency is revealed. The actual generation $\hat{P}_w$ usually deviates from the forecast value. To this end, we assume that the deviation between the forecast and the actual generation $P_w^f := P_w - \hat{P}_w$ is a random variable. The collection of these real-valued random variables is denoted as $P_v := (P^f_w)_{w \in \mathcal{W}}$. We assume that $P_v$ has a distribution $\mathbb{P}$ which is supported on a compact support $\Xi \subset \mathbb{R}^{|\mathcal{W}|}$, where $n_w := |\mathcal{W}|$. We assume that the actual renewable generation as well as the contingency is realized before making the second set of decisions which involves deploying reserves to ensure power balance in a probabilistic sense. The deployment of the reserves is done at a faster timescale to quickly restore the power balance. The total deviation of the actual and the forecast renewable generation in the network is represented by $P^\text{mis}_w := \sum_{w \in \mathcal{W}} P^f_w - \hat{P}_w^f$. For the second-stage, we consider two sets of variables. To balance the mismatch due to renewable uncertainty, we let $\alpha^g_{g_c}, c \in C, g \in \mathcal{G}$ be the affine control policies that determine the level of reserve activation. Given the total power mismatch $P^\text{mis}$, the change in power generation of $g$ using reserves under contingency $c \in C$ is $\alpha^g_{g_c} P^\text{mis}$. Thus, these policies define how $P^\text{mis}$ is handled collectively by generators active under contingency $c$. The second set of variables $\delta^c_{g_c}, c \in C, g \in \mathcal{G}$ tackle the mismatch due to contingency. Specifically, if $c \in C_l$, then there is no net power loss, so we set $\delta^c_{g_c} = 0$ for all $g \in \mathcal{G}$. If $c \in C_g$, then the planned power $P_g^p$ is lost as generator $g_c$ fails, and for this case, the change in power generation is denoted by $\delta^c_{g_c}$. Note that for $g_c$, we set $\delta^c_{g_c} = -P_g^p$. Thus, the net power generated under contingency $c \in C$ by an active generator $g$ is $P_g + P^\text{mis} \alpha^g_{g_c} + \delta^c_{g_c}$. This is a random quantity as the mismatch is uncertain. In our formulation, the power balance holds for this eventually planned generation level if it holds for $P_g$ under forecast renewable generation. However, the line constraints might only hold in a probabilistic manner. We next define our main optimization problem and term it as the chance-constrained N-1 security constrained dispatch (CCSD)

$$\min \sum_{g \in \mathcal{G}} (H_g p_g + H^+_g r^+_g + H^-_g r^-_g + H^\text{con}_g r^\text{con}_g)$$

w. r. t. \{ $p_g, r^+_g, r^-_g, r^\text{con}_g \}_{g \in \mathcal{G}}, \{ \alpha^g_{g_c}, \delta^c_{g_c} \}_{g \in \mathcal{G}, c \in \mathcal{C}}$ \quad (1a)

s. t. \sum_{g \in \mathcal{G}} p_g + \sum_{w \in \mathcal{W}} P_w = \sum_{d \in \mathcal{D}} P_d \quad (1b)

$p_g + r^+_g + r^-_g + r^\text{con}_g \leq P^\text{max}_g, \forall g \in \mathcal{G}$. \quad (1c)

\footnote{The number of elements in a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.}
The objective function (1a) represents the total operational cost of the system, where the first term stands for the cost of planned generation, the second and the third take into account the cost of procuring reserves that handle uncertainty, and the last term quantifies the cost of reserves dedicated for generator contingencies. Constraint (1c) ensures that the planned power generation in the first stage satisfies the load along with the forecast renewable generation levels. Constraints (1d) and (1e) impose power generation limits under maximum up and down regulation provided by reserves. Bounds on the acquired reserve capacities are given in (1f). Using (1g), (1h), and (1i), we restrict the control policies that handle uncertainties to a simplex while making sure that the generator under contingency for any \( c \in C_g \) does not participate in power provision. Constraints (1j) to (1n) balance the loss in generation due to generator contingencies. Here, (1j) enforces the post-contingency power to be zero for the failed generator, (1k) guarantees that the power loss is accommodated by other active generators, (1l) ensures that the reserve activation is bounded by the reserve capacity \( r_{\text{con}} \), (1m) imposes the change in generation due to contingency not to be negative, and (1n) makes reserve activation to be zero when there are only line contingencies. Finally, (1o) is a chance constraint containing three sets of inequalities which must be satisfied jointly with high probability \( 1 - \epsilon \). The first two sets of constraints within the chance constraint are for upper and lower reserves assigned to renewable uncertainty. The third constraint implies that in the presence of renewable uncertainty, the line flow must satisfy line limits, where \( M_{gl}^c, M_{wl}^c, M_{dl}^c \) are Power Transfer Distribution Factors (PTDFs) which respectively translate conventional power injections, renewable energy injections, and load demand into line flows. These matrices can be computed, for example, by implementing the method outlined in [2]. In the chance constraint, the parameter \( P_l^{\text{max}} \) denotes the allowed power flow capacity of the line \( l \in L \).

**Remark 11.1. (Modelling choices):** We consider one joint chance constraint (1o) instead of more in order to simplify the modeling and the exposition of our method. The system operator may separately consider two sets of joint chance constraints, one for line flows and the other for reserve activation. Further, the confidence of satisfying these constraints can be selected independently, for example, \( 1 - \epsilon_1 \) and \( 1 - \epsilon_2 \), respectively. Such splitting can incorporate different levels of importance to reserve limits and power flow constraints. For example, increasing \( \epsilon_1 \) possibly allows more frequent violation of line flows but reduces the incurred cost. Similarly, hard constraints on reserve capacity can be imposed by setting \( \epsilon_2 = 0 \). At the end of the next section, we comment on changes to our method that can accommodate the aforementioned splitting of the chance constraint.

Our aim in this paper is to solve the above chance-constrained optimization problem using data. To this end, we replace the chance constraint in problem (1) with its distributionally robust (DR) counterpart and derive a computationally efficient method for approximating the solution of the resulting optimization problem.

### III. TWO-STEP DISTRIBUTIONALLY ROBUST METHOD FOR SOLVING CCSCD

For the sake of convenience, we rewrite the CCSCD problem (1) in the following compact form:

\[
\begin{align*}
\min_c & \quad c^\top x \\
\text{s. t.} & \quad Ax \leq \beta, \\
& \quad P(e_k^\top x + f_k^\top \xi + x^\top F_k \xi \leq h_k, \forall k \in [K]) \geq 1 - \epsilon,
\end{align*}
\]

where \( x \in \mathbb{R}^{n_x} \) stands for the decision variables and \( \xi \in \mathbb{R}^{n_{\xi}} \) denotes the deviation \( P^d \) between the forecast and the actual generation. We use the notation \( [K] := \{1, 2, \ldots, K\} \). Recall that \( \xi \) has the distribution \( P \) and support \( \Xi \). The constraint (2b) captures the deterministic constraints (1c) to (1n) in the formulation (1). It is worth mentioning that deterministic equality constraints in (1) are treated using two inequality constraints in (2b). Note that here, \( K = 2|G|(|G| + |L|) \) is the number of constraints included in the joint chance-constraint (1o) and vectors \( e_k, f_k, h_k \) and matrix \( F_k \) represent in a general form the \( k \)-th constraint.

