A Distributionally Robust Optimization Model for Unit Commitment Based on Kullback–Leibler Divergence

Article *in* IEEE Transactions on Power Systems · January 2018

DOI: 10.1109/TPWRS.2018.2797069

6 authors, including:

Qinglai Guo
Tsinghua University
204 PUBLICATIONS **1,545** CITATIONS

SEE PROFILE

Hongbin Sun
Tsinghua University
414 PUBLICATIONS **4,720** CITATIONS

SEE PROFILE

Zhengshuo Li
Southern Methodist University
62 PUBLICATIONS **470** CITATIONS

SEE PROFILE

Wenchuan Wu
Tsinghua University
285 PUBLICATIONS **3,057** CITATIONS

SEE PROFILE

A Distributionally Robust Optimization Model for Unit Commitment Based on Kullback–Leibler Divergence

Yuwei Chen, Qinglai Guo, *Senior Member, IEEE*, Hongbin Sun, *Fellow, IEEE*, Zhengshuo Li, *Member, IEEE*, Wenchuan Wu, *Senior Member, IEEE*, and Zihao Li

***Abstract*—This paper proposes a new distance-based distributionally robust unit commitment (DB-DRUC) model via Kullback–Leibler (KL) divergence, considering volatile wind power generation. The objective function of the DB-DRUC model is to minimize the expected cost under the worst case wind distributions restricted in an ambiguity set. The ambiguity set is a family of distributions within a fixed distance from a nominal distribution. The distance between two distributions is measured by KL divergence. The DB-DRUC model is a "min-max-min" programming model; thus, it is intractable to solve. Applying reformulation methods and stochastic programming technologies, we reformulate this "min-max-min" DB-DRUC model into a one-level model, referred to as the reformulated DB-DRUC (RDB-DRUC) model. Using the generalized Benders decomposition, we then propose a two-level decomposition method and an iterative algorithm to address the RDB-DRUC model. The iterative algorithm for the RDB-DRUC model guarantees global convergence within finite iterations. Case studies are carried out to demonstrate the effectiveness, global optimality, and finite convergence of a proposed solution strategy.**

***Index Terms*—Distributionally robust, generalized Benders decomposition, unit commitment.**

NOMENCLATURE

A. Indices and Sets

$\mathbb{D}$	Set of all probability distributions.
$\mathcal{I}^U$	Set of indices of units.
$\mathcal{I}^{Wind}$	Set of indices of wind farms.
j/I_q.	Index/set of feasible cuts.
$k/\mathcal{K}$	Index/set of lines.
l/I_p	Index/set of optimal cuts.
$n/\mathcal{N}$	Index/set of buses.
$\mathbb{P}$	Ambiguity set based on distance.
S_n^U	Set of units located at bus n.
S_n^{wind}	Set of wind farms located at bus n.
$\mathcal{T}$	Set of indices of scheduling periods.
$\mathcal{Z}$	Domain of subproblem.
$\zeta_i/\mathrm{Z_i}$	Index/set of all segments of the generation cost function of unit i.

B. Parameters and Constants

$a_i^{\zeta_i}/b_i^{\zeta_i}$	Linear/constant term of ζ_ith piecewise linear cost function of unit i.
C_i^s	Start-up cost of unit i at period τ.
D_n^τ	Electric load demand at bus n at period τ.
F_k	Transmission capacity limit of line k.
$K_{k,n}$	Shift factor of bus n to line k.
$\mathcal{M}$	Dimension of first-stage variables.
N	Total number of samples.
N_s	Number of observations of scenario s.
$\bar{p}_i/\underline{p}_i$	Maximum/minimum output of unit i.
$\Delta\bar{p}_i/\Delta\underline{p}_i$	Upward/downward ramping capability of unit i.
S	Total number of scenarios.
T_i^u/T_i^d	Minimum up-time/down-time for unit i.
$\breve{u}$	Sample mean.
$\bar{w}_i^\tau$	Expected generation of wind farm i at period τ.
$\breve{z}_{0.95}$	97.5% quantile of the standard normal distribution.
$\breve{\alpha}$	Confidence level of the ambiguity sets.
ρ_s	Nominal probability of scenario s.
ρ_s^{true}	True probability of scenario s.
$\breve{\sigma}$	Standard deviation of the sample.
η	Index of ambiguity.
$\eta_{\breve{\alpha}}$	Divergence tolerance with confidence level $\breve{\alpha}$.
$\chi^2_{S-1,\breve{\alpha}}$	$\breve{\alpha}$ upper quantile of a χ^2 distribution with $S-1$ degrees of freedom.

C. Decision Variables

c_i^τ	Generation cost of unit i at period τ.
p_i^τ	Power output of unit i at period τ.
p_{wi}^τ	Generation of wind farm i at period τ.
v_i^τ	Random variable indicating relative prediction error of wind farm i at period τ.

Manuscript received June 10, 2017; revised November 6, 2017; accepted January 12, 2018. Date of publication January 23, 2018; date of current version August 22, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 51537006, in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 51621065, and in part by the China Postdoctoral Science Foundation under Grant 2016M600091 and Grant 2017T100078. Paper no. TPWRS-00881-2017. *(Corresponding author: Hongbin Sun.)*

Y. Chen, Q. Guo, H. Sun, W. Wu, and Z. Li are with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: 18811362415@163.com; guoqinglai@tsinghua.edu.cn; shb@tsinghua.edu.cn; wuwench@tsinghua.edu.cn; 18811363920@163.com).

Z. Li is with Shenzhen Environmental Science and New Energy Technology Engineering Laboratory, Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua University, Shenzhen 100084, Guangdong China (e-mail: shuozhengli@sina.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2018.2797069

y_i^τ / y_i^h Binary decision variable indicating the unit commitment status of unit i at period τ/h.
$\boldsymbol{y}$ Vector of all first-stage variables.
$\boldsymbol{x}$ Vector of all second-stage variables except θ.
$\boldsymbol{v}$ Vector of all random variables.
z_i^τ Start-up variable of unit i at period τ.
$\boldsymbol{z}$ Vector of all $\boldsymbol{x}_s$ and θ_s.
α Variable derived from duality transformation.
ρ_s' Probability of scenario s.
θ Generation cost of all units.
$(\cdot)_k$ Variables at iteration k.
$(\cdot)_s$ Variables of scenario s.

D. Lagrange Multipliers and Optimal Variables

$\widehat{(\cdot)}$ Lagrange multipliers and optimal variables of subproblem and master problem.
$\widetilde{(\cdot)}$ Lagrange multipliers and optimal variables of second-level subproblem and second-level master problem.
$\overline{(\cdot)}$ Lagrange multipliers and optimal variables of second-level feasibility problem.

I. Introduction

TO COPE with fossil fuel depletion and climate change, large-scale wind generation is being integrated into the energy system [1], [2]. However, due to the intermittent and unpredictable nature of renewable energy, it becomes more difficult to balance the load and generation. Unit commitment (UC) with uncertainties is always a noteworthy topic.

A variety of optimization technologies for UC have been proposed to address uncertainties. Robust optimization [3] is a widely used technology that addresses the uncertainties by calculating the worst-possible value for the objective function, with the uncertain parameters varying in a given uncertainty set. There has also been some research on robust UC [4]–[7], in which the uncertainty set is usually an affine set; thus, decisions can be made without an exact distribution of the uncertain parameters. However, robust optimization sometimes could make conservative decisions because it concerns the worst-case scenario that does not always actually occur [4], [8], [9].

