

# Sample-Adaptive Robust Economic Dispatch With Statistically Feasible Guarantees

Chenbei Lu<sup>ID</sup>, Graduate Student Member, IEEE, Nan Gu<sup>ID</sup>, Graduate Student Member, IEEE,  
Wenqian Jiang<sup>ID</sup>, Graduate Student Member, IEEE, and Chenye Wu<sup>ID</sup>, Member, IEEE

**Abstract**—The high penetration of renewable energy brings significant uncertainty to the power grids. Taking economic dispatch (ED) as an example, the inaccurate prediction of renewable energy generations dramatically increases the dispatch cost and risks the power grid’s reliable operation. The accurate distribution knowledge of the renewable generations enables modeling the ED as stochastic programming with joint chance constraints, which various classical methods can tackle. However, in practice, such distribution knowledge is inaccessible, and we can only observe samples from some unknown distribution. This makes conducting effective ED solely based on the observed samples challenging. It is particularly true when we need to handle the joint chance constraints. To tackle these challenges, we introduce the notions of statistical feasibility and statistically feasible ED to guarantee the satisfaction of the joint chance constraints. Specifically, we first propose a sample-adaptive robust optimization (RO) to decouple the joint constraints. We then identify that the inaccurate uncertainty set leads to RO’s conservativeness, and then reconstruct the constraint-specific uncertainty sets. We design the corresponding sample-adaptive reconstruction-based RO (ReconRO) based on the reconstructed uncertainty sets to further enhance the ED’s effectiveness.

**Index Terms**—Chance-constrained optimization, economic dispatch, robust optimization, sample-based optimization.

## I. INTRODUCTION

ECONOMIC dispatch (ED), a classical procedure of the power grid operation, is very well-investigated, especially with the advances in forecast methods over the past decades. However, the increasing penetration level of renewable energy poses significant challenges to the effective ED. Specifically, the significant prediction error of renewable energy increases the system dispatch cost. It also brings high risks to balancing the geographically distributed loads and maintaining the physical constraints, which are the core tasks of ED.

Manuscript received 7 June 2022; revised 6 January 2023 and 21 February 2023; accepted 7 April 2023. Date of publication 14 April 2023; date of current version 26 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 72271213, and in part by the Shenzhen Science and Technology Program under Grants JCYJ20220530143800001 and RCYX20221008092927070. Paper no. TPWRS-00823-2022. (*Corresponding author: Chenye Wu.*)

Chenbei Lu and Nan Gu are with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China (e-mail: lcb20@ mails.tsinghua.edu.cn; gn19@mails.tsinghua.edu.cn).

Wenqian Jiang and Chenye Wu are with the School of Science and Engineering, Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China (e-mail: wenqianjiang@link.cuhk.edu.cn; chenyewu@yeah.net).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRS.2023.3267097>.

Digital Object Identifier 10.1109/TPWRS.2023.3267097

Although eliminating the prediction error is impossible, it is often viable to estimate the distribution information about renewable energy generation. Such estimation allows us to model the ED as a standard stochastic programming [1]. However, the accurate distribution information may also be hard to obtain in practice, because we can only observe the samples drawn from the distribution. Conducting effective ED solely based on the observed samples can be challenging because ED often involves probabilistic constraints, e.g., chance constraints (CCs) and even joint CCs, whose feasibility can be hard to quantify based on samples.

To overcome these issues, we first introduce the notion of statistical feasibility, which provides a theoretical guarantee of satisfying the joint CCs with samples. Then, we propose a more practical formulation termed as statistically feasible ED built on statistical feasibility. In such a formulation, the remaining hurdle is to handle the joint CCs with a statistical feasibility guarantee. To mitigate this hurdle, we first propose a sample-adaptive robust optimization (RO) which can decouple the joint CCs by generating the corresponding uncertainty sets with provable statistical feasibility. Like most RO, this sample-adaptive RO is still conservative. We identify that the inaccurate uncertainty set construction leads to the conservativeness, which inspires us to reconstruct the constraint-specific uncertainty set and further propose the reconstruction-based robust optimization (ReconRO) to prompt the performance of ED.

## A. Related Works

The conventional method to handle the sample-based ED is scenario generation (SG) [2], [3]. However, the SG is computationally expensive and largely reliant on the quantity and quality of the sample collection. The maturity of RO techniques and chance-constrained methods has promoted power grid operation with uncertainty. In the case of RO, Jiang et al. employ the RO framework to conduct unit commitment (UC) for thermal generators in the day-ahead market and devise a scheme to control the RO’s conservativeness in [4]. Bertsimas et al. adapt the RO framework to the two-stage security-constrained UC problem and develop a practical algorithm in [5].

As for chance-constrained optimization, Wang et al. first propose a two-stage chance-constrained program for the UC problem with wind generation output in [6]. Poze et al. use the chance-constrained framework to incorporate wind and demand uncertainty in a joint energy and reserves scheduling problem

in [7]. Bienstock et al. advance a stylized formulation of chance-constrained optimal power flow and design an efficient algorithm in [8]. For the more challenging joint chance-constrained optimization, the common computational method is to consider the convex and tractable approximations [9], [10]. The most related work is [11], where Van et al. consider a hydrothermal UC problem with two-sided CC, and they modify the classical cutting plane method to obtain an exact solution.

To provide more robustness to the chance-constrained optimization, the distributionally robust optimization (DRO) is proposed for the ED [12]. Just to name a few, Zhang et al. propose a DRO approach to conduct the ED with renewable energy generation in [13]. Duan et al. adopt Wasserstein ball-centered DRO to calculate the optimal power flow in [14]. Xie et al. propose a data-driven distributionally robust chance-constrained optimal power flow model (DRCC-OPF) with an effective second-order cone programming (SOCP) reformulation in [15]. Gu et al. adopt the DRO framework to conduct effective operations of stochastic emission-aware economic dispatch with storage systems in [16]. However, the literature for ED with DRO seldom considers joint CCs since they are more challenging to compute than individual CC [17]. For general literature handling the joint CCs, they can be categorized into three groups. The first type of works use a conservative approximation to build convex approximations of joint CCs for DRO, and produce an approximate solution [18], [19], [20]. The second type of works reformulate joint chance constraints to a mixed-integer conic program/binary bi-linear problem, but makes the original problem complex with additional integer variables [21], [22], [23]. And the third type of works can reformulate the joint chance-constrained problem as a convex program that can be effectively solved. However, these works are effective often under strong assumptions, such as the constraint configuration and the assumptions on the type of ambiguity set [24], [25]. Some of the reformulations also suffer from the curse of dimensionality, i.e., the problem size grows with the dimension of decision variables/random variables [9]. In contrast, our approach can accurately solve the joint CCs in the ED problem by transforming the original problem into an equivalent convex program. Further, our approach is only based on samples without additional assumptions to the random variable's distribution.

We are not the first to consider the satisfaction of joint CCs with samples. Recently, there is a growing body of literature dealing with the joint CCs in the RO scheme. For example, Margellos et al. generate the uncertainty sets in the RO framework through sample-based methods in [26]. Hong et al. further theoretically advance the concept of statistical feasibility in learning the shape of the RO's uncertainty set and practicing the concept by calibrating the uncertainty set's coverage in [27], which inspires our work. In the power sector, there are a few works considering sample-based reformulations of joint CC optimal power flow, such as [28] and [29]. However, they all employ the traditional CC computation method and don't provide the satisfaction guarantees for joint CCs based on samples. That is, the samples may lead to inaccurate prediction for the random variables in some cases, leading to constraint violation and sub-optimal solution. The limitations of the conventional

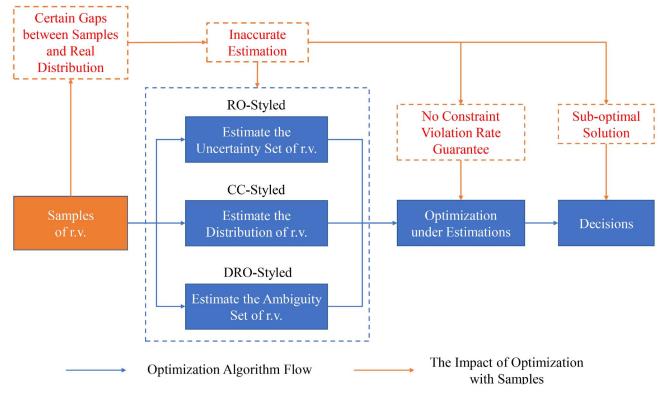


Fig. 1. Conventional Optimization Approaches.

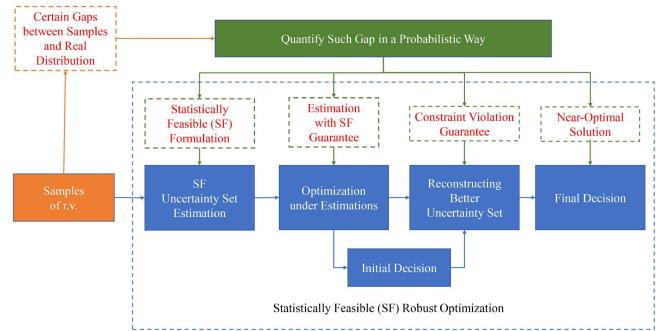


Fig. 2. Framework of Our Approach.

approaches are illustrated in Fig. 1. In contrast, we introduce the notion of statistical feasibility to capture the uncertainty of the samples, and then propose the statistically feasible ED to guarantee the satisfaction of the joint CCs. We also propose a reconstruction-based robust optimization approach to further improve the solution. The framework of our approach is presented in Fig. 2. Specifically, the detailed comparison between our approach and the existing approaches is illustrated in Table I.

## B. Our Contributions

Our major contributions can be summarized as follows:

- *Statistically Feasible ED*: We introduce the notion of statistical feasibility for ED, which can provide a theoretical guarantee for satisfying the joint CCs based on samples, yielding the formulation of statistically feasible ED.
- *Sample-Adaptive RO Customized for ED*: To effectively solve the statistically feasible ED, we first construct the uncertainty set based on samples with statistical feasibility guarantees and then propose the sample-adaptive RO.
- *Sample-Adaptive ReconRO Customized for ED*: To overcome RO's conservativeness, we reconstruct the constraint-specific uncertainty set, yielding the sample-adaptive ReconRO.

The remainder of this paper is organized as follows: Section II introduces the conventional chance-constrained ED. Section III proposes the notion of statistical feasibility and the formulation

TABLE I  
COMPARISON OF DIFFERENT APPROACHES

Approach	Distribution Assumption-Free	Violation Rate Guarantee	Handling Joint CCs	Statistical Feasibility Guarantee[27]
CC[11][6][30][31]	✗	✗	Under Strong Conditions	✗
RO[32][30][5]	✓	✓	✓	✗
DRO[15][13][14][16]	✓	✓	Under Strong Conditions <sup>1</sup>	✗
SG[2][3][33]	✓	✗	✓	✗
Ours	✓	✓	✓	✓

<sup>1</sup> With approximation to the joint CCs, or introduce additional integer variables, or have strong assumptions to the configurations to constraints and the ambiguity set of random variables.

of statistically feasible ED. Section IV proposes a sample average approximation approach to reckon the stochastic objective function and investigates the theoretical accuracy-complexity trade-off of the approximation. To design an effective approach for the statistically feasible ED, Section V introduces the preliminaries of RO and proposes an RO-based formulation to decouple the joint CCs with provable statistical feasibility. Section VI implements a sample-adaptive approach for solving the RO with statistical feasibility guarantees. Section VII identifies the key factors for the conservativeness of RO and then proposes the ReconRO with better economic performance and less conservativeness. Section VIII evaluates the performance of our proposed approaches. Section IX concludes our paper.