Motivated by the real-life situation, we assume that the distribution \( P \) is not entirely known and instead only a finite number of samples of \( \xi \) are available. Our approach then is to approximate the solution of (2) with a DR optimization problem using the available data. To this end, let \( \Xi^N := \{\xi_1, \xi_2, \ldots, \xi_N\} \) be the set of \( N \) samples of the uncertainty and let \( \hat{Q}_N := \frac{1}{N} \sum_{m=1}^{N} \delta_{m} \) be the empirical distribution, where \( \delta_{m} \) is the delta function at the point \( \xi_m \). Given a radius \( \theta \geq 0 \) and the dataset, the Wasserstein ambiguity set is given by

\[
\mathcal{M}^\theta_N := \{Q \in \mathcal{P}(\Xi) \mid d_w(Q, \hat{Q}_N) \leq \theta\},
\]

where \( d_w(Q, \hat{Q}_N) \) is the Wasserstein distance.
where \( \mathcal{P}(\Xi) \) is space of all distributions supported on \( \Xi \) and \( d_w \) stands for the Wasserstein metric. The ambiguity set contains all distributions that are at most \( \theta \) away from the empirical distribution in the Wasserstein metric [39].

\[
d_w(Q, \hat{Q}_N) := \min_{\gamma \in \mathcal{H}(Q, \hat{Q}_N)} \int_{\Xi \in \Xi} ||\xi - \hat{\xi}||_p \gamma(d\xi, d\hat{\xi}),
\]

where \( || \cdot ||_p \) represents a norm in the Euclidean space and \( \mathcal{H}(Q, \hat{Q}_N) \) is the set of distributions on the space \( \Xi \times \Xi \) with marginals \( Q \) and \( \hat{Q}_N \). The DR version of (2) based on the ambiguity set \( \mathcal{M}^0_N \) is given as:

\[
\begin{align}
\min_x & \quad c^T x \\
\text{s.t.} & \quad Ax \leq \beta, \\
& \quad \inf_{Q \in \mathcal{M}^0_N} Q(e_k^T x + f_k^T \xi + x^T F_k \xi \leq h_k, \forall k \in [K]) \\
& \quad \geq 1 - \epsilon. 
\end{align}
\]

We assume that the stochastic parameter is agnostic of the number of data points. This proves to be computationally advantageous, as will be discussed further in the numerical example.

In [19], the same two-step approach was proposed but the \( \Xi_{\text{rob}} \) set was restricted to be a hyper-rectangle. This choice was conservative as it could not take into account the correlation present between various renewables. Next, we provide one way of determining \( \Xi_{\text{rob}} \) which overcomes this limitation. Later, we focus on the tractability of the robust problem.

### A. Constructing \( \Xi_{\text{rob}} \)

We assume that the uncertain generation across the network is correlated in a sparse manner, i.e., renewable generators can be divided into subsets and the correlation exists only among generators belonging to one subset. To make this precise, let \( \mathcal{W} = \{w_1, w_2, \ldots, w_{n_W} \} \) be the set of all renewable generators. We partition these into \( n_W \) number of sets \( \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{n_W} \} \) such that \( \mathcal{W}_i \subseteq \mathcal{W} \) for all \( i \in [n_W] \), \( \mathcal{W}_i \cap \mathcal{W}_j = \emptyset \) for all \( i, j \in [n_W] \), and \( \bigcup_{i=1}^{n_W} \mathcal{W}_i = \mathcal{W} \). We assume that uncertain generations of two generators are correlated only if they belong to one partition. Consequently, we assume that the support of every distribution belonging to the ambiguity set \( \mathcal{M}^0_N \) is given as

\[
\Xi = \prod_{i \in [n_W]} \Xi^i, \quad \Xi^i = \{\xi^i \in \mathbb{R}^{|\mathcal{W}_i|} | \Gamma^i \xi^i \leq \rho \},
\]

where \( \Gamma^i \in \mathbb{R}^{u_i \times |\mathcal{W}_i|} \) and \( \rho \in \mathbb{R}^{u_i} \) define the support of the marginal of the distribution over the partition \( \mathcal{W}_i \). Due to correlations restricted to a partition, we aim to find an appropriate polyhedral set \( \Xi_{\text{rob}} \subseteq \mathbb{R}^{|\mathcal{W}|} \) for each of the partitions by solving a DR problem. Then, we construct \( \Xi_{\text{rob}} := \bigcap_{i=1}^{n_W} \Xi^i \). This procedure is summarized in Algorithm 1. Its detailed description is given below.

**[Informal description of Algorithm 1]:** The procedure admits as input the dataset \( \Xi^N \), violation level \( \epsilon \), radius \( \theta \), and partition \( \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{n_W} \} \). For each sample \( \hat{\Xi}_m \), the components corresponding to the partition \( \mathcal{W}_i \) are collected in the vector (\( \hat{\xi}_m \)). The algorithm starts with computing the empirical covariance matrix \( \text{Cov} \in \mathbb{R}^{|\mathcal{W}| \times |\mathcal{W}|} \) for each partition using the samples (Line 2). Next, eigenvalues \( \{\lambda_{(i,1)}, \ldots, \lambda_{(i,|\mathcal{W}_i|)}\} \) and eigenvectors \( \{v_{(i,1)}, \ldots, v_{(i,|\mathcal{W}_i|)}\} \) of \( \text{Cov} \) are computed (Line 3), where without loss of generality, we assume that the eigenvalues are arranged in an ascending manner. Our idea of constructing \( \Xi_{\text{rob}} \) involves intersecting a number of simple sets, each consisting of a set of points lying between two hyperplanes that share a common normal. These simple sets can be divided into two groups, one where the normal is aligned with one of the axis of the Euclidean space where \( \Xi_{\text{rob}} \) lies, i.e., \( \mathbb{R}^{|\mathcal{W}_i|} \) and the other with the normal being the eigenvectors of \( \text{Cov} \) with the smallest eigenvalues. To this end, we select a number \( 0 \leq \kappa_i < |\mathcal{W}_i| \) determining the number of eigenvector-based simple sets to be constructed. An illustration of this construction in two dimensions is given in Fig. 1. We set \( \kappa_i^{\text{DR}} := \kappa_i + |\mathcal{W}_i| \) as the number of sets that will be intersected to determine \( \Xi_{\text{rob}} \).
$$\Xi_{\text{rob}} = \bigcap_{j=1}^{n_{\text{DR}}} \Xi_{i,j}^{\text{rob}}.$$  

where each set $\Xi_{i,j}^{\text{rob}}$ is of the form

$$\Xi_{i,j}^{\text{rob}} := \left\{ \xi^i \in \mathbb{R}^{|W_i|} \mid b_{i,j} \preceq \zeta_{i,j}^\top \xi^i \preceq \bar{b}_{i,j} \right\}. \quad (10)$$