Stochastic programming [10] is another traditional method to treat data uncertainty in optimization problems. Typical solution methods for stochastic programming includes sample average approximation and stochastic approximation [10]. In these studies [11]–[13], it was demonstrated that the stochastic programming techniques can efficiently improve the expected performance of UC decisions under uncertainties. However, the distribution of uncertain parameters may be inaccurate, which may lead to incorrect decisions.

Distributionally robust optimization (DRO) [14]–[20] is an alternative way of modeling. The objective of DRO is to optimize a problem under the worst case of the distributions in a set (the so-called ambiguity set). The ambiguity set is a family of distributions, and can be formulated in terms of moments (e.g., expectation) or distance from a known distribution. During recent years, DRO, especially moment-based DRO has been applied in several studies in the area of power systems [21]–[27]. Reference [21] developed an inner approximation method to address the distributionally robust congestion management model. This solution strategy is tractable but cannot guarantee accuracy in some cases. In [22]–[27], several more precise reformulations were presented to solve the moment-based DRO with the ambiguity set formulated by the expectation, variance or both of distributions. References [22]–[25] reformulated the moment-based DRO into a semidefinite programming (SDP) problem or a second-order conic programming (SOCP) problem based on some linear assumptions and approximations. References [26], [27] made significant studies on distributionally robust unit commitment (DRUC). Xiong *et al.* formulated a moment-based DRO model for UC under wind uncertainty and applied a linear decision rule approximation to solve this intractable model [26]. Zhao *et al.* studied a moment-based DRO approach for contingency-constrained UC and proposed an iterative algorithm based on a Benders' decomposition [27].

The moment-based DRO uses statistical information such as expectation to formulate the ambiguity set. The distribution knowledge contains more information than moments, but is not fully utilized in the moment-based DRO. In many practical cases, a nominal distribution can be obtained using data fitting, which provides an excellent estimate of the distribution and often contains valuable information about the actual distribution. For instance, wind speeds are often fitted to Weibull distributions, and relative prediction errors are often modeled as normal random variables. Although measurement errors are inevitable, the decision makers may believe that the true distribution is not too "far away" from the nominal distribution derived from analyzing historical data. Motivated by the above facts, references [15]–[19] investigated a new method referred to as "distance-based DRO", where the ambiguity set consisted of the distributions within a fixed distance from the nominal distribution. References [15]–[19] presented detailed reformulations of DRO problems under different measures of distance (e.g., Kullback–Leibler (KL) divergence). However, the objective functions of these reformulated problems are nonlinear, e.g., exponential functions. Specific methods for solving reformulated DRO problems are not covered in [15]–[19]. It is very difficult to solve these nonlinear problems effectively using commercial solvers. Moreover, non-linearity can cause numerical problems. For example, the value of the exponential term may exceed the largest floating-point number that can be represented internally in a computer.

In this paper, we apply the distance-based DRO to the UC problem under wind uncertainty. The ambiguity set is formed via KL divergence. As the traditional robust UC (RUC) model is a three-level "min-max-min" model and the reformulations of distance-based DRO are nonlinear, the distance-based DRUC (DB-DRUC) is difficult to solve using conventional methods. We propose a two-level decomposition method and an iterative algorithm for DB-DRUC, which can overcome several challenges.

The main contribution of this paper can be summarized into the following points.

1) The DB-DRUC model is a "min-max-min" optimization model, which is intractable to solve. After applying reformulation methods for DRO and stochastic programming technologies, we reformulate the DB-DRUC model into a reformulated DB-DRUC (RDB-DRUC) model. The RDB-DRUC model is a one-level optimization model and, thus, is easier to solve than the DB-DRUC model.
2) The RDB-DRUC model is a mixed-integer nonlinear programming (MINLP) problem, which is still a difficult problem. Based on the generalized Benders decomposition (GBD), we propose a two-level decomposition method and an iterative algorithm to solve the RDB-DRUC model. The iterative algorithm for the RDB-DRUC model guarantees global convergence within finite iterations.
3) The uncertainties in RDB-DRUC are represented by scenarios, and thus the problem size grows fast while the system size and the number of scenarios increase. To cope with the possible computational difficulties, the RDB-DRUC problem is decomposed into each scenario by the two-level decomposition method. Parallel computing techniques can be applied to reduce the total computing time.
4) The iterative algorithm takes advantage of the second-level decomposition by solving constrained linear problems and unconstrained nonlinear problems instead of constrained nonlinear problems. Constrained linear problems are easier to solve and less prone to numerical problems than constrained nonlinear problems with exponential terms. Unconstrained nonlinear problems are also easy to solve because they are entirely specified by their objective functions.

The remainder of this paper is organized as follows. Section II presents a traditional two-stage deterministic UC model. In Section III, we formulate a DB-DRUC model and an RDB-DRUC model. Section IV describes the two-level decomposition of the RDB-DRUC model, and proposes an iterative algorithm, which has global optimality and finite convergence. In Section V, case studies are presented to show the effectiveness and global optimality of the proposed model and iterative algorithm. Additionally, the DB-DRUC is compared to the stochastic UC (SUC), the RUC and the moment-based DRUC (MB-DRUC). Conclusions are drawn in Section VI.

II. Deterministic UC Model

A. Formulation of Deterministic UC Model

Traditionally, the day-ahead UC model contains two sets of decisions which are determined in two stages. In the first stage, the on/off statuses of the units are determined for the next day. In the second stage, the generation of all units at each period τ is determined. Assuming that the wind energy generation of the next day is known, we present the deterministic UC mode, which also serves as the nominal model to its distributionally robust counterpart

$$\min \quad \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{I}^U} C_i^s z_i^\tau + \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{I}^U} c_i^\tau \tag{1}$$

$$\text{s.t.} \quad z_i^\tau \geq y_i^\tau - y_i^{\tau-1}, \forall i \in \mathcal{I}^U, \tau \in \mathcal{T} \tag{2}$$

$$-y_i^{\tau-1} + y_i^\tau - y_i^h \leq 0, \forall i \in \mathcal{I}^U, \tau \in \mathcal{T}$$

$$1 \leq h-(\tau-1) \leq T_i^u \tag{3}$$

$$y_i^{\tau-1} - y_i^\tau + y_i^h \leq 1, \forall i \in \mathcal{I}^U, \tau \in \mathcal{T}$$

$$1 \leq h-(\tau-1) \leq T_i^d \tag{4}$$

$$c_i^\tau \geq a_i^{\zeta_i} p_i^\tau + b_i^{\zeta_i} y_i^\tau, \forall \zeta_i \in \mathrm{Z_i}, i \in \mathcal{I}^U, \tau \in \mathcal{T} \tag{5}$$

$$\sum_{i\in\mathcal{I}^T} p_i^\tau + \sum_{i\in\mathcal{I}^{Wind}} p_{wi}^\tau = \sum_{n\in\mathcal{N}} D_n^\tau, \forall \tau \in \mathcal{T} \tag{6}$$

$$-F_k \leq \sum_{n\in\mathcal{N}} K_{k,n}\left(\sum_{i\in S_n^U} p_i^\tau + \sum_{i\in S_n^{Wind}} p_{wi}^\tau - D_n^\tau\right) \leq F_k,$$