## II. SYSTEM MODEL

Consider a set of geographically distributed generators  $\mathcal{N}$  and demands  $\mathcal{M}$  (the renewable generations can be regarded as negative demands) on a power network with a set of transmission lines  $\mathcal{A}$ . The standard ED seeks to conduct load balancing over  $T$  time slots satisfying a set of constraints. The key challenge comes from the stochasticity in the net load. To better characterize such stochasticity, we model the net demand as random variables following certain distributions. The system cost includes the generation cost for each generator and the risk cost from the real-time generation-demand mismatch. The system operator seeks to minimize the system cost by solving the following stochastic optimization:

$$(\mathbf{P1}) \min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + R(G^t, D^t) \right) \quad (1a)$$

$$\text{s.t. } G^t = \sum_{i \in \mathcal{N}} g_i^t, \quad D^t = \sum_{i \in \mathcal{M}} d_i^t, \quad \forall t, \quad (1b)$$

$$S^t = D^t - G^t, \quad \forall t, \quad (1c)$$

$$R(G^t, D^t) = \gamma_1 \mathbb{E}[(S^t)^+] + \gamma_2 \mathbb{E}[(-S^t)^+], \quad \forall t, \quad (1d)$$

$$f^t = \mathbf{H}_g g^t - \mathbf{H}_d d^t, \quad \forall t, \quad (1e)$$

$$\mathbb{P} \left( \bigcap_{t=1}^T \{f_i^t \leq \bar{f}_i\} \right) \geq \rho_{i,1}, \quad \forall i \in \mathcal{A}, \quad (1f)$$

$$\mathbb{P} \left( \bigcap_{t=1}^T \{f_i^t \geq -\bar{f}_i\} \right) \geq \rho_{i,2}, \quad \forall i \in \mathcal{A}, \quad (1g)$$

$$g_i^{\min} \leq g_i^t \leq g_i^{\max}, \quad \forall i \in \mathcal{N}, \forall t, \quad (1h)$$

$$-\text{DR}^i \leq g_i^{t+1} - g_i^t \leq \text{UR}^i, \quad \forall i \in \mathcal{N}, \forall t, \quad (1i)$$

where  $(x)^+ = \max(x, 0)$ .

The decision variables in **(P1)** are  $g_i^t$ 's, the generation of generator  $i$  at time  $t$ , with the vector form  $\mathbf{g}^t = [g_i^t, \forall i \in \mathcal{N}]$ . The other parameters include:

- $G^t$ : total generation at time  $t$ ;
- $D^t$ : total demand at time  $t$ ;
- $S^t$ : generation shortage at time  $t$  ( $S^t < 0$  indicating generation excess);
- $f_i^t$ : power flow of transmission line  $i$  at time  $t$ , with the vector form  $\mathbf{f}^t = [f_i^t, \forall i \in \mathcal{A}]$ ;
- $d_i^t$ : the net demand for demand  $i$  at time  $t$ , with the vector form  $\mathbf{d}^t = [d_i^t, \forall i \in \mathcal{M}]$ ;
- $\gamma_1, \gamma_2$ : the unit generation shortage and excess risk cost;
- $\mathbf{H}_g, \mathbf{H}_d$ : the shift factor matrix for generators and demands, respectively;
- $\bar{f}_i$ : the power flow capacity of transmission line  $i$ , with the vector form  $\bar{\mathbf{f}} = [\bar{f}_i, \forall i \in \mathcal{A}]$ ;
- $\rho_{i,1}, \rho_{i,2}$ : the stability requirements of upper and lower limit capacity on transmission line  $i$ ;
- $g_i^{\min}, g_i^{\max}$ : the minimal and maximal generation capacities of generator  $i$ ;
- $\text{UR}^i, \text{DR}^i$ : the ramp up and down limits of generator  $i$ ;
- $C_i(\cdot)$ : convex generation cost function for generator  $i$ ;
- $R(G^t, D^t)$ : the risk cost with total generation  $G^t$  and total demand  $D^t$ ;

Constraint **(1c)** characterizes the generation shortage; constraint **(1d)** specifies the risk cost which includes the expected generation shortage and excess cost; constraint **(1e)** describes the power flow on the transmission lines; constraint **(1f)** and **(1g)** describe the joint CCs of upper and lower limit capacities on transmission lines during the whole time horizon; constraint **(1h)** characterizes the generation capacity of generators; and constraint **(1i)** characterizes the ramp up and down rate capacity of generators.

This is a classical stochastic ED with joint CCs. The essential difficulty comes from the uncertainty of demands  $d_i^t$ , since both the risk cost  $R(G^t, D^t)$  and the transmission line capacity constraints in **(1f)** and **(1g)** involve  $d_i^t$ . Without full knowledge

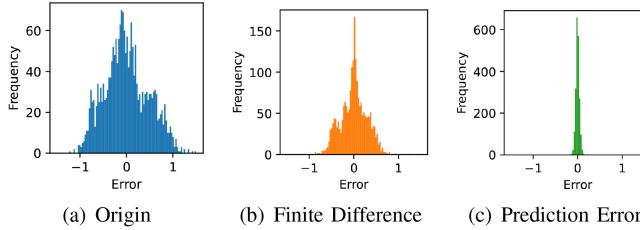


Fig. 3. Power of Prediction Error Decomposition.

of each  $d_i^t$ 's distribution, we cannot accurately evaluate the objective function. Further, the feasibility of the CCs cannot be guaranteed without distribution information. These two issues seriously deteriorate the economic performance and transmission line security.

To tackle these challenges, in the next section, we propose a more practical formulation termed as the statistically feasible ED, which is only dependent on the samples.

### III. STATISTICALLY FEASIBLE ECONOMIC DISPATCH

In this section, we first decompose the prediction errors from  $d_i^t$  to better handle the uncertainty in ED. Then, we illustrate the challenges of the optimization only with samples, and introduce the notion of statistical feasibility. Finally, we customize a more practical statistically feasible ED formulation.

#### A. Prediction Error Decomposition

The uncertainty of  $d_i^t$  is the primary hurdle for solving **(P1)**. However, directly dealing with the uncertainty in  $d_i^t$  can be problematic. First, the realization of  $d_i^t$  can be affected by many factors and always exhibits time-varying characteristics. Further, the variance of  $d_i^t$  is significant, which can be observed from the blue histogram in Fig. 3(a). This often makes the CCs very conservative. To eliminate the temporal dependency, the most common approach is to take the difference of the time series, the histogram of which is illustrated in Fig. 3(b). We can observe that the variance of the new difference series is well reduced. An even better approach [16] is to use predictive models (i.e., neural networks) to capture the time-varying factors directly, and then all the uncertainties come from the prediction errors, i.e.,  $d_i^t = \bar{d}_i^t + \xi_i^t$ , where  $\bar{d}_i^t$  and  $\xi_i^t$  denote the predicted demand and prediction error for demand  $i$  at time  $t$ , respectively. From Fig. 3(c), we can find that  $\xi_i^t$  exhibits a very small variance.

Such a prediction error decomposition transforms **(P1)** into the following equivalent problem **(P1e)**:

**(P1e)**

$$\min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + R(G^t, D^t) \right) \quad (2a)$$

$$\text{s.t. } f^t(g^t; \xi^t) = H_g g^t - H_d(\bar{d}^t + \xi^t), \quad \forall t, \quad (2b)$$

$$\mathbb{P} \left( \bigcap_{t=1}^T \{f_i^t(g^t; \xi^t) \leq \bar{f}_i\} \right) \geq \rho_{i,1}, \quad \forall i \in \mathcal{A}, \quad (2c)$$

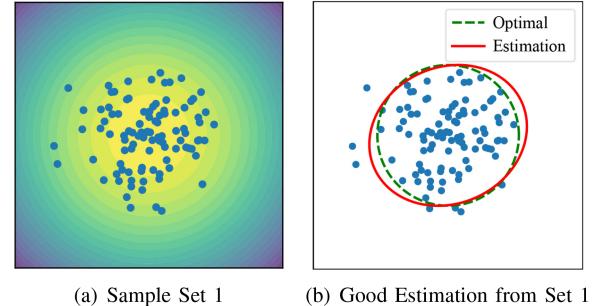


Fig. 4. Estimation from different Sample Sets.

$$\mathbb{P} \left( \bigcap_{t=1}^T \{f_i^t(g^t; \xi^t) \geq -\bar{f}_i\} \right) \geq \rho_{i,2}, \quad \forall i \in \mathcal{A},$$

Constraints (1b)–(1d), (1h)–(1i), (2d)

where  $\bar{d}^t$  is the vector form of the predicted demands, i.e.,  $\bar{d}^t = [\bar{d}_i^t, \forall i \in \mathcal{M}]$ ;  $\xi^t$  is the vector form of the prediction errors at time  $t$ , i.e.,  $\xi^t = [\xi_i^t, \forall i \in \mathcal{M}]$ . We also denote  $\xi = [\xi^t, \forall t]$ . Note that, we adopt the notation  $f^t(g^t; \xi^t)$  to highlight that the power flow  $f^t$  is the function of decision variables  $g^t$  and random variables  $\xi^t$ .

It is clear that, the uncertainty in **(P1e)** only exists in  $\xi$ . Moreover,  $D_t$  in the objective function also involves  $\xi$  implicitly.

In practice, the exact distribution of  $\xi^t$  is unknown. we can only observe a set of samples of  $\xi^t$ , i.e.,  $\mathcal{D}_\xi = \{\xi^{(i)}, \forall i \in \mathcal{S}\}$ , where  $\xi^{(i)}$  is an error sample including all errors in an ED process, i.e.,  $\xi^{(i)} = \{\xi_j^{(i),t}, \forall t \leq T, \forall j \in \mathcal{M}\}$ .  $\mathcal{S}$  is the index set of error samples. Generally, the prediction errors has time-invariant property [34], thus we can assume that every single  $\xi^{(i)}$  is randomly sampled from  $\xi$ 's actual distribution  $\mathcal{F}_\xi$ .

#### B. Statistically Feasible ED

However, it is challenging to estimate the real distribution  $\mathcal{F}_\xi$  only with samples. We provide an example to illustrate this challenge in Fig. 4, where we want to conduct the distribution estimation. Consider the case that all samples are independently drawn from a standard normal distribution. Fig. 4(a) and (c) illustrate two possible sample sets.

The two sample sets will lead to different estimations of random variable  $\xi$ : Fig. 4(b) and (d) illustrate the ellipsoid based on the estimated mean and covariance matrix from the sample sets 1 and 2, respectively. We can observe that the estimation

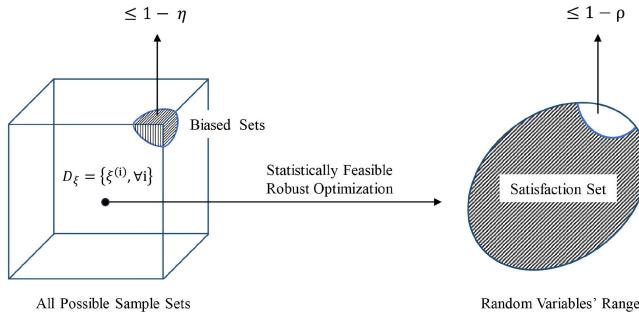


Fig. 5. Statistically Feasible Optimization.

from set 1 is closer to the optimal shape, while the estimation from set 2 is dramatically different from the optimal one.