For the above set, the vector $\zeta_{i,j} \in \mathbb{R}^{|W_i|}$ determines the normal to the set of hyperplanes $\zeta_{i,j}^\top \xi^i = b_{i,j}$ and $\zeta_{i,j}^\top \xi^i = \bar{b}_{i,j}$ that construct $\Xi_{i,j}^{\text{rob}}$. If $1 \leq j \leq |W_i|$, then we set $\zeta_{i,j} = \zeta_{(i,j)}$, and otherwise $\zeta_{i,j} = \zeta_{(i,j-|W_i|)}$ is the eigenvector of the covariance matrix corresponding to the eigenvalue $\lambda_{(i,j-|W_i|)}$. The bounds $b_{i,j}, \bar{b}_{i,j}$ are obtained as the solutions of the DR optimization problem (11). Finally, we set

$$\Xi_{\text{rob}} = \prod_{i=1}^{n_W} \Xi_{i,1}^{\text{rob}},$$

as the intersection of the half-spaces defining each $\Xi_{i,1}^{\text{rob}}$. A given partition $i \in \{1, \ldots, n_W\}$ and normal vector $\zeta_{i,j}$ consider the following DR problem

$$\min_{(\bar{b}, \bar{b})} \bar{b} - \bar{b}.$$  

s. t.  

$$b_{i,j} \preceq \bar{b}_{i,j}, \quad \zeta_{i,j}^\top \bar{b}_{i,j} \preceq \bar{b}_{i,j} \preceq \zeta_{i,j}^\top \bar{b}_{i,j}. \quad (11a)$$

$$\inf_{Q \in \mathcal{M}_N^\theta} Q \left( \max \left\{ \zeta_{i,j}^\top \xi^i - \bar{b}_{i,j} - \zeta_{i,j}^\top \xi^i \right\} \leq 0 \right) \geq 1 - \frac{\epsilon}{n_{\text{DR}}}. \quad (11b)$$

$$\inf_{Q \in \mathcal{M}_N^\theta} Q \left( \max \left\{ \zeta_{i,j}^\top \xi^i - \bar{b} - \zeta_{i,j}^\top \xi^i \right\} \leq 0 \right) \geq 1 - \frac{\epsilon}{n_{\text{DR}}}. \quad (11c)$$

In the above problem, the constraint (11c) enforces that the probability mass placed on the set $\{ \xi^i \mid \bar{b} \preceq \zeta_{i,j}^\top \xi^i \preceq \bar{b} \}$ for any feasible point $(\bar{b}, \bar{b})$ is no less than $1 - \epsilon/n_{\text{DR}}$ under any distribution in $\mathcal{M}_N^\theta$. This constraint eventually ensures that we identify the right set $\Xi_{\text{rob}}$. To give some intuition, the uncertainty set will lie between the two parallel hyperplanes described as $\zeta_{i,j}^\top \xi^i = \bar{b}$ and $\zeta_{i,j}^\top \xi^i = b$. The above problem is difficult to solve in practice as is. Below we provide a reformulation that helps in computation. Subsequently, we provide a heuristic to find an approximate optimizer of the problem.

**Proposition III.1.** (Finite-dimensional reformulation of (11)): Assume that the support of each distribution in $\mathcal{M}_N^\theta$ is of the form (8). Consider problem (11) written for some partition $i \in [n_W]$ and simple set $j \in [n_{i,\text{DR}}]$ of a given normal $\zeta \in \mathbb{R}^{|W_i|}$ and a constant $b \in \mathbb{R}$, we define the notation

$$\mathcal{H}_i(\zeta, b) := \Xi \cap \{ \xi^i \mid \zeta^\top \xi^i = b \}. \quad (12)$$

as the intersection of $\Xi$ and the hyperplane formed with $(\zeta, b)$. Then, the optimization problem (11) is equivalent to

$$\min_{\bar{b}, \bar{b}} \bar{b} - \bar{b} \quad \text{w. r. t.} \quad \bar{b}, \bar{b}, \lambda \geq 0, \left\{ s_m \geq 0 \right\}_{m=1}^N \quad (13a)$$

$$s_\zeta = \zeta \leq \bar{b}, \quad \text{s. t.} \quad \bar{b} - \bar{b}, \lambda \geq 0, \left\{ s_m \geq 0 \right\}_{m=1}^N \quad (13b)$$

$$\lambda \theta + 1/N \sum_{m=1}^N s_m \leq \epsilon/n_{\text{DR}}, \quad (13c)$$

for all $m \in [N]$:

$$s_m \geq 1 - \lambda \text{dist} \left( \left( \xi_m \right), \mathcal{H}_i(\zeta_{i,j}, \bar{b}) \right), \quad (13e)$$

$$s_m \geq 1 - \lambda \text{dist} \left( \left( \xi_m \right), \mathcal{H}_i(\zeta_{i,j}, b) \right), \quad (13f)$$

where $\text{dist} \left( \left( \xi_m \right), \mathcal{H}_i(\zeta_{i,j}, \bar{b}) \right)$ is the distance between the sample $(\xi_m)$ and the set $\mathcal{H}_i(\zeta_{i,j}, \bar{b})$ under the norm $\| \cdot \|$ and the

**Algorithm 1: Constructing $\Xi_{\text{rob}}$.**

**Data:** Data $\Xi \equiv \{ \xi^1, \xi^2, \ldots, \xi^N \}$, partition $\{ W_1, W_2, \ldots, W_n \}$, Wasserstein radius $\theta$, and tolerance $\epsilon$

1. for $i \in [n_W]$ do

2. Compute the empirical covariance matrix $\hat{\text{Cov}}^i$ for each partition $W_i$.

3. Compute eigenvectors and eigenvalues of $\hat{\text{Cov}}^i$.

4. Select $\kappa_i \in \{0, \ldots, |W_i| - 1\}$ to be the number of eigenvectors to be used to find hyperplane sets.

5. Set $\kappa_i^{\text{DR}} := \kappa_i + |W_i|$.

6. end

7. for $i \in [n_W]$ do

8. Construct $\Xi_{i,1}^{\text{rob}}$ as (9) using definition (10) and solving (11).

9. end

10. Construct $\Xi_{\text{rob}} = \bigcap_{i=1}^{n_W} \Xi_{i,1}^{\text{rob}}$. 


notation \( \text{dist}(\hat{\xi}_m)^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b}) \) is analogous. These distances can be computed by solving the following optimization problem

\[
\text{dist} \left( \hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b}) \right) = \left\{ \begin{array}{ll}
\max_{y,z} & (b - \zeta_{i,j}^i)^\top y - (\rho^i - \Gamma^i(\hat{\xi}_m)^i)^\top z \\
\text{s.t.} & \|\zeta^i - (\Gamma^i)^\top z\|_\infty \leq 1, \\
& y \geq 0, z \geq 0, \end{array} \right.
\]  

(14)

where \( \| \cdot \|_\infty \) is the dual norm of \( \| \cdot \|_p \) used in (4).