$$\forall k \in \mathcal{K}, \tau \in \mathcal{T} \tag{7}$$

$$\underline{p}_i y_i^\tau \leq p_i^\tau \leq \bar{p}_i y_i^\tau, \forall i \in \mathcal{I}^U, \tau \in \mathcal{T} \tag{8}$$

$$0 \leq p_{wi}^\tau \leq \bar{w}_i^\tau (1+v_i^\tau), \forall i \in \mathcal{I}^{Wind}, \tau \in \mathcal{T} \tag{9}$$

$$p_i^{\tau-1} - p_i^\tau \leq \Delta\underline{p}_i y_i^\tau + \bar{p}_i(1-y_i^\tau)$$

$$p_i^\tau - p_i^{\tau-1} \leq \Delta\bar{p}_i y_i^{\tau-1} + \bar{p}_i(1-y_i^{\tau-1})$$

$$\forall i \in \mathcal{I}^U, \tau \in \mathcal{T} \tag{10}$$

$$y_i^\tau, z_i^\tau = \{0,1\}, \forall i \in \mathcal{I}^U, \tau \in \mathcal{T}. \tag{11}$$

In the above formulations, the objective function (1) contains the start-up cost and generation cost of all units. The turn-on operation of the units is constrained by (2). Constraints (3) and (4) are the minimum up-time and down-time constraints, respectively. The piecewise linear generation cost of the units is denoted by (5). Constraints (6) enforce the electricity power balance between generation supply and demand. Constraints (7) restrict the electricity power transmission within the transmission capacity. Constraints (8) define the maximum and minimum capacity of power output. Constraints (9) denote that the wind power generation cannot exceed the available wind power, where relative prediction errors v_i^τ are fixed. The downward and upward ramping limits of power output are expressed by (10).

B. Matrix Form of Deterministic UC Model

The deterministic UC model above can be written in a general matrix form. We use the vector $\boldsymbol{y}$ to denote all first-stage decision variables including y_i^τ and z_i^τ, and use $\mathcal{M}$ to denote the dimension of $\boldsymbol{y}$. All second-stage decision variables are denoted by the vector $\boldsymbol{x}$. The vector of all relative prediction errors v_i^τ is denoted by $\boldsymbol{v}$, which is a given parameter in this deterministic

UC model. Hence the first-stage problem can be expressed as

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & \boldsymbol{c}^T \boldsymbol{y} + H(\boldsymbol{y}, \boldsymbol{v}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b} \leq 0, \\ & \boldsymbol{y} \in \{0,1\}^{\mathcal{M}} \end{aligned} \tag{12}$$

where the function $H(\boldsymbol{y}, \quad \boldsymbol{v})$ is the optimal value of the second-stage problem

$$\begin{aligned} H(\boldsymbol{y}, \boldsymbol{v}) = \min_{\boldsymbol{x}} \ & \boldsymbol{q}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{v}) \leq 0 \end{aligned} \tag{13}$$

where the symbol $\boldsymbol{h}(\boldsymbol{v})$ means that the vector $\boldsymbol{h}$ is affected by the parameter $\boldsymbol{v}$.

To facilitate the discussion in Section IV, we add a variable θ to represent the generation cost of all units. Moreover, a new constraint $\theta - \boldsymbol{q}^T \boldsymbol{x} = 0$ is added. The new form of the second-stage problem is as follows:

$$\begin{aligned} H(\boldsymbol{y}, \boldsymbol{v}) = \min_{\boldsymbol{x}, \theta} \ & \theta \\ \text{s.t.} \quad & \theta - \boldsymbol{q}^T \boldsymbol{x} = 0 \\ & \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{v}) \leq 0. \end{aligned} \tag{14}$$

Based on the first-stage problem (12) and second-stage problem (14), their distributionally robust counterpart will be discussed in the next section.

III. DB-DRUC and RDB-DRUC Models

We first give a distributionally robust counterpart for the deterministic UC model in Section III-A, which is based on KL divergence and referred to as the DB-DRUC model. In Section III-B, we describe the methods for constructing the ambiguity sets. Then in Section III-C, we reformulate this counterpart into a stochastic programming problem with a fixed probability distribution. Finally, in Section III-D, stochastic scenario generation methods and reduction methods are applied to formulate the RDB-DRUC problem, which is a MINLP problem.

The KL divergence is well-known in the field of information theory and is now widely used as a good measure of distance between two probability distributions. The KL divergence for P to P_0 is defined as

$$D_{KL}(P||P_0) = \int_{\Omega} P(\theta) \log \frac{P(\theta)}{P_0(\theta)} d\theta \tag{15}$$

where P and P_0 are distribution functions in measure space Ω.

The DRO based on KL divergence has been studied by many scholars [15]–[18]. According to the reformulation methods in these research, the DRUC based on KL divergence can be solved. The solution strategy will be discussed in Section IV. In this paper, we use the KL divergence as a starting point for studying the DB-DRUC.

A. DB-DRUC Model

In the DRO model, the uncertainty of random variables is captured by an ambiguity set $\mathbb{P}$. The distance-based distributionally robust counterpart of the first-stage problem is as follows:

$$\begin{aligned} \text{DB - DRUC}: \min \quad & \boldsymbol{c}^T \boldsymbol{y} + \max_{P \in \mathbb{P}} \mathbb{E}_P[H(\boldsymbol{y}, \boldsymbol{v})] \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b} \leq 0 \\ & \boldsymbol{y} \in \{0,1\}^{\mathcal{M}} \end{aligned} \tag{16}$$

where $\mathbb{E}_P[H(\boldsymbol{y}, \quad \boldsymbol{v})]$ denotes the expected value with respect to a distribution P of random variables $\boldsymbol{v}$. The ambiguity set $\mathbb{P}$ is constructed by the KL divergence, expressed by

$$\mathbb{P} = \{P \in \mathbb{D} : D_{KL}(P||P_0) \leq \eta\} \tag{17}$$

where the divergence tolerance η is also the index of ambiguity, since it controls the size of the ambiguity set. P_0 is the nominal distribution which is derived from historical data. The conservatism of the DB-DRUC can be adjusted by varying η.

Compared with the deterministic UC model, the second-stage problem of the DB-DRUC does not change, and is still expressed by (14).

The construction of the ambiguity sets, which affect the accuracy and conservatism of the dispatching plan, is very important to the DB-DRUC. The step-by-step procedure for constructing an ambiguity set will be described in Section III–B.

B. Ambiguity Set Construction

We first briefly introduce the construction of the uncertainty sets in typical UC research. The ideas behind this construction can be applied to the construction of the ambiguity sets used in the DB-DRUC. Several approaches to the construction of uncertainty sets for typical UC uncertainties are presented in [28]. The most basic uncertainty set is a confidence interval:

$$[\breve{u} - \breve{z}_{0.95}\breve{\sigma}, \breve{u} + \breve{z}_{0.95}\breve{\sigma}]. \tag{18}$$

This confidence interval has been used in many studies, such as [29]–[31]. In (18), $\breve{u}$ is the sample mean. $\breve{\sigma}$ is the standard deviation of the sample, and $\breve{z}_{0.95}$ is the 97.5% quantile of the standard normal distribution, in which case the confidence level of interval (18) is 95%. In other words, the proportion of interval (18) that contains the true value of the uncertainty is 95%. The conservatism of (18) can be adjusted by replacing quantile $\breve{z}_{0.95}$ with other quantiles.