It indicates that different samples will lead to different estimations. When we are provided with sample sets that cause large estimation errors, it is often difficult to guarantee the satisfaction of constraints, as well as the economic performance of optimization algorithms. To quantify such uncertainty from the dataset and its influence on the optimization, we introduce the notion of statistical feasibility.

We first introduce the notion of statistical feasibility, and then introduce formulate our statistically feasible ED:

**Definition 1 (Statistical Feasibility [27]):** For any given dataset  $\mathcal{D}_\xi$ , an approach will produce a solution  $\hat{g}$ . An approach has the statistical feasibility guarantees if the solution  $\hat{g}$  it provides is feasible for the constraints with confidence level  $\eta$  for any given dataset  $\mathcal{D}_\xi$ .

Specifically, for a CC  $\mathbb{P}_\xi(f(\hat{g}; \xi) \leq b) \geq \rho$ , the statistical feasibility is defined as:

$$\mathbb{P}_{\mathcal{D}_\xi}(\mathbb{P}_\xi(f(\hat{g}(\mathcal{D}_\xi); \xi) \leq b) \geq \rho) \geq \eta. \quad (3)$$

Since the solution  $\hat{g}$  is based on dataset  $\mathcal{D}_\xi$ , we use symbol  $\hat{g}(\mathcal{D}_\xi)$  to demonstrate that  $\hat{g}$  is influenced by  $\mathcal{D}_\xi$ . Moreover, since the samples of the dataset are randomly selected from the distribution  $\mathcal{F}_\xi$ , the solution  $\hat{g}$  also follows a certain distribution. The inside constraint in (3) is the conventional CC. And the outside constraint in (3) further guarantees that for all resulted solutions  $\hat{g}$  (essentially, for different sample sets  $\mathcal{D}_\xi$ ), the inside CC can be satisfied with a probability no smaller than  $\eta$ .

Fig. 5 illustrates the main idea of our proposed notion. Specifically, the cube on the left characterizes all possible sample sets. Any point in the cube is a possible sample set with a number of samples, i.e.,  $\mathcal{D}_\xi = \{\xi^{(i)}, \forall i\}$ . In our statistically feasible robust optimization, we guarantee that except for a small part of biased sets ( $\leq \eta$ , potentially brings huge estimation error), for any other sample set, we can accurately guarantee that based on the sample set, the CC can be satisfied.

**Remark:** Note that, although with similar names, the notion of statistical feasibility is quite different from the statistical guarantee for DRO. The statistical feasibility characterizes the ability of an approach (algorithm) to provide feasible solutions for CCs with different sample sets. While the statistical guarantee for DRO refers to its ability to make robust decisions against a set of possible distributions (the ambiguity set) [21].

Now we adopt the notion of statistical feasibility to define the statistically feasible ED as follows:

$$(\mathbf{P2}) \min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + R(G^t, D^t) \right) \quad (4a)$$

$$\text{s.t. } \mathbb{P}_{\mathcal{D}_\xi}(\mathbb{P}_\xi \left( \bigcap_{t=1}^T \{f_i^t(g^t; \xi^t) \leq \bar{f}_i\} \right) \geq \rho_{i,1}) \geq \eta_{i,1}, \\ \forall i \in \mathcal{A}, \quad (4b)$$

$$\mathbb{P}_{\mathcal{D}_\xi}(\mathbb{P}_\xi \left( \bigcap_{t=1}^T \{f_i^t(g^t; \xi^t) \geq -\bar{f}_i\} \right) \geq \rho_{i,2}) \geq \eta_{i,2}, \\ \forall i \in \mathcal{A}, \quad (4c)$$

where  $\eta_{i,1}$  and  $\eta_{i,2}$  denote the required confidence levels for the inside CCs. Note that, each  $f_i^t$  is a function of  $g^t$  and  $\xi^t$ , and  $g^t$  is dependant on dataset  $\mathcal{D}_\xi$ . The statistically feasible constraint guarantees that the inside CC can be satisfied with a probability  $\eta$  for any dataset  $\mathcal{D}_\xi$ . Essentially, this is a more practical formulation that focuses on the samples with theoretical guarantees. In the subsequent analysis, we focus on solving (P2).

However, there are two main hurdles to solving this problem. First, the objective function in (4a) implicitly involves random variable  $\xi$  (in  $D^t$ ), which still requires the distribution information. Second, it is not clear how to construct the solution satisfying CCs (4b) and (4c).

#### IV. SAMPLE AVERAGE APPROXIMATION

This section proposes a sample-oriented approach to approximate the objective function. We also theoretically characterize the trade-off between the approximation accuracy and approximation complexity.

##### A. Sample Average Approximation

We focus on the risk cost  $R(G^t, D^t)$  as it is the only term with uncertainty in the objective function of (P2). The specific form of  $R(G^t, D^t)$  is as follows:

$$R(G^t, D^t) = \gamma_1 \mathbb{E}[(S^t)^+] + \gamma_2 \mathbb{E}[(-S^t)^+] \\ = \gamma_1 \mathbb{E} \left[ \left( \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^t) - G^t \right)^+ \right] \\ + \gamma_2 \mathbb{E} \left[ \left( G^t - \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^t) \right)^+ \right]. \quad (5)$$

Without the knowledge of  $\xi$ 's distribution, we cannot obtain the accurate value of  $R(G^t, D^t)$ . To tackle this issue, we adopt the notion of sample average approximation (SAA), which approximates the expectation by sample average. Specifically, we first randomly sample a subset  $\mathcal{D}_\xi^s$  with size  $n_s$  from  $\mathcal{D}_\xi$ . The sample average of  $\hat{R}(G^t, D^t)$  is calculated as follows:

$$\hat{R}(G^t, D^t) = \frac{1}{n_s} \sum_{\xi^{(j)} \in \mathcal{D}_\xi^s} \hat{R}(G^t, \xi^{(j), t}), \quad (6)$$

where

$$\begin{aligned}\tilde{R}(G^t, \xi^{(j),t}) &= \gamma_1 \left( \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^{(j),t}) - G^t \right)^+ \\ &\quad + \gamma_2 \left( G_t - \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^{(j),t}) \right)^+.\end{aligned}\quad (7)$$

Since a linear transformation can easily remove the  $(\cdot)^+$  operator, the derived sample average  $\hat{R}(G^t, D^t)$  is essentially a linear function of the decision variables  $g_i^t$ .

### B. Accuracy-Complexity Trade-off

Intuitively, with more samples for SAA, the approximation should be more accurate. However, it will also complicate the approximated function, which directly affects the computational complexity. To validate this intuition, we derive the following theorem to characterize the approximation error:

*Theorem 1:* Given sample size  $n_s$ , the approximation error  $e(n_s)$  between the real objective function and its approximation is bounded as follows:

$$\mathbb{P}(|e(n_s)| \geq \Delta) \leq 2 \exp \left( -\frac{2n_s \Delta^2}{T^2(\gamma_1^2 + \gamma_2^2)(\bar{\xi} - \underline{\xi})^2} \right), \quad (8)$$

where  $\bar{\xi}$  and  $\underline{\xi}$  denote the upper and lower bounds for  $\xi_i^t$ , respectively.

This result indicates that the approximation error reduces exponentially fast with increasing sample size  $n_s$ . Further, a longer time horizon  $T$ , a higher risk cost, and a more significant error range  $(\bar{\xi} - \underline{\xi})$  all contribute to a more significant approximation error, which coincides with our intuition. The detailed proof is deferred to Appendix A.

The computational complexity refers to the computation time for the optimization with sample size  $n_s$ . Although the objective function only contains the original decision variables  $g_i^t$ , transforming the term with  $(\cdot)^+$  operator implicitly includes more decision variables. Specifically, the number of terms with  $(\cdot)^+$  operator is  $2n_s T$ .

Therefore, we can derive the following corollary to characterize the approximation accuracy-complexity trade-off:

*Corollary 1:* Given time horizon  $T$  and sample size  $n_s$ , the approximation error is  $O(\exp(-n_s/T^2))$  and the computational complexity is  $O(n_s T)$ .

This corollary demonstrates that with the increasing  $n_s$ , the approximation error decreases exponentially whereas the computational complexity increases linearly. A detailed proof is provided in Appendix B.

## V. SAMPLE-ADAPTIVE RO: THE BASIS

In this section, we further propose an RO-based framework to solve the original problem (P2), which can effectively decouple the joint CCs while satisfying the requirement of statistical feasibility. Specifically, we first revisit the traditional RO for more intuitions, which shed light on solving statistically feasible RO.

### A. Revisit Robust Optimization

Consider a typical chance constrained optimization as follows:

$$(\text{CC}) \quad \min f(\mathbf{x}; \xi) \quad (9a)$$

$$\text{s.t. } h_i(\mathbf{x}) \leq a_i, \forall i, \quad (9b)$$

$$\mathbb{P}_{\xi} (g(\mathbf{x}; \xi) \leq b) \geq \rho, \quad (9c)$$

where  $\mathbf{x}$  denotes the decision variables, and  $\xi$  denotes the random variables. Equation (9b) denotes the deterministic constraints, and (9c) denotes the CCs. Specifically, it requires that the condition  $g(\mathbf{x}; \xi) \leq b$  need to be satisfied with a probability of at least  $\rho$  over  $\xi$ .

Solving a chance-constrained optimization requires the distribution information about  $\xi$ . In contrast, the RO requires less information about  $\xi$  and can produce an approximate solution of (CC) while satisfying the CCs. Specifically, the corresponding robust optimization (RO) is as follows:

$$(\text{RO}) \quad \min f(\mathbf{x}; \xi) \quad (10a)$$

$$\text{s.t. } h_i(\mathbf{x}) \leq a_i, \forall i, \quad (10b)$$

$$g(\mathbf{x}; \xi) \leq b, \forall \xi \in \mathcal{U}, \quad (10c)$$

where  $\mathcal{U}$  is the uncertainty set, connecting (RO) and (CC). The following fact identifies the requirement for the uncertainty set  $\mathcal{U}$ :

*Fact 1:* Any feasible solution to (RO) is also feasible to (CC) if the following condition holds for  $\mathcal{U}$ :

$$\mathbb{P}_{\xi}(\xi \in \mathcal{U}) \geq \rho. \quad (11)$$

This fact indicates that, if the uncertainty set  $\mathcal{U}$  can cover  $\rho$  of  $\xi$ 's range, the solution of (RO) will be feasible to (CC). Intuitively, if  $\mathcal{U}$  is large enough, constraint  $g(\mathbf{x}; \xi) \leq b$  can be satisfied for  $\rho$  part of  $\xi$ , which satisfies the original CC (9c).

More importantly, the RO only requires an uncertainty set to approximately cover the range of  $\xi$ , instead of  $\xi$ 's specific distribution. This makes it possible to generate uncertainty sets based on samples.