**Proof:** The key step involves reformulating (11c). Recall that for a point \( \xi \in \mathbb{R}^n \), we use the notation \( \xi = (\xi^1, \xi^2, \ldots, \xi^n) \) to represent the components belonging to each of the partitions. That is, \( \xi^i \in \mathbb{R}^{w_i} \) consists of components corresponding to the partition \( i \). Denote \( \hat{b} = (\hat{b}^i, \hat{b}) \) and we consider \( Z(\hat{b}, \xi) := \max\{\xi_{i,j}^i - \hat{b}^i, \hat{b} - \xi_{i,j}^i \} \). Then, we compute

\[
\sup_{\hat{b} \in \mathbb{M}_R^N} Q(Z(\hat{b}, \xi) > 0) = \sup_{\hat{b} \in \mathbb{M}_R^N} E_{\xi} \left\{ \hat{\xi} \in \Xi | Z(\hat{b}, \xi) > 0 \right\}
\]

where the equality follows from [40, Proposition 4]. Following from [40, Proposition 1], it is equivalent to

\[
\inf_{\lambda \geq 0, s} \frac{1}{N} \sum_{m=1}^{N} s_m
\]

s.t. 
\[
s_m \geq \sup_{\xi \in \Xi_1} \left\{ \sup_{\xi \in \Xi_2} \left( 1 - \|\xi - \xi_m\| \right) + \sup_{\xi \in \Xi_1} \left( -\|\xi - \hat{\xi}_m\| \right) \right\}.
\]  

(15)

Note that since the function \( Z \) only depends on the component \( \xi^i \) of the vector \( \xi \), we can write

\[
\Xi_1 = \prod_{k=1}^{i-1} \Xi^k \times \Xi^i_1 \times \prod_{k=i+1}^{n_i} \Xi^k,
\]

\[
\Xi_2 = \prod_{k=1}^{i-1} \Xi^k \times (\Xi^i \setminus \Xi^i_1) \times \prod_{k=i+1}^{n_i} \Xi^k,
\]  

(16)

where we use the notation

\[
\Xi^i_1 := \{ \xi^i \in \Xi^i | Z(\hat{b}, \xi^i) > 0 \},
\]

with a slight abuse of notation that now \( Z(\hat{b}, \xi^i) = \max\{\xi_{i,j}^i - \hat{b}^i, \hat{b} - \xi_{i,j}^i \} \). Using the decomposition (16) in (15), we get

\[
s_m \geq \max\left\{ \sup_{\xi \in \Xi_1} \left( 1 - \|\xi^i - \hat{\xi}_m^i\| \right), \sup_{\xi \in \Xi_2} \left( -\|\xi^i - \hat{\xi}_m^i\| \right) \right\}.
\]  

(17)

To handle the terms inside the \max{} operator, note that if \( \hat{\xi}_m^i \in \Xi_1^{i} \) the first term is 1 while the second term is non-positive, hence \( s_m = 1 \). Similarly, if \( \hat{\xi}_m^i \in \Xi_2^{i} \), the second term is zero. Thus, we can rewrite (17) as \( s_m \geq \max\{0, 1 - \inf_{\xi \in \Xi_1^{i}} \lambda \|\xi^i - \hat{\xi}_m^i\| \} \).

Following the definition of \( \Xi_1 \), we have

\[
\inf_{\xi \in \Xi_1^{i}} \lambda \|\xi^i - \hat{\xi}_m^i\| \leq \min \{ \text{dist}(\hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b})), \text{dist}(\hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b})) \},
\]

where \( \mathcal{H}_i \) is given in (12). Using the above relation, we write (18) equivalently as \( s_m \geq 0, s_m \geq 1 - \lambda \text{dist}(\hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b})) \), and \( s_m \geq 1 - \lambda \text{dist}(\hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b})) \). This shows the equivalence between (11) and (13). The expression for \( \text{dist}(\hat{\xi}_m^i, \mathcal{H}_i(\zeta_{i,j}, \hat{b})) \) for a given \( \xi \) as given in (14) is established in [29, Lemma 1]. This completes the proof. □

**Remark 3.2.** (Approximating the solution of (11) using Proposition 3.1): While Proposition 3.1 provides a finite-dimensional exact reformulation of (11), the resulting optimization (13) is difficult to solve as (13e) and (13f) are bilinear in nature if one uses the formulation (14). Therefore, we provide a heuristic way of solving (13). We implement two algorithms. The first algorithm consists of finding lower and upper bounds for both \( \hat{b} \) and \( \hat{b} \) such that two conditions hold: (a) collectively the lower bound \( \text{lb}_{\text{feas}} \) of \( \hat{b} \) and upper bound \( \text{ub}_{\text{feas}} \) of \( \hat{b} \) are feasible for (13), and (b) the upper bound \( \text{lb}_{\text{feas}} \) of \( \hat{b} \) and lower bound \( \text{ub}_{\text{feas}} \) of \( \hat{b} \) are infeasible for (13). The second algorithm then performs a bisection search independently over intervals \((\text{lb}_{\text{feas}}, \text{ub}_{\text{feas}})) \) and \((\text{ub}_{\text{feas}}, \text{lb}_{\text{feas}})) \) to find values \( \hat{b} \) and \( \hat{b} \), respectively. The bisection procedure involves checking the midpoint of an interval for feasibility of constraints (13d) to (13f). If the point is feasible or infeasible, then the interval is contracted by upper or lower value of the interval. Note that for any given points \( \hat{b} \) and \( \hat{b} \), checking for feasibility of constraints (13d) to (13f) involves solving a convex program. This follows from the fact that all distances in (13e) and (13f) can be computed by solving problem (14). Note that instead of appealing to the above heuristic, one can use (14) in constraints (13e) and (13f) to form a bilinear optimization problem and solve it as is using methods tailor-made for bilinear problems. In the interest of space, we do not provide further details regarding it.

The following result summarizes the property of \( \Xi_{\text{rob}} \).

**Proposition 3.3.** (Guarantees for \( \Xi_{\text{rob}} \)): Let \( \Xi_{\text{rob}} \) be the output of Algorithm 1. Then, we have

\[
\inf_{Q \in \mathcal{M}_B^N} Q(\xi \in \Xi_{\text{rob}}) \geq 1 - \epsilon.
\]  

(19)

**Proof:** Considering \( \Xi_{\text{rob}} = \bigcap_{j=1}^{n_i} \bigcup_{j=1}^{d_{\text{rob}}} \Xi_{i,j} \) and the fact that the complement of the intersection of sets is the same as the union of their complements, we obtain the bound:

\[
\inf_{Q \in \mathcal{M}_B^N} Q(\xi \in \Xi_{\text{rob}}) = 1 - \sup_{Q \in \mathcal{M}_B^N} Q(\xi \notin \Xi_{\text{rob}}) = 1 - \sup_{Q \in \mathcal{M}_B^N} Q(\xi \notin \bigcup_{j=1}^{n_i} \bigcup_{j=1}^{d_{\text{rob}}} \Xi_{i,j}^{i,j}).
\]
In the above expression, (a) follows from the union bound, and (b) is due to (11c).