The ambiguity set (17) can be constructed in a similar manner, as a confidence set:

$$\mathbb{P} = \{P \in \mathbb{D} : D_{KL}(P||P_0) \leq \eta_{\breve{\alpha}}\} \tag{19}$$

where $\breve{\alpha}$ is the confidence level and $\eta_{\breve{\alpha}}$ is the divergence tolerance with confidence level $\breve{\alpha}$. The constraint describing this confidence set $\mathbb{P}$ guarantees that the true distribution is within $\mathbb{P}$ with $\breve{\alpha}$ confidence level.

Two key steps in the construction of an ambiguity set are the determination of P_0 and η. The nominal distribution P_0 can be obtained using both parametric and nonparametric methods. The methods for determining η vary, depending on whether P_0 is parametric or nonparametric. In detail, the steps for constructing parametric and nonparametric ambiguity sets are as follows:

1) Parametric Ambiguity Set:

Step 1: Estimate parametric P_0.

Parametric estimation methods assume that the uncertain variable comes from a predefined distribution (e.g., the relative prediction error follows a normal distribution). The parameters (e.g., the mean and variance of a normal distribution) of the assumed distribution can be obtained from historical data. For detailed descriptions of parametric estimation methods such as the maximum likelihood method, please refer to [32]. Assuming we take random samples $(X_1, \ldots, X_N)$ from a normal distribution, the maximum likelihood estimators of the mean and variance are

$$\frac{1}{N}\sum_{i=1}^{N} X_i \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \frac{1}{N}\sum_{i=1}^{N} X_i\right)^2.$$

Step 2: Determine $\eta_{\breve{\alpha}}$ with confidence level $\breve{\alpha}$.

The divergence tolerance $\eta_{\breve{\alpha}}$, depends on the assumed distribution and can be calculated using Monte Carlo simulations. The divergence tolerances for the normal, exponential, uniform and Laplace distributions are presented in [33]. For instance, if the sample $(X_1, \ldots, X_N)$ is assumed to be normally distributed, $N = 50$ and $\breve{\alpha} = 0.95$, then $\eta_{\breve{\alpha}} = 0.0589$.

2) Nonparametric Ambiguity Set:

Step 1: Estimate nonparametric P_0.

Nonparametric estimation methods make no assumptions about the distribution. Hence, they are often used to handle data-driven problems. Assuming we have a large number of samples $(X_1, \ldots, X_N)$ that are assigned to S scenarios according to a scenario reduction method such as the fast-forward selection method [34]. One widely used empirical distribution is as follows:

$$\rho_s = \frac{N_s}{N}, s = 1, \ldots, S, \tag{20}$$

where ρ_s is the nominal probability of scenario s; N_s is the number of observations of scenario s and $\sum_{s=1}^{S} N_s = N$; S is the total number of scenarios.

Step 2: Determine $\eta_{\breve{\alpha}}$ with confidence level $\breve{\alpha}$.

ρ_s^{true} is used to denote the true probability of scenario s. According to [35, Th. 3.1], the statistic $2N\sum_{s=1}^{S} \rho_s^{\text{true}} \log(\rho_s^{\text{true}}/\rho_s)$ converges in distribution to a χ^2 distribution with $S-1$ degrees of freedom. Hence the distance level $\eta_{\breve{\alpha}}$ with confidence level $\breve{\alpha}$ can be expressed as:

$$\eta_{\breve{\alpha}} = \frac{1}{2N}\chi^2_{S-1,\breve{\alpha}}, \tag{21}$$

where $\chi^2_{S-1,\breve{\alpha}}$ is the $\breve{\alpha}$ upper quantile of a χ^2 distribution with $S-1$ degrees of freedom.

C. Reformulation of the First-Stage Problem

The inner maximization problem in (16) is as follows:

$$\max_{P\in\mathbb{P}} \quad \mathbb{E}_P[H(\boldsymbol{y}, \boldsymbol{v})]. \tag{22}$$

According to [15] and [18], problem (22) is equivalent to

$$\min_{\alpha\geq 0} \quad \alpha \log \mathbb{E}_{P_0}[e^{H(\boldsymbol{y},\boldsymbol{v})/\alpha}] + \alpha\eta, \tag{23}$$

where α is a new dual variable derived by implementing the duality theory, with details provided in [15] and [18]. The maximization problem (22) is reformulated as a minimization problem (23). The above transformation was also applied in [36]. The expectation in (23) is taken with respect to nominal distribution P_0. In the case where P is discrete, we will present a brief proof of the above transformation based on the derivation in [15]. When P is discrete, the special case of (22) can be expressed as

$$\begin{aligned} \max_{\rho'_s, s=1,2,\ldots,S} \quad & \sum_{s=1}^{S} \rho'_s H_s \\ \text{s.t.} \quad & \sum_{s=1}^{S} \rho'_s \log(\rho'_s/\rho_s) \leq \eta \\ & \sum_{s=1}^{S} \rho'_s = 1, \rho'_s \geq 0, \forall s = 1, 2, \ldots, S \end{aligned} \tag{24}$$

where H_s is shorthand for $H(\boldsymbol{y}, \boldsymbol{v}_s)$. The dual objective function of (24) is denoted by $g(\alpha, \upsilon)$ and satisfies

$$\begin{aligned} g(\alpha,\upsilon) &= \alpha\eta + \upsilon + \sum_{s=1}^{S} \max_{\rho'_s\geq 0} \rho'_s \left(H_s - \upsilon - \alpha \log\frac{\rho'_s}{\rho_s}\right) \\ &= \alpha\eta + \upsilon + \sum_{s=1}^{S} \rho_s \max_{t\geq 0} t\,(H_s - \upsilon - \alpha\log t) \\ &= \alpha\eta + \upsilon + \alpha\sum_{s=1}^{S} \rho_s e^{(H_s-\upsilon-\alpha)/\alpha} \end{aligned} \tag{25}$$

where υ is the dual variable associated with $\sum_{s=1}^{S} \rho'_s = 1$.

Thus, the dual problem of (24) is

$$\min_{\alpha\geq 0,\upsilon} \quad \alpha\eta + \upsilon + \alpha\sum_{s=1}^{S} \rho_s e^{(H_s-\upsilon-\alpha)/\alpha}. \tag{26}$$

υ^* is used to denote the optimal point of υ. It can be obtained directly by Fermat's theorem (stationary points) and satisfies the following:

$$\sum_{s=1}^{S} \rho_s e^{(H_s-\upsilon^*-\alpha)/\alpha} = 1 \tag{27}$$

$$\upsilon^* = \alpha \log \sum_{s=1}^{S} \rho_s e^{H_s/\alpha} - \alpha. \tag{28}$$

We substitute (27) and (28) into (26) to obtain

$$\min_{\alpha\geq 0} \quad \alpha \log \sum_{s=1}^{S} \rho_s e^{H_s/\alpha} + \alpha\eta. \tag{29}$$

So far, we have proved that (22) can be transformed into (23) when P is discrete. There is a more general proof that holds for continuous, discrete and mixed distributions in [18].