### B. RO With Statistical Feasibility

Similarly, we propose an RO with statistical feasibility to approximate the solution of (P2) as follows:

$$(\text{P3}) \quad \min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + \hat{R}(G^t, D^t) \right) \quad (12a)$$

$$\text{s.t. } \bigcap_{t=1}^T \{f_i^t(\mathbf{g}^t; \xi^t) \leq \bar{f}_i\}, \forall \xi \in \mathcal{U}_{i,1}(\mathcal{D}_{\xi}), \quad (12b)$$

$$\bigcap_{t=1}^T \{f_i^t(\mathbf{g}^t; \xi^t) \geq -\bar{f}_i\}, \forall \xi \in \mathcal{U}_{i,2}(\mathcal{D}_{\xi}),$$

$$\text{Constraints (1b)–(1d), (2b), (1h)–(1i),} \quad (12c)$$

where  $\mathcal{U}_{i,1}^t(\mathcal{D}_{\xi})$ ,  $\mathcal{U}_{i,2}^t(\mathcal{D}_{\xi})$  are the uncertainty sets associated with the dataset  $\mathcal{D}_{\xi}$ .

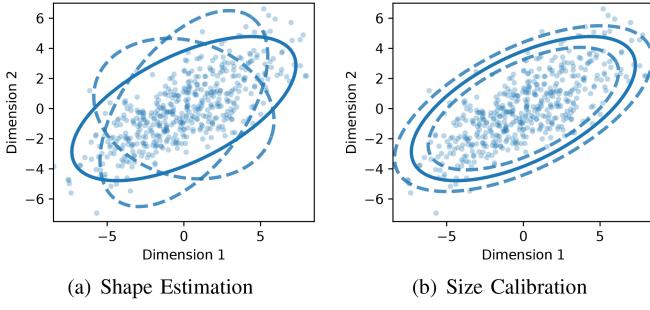


Fig. 6. Uncertainty Set Construction.

Similar to Fact 1, we can derive the following theorem to bridge the gap between (P3) and (P2):

**Theorem 2:** Any feasible solution to (P3) is also feasible to (P2) if the following conditions for the uncertainty sets  $\mathcal{U}_{i,1}^t(\mathcal{D}_\xi), \mathcal{U}_{i,2}^t(\mathcal{D}_\xi)$  hold:

$$\mathbb{P}_{\mathcal{D}_\xi}(\mathbb{P}_\xi(\xi \in \mathcal{U}_{i,1}(\mathcal{D}_\xi)) \geq \rho_{i,1}) \geq \eta_{i,1}, \quad \forall i \in \mathcal{A}, \quad (13)$$

$$\mathbb{P}_{\mathcal{D}_\xi}(\mathbb{P}_\xi(\xi \in \mathcal{U}_{i,2}(\mathcal{D}_\xi)) \geq \rho_{i,2}) \geq \eta_{i,2}, \quad \forall i \in \mathcal{A}. \quad (14)$$

This theorem is obvious since the original constraint (4b) and (4c) in (P2) can be derived by combining constraints (12b) and (12c) with (13) and (14) in (P3). Intuitively, it requires that for dataset  $\mathcal{D}_\xi$ , the desired uncertainty sets should cover  $\rho$  part of  $\xi$  with confidence level  $\eta$ . We next solve problem (P3) to obtain a feasible solution to (P2).

## VI. SAMPLE-ADAPTIVE RO: IMPLEMENTATION

In this section, we propose an approach to construct the uncertainty sets  $\mathcal{U}_{i,1}(\mathcal{D}_\xi), \mathcal{U}_{i,2}(\mathcal{D}_\xi)$  for a given dataset  $\mathcal{D}_\xi$  while satisfying the conditions (13) and (14). Then, we solve the RO with the constructed uncertainty sets.

### A. Sample-Adaptive Uncertainty Set Generation

In the subsequent analysis, we only construct  $\mathcal{U}_{i,1}$  because the process to construct  $\mathcal{U}_{i,2}$  follows exactly the same routine. For better tractability, we design  $\mathcal{U}_{i,1}$  with the shape of ellipsoid:

$$\mathcal{U}_{i,1} = \{\xi | (\xi - \mu)^\top M^{-1}(\xi - \mu) \leq s_{i,1}\}, \quad (15)$$

where  $\mu$  is a vector with the same size of  $\xi$ ,  $M$  is a symmetric matrix, and  $s_{i,1}$  is a scalar. Specifically,  $\mu$  and  $M$  characterize the shape of the ellipsoid. Parameter  $s_{i,1}$  characterizes the size of the ellipsoid and varies for different uncertainty set  $\mathcal{U}_{i,1}$ .

To estimate these parameters, we first divide the dataset  $\mathcal{D}_\xi$  into two parts  $\mathcal{D}_{\xi,1}$  and  $\mathcal{D}_{\xi,2}$  with sizes  $n_1$  and  $n_2$ .  $\mathcal{D}_{\xi,1}$  is used to estimate the shape parameters  $\mu$  and  $M$ , and  $\mathcal{D}_{\xi,2}$  is used to estimate the size parameter  $s_{i,1}$ . Specifically,  $\mathcal{D}_{\xi,1} = \{\xi^{(i)}, \forall i \in \mathcal{S}_1\}$ ,  $\mathcal{D}_{\xi,2} = \{\xi^{(i)}, \forall i \in \mathcal{S}_2\}$ , where  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are the index sets of error samples in  $\mathcal{D}_{\xi,1}$  and  $\mathcal{D}_{\xi,2}$ . Each sample  $\xi^{(i)} = \{\xi_j^{(i),t}, \forall t, \forall j \in \mathcal{M}\}$ . The two stages of uncertainty set construction are visualized in Fig. 6.

1) *Shape Estimation:* We estimate  $\mu$  by the sample mean of  $\xi^{(i)}$  in  $\mathcal{D}_{\xi,1}$ :

$$\mu = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1} \xi^{(i)}. \quad (16)$$

Further, we use the sample covariance matrix to estimate  $M$ , since it can well characterize the geometric distribution of  $\xi$ . Mathematically,

$$M = \frac{1}{n_1 - 1} \sum_{i \in \mathcal{S}_1} (\xi^{(i)} - \mu)(\xi^{(i)} - \mu)^\top. \quad (17)$$

2) *Size Calibration:*  $s_{i,1}$  determines the size of the ellipsoid. Intuitively, an uncertainty set with a large  $s_{i,1}$  can cover more sample points, yet makes RO more conservative. While a smaller  $s_{i,1}$  leads to a less conservative RO, but may not satisfy the requirement in (13). Therefore, we seek to decide the smallest possible  $s_{i,1}$  satisfying (13) to get the optimal solution.

To obtain such a  $s_{i,1}$ , we first define a dimension-collapsing transformation function  $F(\xi)$  as follows:

$$F(\xi) = (\xi - \mu)^\top M^{-1}(\xi - \mu). \quad (18)$$

By the transformation function  $F(\xi)$ , each sample vector  $\xi$  can be transformed into a scalar. We transform all  $\xi^{(i)} \in \mathcal{D}_{\xi,2}$  with  $F$ , and then sort all these  $F(\xi^{(i)})$  into a new set  $\{\hat{s}_k, 1 \leq k \leq n_2\}$  in ascending order, i.e.,  $\hat{s}_k \leq \hat{s}_{k+1}$ .

The desired  $s_{i,1}$  is estimated by:

$$s_{i,1} = \hat{s}_{k^*}, \quad (19)$$

where the index  $k^*$  satisfies:

$$k^* = \min \left\{ r : \sum_{k=0}^{r-1} C_{n_2}^k (\rho_{i,1})^k (1 - \rho_{i,1})^{n_2 - k} \geq \eta_{i,1} \right\}. \quad (20)$$

We can prove the following theorem for the constructed uncertainty set:

**Theorem 3 (Statistical Feasibility Guarantee, Theorem 1 in [27]):** The ellipsoid uncertainty set  $\mathcal{U}_{i,1}$  with parameters  $\mu$ ,  $M$ , and  $s_{i,1}$  from the above estimation satisfies the requirements in (13), (14) if

$$n_2 \geq \frac{\ln(1 - \rho_{i,1})}{\ln \eta_{i,1}}, \quad \forall i \in \mathcal{A}. \quad (21)$$

Intuitively, this theorem indicates a minimal dataset size adopted for size calibration. When it requires larger  $\rho$  and  $\eta$ , the required amount of samples becomes larger. It is also remarkable that the required sample size doesn't depend on the dimension  $d$  of the random variable  $\xi$ , in contrast with SG, which requires the sample size of  $O(\frac{d}{1-\rho} \log \frac{1}{1-\rho})$  to guarantee the desired stability requirements [35].

Therefore, based on the constructed uncertainty set, we can solve (P3) to get a solution with a statistical feasibility guarantee.

### B. Solving RO by Linear Transformation

Robust constraints with uncertainty sets are difficult to tackle since such constraints are often not linear or even non-convex. Here we prove that, the robust constraints in (12b) and (12c) can be transformed into linear constraints based on our designed

uncertainty sets. **(P3)** can also be transformed into convex programming.

Specifically, for joint constraints **(12b)** and **(12c)**, they can be naturally decoupled as follows:

$$f_i^t \leq \bar{f}_i, \forall \xi \in \mathcal{U}_{i,1}, \forall i \in \mathcal{A}, \forall t, \quad (22)$$

$$f_i^t \geq -\bar{f}_i, \forall \xi \in \mathcal{U}_{i,2}, \forall i \in \mathcal{A}, \forall t. \quad (23)$$

These constraints can be explicitly expressed in terms of  $\mathbf{g}^t$  and  $\xi^t$  in the following form:

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i (\bar{\mathbf{d}}^t + \xi^t) \leq \bar{f}_i, \forall \xi \in \mathcal{U}_{i,1}, \forall i \in \mathcal{A}, \forall t, \quad (24)$$

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i (\bar{\mathbf{d}}^t + \xi^t) \geq -\bar{f}_i, \forall \xi \in \mathcal{U}_{i,2}, \forall i \in \mathcal{A}, \forall t, \quad (25)$$

where  $\mathbf{h}_g^i, \mathbf{h}_d^i$  denote the vector in the  $i$ th row of  $\mathbf{H}_g$  and  $\mathbf{H}_d$ ,  $\bar{\mathbf{d}}^t$  denotes the vector of predicted demand at time  $t$ , i.e.,  $\bar{\mathbf{d}}^t = [\bar{d}_i^t, \forall i \in \mathcal{A}]$ .

Such constraints can be further transformed into linear constraints as follows:

*Theorem 4:* The robust constraints **(24)** and **(25)** can be equivalently transformed into the following linear constraints:

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i \bar{\mathbf{d}}^t \leq \bar{f}_i - \alpha_{i,1}^t, \forall i \in \mathcal{A}, \forall t, \quad (26)$$

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i \bar{\mathbf{d}}^t \geq -\bar{f}_i + \alpha_{i,2}^t, \forall i \in \mathcal{A}, \forall t, \quad (27)$$

where  $\alpha_{i,1}^t, \alpha_{i,2}^t$  are scalars satisfying:

$$\alpha_{i,1}^t = -\mathbf{h}_d^i \boldsymbol{\mu}^t + \sqrt{s_{i,1}} \|(\mathbf{M}^t)^{\frac{1}{2}} (\mathbf{h}_d^i)^T\|_2, \quad (28)$$

$$\alpha_{i,2}^t = \mathbf{h}_d^i \boldsymbol{\mu}^t + \sqrt{s_{i,2}} \|(\mathbf{M}^t)^{\frac{1}{2}} (\mathbf{h}_d^i)^T\|_2. \quad (29)$$

Note that,  $\mathbf{M}^t$  denotes the submatrix of  $\mathbf{M}$  corresponding to dimension  $\xi^t$  in  $\xi$ . The detailed proof is deferred to Appendix C.