B. Robust Optimization Using \( \Xi_{\text{rob}} \)

Having constructed the set \( \Xi_{\text{rob}} \) in the previous section, we propose solving the following robust optimization problem to approximate the solution of the DR problem (5):

\[
\begin{align*}
\min_{x} & \quad c^\top x \\
\text{s.t.} & \quad Ax \leq \beta, \\
& \quad e_k^\top x + f_k^\top \xi + x^\top F_k \xi \leq h_k, \quad \forall \xi \in \Xi_{\text{rob}}, k \in [K].
\end{align*}
\]

Proposition III.4. (Reformulation of (20) with polyhedral uncertainty set): The optimization problem (20) with the uncertainty set given in (21) is equivalent to the following problem

\[
\begin{align*}
\min_{x, \{y_k\}_{k \in [K]}} & \quad c^\top x \\
\text{s.t.} & \quad Ax \leq \beta, \\
& \quad g_k^\top y_k + e_k^\top x \leq h_k, \quad \forall k \in [K], \\
& \quad F_k^\top x - G_k^\top y_k + f_k = 0, \quad \forall k \in [K], \\
& \quad y_k \in \mathbb{R}^q_{\geq 0}, \quad \forall k \in [K].
\end{align*}
\]

Proposition III.5. (Approximating (5) with (20)): Let \( x^* \) be an optimal solution of problem (20), where \( \Xi_{\text{rob}} \) is constructed using Algorithm 1. Then, \( x^* \) is a feasible solution of the DR problem (5).

Proof: Note that from Lemma III.3, the set \( \Xi_{\text{rob}} \) has the property (19). Trivially, \( x^* \) satisfies (20c) for all \( \xi \in \Xi_{\text{rob}} \). Thus, constraint (5c) is satisfied, which means \( x^* \) is a feasible solution of the DR problem (5).

As a consequence of the above result, if the true probability distribution \( \mathbb{P} \) belongs to the ambiguity set \( \mathcal{M}_N^\theta \) with at least probability \( p \) for some \( p \in (0, 1) \), then \( x^* \) satisfies the joint chance-constraint (10) and so is feasible for problem (1) with at least probability \( p \). When samples are i.i.d, then the radius can be tuned to achieve a certain confidence \( p \). Next, we present such a relationship. We note that the most popular result of this kind in the literature, see [39] are written for general distributions and involve constants that are difficult to compute. Our result is based on [41, Corollary 25] that uses the fact that the support of distributions is compact.

Proposition III.6. (Confidence of satisfying constraints of (1) by a solution of (20)): Assume that the samples in \( \Xi_N^\gamma \) are drawn in an i.i.d manner from \( \mathbb{P} \) and \( \eta_q > 2p \), where \( p \) defines the norm used in (4). Let \( \gamma \in (0, 1) \) define a confidence level and the radius of the ambiguity set be selected as

\[
\theta_N(\gamma) := 2\text{diam}(\Xi) \left( \frac{\ln(C* \gamma^{-1})}{C*} \right)^{\frac{1}{2}} N^{-\frac{1}{2}},
\]

where \( \text{diam}(\Xi) \) is the diameter of the support \( \Xi \) in \( \infty \)-norm given by \( \text{diam}(\Xi) = \max_{x,y \in \Xi} \|x - y\|_{\infty} \) and \( C* := C \frac{d^d}{d^d} \) with

\[
C := \sqrt{d(2d-2)/(2p)} \left( \frac{1}{1 - 2(p-\eta_q)^2/2} + \frac{1}{1 - 2p} \right)^{1/p}.
\]

Let \( x^* \) be an optimal solution of problem (20), where \( \Xi_{\text{rob}} \) is constructed using Algorithm 1. Then, the solution \( x^* \) satisfies constraint (2b) almost surely and the chance-constraint (2c) with probability \( 1 - \gamma \).

In reality, the radius that we obtain from the above guaranteed confidence bound might be too large. This is due to the limitation of the theoretical bound. Moreover, the samples might not be drawn in an i.i.d manner from the distribution. Therefore, to overcome these factors, it would be better to tune the radius in an empirical way by studying the trade-off between cost and constraint satisfaction for different radii. We demonstrate this in the next section.

We next comment about using our method for a more general dispatch problem. Referring back to Remark II.1, we note that if the reserve capacity and flow constraints are split into two different joint chance-constraints with confidences \( 1 - \epsilon_g \) and \( 1 - \epsilon_f \), respectively, then one would construct two sets \( \Xi_{\text{rob}}(\epsilon_g) \) and \( \Xi_{\text{rob}}(\epsilon_f) \) in the first step of our method. For constructing these sets, \( \epsilon_g \) and \( \epsilon_f \) are used as parameters in Algorithm 1, respectively. In the second step of the method, the constraints using this equivalence in (23c) yields the conclusion. □

The following result summarizes the properties of our proposed method.

Using this equivalence in (23c) yields the conclusion. □

The following result summarizes the properties of our proposed method.

Proposition III.5. (Approximating (5) with (20)): Let \( x^* \) be an optimal solution of problem (20), where \( \Xi_{\text{rob}} \) is constructed using Algorithm 1. Then, \( x^* \) is a feasible solution of the DR problem (5).

Proof: Note that from Lemma III.3, the set \( \Xi_{\text{rob}} \) has the property (19). Trivially, \( x^* \) satisfies (20c) for all \( \xi \in \Xi_{\text{rob}} \). Thus, constraint (5c) is satisfied, which means \( x^* \) is a feasible solution of the DR problem (5).

As a consequence of the above result, if the true probability distribution \( \mathbb{P} \) belongs to the ambiguity set \( \mathcal{M}_N^\theta \) with at least probability \( p \) for some \( p \in (0, 1) \), then \( x^* \) satisfies the joint chance-constraint (10) and so is feasible for problem (1) with at least probability \( p \). When samples are i.i.d, then the radius can be tuned to achieve a certain confidence \( p \). Next, we present such a relationship. We note that the most popular result of this kind in the literature, see [39] are written for general distributions and involve constants that are difficult to compute. Our result is based on [41, Corollary 25] that uses the fact that the support of distributions is compact.

Proposition III.6. (Confidence of satisfying constraints of (1) by a solution of (20)): Assume that the samples in \( \Xi_N^\gamma \) are drawn in an i.i.d manner from \( \mathbb{P} \) and \( \eta_q > 2p \), where \( p \) defines the norm used in (4). Let \( \gamma \in (0, 1) \) define a confidence level and the radius of the ambiguity set be selected as

\[
\theta_N(\gamma) := 2\text{diam}(\Xi) \left( \frac{\ln(C* \gamma^{-1})}{C*} \right)^{\frac{1}{2}} N^{-\frac{1}{2}},
\]

where \( \text{diam}(\Xi) \) is the diameter of the support \( \Xi \) in \( \infty \)-norm given by \( \text{diam}(\Xi) = \max_{x,y \in \Xi} \|x - y\|_{\infty} \) and \( C* := C \frac{d^d}{d^d} \) with

\[
C := \sqrt{d(2d-2)/(2p)} \left( \frac{1}{1 - 2(p-\eta_q)^2/2} + \frac{1}{1 - 2p} \right)^{1/p}.
\]

Let \( x^* \) be an optimal solution of problem (20), where \( \Xi_{\text{rob}} \) is constructed using Algorithm 1. Then, the solution \( x^* \) satisfies constraint (2b) almost surely and the chance-constraint (2c) with probability \( 1 - \gamma \).