Based on (22) and (23), the first-stage problem (16) can be reformulated as

$$\begin{aligned}\min_{\alpha\geq 0,\boldsymbol{y}}\ & \boldsymbol{c}^T\boldsymbol{y}+\alpha\log\mathbb{E}_{P_0}[e^{H(\boldsymbol{y},\boldsymbol{v})/\alpha}]+\alpha\eta\\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}\leq 0\\ & \boldsymbol{y}\in\{0,1\}^{\mathcal{M}}.\end{aligned}\tag{30}$$

The problem (30) is a stochastic programming problem with fixed distribution P_0; thus, we can apply the traditional stochastic optimization methods.

D. RDB-DRUC Model

It was assumed that the historical relative prediction errors could be obtained. If we construct a parametric ambiguity set, the historical relative prediction errors can be fitted into the predefined distributions. Based on the distributions of uncertainty, the scenarios can be generated by the Monte Carlo method. We then reduce the number of scenarios to S by applying the fast-forward selection method. On the other hand, the fast-forward selection method is applied during the construction of nonparametric ambiguity sets. Hence, the nominal distribution is already a discrete distribution (20) with S scenarios. This is how we generate S scenarios from historical data.

As per the above discussion, it is assumed that S scenarios are obtained after scenario reduction. $\{\boldsymbol{v}_1,\boldsymbol{v}_1,\ldots,\boldsymbol{v}_S\}$ are the uncertainty variables in these scenarios with respective probabilities $\{\rho_1,\rho_2,\ldots,\rho_S\}$. The probabilities $\{\rho_1,\rho_2,\ldots,\rho_S\}$ is obtained together with reduced scenarios $\{\boldsymbol{v}_1,\boldsymbol{v}_1,\ldots,\boldsymbol{v}_S\}$ by the fast-forward selection method. This method is a heuristic algorithm for calculating the optimal reduced distribution that is closest to the initial distribution in terms of a Kantorovich functional metric.

The RDB-DRUC problem is formulated as (31) for both parametric and nonparametric ambiguity sets, which is a one-level problem and, hence, is much easier to solve

$$\begin{aligned}\text{RDB - DRUC}:\min\ & \boldsymbol{c}^T\boldsymbol{y}+\alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\theta_s/\alpha}\right)+\alpha\eta\\ \text{s.t.}\ & \theta_s-\boldsymbol{q}_s^T\boldsymbol{x}_s=0,\forall s=1,2,\ldots,S\\ & \boldsymbol{T}\boldsymbol{x}_s+\boldsymbol{W}\boldsymbol{y}-\boldsymbol{h}_s\leq 0,\forall s=1,2,\ldots,S,\\ & \boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}\leq 0\\ & \boldsymbol{y}\in\{0,1\}^{\mathcal{M}},\alpha\geq 0\end{aligned}\tag{31}$$

where the subscript s represents the index of scenarios; and the vector $\boldsymbol{h}_s$ is the abbreviation of $\boldsymbol{h}(\boldsymbol{v}_s)$.

The RDB-DRUC problem (31) is a large-scale MINLP problem which is difficult to solve. The details of solution strategy will be introduced in the next section.

IV. Solution Strategy

In this section, we propose a two-level decomposition method for RDB-DRUC based on the GBD.

A. First-Level Decomposition of RDB-DRUC Problem

For convenience, $\boldsymbol{z}$ is used to denote all of the variables in (31) except $\boldsymbol{y}$. The objective function and constraint functions of the RDB-DRUC problem (31) have following properties:

1) the objective function is convex on $\boldsymbol{z}$ when $\alpha\geq 0$ for each fixed $\boldsymbol{y}$ [40, pp. 72–74, 79–0, 89–90], and
2) all of the constraint functions are linear on $\boldsymbol{z}$ for each fixed $\boldsymbol{y}$.

Thus, we can apply the GBD [37] to calculate the optimal solution of (31). Based on the GBD, the RDB-DRUC problem (31) is decomposed into the subproblem (SP) and master problem (MP). The subscript k is the iteration number:

$$\begin{aligned}\text{SP}:\min_{\boldsymbol{z}\in\mathcal{Z}}\ & \alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\theta_s/\alpha}\right)+\alpha\eta\\ \text{s.t.}\ & \theta_s-\boldsymbol{q}_s^T\boldsymbol{x}_s=0,\forall s=1,2,\ldots,S\\ & \boldsymbol{T}\boldsymbol{x}_s+\boldsymbol{W}\boldsymbol{y}_k-\boldsymbol{h}_s\leq 0,\forall s=1,2,\ldots,S\end{aligned}\tag{32}$$

where the term $\boldsymbol{c}^T\boldsymbol{y}_k$ in the objective function of the SP is omitted, as this term is not related to $\boldsymbol{z}$; $\mathcal{Z}$ is the domain of the SP, where $\mathcal{Z}=\{\alpha,\theta_s,\boldsymbol{x}_s,s=1,2,\ldots,S|\alpha\geq 0\}$.

The MP is as follows:

$$\begin{aligned}\text{MP}:\min_{\boldsymbol{y}}\ & \boldsymbol{c}^T\boldsymbol{y}+\mu_B\\ \text{s.t.}\ & \mu_B\geq\xi_l(\boldsymbol{y},\hat{\mu}_l,\hat{\boldsymbol{\lambda}}_l)=\inf_{\boldsymbol{z}\in\mathcal{Z}}L(\boldsymbol{z},\boldsymbol{y},\hat{\mu}_l,\hat{\boldsymbol{\lambda}}_l),l\in I_p,\\ & 0\geq\bar{\xi}_j(\boldsymbol{y},\bar{\boldsymbol{\lambda}}_j)=\inf_{\boldsymbol{z}\in\mathcal{Z}}\bar{L}(\boldsymbol{z},\boldsymbol{y},\bar{\boldsymbol{\lambda}}_j),j\in I_q\\ & \boldsymbol{y}\in\{0,1\}^{\mathcal{M}}\end{aligned}\tag{33}$$

where I_p and I_q are two index sets; $L(\cdot)$, $\bar{L}(\cdot)$ are Lagrange functions, and $\hat{\mu}_l$, $\hat{\boldsymbol{\lambda}}_l$, $\bar{\boldsymbol{\lambda}}_j$ are Lagrange multipliers; and $\xi_l(\cdot)$, $\bar{\xi}_j(\cdot)$ are support functions.

B. Second-Level Decomposition of Subproblem

The SP (32) is a large-scale convex programming problem. Solving the SP directly may take a long time. Thus, we propose a decomposition method for the SP. This method reduces the size of each programming problem, and can be combined with parallel computing techniques to reduce the computing time.

Considering the objective function of SP, we find that

1) the objective function is monotonically increasing with respect to $\theta_s,\forall s=1,2,\ldots,S$, and
2) θ_s is only constrained by constraints $\theta_s-\boldsymbol{q}_s^T\boldsymbol{x}_s=0$ and $\boldsymbol{T}\boldsymbol{x}_s+\boldsymbol{W}\boldsymbol{y}_k-\boldsymbol{h}_s\leq 0,\forall s=1,2,\ldots,S$.