Intuitively, constants  $\alpha_{i,1}^t, \alpha_{i,2}^t$  are included to make the constraints tighter than the case without  $\xi$ .

By such a transformation, we can derive the following tractable RO:

$$(P4) \min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + \hat{R}(G^t, D^t) \right) \quad (30a)$$

$$\text{s.t. } \mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i \bar{\mathbf{d}}^t \leq \bar{f}_i - \alpha_{i,1}^t, \forall i \in \mathcal{A}, \forall t, \quad (30b)$$

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i \bar{\mathbf{d}}^t \geq -\bar{f}_i + \alpha_{i,2}^t, \forall i \in \mathcal{A}, \forall t,$$

$$\text{Constraints (1b)–(1d), (1h)–(1i).} \quad (30c)$$

Clearly, as long as  $C_i(\cdot)$  is convex, **(P4)** is a convex programming, which can be efficiently solved. By solving **(P4)**, we can obtain the solution to **(P2)**. Algorithm 1 summarizes the whole process.

## VII. SAMPLE-ADAPTIVE RO: RECONSTRUCTION

In this section, we introduce a solution reconstruction approach to further enhance the performance of RO. Specifically, we first demonstrate why RO is conservative and provide the intuition for the reconstruction approach. Then, we reconstruct

---

### Algorithm 1: Robust Optimization (RO).

---

**Input**  $\mathcal{D}_\xi, \rho_{i,1}, \rho_{i,2}, \eta_{i,1}, \eta_{i,2}, n_1, n_2, n_s$ ;

**Output** Solution  $\mathbf{g}$  with statistical feasibility guarantees;

- 1: Divide  $\mathcal{D}_\xi$  into  $\mathcal{D}_{\xi,1}$  and  $\mathcal{D}_{\xi,2}$  with size  $n_1$  and  $n_2$ ;
- 2: Calculate the approximated objective function according to **(6)**;

#### Uncertainty Set Construction:

- 1: Estimate shape parameters  $\boldsymbol{\mu}$  and  $\mathbf{M}$  based on  $\mathcal{D}_{\xi,1}$  according to **(16)**, **(17)**;

2: Construct the dimension-collapsing transformation function  $F(\xi^{(i)})$ ;

3: Calculate  $F(\xi^{(i)})$  for all  $\xi^{(i)} \in \mathcal{D}_{\xi,2}$ ;

4: Sort all  $F(\xi^{(i)})$  in ascending order to get the list  $\hat{\mathcal{S}} = \{\hat{s}_k, 1 \leq k \leq n_2\}$ ;

5: **for**  $i \in \mathcal{A}$  **do**

6:     Calculate the index  $k^*$  by **(20)** for  $\mathcal{U}_{i,1}$  and  $\mathcal{U}_{i,2}$ ;

7:     Obtain  $s_{i,1}, s_{i,2}$  based on  $k^*$  and  $\hat{\mathcal{S}}$ ;

8: **end for**

#### Optimization:

1: **for**  $t = 1, 2, \dots, T$  **do**

2:     **for**  $i \in \mathcal{A}$  **do**

3:         Calculate  $\alpha_{i,1}^t$  and  $\alpha_{i,2}^t$  based on **(28)** and **(29)**;

4:     **end for**

5: **end for**

6: Solve **(P4)** to get the solution  $\mathbf{g}$ ;

---

the uncertainty set. Finally, we solve the RO based on the reconstructed uncertainty set.

### A. Intuitions

The solution to **(P4)** is still conservative and can be further improved. The conservativeness comes from the generated uncertainty sets  $\mathcal{U}_{i,1}$  and  $\mathcal{U}_{i,2}$ . For example, we determine that all the uncertainty sets are approximated by ellipsoids, which may not be optimal. Further, the constructed uncertainty sets are independent of the specific forms of the constraints. Therefore, there is a gap between the generated uncertainty set and the optimal uncertainty set. Specifically, for uncertainty set  $\mathcal{U}_{i,1}$  and the corresponding solution  $\mathbf{g}$ , the following condition holds:

$$\mathbb{P}_\xi (\xi \in \mathcal{U}_{i,1}) \leq \mathbb{P}_\xi \left( \bigcap_{t=1}^T \{f_i^t(\mathbf{g}^t; \xi^t) \leq \bar{f}_i\} \right), \forall i, \forall t. \quad (31)$$

This gap comes from the RO framework itself, even with  $\mathbb{P}_\xi (\xi \in \mathcal{U}_{i,1}) = \rho_{i,1}$ , such a gap still exists and makes our solution conservative. To handle this problem, we make an important observation:

*Fact 2:* For any feasible solution  $\mathbf{g} = [\mathbf{g}^t, \forall t]$ , it holds:

$$\mathbb{P}_\xi (\xi \in \mathcal{U}_{i,1}) = \mathbb{P}_\xi \left( \bigcap_{t=1}^T \{f_i^t(\mathbf{g}^t; \xi^t) \leq \bar{f}_i\} \right), \quad (32)$$

if

$$\mathcal{U}_{i,1} = \{\xi \mid \bigcap_{t=1}^T \{f_i^t(\mathbf{g}^t; \xi^t) \leq \bar{f}_i\}\}, \forall i. \quad (33)$$

This observation inspires us to construct a better uncertainty set with the shape closely related to the solution  $\mathbf{g}$ . Specifically, when we obtain uncertainty sets  $\mathcal{U}_{i,1}, \mathcal{U}_{i,2}$  and get the solution  $\mathbf{g}$ , to reduce the gap in (31), we can further reconstruct the uncertainty set with the shape specified by (33) based on  $\mathbf{g}$ .

### B. Uncertainty Set Reconstruction

To implement this idea, we estimate the following uncertainty sets:

$$\mathcal{U}_{i,1} = \left\{ \boldsymbol{\xi} \mid \bigcap_{t=1}^T \{f_i^t(\hat{\mathbf{g}}^t; \boldsymbol{\xi}^t) \leq \bar{f}_i + s_{i,1}\} \right\}, \quad (34)$$

$$\mathcal{U}_{i,2} = \left\{ \boldsymbol{\xi} \mid \bigcap_{t=1}^T \{f_i^t(\hat{\mathbf{g}}^t; \boldsymbol{\xi}^t) \geq -\bar{f}_i - s_{i,2}\} \right\}. \quad (35)$$

We again split the dataset into  $\mathcal{D}_{\xi,1}, \mathcal{D}_{\xi,2}$  with sizes  $n_1$  and  $n_2$ .  $\mathcal{D}_{\xi,1}$  is adopted for estimating  $\hat{\mathbf{g}}$ , and  $\mathcal{D}_{\xi,2}$  is adopted for estimating  $s_{i,1}, s_{i,2}$ . Again, we only construct  $\mathcal{U}_{i,1}$  as the process to construct  $\mathcal{U}_{i,2}$  following exactly the same routine.

1) *Shape Estimation*: We can run the RO algorithm straightforwardly based on  $\mathcal{D}_{\xi,1}$  to get the initial solution  $\hat{\mathbf{g}}$ .

2) *Size Calibration*: This calibration for  $s_{i,1}$  is similar to that in RO. The only difference comes from the joint constraints in the uncertainty set. We design a customized dimension-collapsing transformation function  $F_{Recon,1}$  as follows:

$$F_{Recon,1}(\boldsymbol{\xi}) = \max_t f_i^t(\hat{\mathbf{g}}^t; \boldsymbol{\xi}^t) - \bar{f}_i. \quad (36)$$

Intuitively, for any sample  $\boldsymbol{\xi}$ ,  $F_{Recon}$  produces a minimal  $s_{i,1}$  to make sure the constructed uncertainty set can cover this sample. Similar to the RO, we transform each sample vector  $\boldsymbol{\xi}^{(i)}$  in  $\mathcal{D}_{\xi,1}$  into a scalar with  $F_{Recon}(\boldsymbol{\xi}^{(i)})$ , and then sort all these  $F_{Recon}(\boldsymbol{\xi}^{(i)})$  into a new set  $\{\hat{s}_k, 1 \leq k \leq n_2\}$  ascendingly, i.e.,  $\hat{s}_k \leq \hat{s}_{k+1}$ .

The desired  $s_{i,1}$  is estimated by:

$$s_{i,1} = \hat{s}_{k^*}, \quad (37)$$

where the index  $k^*$  satisfies:

$$k^* = \min \left\{ r : \sum_{k=0}^{r-1} C_{n_2}^k (\rho_{i,1})^k (1 - \rho_{i,1})^{n_2-k} \geq \eta_{i,1} \right\}. \quad (38)$$

For  $s_{i,2}$ , the only difference is that, the dimension-collapsing transformation function  $F_{Recon,2}$  should be:

$$F_{Recon,2}(\boldsymbol{\xi}) = \max_t -f_i^t(\hat{\mathbf{g}}^t; \boldsymbol{\xi}^t) - \bar{f}_i. \quad (39)$$

Similarly, the statistical feasibility of the constraints can be guaranteed as long as condition (21) holds.

### C. Solving ReconRO by Linear Transformation

Similarly, by solving the robust constraints with our estimated uncertainty set, the original uncertainty set

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i (\bar{\mathbf{d}}^t + \boldsymbol{\xi}^t) \leq \bar{f}_i, \forall \boldsymbol{\xi} \in \mathcal{U}_{i,1}, \forall i \in \mathcal{A}, \forall t, \quad (40)$$

$$\mathbf{h}_g^i \mathbf{g}^t - \mathbf{h}_d^i (\bar{\mathbf{d}}^t + \boldsymbol{\xi}^t) \geq -\bar{f}_i, \forall \boldsymbol{\xi} \in \mathcal{U}_{i,2}, \forall i \in \mathcal{A}, \forall t, \quad (41)$$

---

### Algorithm 2: Reconstructed Robust Optimization (ReconRO).

---

**Input**  $\mathcal{D}_{\xi}, \rho_{i,1}, \rho_{i,2}, \eta_{i,1}, \eta_{i,2}, n_1, n_2, n_s$ ;

**Output** Solution  $\mathbf{g}$  with statistical feasibility guarantees;

1: Divide  $\mathcal{D}_{\xi}$  into  $\mathcal{D}_{\xi,1}$  and  $\mathcal{D}_{\xi,2}$  with size  $n_1$  and  $n_2$ ;

2: Calculate the approximate objective function according to (6);

#### Uncertainty Set Construction:

1: Calculate the approximated solution  $\hat{\mathbf{g}}$  based on Algorithm 1 and  $\mathcal{D}_{\xi,1}$ ;

2: Construct the dimension-collapsing transformation function  $F_{Recon,1}, F_{Recon,2}$ ;

3: **for**  $i \in \mathcal{A}$  **do**

4: Calculate  $F_{Recon,1}(\boldsymbol{\xi}^{(i)}), F_{Recon,2}(\boldsymbol{\xi}^{(i)})$  for all  $\boldsymbol{\xi}^{(i)} \in \mathcal{D}_{\xi,2}$ ;

5: Sort all  $F_{Recon,1}(\boldsymbol{\xi}^{(i)}), F_{Recon,2}(\boldsymbol{\xi}^{(i)})$  in ascending order to get the lists  $\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2$ ;

6: Calculate the index  $k^*$  by (38) for  $\mathcal{U}_{i,1}$  and  $\mathcal{U}_{i,2}$ ;

7: Obtain  $s_{i,1}, s_{i,2}$  based on  $k^*, \hat{\mathcal{S}}_1$ , and  $\hat{\mathcal{S}}_2$ ;

8: **end for**

#### Optimization:

1: Solve (P5) to get the solution  $\mathbf{g}$ ;

---

can be transformed into the linear constraints:

$$\mathbf{h}_g^i (\mathbf{g}^t - \hat{\mathbf{g}}^t) + s_{i,1} \leq 0, \forall i \in \mathcal{A}, \forall t, \quad (42)$$

$$\mathbf{h}_g^i (\mathbf{g}^t - \hat{\mathbf{g}}^t) - s_{i,2} \geq 0, \forall i \in \mathcal{A}, \forall t. \quad (43)$$

Through such a transformation, we can derive the following tractable RO:

$$(P5) \min \sum_{t=1}^T \left( \sum_{i \in \mathcal{N}} C_i(g_i^t) + \hat{R}(G^t, D^t) \right) \quad (44a)$$

s.t. Constraints (42), (43),

Constraints (1b)–(1d), (1h)–(1i).  $(44b)$

(P5) is also a convex programming and can be efficiently solved. The whole process is provided in Algorithm 2.