In reality, the radius that we obtain from the above guaranteed confidence bound might be too large. This is due to the limitation of the theoretical bound. Moreover, the samples might not be drawn in an i.i.d manner from the distribution. Therefore, to overcome these factors, it would be better to tune the radius in an empirical way by studying the trade-off between cost and constraint satisfaction for different radii. We demonstrate this in the next section.
pertaining to reserves and flows would be imposed in a robust manner for sets $\Xi_{\text{rob}}(\epsilon_g)$ and $\Xi_{\text{rob}}(\epsilon)$, respectively. Another generalization of the problem is to consider dispatch over a time horizon $T = \{t_1, \ldots, t_T\}$. In this case the size of the uncertainty will be the product of the number of renewables and the length of the horizon $T$. Consequently, assuming that the generation of a renewable across the horizon belongs to the same partition, one can follow the same two-steps as in the single-shot case. Computationally, we note that the number of distributionally robust problems to be solved in the first-step, the size of each DR problem, and the size of the robust optimization in the second step all will get multiplied by the factor $|T|$.

IV. NUMERICAL RESULTS

We test our approach on a modified version of the IEEE RTS 24-bus reference case system [42] with 12 conventional generation units (see Fig. 2). Note that at node 15 and 23 two generators are located but for the sake of simplicity they are not depicted separately. The network also includes six wind generation units with each having capacity of 200 MW. The total demand in the system is 2500 MW which is distributed across 17 locations with the distribution factors being same as those of the first time interval in the test case [42]. The system operator aims to satisfy the demand using conventional and wind generation units in a cost-optimal manner while satisfying (with high probability) all capacity and flow constraints. That is, we wish to solve problem (1). We consider 45 contingencies including all possible line and generator contingencies except the contingency of the line that connects bus 7 and 8 as it disconnects the network. The cost coefficients, the minimum and maximum generation capacities, and maximum reserve capacities of the conventional generators are given in Table I. The reactances and capacities of power lines ($P_l^{\text{MAX}}$) are taken as the values given for the standard IEEE 24 bus system [42]. The PTDF matrices ($M_{\text{gl}}, M_{\text{wl}}, M_{\text{dl}}$) are calculated based on [43]. We aim to achieve less than 5% violation of the power flow and reserve capacity constraints, so we set $\epsilon = 0.05$. In this case wind generators are divided into two different partitions. Wind generators located at bus 3, 5, and 7 belong to one partition, and wind generators located at bus 16, 21, and 23 are in the other partition. The forecast for each generator is 100 MW. The random variable, that is the deviation between the forecast and the actual generation, is assumed to be sampled from a truncated multivariate normal distribution with zero mean and truncation interval being $[-20, 20]$. For each partition, the covariance matrix consists of all diagonal entries being 20 and all off-diagonal entries as 16. We assume that the dispatch decision is made around 15 min before the actual time for which the demand is to be met. This justifies the assumption of the bounded support of the deviation between the forecast and the actual realization. Note that technically wind generators are capable of generating 0 MW to 200 MW which is regarded as their physical bounds. The above data fully specifies the optimization problem (1).

We assume that the system operator has access to the samples of the uncertainty in form of historical data. We compare our proposed method with four others that are employed to approximate the solution of the CCSCD problem (1). These are explained below:

1) Worst-Case (WC): For the worst-case, we solve the CCSCD problem (1) with the chance-constraint (1o) replaced with robust constraint, where the constraints in (1o) are required to hold for all values of the uncertainty.
2) Scenario: This is the well known scenario method [11], where the constraints in (1o) are imposed for a set of samples of the uncertainty.
3) DR optimization using CVaR approximation (CVaR): Here we form a convex inner-approximation of DR chance constraints using DR CVaR constraints. The problem is solved by reformulating it into a finite-dimensional problem using the ideas in [19].

TABLE I

| Unit # | Node | $P_l^{\text{MAX}}$ | $P_l^{\text{MIN}}$ | $M_{\text{gl}}$ | $M_{\text{wl}}$ | $M_{\text{dl}}$ | $H_{\text{gl}}$ | $H_{\text{wl}}$ | $H_{\text{dl}}$ |
|-------|------|-------------------|-------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1     | 1    | 150              | 10               | 0.4            | 40             | 60             | 13.32         | 15             | 11             |
| 2     | 2    | 152              | 10               | 0.4            | 40             | 60             | 15.32         | 15             | 11             |
| 3     | 3    | 350              | 75               | 70             | 20.7           | 24             | 16             |                |                |
| 4     | 4    | 591              | 206.85           | 180            | 20.93          | 25             | 17             |                |                |
| 5     | 5    | 60               | 12               | 60             | 26.11          | 28             | 23             |                |                |
| 6     | 6    | 155              | 54.25            | 30             | 10.52          | 16             | 7              |                |                |
| 7     | 7    | 16               | 54.25            | 30             | 10.52          | 16             | 7              |                |                |
| 8     | 8    | 400              | 100              | 0              | 6.02           | 0              | 0              |                |                |
| 9     | 9    | 21               | 400              | 0              | 5.47           | 0              | 0              |                |                |
| 10    | 10   | 22               | 300              | 0              | 0              | 0              | 0              |                |                |
| 11    | 11   | 23               | 310              | 108.5          | 60             | 10.32          | 14             | 8              |                |
| 12    | 12   | 23               | 350              | 140            | 40             | 10.89          | 16             | 8              |                |

Fig. 2. The modified IEEE RTS 24-bus system with 12 conventional generators and 6 wind generation units illustrated in light blue. The 17 load locations are represented by arrows.
4) Robust Optimization with DR Box Uncertainty Set (RO+BoxSet): This is the approach proposed in [19] where the set $\Xi_{\text{rob}}$ used in the robust optimization (20) is restricted to be a hyper-rectangle. This is a special case of our method where the correlation among the wind generators is neglected.

5) Robust Optimization with DR Polyhedral Uncertainty Set (RO+PolySet): This is our proposed method, where $\Xi_{\text{rob}}$ is constructed using Algorithm 1.

For the above methods, we compare the cost incurred at the obtained solution and the probability of violating the constraints. The later quantity is obtained by considering a large validation dataset of uncertainty and computing the fraction of samples for which at least one constraint is violated. Note that in the worst-case method, since constraints need to hold for all values of the uncertainty, the violation frequency of the obtained optimal solution is zero. We consider the worst-case cost as a benchmark to compare the costs in different methods. All computations are carried out in MATLAB using CVX and MOSEK, on a personal computer with 16 GB of memory.

In our experiments, we draw 50 samples of the uncertainty from the specified distribution in an i.i.d manner and solve the problem (1) considering two different settings. In the first setting, we neglect all the contingencies to reduce the problem size and we consider 1-norm to define the Wasserstein distance. This allows us to draw a comparison with all methods, including the CVaR. In the second setting, where we consider contingencies and 2-norm norm for Wasserstein distance, the CVaR method could not be implemented as the memory requirements were beyond the allowable limit.