According to the above two points, the SP can be further decomposed into the following second-level subproblem (SL-SP) and second-level master problem (SL-MP), as shown in (34) and (35), respectively. The solution procedure of the SP can be divided into two steps. The first is to solve each SL-SP and get the optimal solution $\tilde{\theta}_{s,k},\forall s=1,\ldots,S$. The calculation time of solving these SL-SPs can be reduced by parallel computing techniques. Then, for fixed $\tilde{\theta}_{s,k}$, the SL-MP is solved to get its

optimal value, which is also the optimal value of the SP:

$$\begin{aligned}\text{SL - SP}: \tilde{\theta}_{s,k} = \min_{\boldsymbol{x}_s,\theta_s}\ & \theta_s \\ \text{s.t.}\quad & \theta_s - \boldsymbol{q}_s^T \boldsymbol{x}_s = 0 \\ & \boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s \le 0 \end{aligned} \tag{34}$$

$$\text{SL - MP}: \min_{\alpha\ge 0}\ \alpha \log\left(\sum_{s=1}^{S} \rho_s e^{\tilde{\theta}_{s,k}/\alpha}\right) + \alpha\eta. \tag{35}$$

C. Feasible Subproblem

The SL-MP has a feasible solution for any $\tilde{\theta}_{s,k}$, as it is an unconstrained convex programming problem on its domain. Therefore, the SP is feasible if and only if all of the SL-SPs are feasible.

We can get the Lagrange multipliers of each SL-SP after solving all the SL-SPs. We use $\tilde{\mu}_{s,k}$ to denote the multiplier associated with $\theta_s - \boldsymbol{q}_s^T \boldsymbol{x}_s = 0$, and $\tilde{\boldsymbol{\lambda}}_{s,k}$ to denote the multiplier vector associated with $\boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s \le 0$. Symbols that are marked with a tilde ($\sim$) denote the multipliers of the SL-SPs.

Symbols marked with a hat ($\hat{}$) are used to denote the multipliers of the SP, and the Lagrange function $L(\cdot)$ of the SP is as follows:

$$\begin{aligned} L(\boldsymbol{z},\boldsymbol{y},\hat{\mu}_l,\hat{\boldsymbol{\lambda}}_l) &= \alpha \log\left(\sum_{s=1}^{S} \rho_s e^{\theta_s/\alpha}\right) + \alpha\eta \\ &+ \sum_{s=1}^{S}\left(\hat{\mu}_{s,k}\left(\theta_s - \boldsymbol{q}_s^T\boldsymbol{x}_s\right) + \hat{\boldsymbol{\lambda}}_{s,k}^T\left(\boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s\right)\right) \end{aligned} \tag{36}$$

where $\hat{\mu}_l = \hat{\mu}_k = (\hat{\mu}_{1,k},\ldots,\hat{\mu}_{S,k})^T$ and $\hat{\boldsymbol{\lambda}}_l = \hat{\boldsymbol{\lambda}}_k = (\hat{\boldsymbol{\lambda}}_{1,k}^T, \ldots, \hat{\boldsymbol{\lambda}}_{S,k}^T)^T$.

The difficulty is how to obtain the multipliers of the SP, as only the SL-SPs and the SL-MP are solved. In Appendix I, we present a more general proof to show the relationship between $\hat{\mu}_{s,k}, \hat{\boldsymbol{\lambda}}_{s,k}$ and $\tilde{\mu}_{s,k}, \tilde{\boldsymbol{\lambda}}_{s,k}$. The results are as follows:

$$\hat{\mu}_{s,k} = \frac{\rho_s e^{\tilde{\theta}_{s,k}/\tilde{\alpha}_k}}{\sum_{s'=1}^{S}\rho_{s'} e^{\tilde{\theta}_{s',k}/\tilde{\alpha}_k}}\tilde{\mu}_{s,k} \tag{37}$$

$$\hat{\boldsymbol{\lambda}}_{s,k} = \frac{\rho_s e^{\tilde{\theta}_{s,k}/\tilde{\alpha}_k}}{\sum_{s'=1}^{S}\rho_{s'} e^{\tilde{\theta}_{s',k}/\tilde{\alpha}_k}}\tilde{\boldsymbol{\lambda}}_{s,k} \tag{38}$$

where $\tilde{\alpha}_k$ is the optimal solution of the SL-MP. The summation is denoted by the subscript s', which is distinct from the subscript s.

D. Infeasible Subproblem

If there is one SL-SP that is infeasible, the whole SP is infeasible. When this occurs, we solve the second level feasibility problem (SL-FP):

$$\begin{aligned}\text{SL - FP}: \min_{\boldsymbol{x}_s,\beta}\ & 1^T\beta \\ \text{s.t.}\quad & \boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s \le \beta, \\ & 0 \le \beta \end{aligned} \tag{39}$$

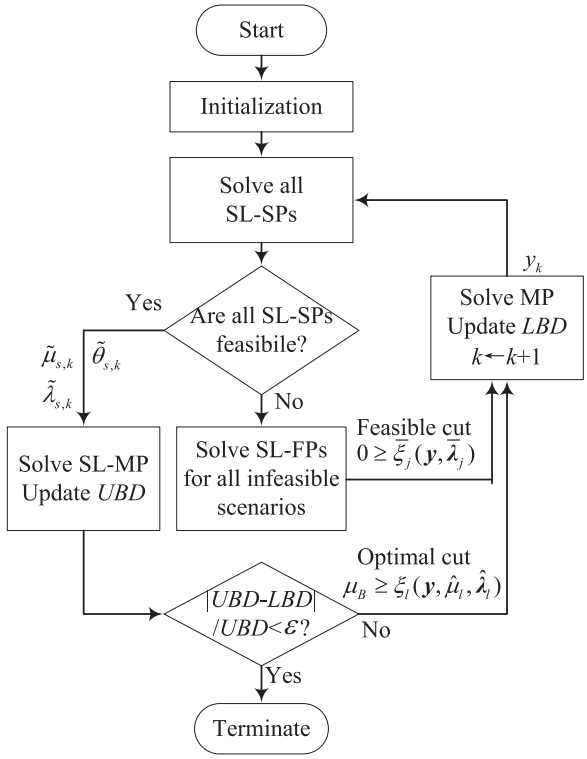

Fig. 1. Flow chart of the algorithm described in Section IV-F.

where the constraint $\theta_s - \boldsymbol{q}_s^T\boldsymbol{x}_s = 0$ is omitted, because it is always satisfied by some θ_s. Symbols marked with an over-bar (-) are used to denote the multipliers of the SL-FP. The Lagrange function of the SL-FP is given by

$$\bar{L}(\boldsymbol{z},\boldsymbol{y},\bar{\boldsymbol{\lambda}}_j) = \bar{\boldsymbol{\lambda}}_{s,k}^T(\boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s) \tag{40}$$

where $\bar{\boldsymbol{\lambda}}_j = \bar{\boldsymbol{\lambda}}_{s,k}$.

E. Cutting Planes in Master Problem

The inequality constraints in the MP are also called the cutting planes. In (33), the cutting plane expressions contain 'inf', and thus the MP cannot be solved directly. In Appendix II, we present the detailed process of reformulating these cutting planes. The results are shown as

$$\inf_{\boldsymbol{z}\in\mathcal{Z}} L(\boldsymbol{z},\boldsymbol{y},\hat{\mu}_l,\hat{\boldsymbol{\lambda}}_l) = \sum_{s=1}^{S}\hat{\boldsymbol{\lambda}}_{s,k}^T\left(\boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s\right) \tag{41}$$

$$\inf_{\boldsymbol{z}\in\mathcal{Z}} \bar{L}(\boldsymbol{z},\boldsymbol{y},\bar{\boldsymbol{\lambda}}_j) = \bar{\boldsymbol{\lambda}}_{s,k}^T(\boldsymbol{W}\boldsymbol{y}_k - \boldsymbol{h}_s). \tag{42}$$

F. Algorithm

Based on the discussions provided above, the main steps of the algorithm are summarized in the flow chart in Fig. 1, and an iteration procedure is given as follows.