*Remark*: Our approach can be naturally extended to the unit commitment economic dispatch (UCED) problem, since the estimation process just relies on the samples. After obtaining the uncertainty set, the statistically robust optimization for UCED will be mixed-binary linear programming (due to the startup cost and UC binary decision variables).

## VIII. NUMERICAL STUDIES

We evaluate the performance of our proposed RO and ReconRO algorithm in terms of the dispatch cost, the constraint violation rate, and the computational time on a 4-bus system and the IEEE 118-bus system. The numerical study is performed by Gurobi 9.5.0 [36] on a PC with Intel Core i5-11400F CPU and 16G RAM.

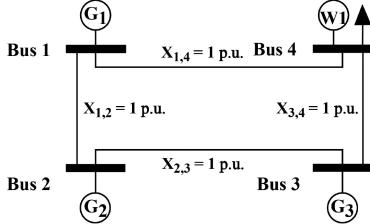


Fig. 7. Network of the 4-bus case (modified from [38]).

TABLE II  
4-BUS CASE: GENERATOR INFORMATION

Unit	Min (MW)	Max (MW)	$u_1(\$/\text{MW})$	$u_2(\$/\text{MW})$	$u_3(\$/\text{MW})$	$v_1 (\$)$	$v_2 (\$)$	$v_3 (\$)$	Ramp Rate
G1	0	25.0	40.0	45.0	60.0	0	-50	-350	0.1
G2	0	20.0	50.0	55.0	70.0	0	-50	-250	0.15
G3	0	15.0	60.0	65.0	90.0	0	-50	-360	0.2

### A. Data Description and Experimental Settings

The renewable energy data are extracted from open data provided by the Bonneville Power Administration (BPA) [37]. Specifically, it provides historical data on the predicted and actual total outputs of wind turbines in the BPA control area with a resolution of 5 minutes. The first numerical case we adopt is modified from the 4-bus case in [38]. Fig. 7 illustrates the network structure and line parameters. The parameters of generators are listed in Table II, where the generation cost is a piece-wise linear function, specified by slope  $u_i(\$/\text{MW})$  and the intercept  $v_i(\$)$ ,  $i = 1, 2, 3$ .

Another numerical study is based on the IEEE 118-bus System. The parameters for traditional generators (i.e., maximum and minimum output, and up and down ramp limit) and line capacities follow the settings in [39]. To contrive the extreme case of network congestion, we reduce the line capacities to 30% of their original values.

We set the shortage and excess risk costs  $\gamma_1$  and  $\gamma_2$  to be \$300/MWh and \$20/MWh. All the stability requirement  $\rho$  and confidence level  $\eta$  are set to be 0.95.

### B. Competing Methods

We compare the following three categories of five different approaches for ED:

1) *Chance-Constrained Optimization*: Chance constrained optimization first estimates the distribution of  $\xi$  based on the samples, and then obtains the explicit form of the joint CCs. Further, it decomposes the joint CCs with the Bonferroni correction [40], and then solves the decomposed problem. We compare two variants of chance constrained optimization by assuming the error  $\xi$  follows Gaussian distribution (CC-G) and Laplace distribution (CC-L).

The chance-constrained optimization requires the distribution type information of the random variables. With a better distribution type assumption, the performance of CC should be better. Specifically, Fig. (8) illustrates the net load forecast error.

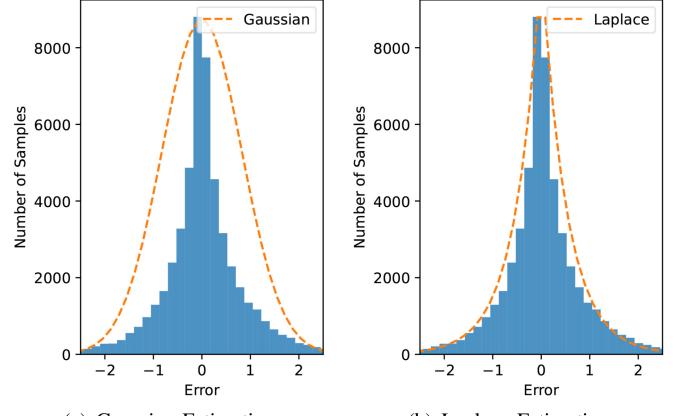


Fig. 8. Gaussian and Laplace Estimations of Net Load Forecast Error.

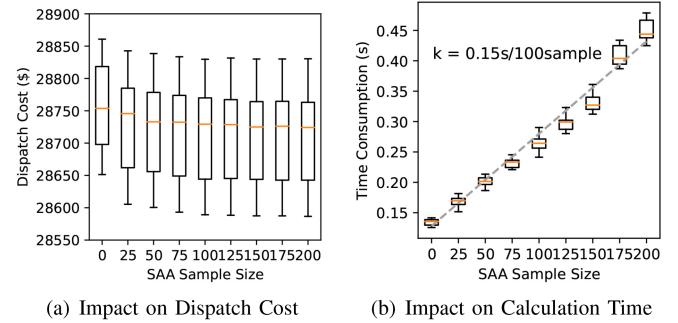


Fig. 9. Objective Function Approximation Performance.

We can observe that the Laplace distribution has a better fitting performance than the Gaussian distribution.

2) *Robust Optimization (RO)*: We compare two RO algorithms proposed in this paper: the basic robust optimization algorithm (RO) and ReconRO.

3) *Scenario Generation*: SG is an approach to tackle the chance-constrained optimization based on sampling the CCs.

### C. Evaluation on 4-Bus System

We first evaluate the objective function approximation based on our proposed ReconRO approach. Fig. 9 illustrates how the dispatch cost and the running time change with increasing SAA sample size. We can observe that the dispatch cost first goes down and then flattens out, while the running time increases linearly with a slope  $k = 0.15 \text{ s}/100 \text{ samples}$ . This coincides with our theoretical trade-off results. We choose the SAA sample size to be 50 in the subsequent experiments based on this trade-off.

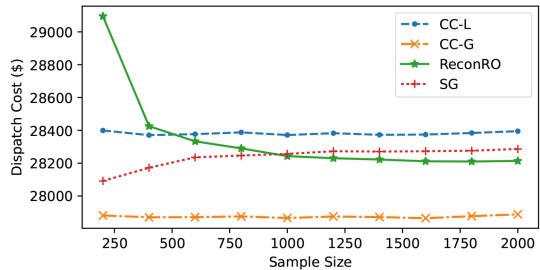
We perform 300 trials for the five different approaches in limited sample case ( $n = 200$ ) and sufficient sample case ( $n = 1000$ ), and each trial involves a randomly generated dataset of samples. The experimental results are provided in Tables III and IV. We can observe that, with limited samples, the empirical violation rates  $1 - \hat{\rho}$  and statistical violation rate  $1 - \hat{\eta}$  of CC-G and SG are larger than 0.05, which do not satisfy the requirement.

TABLE III  
PERFORMANCE ON 4-BUS SYSTEM WITH LIMITED SAMPLES

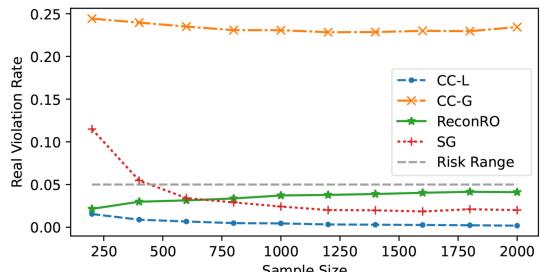
	Approaches ( $n = 200$ , $1 - \rho \leq 0.05$ , $1 - \eta \leq 0.05$ )				
	CC-L	CC-G	RO	ReconRO	SG
Cost (\$)	<b>28399.1</b>	27871.6	35480.0	29095.1	28091.4
$1 - \hat{\rho}$	0.015	0.243	0.0	0.022	0.115
$1 - \hat{\eta}$	0.013	0.997	0.0	0.040	0.983

TABLE IV  
PERFORMANCE ON 4-BUS SYSTEM WITH SUFFICIENT SAMPLES

	Approaches ( $n = 1000$ , $1 - \rho \leq 0.05$ , $1 - \eta \leq 0.05$ )				
	CC-L	CC-G	RO	ReconRO	SG
Cost (\$)	28375.5	27870.0	32593.7	<b>28242.8</b>	28256.3
$1 - \hat{\rho}$	0.004	0.233	0.0	0.037	0.024
$1 - \hat{\eta}$	0.0	1.0	0.0	0.043	0.027



(a) Impact on Dispatch Cost

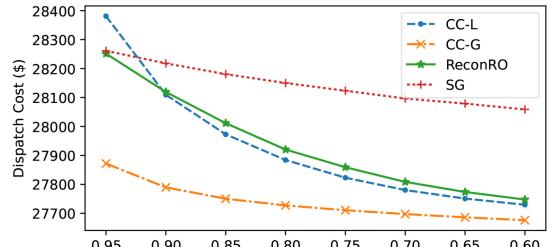


(b) Impact on Violation Rate

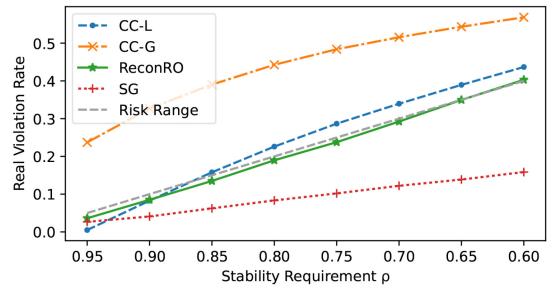
Fig. 10. Impact of Sample Size.

RO is too conservative, yielding a significantly higher dispatch cost. ReconRO is less conservative than RO, yielding a significantly lower dispatch cost. CC-L enjoys the best performance while satisfying the requirements for risk rates. This is because CC-L can approximate the prediction error very well. On the other hand, with sufficient samples ( $n = 1000$ ), only CC-G violates the risk rate constraints. Among the other approaches, RO is still too conservative; ReconRO and SG enjoy the lowest generation cost. The detailed calculation process is provided in Appendix D.