In the first setting, we consider 100 MW extra generation capacity and 10 MW extra reserve capacity for all the conventional generators compared to the values reported in Table I. This adjustment is required to make the CVaR approach feasible for larger radii. We consider the same set of samples and different radii of ambiguity set when implementing the CVaR, RO+BoxSet, and RO+PolySet. We performed this experiment 50 times. The average values and variances for cost and violation frequency are reported in Fig. 3. We observe that for both cost and violation frequency, RO+BoxSet and RO+PolySet have similar performance. Fig. 3(a) shows that for the smaller $\theta$ values, CVaR costs less than RO+BoxSet and RO+PolySet while, based on Fig. 3(b), the violation frequency is higher. Larger values of $\theta$ reduce the violation frequency at the expense of higher cost for all three methods, while CVaR suffers more. All the methods can be compared regardless of the radius using the Pareto plot given in Fig. 4. This demonstrates the trade-off between the cost and violation and can be used as a lookup table by the decision maker to tune the radius $\theta$. As is evident, for the same violation frequency, the CVaR incurs a higher cost as compared to the other two methods.

In the second setting, we solve problem (1) with all possible contingencies using Scenario, RO+BoxSet, and RO+PolySet methods, where for the latter two methods different radii of ambiguity sets are considered for the same set of samples.

Fig. 3. Operational cost and violation frequency of constraints for the first setting where all contingencies are neglected. The ambiguity sets are defined for $\theta = \{0.0001, 0.005, 0.01, 0.05\}$ and the experiment is repeated 50 times. For each run, the problem is solved with a set of 50 correlated samples. The lines and the shaded regions represent the average values and the standard deviation, respectively.

Fig. 4. Pareto front depicting the trade-off between cost and constraint violation for the first setting where no contingencies are considered. Different points stand for different radii of the ambiguity set. The larger value of the radius results in a higher cost and lower violation. The Scenario method does not have such a tuning parameter and hence is depicted as a point.
The shaded regions represent the standard deviation of the results around the average results.

Fig. 6. Pareto front for the second setting. Each line plots the cost and the violation frequency at different radii. Larger value of the radius results in higher cost and lower violation. The Scenario method does not have such a tuning parameter and hence is depicted as a point.

We repeat this experiment 50 times. The average values and variances for cost and violation frequency are reported in Fig. 5. Fig. 5(a) shows that for smaller values of $\theta$, our method RO+PolySet costs less compared to RO+BoxSet while the violation frequency, as shown in Fig. 5(b), does not change significantly. Therefore, it shows clearly that keeping the safety guarantees same, the our method is less conservative than RO+BoxSet for small values of $\theta$. It also shows that both these DR methods result in additional cost and less violation frequency as compared to the Scenario approach, as expected. The advantage of our method is even more clear in Fig. 6 where we plot the trade-off between cost and violation frequency for different value of radii as a Pareto plot. This figure indicates that with the same cost, RO+PolySet guarantees lesser violation frequency as compared to RO+BoxSet. This plot gives the system operator a practical guideline of tuning the radius of the ambiguity set based on the level of importance given to cost-optimality and constraint satisfaction. Increasing $\theta$ results in higher costs but safer dispatch decisions. In Fig. 7, we show the Pareto front for two other representative days. For the first one, we consider the total demand as 22500 MW and the forecast values as 130 MW. For the second representative day, we consider 2500 MW of total demand with 70 MW forecast values. Both cases behave...
the same way, and the trade-off between costs and violation frequency is apparent.

We next comment about the computational burden incurred by each method, as outlined in Table II. For smaller number of samples the Scenario approach performs better in terms of computational time compared to the other approaches. However, we recall from earlier discussion that this method performs poorly in terms of constraint violation. For RO+BoxSet and RO+PolySet the variables and constraints mentioned in the table correspond to the robust optimization problem, however, the runtime consists of the time taken to solve the robust as well as all the DR problems. For 50 samples, the runtime for both DR methods is of the same order as the Scenario. However, for 200 samples, the runtime of Scenario is higher than the two DR approaches. Comparing RO+BoxSet and RO+PolySet, the former has lower number of variables and constraints and so the runtime is lower than the latter. This computational benefit that RO+BoxSet enjoys comes at the cost of the solution being more conservative when compared to the solution of RO+PolySet, as explained previously. Finally, we comment that the DR problem (13) that is solved in a heuristic way and computational times depends on the heuristic algorithm.

V. CONCLUSION

We studied chance-constrained security-constrained dispatch and developed an algorithm to solve its DR counterpart where Wasserstein ambiguity sets were used. An attractive feature of our algorithm is the computational ease of approximating a solution, even for large networks. This was facilitated by adopting a two-step approach of solving the DR problem. We demonstrated the cost-robustness trade-offs of our results on stylized IEEE test case. In future, we wish to explore distributed algorithms to solve the DR problems formulated here. We also aim to reduce the conservativeness of our approach by adopting a one-step procedure that combines uncertainty quantification and robust optimization.

ACKNOWLEDGMENT

The last author would like to thank Dr. B.K. Poolla, Dr. A.R. Hota, and Dr. S. Bolognani for useful discussions regarding the optimization problem and the algorithmic framework.

REFERENCES

[1] P. Kundur et al., “Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions,” IEEE Trans. Power Syst., vol. 19, no. 3, pp. 1387–1401, Aug. 2004.

[2] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, “A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration,” IEEE Trans. Power Syst., vol. 28, no. 4, pp. 3885–3896, Nov. 2013.

[3] K. Sundar, H. Nagaregian, L. Roald, S. Misra, R. Bent, and D. Bienstock, “Chance-constrained unit commitment with N-1 security and wind uncertainty,” IEEE Trans. Control Netw. Syst., vol. 6, no. 3, pp. 1062–1074, Sep. 2019.

[4] H. Sharifzadeh, N. Amjadi, and H. Zareipour, “Multi-period stochastic security-constrained OPF considering the uncertainty sources of wind power, load demand and equipment unavailability,” Electric Power Syst. Res., vol. 146, pp. 33–42, 2017.

[5] M. Vrakopoulou et al., “A unified analysis of security-constrained OPF formulations considering uncertainty, risk, and controllability in single and multi-area systems,” in Proc. IEEE IREP Symp. Bulk Power Syst. Dyn. Control – IX Optim., Secur. Control Emerg. Power Grid, 2013, pp. 1–19.

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

[7] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, “Analytical reformulation of security constrained optimal power flow with probabilistic constraints,” in Proc. IEEE Grenoble Conf., 2013, pp. 1–6.

[8] Y. Ding, T. Mostyn, and M. D. McCalloch, “Distributionally robust joint chance-constrained optimization for networked microgrids considering contingencies and renewable uncertainty,” IEEE Trans. Smart Grid, vol. 13, no. 3, pp. 2467–2478, May 2022.