G. Convergence and Optimality

The SP satisfies the assumptions of v3-GBD in [38], which guarantees the global optimality and finite convergence of the GBD. Additionally, all transformations in this section are strictly

Algorithm for solving the RDB-DRUC.

Step 1. Initialization.

- Set the counters $k = 1$ for iteration, $l = 0$ for optimal cut and $j = 0$ for feasible cut. Set the sets $I_p = \emptyset$ and $I_q = \emptyset$. Initialize the upper bound $UBD = +\infty$, and the lower bound $LBD = 0$. Select the convergence tolerance ε. Initialize a feasible $\boldsymbol{y}_0$ (e.g., $y_i^\tau = 1$, $z_i^\tau = 0, \forall i, \tau$).

Step 2. Solve all the SL-SPs for fixed $\boldsymbol{y}_k$.

Step 2a. Some SL-SPs are infeasible.

- Let Λ_k be the set of indices of infeasible SL-SP. For any $s \in \Lambda_k$, solve the SL-FP and obtain the Lagrange multipliers $\bar{\boldsymbol{\lambda}}_{s,k}$.
- Update $j = j + 1$ and add j to the set I_q. Let $\bar{\lambda}_j = \bar{\boldsymbol{\lambda}}_{s,k}$, and then generate the support function $\bar{\xi}_j(\boldsymbol{y}, \bar{\boldsymbol{\lambda}}_j)$.
- Traverse all the elements in Λ_k and then proceed to Step 4.

Step 2b. All the SL-SPs are feasible.

- Each SL-SP has the objective value $\tilde{\theta}_{s,k}$ finite with the optimal multipliers $\tilde{\mu}_{s,k}$ and $\tilde{\boldsymbol{\lambda}}_{s,k}$. Continue to Step 3.

Step 3. Solve the SL-MP for fixed $\tilde{\theta}_{s,k}$.

- Solve the SL-MP and obtain the optimal solution $\tilde{\alpha}_k$ and the optimal value v_k.
- Update the upper bound $UBD = \min\{UBD, v_k + \boldsymbol{c}^T y_k\}$.
- If $|\text{UBD} - \text{LBD}|/\text{UBD} < \varepsilon$, then terminate. Otherwise, calculate $\hat{\mu}_{s,k}$ and $\hat{\boldsymbol{\lambda}}_{s,k}$ based on (37)–(38). Update $l = l + 1$ and add l to the set I_p. Let $\hat{\mu}_l = \hat{\mu}_k$ and $\hat{\boldsymbol{\lambda}}_l = \hat{\boldsymbol{\lambda}}_k$, and then generate the support function $\xi_l(\boldsymbol{y}, \hat{\mu}_l, \hat{\boldsymbol{\lambda}}_l)$.

Step 4. Solve the MP.

- Update the lower bound *LBD* as the objective value of the MP.
- Create variable y_k using the solution of the MP. Update $k = k + 1$ and return to Step 2.

equivalent. Thus, our algorithm can converge to the global optimal solution of (31) in a finite number of iterations. A trivial upper bound for the number of iterations is $2^{\mathcal{M}}$. The global optimality and the finite convergence of our algorithm will also be demonstrated in the case studies.

The number of iterations of the GBD depends on the specific problems, and it can be reduced by adding reasonable heuristic constraints to the MPs, such as

$$\sum_{n\in\mathcal{N}} D_n^\tau \le \sum_{i\in\mathcal{I}^U} \bar{p}_i y_i^\tau + \min_{s=1,..,S} \sum_{i\in\mathcal{I}^{Wind}} \bar{w}_i^\tau (1 + v_{i,s}^\tau), \forall \tau \in \mathcal{T}. \tag{43}$$

Constraints (43) are relaxations of the power balance constraints, where $\sum_{n\in\mathcal{N}} D_n^\tau$ is the total electric load demand at period τ;$\sum_{i\in\mathcal{I}^U} \bar{p}_i y_i^\tau$ denotes the sum of the maximum output of all units running at period τ;$\min_{s=1,..,S} \sum_{i\in\mathcal{I}^{Wind}} \bar{w}_i^\tau (1 + v_{i,s}^\tau)$ is a known parameter that denotes the minimum wind energy generation at period τ in all scenarios.

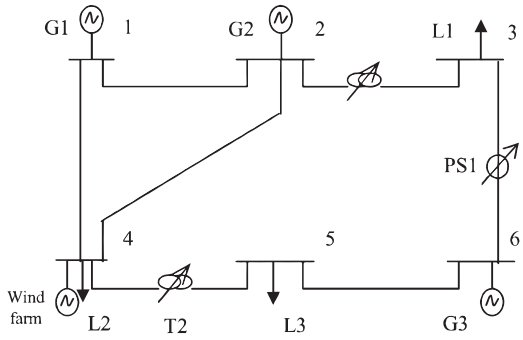

Fig. 2. Configuration of the six-bus system [41].

V. Case Studies

In this section, case studies are conducted to test the proposed model and solution strategy. All tests are performed on a computer with a quad-core processor running at 2.40 GHz with 4 GB of memory. Programs are coded using MATLAB R2013b. All MPs, SL-SPs and SL-FPs are solved by CPLEX [39]. All SL-MPs are solved by Newton's method [40, pp. 484–496]. The objective function of the SL-MP contains an exponential term. In order to reduce the numerical problems, the initial value of α in the Newton's method is set to the mean of θ_s and a numerical calculation method is proposed with details in Appendix III.

We conduct experiments on three systems of different scales. First, a six-bus system is used for testing the effectiveness and global optimality of the proposed model and iterative algorithm. The proposed RDB-DRUC model is compared with SUC, RUC, MB-DRUC models. Then, a modified 118-bus system is examined in case studies with different numbers of scenarios. The calculation times are listed to illustrate the effect of parallelization. Finally, we investigate a practical 319-bus large-scale system with a large number of scenarios.

A. Six-Bus System

The configuration of the six-bus system is depicted in Fig. 2. This six-bus system consists of three units, one wind farm, three loads and seven transmission lines, with detailed data in [41].

We construct the ambiguity set using the nonparametric method. For simplicity, we assume that the true distribution of the relative wind prediction errors is the normal distribution with mean 0 and standard deviation 0.1. The Monte Carlo method is used to generate 3,000 scenarios as hypothetical historical data. The fast-forward selection method is used to reduce the number of scenarios to 10 and generate the nominal distribution.

1) Stochastic UC: A stochastic UC model can be obtained by replacing the objective function of RDB-DRUC model with

$$\text{SUC}: \min \ \boldsymbol{c}^T \boldsymbol{y} + \sum_{s=1}^{S} \rho_s \theta_s. \tag{44}$$

The above SUC model is solved using the L-shaped method [10]. The optimal value of SUC is \$102,768. The optimal UC decisions of three units are shown in Table I.