Further, Fig. 10 illustrates how the dispatch cost and violation rate of different approaches change with the sample size. Interestingly, the performance trend varies significantly between different methods. Specifically, the dispatch cost and violation rate of CC-L and CC-G do not change much with increasing sample size. The reason is that these two approaches are mainly



(a) Impact on Dispatch Cost



(b) Impact on Violation Rate

Fig. 11. Impact of Stability Requirement  $\rho$ .

based on parametric estimation of the sample distribution, i.e., the mean and variance. The law of large numbers guarantees that it only requires much fewer samples for such parametric estimations, so more samples hardly lead to improvement. For SG, the dispatch cost increases with sample size, but the violation rate decreases. This is because SG generates the constraints based on samples. When the sample size is limited, the amount of the constraints is also limited, which leads to a solution with low cost but often infeasible. On the other hand, with more samples, SG can generate enough constraints to make the solution more conservative yet feasible to the CCs. In contrast, with the increase in sample size, our ReconRO can better exploit the flexibility reserve from the CCs and provide a less conservative solution with an increasing but condition-satisfying violation rate. It also outperforms all the other feasible approaches when the sample size  $n \geq 1000$ , demonstrating its remarkable performance.

Fig. 11 presents the performance of different approaches when the stability requirement  $\rho$  decreases. The sample size  $n = 1000$ . We can observe that the solution CC-G is always infeasible (the curve is above the grey risk range curve in Fig. 11(b)), and when  $\rho < 0.85$ , the solution of CC-L is always infeasible. This indicates that CC methods are not always guaranteed to satisfy the CCs. On the other hand, the solution of SG is always feasible and performs more economically when  $\rho$  decreases. In contrast, it is interesting to find that the violation rate curve of ReconRO is always lower but closely approaches the allowed risk range curve, which indicates that it can exploit the flexibility reserve provided by the CCs. The dispatch cost of ReconRO also drops quickly when  $\rho$  decreases.

In summary, these approaches have the following properties:

- **CC:** First, CC often strongly relies on the distribution assumption. With a more accurate distribution assumption,

TABLE V  
PERFORMANCE ON 118-BUS SYSTEM

	Approaches ( $n = 2000, 1 - \rho \leq 0.05, 1 - \eta \leq 0.05$ )				
	CC-L	CC-G	RO	ReconRO	SG
Extra Cost (\$)	9315.2	9214.1	17396.7	<b>9086.2</b>	9253.3
$1 - \hat{\rho}$	0.0001	0.0013	0.0	0.0361	0.0006

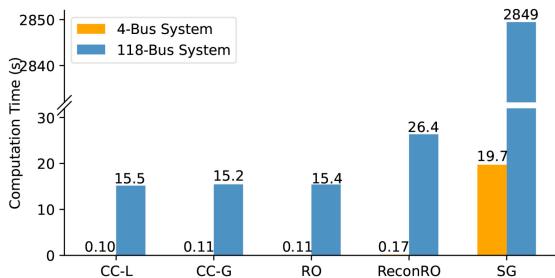


Fig. 12. Time Consumption.

it can enjoy a good economic performance (like CC-L). Second, it is not sensitive to the sample size. Although it can perform well with limited data, the performance doesn't change when the sample size increases. Finally, it cannot always guarantee the violation rate requirements.

- **RO:** RO is the most conservative approach with the highest cost in almost all cases. However, it is also the safest approach and never violates the constraints.
- **ReconRO:** Our ReconRO can stably guarantee the violation rate requirements. When the sample size increases, its economic performance gradually improves. In most cases, it outperforms the other constraint-satisfying approaches.
- **SG:** SG is strongly dependent on the sample size. When the sample size is limited, it cannot satisfy the violation rate requirement. And when the samples are enough, it can achieve an economic performance close to ReconRO with a significantly longer running time.

#### D. Evaluation on IEEE 118-Bus System

We evaluate the performance of our algorithm based on the IEEE 118-bus system. Table V illustrates the extra cost (refer to the gap to the optimal dispatch with accurate prediction) and empirical violation rate for different approaches. We can observe that the violation rates of all approaches meet the requirements. RO is still the most conservative approach. Our ReconRO achieves the lowest dispatch cost among all methods and reduces the extra cost by 1.39% compared with the optimal approach.

For the computational efficiency, Fig. 12 illustrates the running times of different approaches on the 4-bus system and IEEE 118-bus system. CC-G, CC-L and RO enjoy the shortest computation time. The ReconRO takes 70% more computation time since it contains two ROs in the solution process. The ReconRO can be performed on the IEEE-118 bus system within 30s. The SG is the most time-consuming approach, which takes roughly 100 times more running time than ReconRO. This

demonstrates that our ReconRO simultaneously achieves good economic performance and high computational efficiency.

## IX. CONCLUSION

In this paper, we introduce the notion of statistical feasibility and the formulation of statistically feasible ED. To tackle the statistically feasible ED, we propose a robust sample-adaptive optimization with statistical feasibility guarantees. To overcome the conservativeness of RO, we further propose a reconstruction approach to generate the constraint-specific uncertainty sets and then design the corresponding ReconRO for effective ED. The numerical studies based on a 4-bus system and the IEEE 118-bus system verify our methods' performance and sample adaptiveness.

## APPENDIX

### A. Proof for Theorem 1

The approximation error  $e(n_s)$  with sample size  $n_s$  satisfies:

$$e(n_s) = \sum_{t=1}^T \left( \hat{R}(G^t, D^t) - R(G^t, D^t) \right). \quad (45)$$

Combining with (6) yields:

$$e(n_s) = \sum_{t=1}^T \left( \frac{1}{n_s} \sum_{\xi^{(j)} \in \mathcal{D}_\xi^s} \left( \tilde{R}(G^t, \xi^{(j),t}) - R(G^t, D^t) \right) \right). \quad (46)$$

Denote  $\sum_{t=1}^T \tilde{R}(G^t, \xi^{(j),t})$  as  $A(\xi^{(j)})$ . Standard manipulations simplify  $e(n_s)$  as follows:

$$e(n_s) = \frac{1}{n_s} \sum_{\xi^{(j)} \in \mathcal{D}_\xi^s} \left( A(\xi^{(j)}) - \bar{A}(\xi) \right), \quad (47)$$

where  $A(\xi^{(j)})$  is i.i.d. for different  $j$ , and  $\bar{A}(\xi)$  denotes the expectation of  $A(\xi^{(j)})$  for all  $j$ .

Assuming each element in  $\xi$  can be bounded by  $[\underline{\xi}, \bar{\xi}]$ , we can derive the following condition for  $A(\xi)$ :

$$\max_{\xi} A(\xi) - \min_{\xi} A(\xi) \leq T \max(\gamma_1, \gamma_2)(\bar{\xi} - \underline{\xi}), \quad (48)$$

$$\leq T \sqrt{\gamma_1^2 + \gamma_2^2} (\bar{\xi} - \underline{\xi}). \quad (49)$$

Plugging (49) and the Hoeffding's inequality [41] into (47) immediately yields our results.

### B. Proof for Corollary 1

For the approximation error, Theorem 1 theoretically addresses the error as follows:

$$\mathbb{P}(|e(n_s)| \geq \Delta) \leq 2 \exp \left( - \frac{2n_s \Delta^2}{T^2(\gamma_1^2 + \gamma_2^2)(\bar{\xi} - \underline{\xi})^2} \right), \quad (50)$$

where  $n_s$  denotes the sample size.

It shows that given an upper limit of error  $\Delta$ , the probability that the absolute value of approximation error is larger than  $\Delta$

is smaller than  $2 \exp(-\frac{2n_s \Delta^2}{T^2(\gamma_1^2 + \gamma_2^2)(\bar{\xi} - \underline{\xi})^2})$ . Eliminating the independent parameters  $\gamma_1$ ,  $\gamma_2$ ,  $\bar{\xi}$ ,  $\underline{\xi}$ , and  $\Delta$ , we can directly obtain that such a probability decreases at the rate of  $O(\exp(-n_s/T^2))$  with the increase of sample size  $n_s$ .

For the computational complexity, we use sample average approximation (SAA) to approximate the risk cost  $R(G^t, D^t)$  as follows:

$$\hat{R}(G^t, D^t) = \frac{1}{n_s} \sum_{\xi^{(j)} \in \mathcal{D}_\xi^s} \tilde{R}(G^t, \xi^{(j), t}), \quad (51)$$

where

$$\begin{aligned} \tilde{R}(G^t, \xi^{(j), t}) &= \gamma_1 \left( \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^{(j), t}) - G^t \right)^+ \\ &\quad + \gamma_2 \left( G_t - \sum_{i \in \mathcal{M}} (\bar{d}_i^t + \xi_i^{(j), t}) \right)^+. \end{aligned} \quad (52)$$

and

$$(x)^+ = \max\{x, 0\}. \quad (53)$$

Since a linear transformation can easily remove the  $(\cdot)^+$  operator, the derived sample average  $\hat{R}(G^t, D^t)$  is essentially a linear function of the decision variable  $g_i^t$ . However, removing the term with  $(\cdot)^+$  operator implicitly includes one additional decision variable, i.e., to represent a variable  $(x)^+$ . Hence, we need to include at least one more variable  $y$  with the following constraints<sup>1</sup>:

$$y \geq x, \quad \text{and} \quad y \geq 0. \quad (54)$$

Specifically, the number of terms with  $(\cdot)^+$  operator is  $2n_s T$ . Therefore, the number of implicit decision variables is  $O(n_s T)$ . This concludes our proof.

### C. Proof for Theorem 4

Consider the following robust constraints:

$$\mathbf{h}_g^i g^t - \mathbf{h}_d^i (\bar{d}^t + \xi^t) \leq \bar{f}_i, \forall \xi \in \mathcal{U}_{i,1}, \forall i \in \mathcal{A}, \forall t, \quad (55)$$

$$\mathbf{h}_g^i g^t - \mathbf{h}_d^i (\bar{d}^t + \xi^t) \geq -\bar{f}_i, \forall \xi \in \mathcal{U}_{i,2}, \forall i \in \mathcal{A}, \forall t. \quad (56)$$

We can observe that the random variable  $\xi^t$  is with the additive relation to the decision variables  $g^t$ . Therefore, constraints (55), (56) can be transformed into the following form:

$$\mathbf{h}_g^i g^t - \mathbf{h}_d^i \bar{d}^t \leq \bar{f}_i - \alpha_{i,1}^t, \forall i \in \mathcal{A}, \forall t, \quad (57)$$

$$\mathbf{h}_g^i g^t - \mathbf{h}_d^i \bar{d}^t \geq -\bar{f}_i + \alpha_{i,2}^t, \forall i \in \mathcal{A}, \forall t, \quad (58)$$

where  $\alpha_{i,1}^t$  and  $\alpha_{i,2}^t$  satisfy:

$$\alpha_{i,1}^t = -\min \mathbf{h}_d^i \xi^t \quad (59)$$

<sup>1</sup>Essentially, such a transformation only guarantees  $y \geq (x)^+$  instead of  $y = (x)^+$ . To strictly enforce  $y = (x)^+$ , we need to include integer variables. However, it can be avoided in our problem, since we target to minimize the total cost, and a larger  $y$  will lead to a higher risk cost. Therefore, all terms with  $y$  are minimized in the solution, yielding  $y = (x)^+$ . This is also a commonly adopted transformation for problems with such a property.

$$\text{s.t. } (\xi - \mu)^\top M^{-1} (\xi - \mu) \leq s_{i,1}, \quad (60)$$

$$\alpha_{i,2}^t = \max \mathbf{h}_d^i \xi^t \quad (61)$$

$$\text{s.t. } (\xi - \mu)^\top M^{-1} (\xi - \mu) \leq s_{i,2}. \quad (62)$$

The Slater's condition [42] further guarantees the strong duality condition, yielding:

$$\begin{aligned} \alpha_{i,1}^t &= -\max_{\lambda} \min_{\xi} \mathbf{h}_d^i \xi^t + \lambda ((\xi - \mu)^\top M^{-1} (\xi - \mu) - s_{i,1}) \\ \text{s.t. } \lambda &\geq 0, \end{aligned}$$

$$\begin{aligned} \alpha_{i,2}^t &= \max_{\lambda} \min_{\xi} -\mathbf{h}_d^i \xi^t + \lambda ((\xi - \mu)^\top M^{-1} (\xi - \mu) - s_{i,2}) \\ \text{s.t. } \lambda &\geq 0. \end{aligned}$$

By the positive definitiveness of the covariance matrix  $M$ , we can directly solve the quadratic programming and derive the objective explicitly to characterize  $\alpha_{i,1}^t$ ,  $\alpha_{i,2}^t$ .