[9] Y. Yu et al., “Transmission contingency-constrained unit commitment with high penetration of renewables via interval optimization,” IEEE Trans. Power Syst., vol. 32, no. 2, pp. 1410–1421, Mar. 2017.

[10] L. Yang, C. Zhang, and J. Jian, “A parallel method for solving the DC security constrained optimal power flow with demand uncertainties,” Int. J. Elect. Power Energy Syst., vol. 102, pp. 171–178, 2018.

[11] G. C. Calafiore and M. C. Campi, “The scenario approach to robust control design,” IEEE Trans. Autom. Control, vol. 51, no. 5, pp. 742–753, May 2006.

[12] F. Capitanescu, S. Fliscounakis, P. Panciatici, and L. Wohenkel, “Cautious operation planning under uncertainties,” IEEE Trans. Power Syst., vol. 27, no. 4, pp. 1859–1869, Nov. 2012.

[13] F. Bouffard and F. D. Galiana, “Stochastic security for operations planning with significant wind power generation,” IEEE Trans. Power Syst., vol. 23, no. 2, pp. 306–316, May 2008.

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

[15] L. Roald and G. Andersson, “Chance-constrained AC optimal power flow: Reformulations and efficient algorithms,” IEEE Trans. Power Syst., vol. 33, no. 3, pp. 2906–2918, May 2018.

[16] A. Venzke, L. Halilbaši, A. Barré, L. Roald, and S. Chatzivasilisidis, “Chance-constrained AC optimal power flow integrating HVDC lines and controllability,” Int. J. Elect. Power Energy Syst., vol. 116, 2020, Art. no. 105522.

[17] G. Chen, H. Zhang, and Y. Song, “Chance-constrained DC optimal power flow with non-Gaussian distributed uncertainties,” in Proc. IEEE Power Energy Soc. Gen. Meeting, 2022, pp. 1–5.

[18] Z. Wang, Q. Bian, H. Xin, and D. Gan, “A distributionally robust co-ordinated reserve scheduling model considering CVaR-based wind power reserve requirements,” IEEE Trans. Sustain. Energy, vol. 7, no. 2, pp. 635–636, Apr. 2016.

[19] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, “Wasserstein distributionally robust look-ahead economic dispatch,” IEEE Trans. Power Syst., vol. 36, no. 3, pp. 2010–2022, May 2021.

[20] Y. Zhang, S. Shen, and J. L. Mathieu, “Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads,” IEEE Trans. Power Syst., vol. 32, no. 3, pp. 1378–1388, Mar. 2017.

[21] Y. Chen, Q. Guo, H. Sun, Z. Li, W. Wu, and Z. Li, “A distributionally robust optimization model for unit commitment based on Kullback–Leibler divergence,” IEEE Trans. Power Syst., vol. 33, no. 5, pp. 5147–5160, Sep. 2018.

[22] C. Zhao and Y. Guan, “Data-driven stochastic unit commitment for integrating wind generation,” IEEE Trans. Power Syst., vol. 31, no. 4, pp. 2587–2596, Jul. 2016.

[23] Z. Shi, H. Liang, and V. Dinavahi, “Data-driven distributionally robust chance-constrained unit commitment with uncertain wind power,” IEEE Access, vol. 7, pp. 135087–135098, 2019.
[24] T. Ding et al., “Duality-free decomposition based data-driven stochastic security-constrained unit commitment,” *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 82–93, Jan. 2019.

[25] X. Lu, K. W. Chan, S. Xia, B. Zhou, and X. Luo, “Security-constrained multiperiod economic dispatch with renewable energy utilizing distributionally robust optimization,” *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 768–779, Apr. 2019.

[26] X. Zheng and H. Chen, “Data-driven distributionally robust unit commitment with Wasserstein metric: Tractable formulation and efficient solution method,” *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4940–4943, Nov. 2020.

[27] A. Zhou, M. Yang, M. Wang, and Y. Zhang, “A linear programming approximation of distributionally robust chance-constrained dispatch with Wasserstein distance,” *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3366–3377, Sep. 2020.

[28] R. Zhu, H. Wei, and X. Bai, “Wasserstein metric based distributionally robust approximate framework for unit commitment,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2991–3001, Jul. 2019.

[29] A. Arrigo, C. Ordoudis, J. Kazempour, Z. D. Grève, J. Toubeau, and F. Vallée, “Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation,” *Eur. J. Oper. Res.*, vol. 296, no. 1, pp. 304–322, 2022.

[30] Z. Chen, D. Kuhn, and W. Wiesemann, “Data-driven chance constrained programs over Wasserstein balls,” *Operations Res.*, 2022, doi: 10.1287/opre.2022.2330.

[31] W. Xie, “On distributionally robust chance constrained programs with Wasserstein distance,” *Math. Program.*, vol. 186, no. 1, pp. 115–155, 2021.

[32] C. Oordoudis, V. Nguyen, D. Kuhn, and P. Pinson, “Energy and reserve dispatch with distributionally robust joint chance constraints,” *Operations Res. Lett.*, vol. 49, no. 3, pp. 291–299, 2021.

[33] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, “Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric,” *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4924–4936, Sep. 2018.

[34] R. A. Jabr, “Distributionally robust CVaR constraints for power flow optimization,” *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3764–3773, Sep. 2020.

[35] N. Jiang and W. Xie, “ALSO-X and ALSO-X: Better convex approximations for chance constrained programs,” *Oper. Res.*, vol. 70, pp. 3035–3628, 2022, doi: 10.1287/opre.2021.2225.

[36] A. Esteban-Pérez and J. M. Morales, “Distributionally robust optimal power flow with contextual information,” *Eur. J. Oper. Res.*, vol. 306, no. 3, pp. 1047–1058, 2023.

[37] A. Arrigo, J. Kazempour, Z. D. Grève, J.-F. Toubeau, and F. Vallée, “Embedding dependencies between wind farms in distributionally robust optimal power flow,” *IEEE Trans. Power Syst.*, early access, Nov. 10, 2022, doi: 10.1109/TPWRS.2022.3221348.

[38] F. Golestaneh, P. Pinson, and H. Gooi, “Polyhedral predictive regions for power system applications,” *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 693–704, Jan. 2019.

[39] P. M. Esfahani and D. Kuhn, “Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations,” *Math. Program.*, vol. 171, no. 1/2, pp. 115–166, 2018.

[40] R. Gao and A. Kleywegt, “Distributionally robust stochastic optimization with Wasserstein distance,” *Math. Operations Res.*, 2022, doi: 10.1287/moor.2022.1275.

[41] D. Boskos, J. Cortés, and S. Martínez, “High-confidence data-driven ambiguity sets for time-varying linear systems,” 2021, arXiv:2102.01142.

[42] C. Oordoudis, P. Pinson, J. M. González, and M. Zugno, “An updated version of the IEEE RTS 24-Bus system for electricity market and power system operation studies,” *Tech. Univ. Denmark, Lyngby, Denmark, Tech. Rep.*, 2016.

[43] R. D. Christie, B. F. Wollenberg, and I. Wangensteen, “Transmission management in the deregulated environment,” *Proc. IEEE*, vol. 88, no. 2, pp. 170–195, Feb. 2000.