### D. Details of the Experimental Calculation

We divide the samples into two equal parts for estimation (Set  $\mathcal{X}$ ) and for evaluation (Set  $\mathcal{Y}$ ), respectively. For calculating the items in Tables II and III in our work, we first randomly generate 300 sample sets with size 200 (Table II) and size 1000 (Table III) by sampling the data from Set  $\mathcal{X}$  for 300 times. Based on each generated sample set, we conduct the estimations for different approaches (e.g., uncertainty set, distribution estimation), and then conduct the economic dispatch with different approaches based on the predicted netload (consisting of the basic load from the system parameter and the turbine's output) and the samples of the prediction error of Set  $\mathcal{Y}$  on the system.

Specifically, the average cost, average empirical violation rate  $1 - \hat{\rho}$  and the average statistical violation rate  $1 - \hat{\eta}$  in the table are calculated as follows:

$$\text{Cost} = \frac{1}{300 \cdot ||\mathcal{Y}||} \sum_{\mathcal{D}_\xi} \sum_{\xi \in \mathcal{Y}} \text{Cost}(\mathcal{D}_\xi, \xi), \quad (63)$$

$$1 - \hat{\rho} = \frac{1}{300 \cdot ||\mathcal{Y}||} \sum_{\mathcal{D}_\xi} \sum_{\xi \in \mathcal{Y}} I_{vio}(\mathcal{D}_\xi, \xi), \quad (64)$$

$$1 - \hat{\eta} = \frac{1}{300} \sum_{\mathcal{D}_\xi} I \left( \frac{1}{||\mathcal{Y}||} \sum_{\xi \in \mathcal{Y}} I_{vio}(\mathcal{D}_\xi, \xi) > 0.05 \right), \quad (65)$$

where  $\text{Cost}(\mathcal{D}_\xi, \xi)$  represents the dispatch cost with the sample set  $\mathcal{D}_\xi$  and real prediction error  $\xi$ ;  $I_{vio}(\mathcal{D}_\xi, \xi)$  denotes the indicator function that the solution violates any constraint (1) or not (0) with the sample set  $\mathcal{D}_\xi$  and real prediction error  $\xi$ ;  $I(\frac{1}{||\mathcal{Y}||} \sum_{\xi \in \mathcal{Y}} I_{vio}(\mathcal{D}_\xi, \xi) > 0.05)$  is the indicator function that indicates whether the average violation rate exceeds 0.05 (1) or not (0) with  $\mathcal{D}_\xi$ .

## REFERENCES

- [1] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Berlin, Germany, Springer, 2011.
- [2] X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang, "Scenario generation of wind power based on statistical uncertainty and variability," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 894–904, Oct. 2013.

- [3] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 422–433, Apr. 2015.
- [4] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 800–810, May 2012.
- [5] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb. 2013.
- [6] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 206–215, Feb. 2012.
- [7] D. Pozo and J. Contreras, "A chance-constrained unit commitment with an  $n - k$  security criterion and significant wind generation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2842–2851, Aug. 2013.
- [8] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Rev.*, vol. 56, no. 3, pp. 461–495, 2014.
- [9] G. A. Hanusanto, V. Roitch, D. Kuhn, and W. Wiesemann, "Ambiguous joint chance constraints under mean and dispersion information," *Oper. Res.*, vol. 65, no. 3, pp. 751–767, 2017.
- [10] A. Geletu, A. Hoffmann, and P. Li, "Analytic approximation and differentiability of joint chance constraints," *Optimization*, vol. 68, no. 10, pp. 1985–2023, 2019.
- [11] W. van Ackooij, E. C. Finardi, and G. M. Ramalho, "An exact solution method for the hydrothermal unit commitment under wind power uncertainty with joint probability constraints," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6487–6500, Nov. 2018.
- [12] X. Geng and L. Xie, "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization," *Annu. Rev. Control*, vol. 47, pp. 341–363, 2019.
- [13] 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. 2, pp. 1378–1388, Mar. 2017.
- [14] 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.
- [15] W. Xie and S. Ahmed, "Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1860–1867, Mar. 2018.
- [16] N. Gu, H. Wang, J. Zhang, and C. Wu, "Bridging chance-constrained and robust optimization in an emission-aware economic dispatch with energy storage," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1078–1090, Mar. 2022.
- [17] S. Ghosal and W. Wiesemann, "The distributionally robust chance-constrained vehicle routing problem," *Oper. Res.*, vol. 68, no. 3, pp. 716–732, 2020.
- [18] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4/part-1, pp. 902–917, 2010.
- [19] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2007.
- [20] W. Xie, S. Ahmed, and R. Jiang, "Optimized bonferroni approximations of distributionally robust joint chance constraints," *Math. Program.*, vol. 191, no. 1, pp. 79–112, 2022.
- [21] W. Xie, "On distributionally robust chance constrained programs with Wasserstein distance," *Math. Program.*, vol. 186, no. 1/2, pp. 115–155, 2021.
- [22] Z. Chen, D. Kuhn, and W. Wiesemann, "Data-driven chance constrained programs over Wasserstein balls," *Oper. Res.*, early access, 2022, doi: [10.1287/opre.2022.2330](https://doi.org/10.1287/opre.2022.2330).
- [23] S. Wang, J. Li, and S. Mehrotra, "A solution approach to distributionally robust joint-chance-constrained assignment problems," *INFORMS J. Optim.*, vol. 4, no. 2, pp. 125–147, 2022.
- [24] R. Jagannathan, "Chance-constrained programming with joint constraints," *Oper. Res.*, vol. 22, no. 2, pp. 358–372, 1974.
- [25] W. Xie and S. Ahmed, "On deterministic reformulations of distributionally robust joint chance constrained optimization problems," *SIAM J. Optim.*, vol. 28, no. 2, pp. 1151–1182, 2018.
- [26] 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.
- [27] L. J. Hong, Z. Huang, and H. Lam, "Learning-based robust optimization: Procedures and statistical guarantees," *Manage. Sci.*, vol. 67, no. 6, pp. 3447–3467, 2021.
- [28] 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.
- [29] A. Peña-Ordieres, D. K. Molzahn, L. A. Roald, and A. Wächter, "DC optimal power flow with joint chance constraints," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 147–158, Jan. 2021.
- [30] Z. Wang, C. Shen, F. Liu, X. Wu, C.-C. Liu, and F. Gao, "Chance-constrained economic dispatch with non-Gaussian correlated wind power uncertainty," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4880–4893, Nov. 2017.
- [31] Y. Yang, W. Wu, B. Wang, and M. Li, "Chance-constrained economic dispatch considering curtailment strategy of renewable energy," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5792–5802, Nov. 2021.
- [32] Y. Zhang, X. Ai, J. Wen, J. Fang, and H. He, "Data-adaptive robust optimization method for the economic dispatch of active distribution networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3791–3800, Jul. 2019.
- [33] M. A. Velasquez, N. Quijano, A. I. Cadena, and M. Shahidehpour, "Distributed stochastic economic dispatch via model predictive control and data-driven scenario generation," *Int. J. Elect. Power Energy Syst.*, vol. 129, 2021, Art. no. 106796.
- [34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [35] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.
- [36] Gurobi, "Gurobi solver 9.5.0," 2021. Accessed: May 22, 2022. [Online]. Available: <https://www.gurobi.com>
- [37] B. P. Administration, "Wind generation and total load in the bonneville power administration balancing authority," 2020. Accessed: May 17, 2022. [Online]. Available: <https://transmission.bpa.gov/Business/Operations/Wind/>
- [38] C. Lu, W. Jiang, and C. Wu, "Effective end-to-end learning framework for economic dispatch," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2673–2683, Jul./Aug. 2022.
- [39] I. I. of Technology Power Group, "IEEE 118-bus test system," 2015. Accessed: May 1, 2022. [Online]. Available: <http://motor.ece.iit.edu/data/>
- [40] X. Chen, M. Sim, and P. Sun, "A robust optimization perspective on stochastic programming," *Oper. Res.*, vol. 55, no. 6, pp. 1058–1071, 2007.
- [41] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [42] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.



**Chenbei Lu** (Graduate Student Member, IEEE) received the bachelor's degree from the School of Software Engineering, Huazhong University of Science and Technology, Wuhan, China. He is currently working toward the Ph.D. degree in computer science and technology with the Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, China, advised by Professor Chenye Wu. He was the recipient of the National Scholarship in 2017 and Excellent Graduate of Huazhong University of Science and Technology in 2020. He is currently working on the optimal design and operation of power systems.



**Nan Gu** (Graduate Student Member, IEEE) received the bachelor's degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China. He is currently working toward the Ph.D. degree in computer science and technology at IIIS, Tsinghua University, advised by Professor Chenye Wu. Her research interests include market design and optimization in the power system. She was awarded Excellent Graduate of Tsinghua University in 2019.



**Wenqian Jiang** (Graduate Student Member, IEEE) received the master's degree from China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan, China. He is currently working toward the Ph.D. degree in computer and information engineering with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China, advised by Professor Chenye Wu. She was awarded Excellent Graduate of Huazhong University of Science and Technology in 2021. He is working on economic analysis and optimization of power systems.



**Chenyen Wu** (Member, IEEE) received the Ph.D. degree in computer science and technology from IIIS, Tsinghua University, Beijing, China, in July 2013. He is currently an Assistant Professor with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen (CUHK Shenzhen), China. Before joining CUHK Shenzhen, he was an Assistant Professor at IIIS, Tsinghua University. He was with ETH Zurich as a Wiss. Mitarbeiter (Research Scientist), working with Professor Gabriela Hug, in 2016. Before that, Professor Kameshwar Poolla and Professor Pravin Varaiya hosted. He is a Postdoctoral Researcher at UC Berkeley for two years. During 2013–2014, he spent one year at Carnegie Mellon University as a Postdoc Fellow, hosted by Professor Gabriela Hug and Professor Soummya Kar. His Ph.D. Advisor is Professor Andrew Yao, the laureate of the A.M. Turing Award in the year of 2000. Dr. Wu was the best paper award co-recipients of IEEE SmartGridComm 2012, IEEE PES General Meeting 2013, IEEE PES General Meeting 2020. He is working on economic analysis, optimal control and operation of power systems.