TABLE I
OPTIMAL UC DECISIONS OF SUC

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
G1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0
G3	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	0	1	1	1	1	0	0

2) Robust UC: To compare all UC methods fairly, the objective function of robust UC is to minimize the total cost under the worst scenario:

$$\text{RUC}: \min \quad \boldsymbol{c}^T \boldsymbol{y} + \max_{s\in\{1,2,\ldots,S\}} \theta_s. \tag{45}$$

The above RUC model is a basic RUC model with a very simple uncertainty set and no control over the conservativeness. Problem (45) can be reformulated as follows:

$$\begin{aligned} \text{Basic RUC}: \min \quad & \boldsymbol{c}^T \boldsymbol{y} + \theta_{\max} \\ \text{s.t.} \quad & \theta_{\max} \geq \theta_s, \forall s = 1, 2, \ldots, S \\ & \theta_s - \boldsymbol{q}_s^T \boldsymbol{x}_s = 0, \forall s = 1, 2, \ldots, S \\ & \boldsymbol{T}\boldsymbol{x}_s + \boldsymbol{W}\boldsymbol{y} - \boldsymbol{h}_s \leq 0, \forall s = 1, 2, \ldots, S \\ & \boldsymbol{A}\boldsymbol{y} - \boldsymbol{b} \leq 0, \boldsymbol{y} \in \{0,1\}^{\mathcal{M}}. \end{aligned} \tag{46}$$

The model (46) is a mixed-integer linear programming model and can be solved by CPLEX [39]. The optimal value of RUC is \$106,178.

3) RDB-DRUC: First, we solve the RDB-DRUC model with the index of ambiguity $\eta = 0$. The optimal value of the RDB-DRUC is \$102,768, which is the same as that of the SUC. And the optimal UC decisions of the RDB-DRUC are also the same as those in Table I. In fact, when $\eta = 0$, only the nominal distribution is contained in the ambiguity set. The DB-DRUC model (16) is equivalent to the SUC model. The results noted above illustrate that the transformation from problem (16) to problem (30) is strictly equivalent.

For fixed $\boldsymbol{y}_k$, the SP (32) is a convex problem. The global optimal solution of the SP can also be obtained by the primal-dual interior-point method [40, pp. 609–615]. Thus, the global optimal solution of the RDB-DRUC (31) can be obtained by enumerating all UC decisions. For different η, we solve the RDB-DRUC model by the enumeration method and the iterative algorithm in Section IV respectively, with results shown in Table II. The convergence tolerance of the iterative algorithm is 10^{-4}. Constraints (43) are added to the MPs. All SL-MPs and SL-FPs are calculated serially.

For each η, the optimal value derived from the iterative algorithm is the same as that derived from the enumeration method. Thus, our iterative algorithm for the RDB-DRUC model guarantees global convergence within finite iterations. When $\eta = 0.1$, the number of iterations is 24. However, when constraint (43) is omitted, the number of iterations is more than 100. Hence, the convergence performance is improved significantly by the addition of reasonable heuristic constraints to the MPs.

TABLE II
OPTIMAL VALUES OF RDB-DRUC WITH DIFFERENT INDEX OF AMBIGUITY

Index of ambiguity	Optimal value (iterative algorithm)	Calculating time (iterative algorithm)	Number of iterations (iterative algorithm)	Optimal value (enumeration method)
0	\$102,768	26.83 s	19	\$102,768
0.05	\$103,338	30.14 s	22	\$103,338
0.1	\$103,577	28.59 s	24	\$103,577
0.3	\$104,153	26.05 s	19	\$104,153
0.5	\$104,546	18.03 s	15	\$104,546
0.7	\$104,859	14.42 s	12	\$104,859
1	\$105,239	15.66 s	13	\$105,239
1.5	\$105,720	16.86 s	14	\$105,720

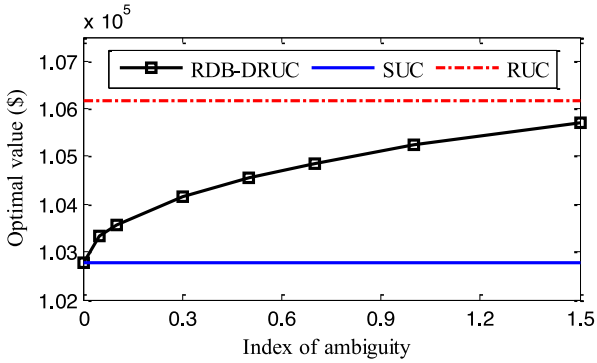

Fig. 3. Optimal values of SUC, RUC and RDB-DRUC.

4) Comparison of UC Methods: The optimal values of SUC, RUC and RDB-DRUC are depicted in Fig. 3.

a) Comparison of RDB-DRUC and SUC: In Fig. 3, the curve RDB-DRUC is higher than the curve SUC. When the index of ambiguity $\eta = 0$, the RDB-DRUC is equivalent to the SUC. For any $\eta > 0$, the RDB-DRUC is more conservative than the SUC, because the objective function of RDB-DRUC is to minimize the total cost under the worst case of the distributions, and the nominal distribution in the SUC is only an element in the ambiguity set.

b) Comparison of RDB-DRUC and RUC: Below we make a simplistic comparison between the RDB-DRUC and the basic RUC. In Fig. 3, the curve RDB-DRUC is lower than the curve RUC. As the index of ambiguity η approaches infinity, the worst case of the distributions has the property that the probability of worst scenario is equal to 1. Thus, as η approaches infinity, the RDB-DRUC is equivalent to the basic RUC. For any $\infty > \eta > 0$, the RDB-DRUC is less conservative than the basic RUC (46).

B. Comparison of DB-DRUC and MB-DRUC

The MB-DRUC is difficult to solve precisely, and Xiong *et al.* only presented an approximate solution of MB-DRUC in [26]. Thus, we only present a qualitative comparison between the performance of DB-DRUC and that of MB-DRUC.

First, it has been shown elsewhere that MB-DRUC is less conservative than RUC [26]. It is also known that MB-DRUC is more conservative than SUC, because the nominal

TABLE III
OPTIMAL VALUES OF RDB-DRUC WITH DIFFERENT NUMBER OF SCENARIOS AND DIFFERENT NUMBER OF WORKERS

Number of scenarios	Index of ambiguity	Number of workers	Optimal value	Calculating time	Number of iterations
20	0.3	1	$1,837,000	57.59 s	8
20	0.3	2	$1,837,000	39.77 s	8
20	0.3	4	$1,837,000	33.12 s	8
40	0.3	1	$1,836,424	96.85 s	8
40	0.3	2	$1,836,424	63.22 s	8
40	0.3	4	$1,836,424	47.61 s	8
60	0.3	1	$1,836,640	151.54 s	8
60	0.3	2	$1,836,640	95.21 s	8
60	0.3	4	$1,836,640	67.34 s	8
80	0.3	1	$1,837,011	215.72 s	8
80	0.3	2	$1,837,011	138.28 s	8
80	0.3	4	$1,837,011	76.54 s	8
100	0.3	1	$1,837,061	267.28 s	8
100	0.3	2	$1,837,061	165.10 s	8
100	0.3	4	$1,837,061	89.82 s	8

distribution in SUC is only an element in the ambiguity set. Thus, the conservatism of MB-DRUC is between those of RUC and SUC. Furthermore, the conservatism of MB-DRUC is hardly adjustable because the moments that are derived from historical data are fixed.

According to the discussion in Section V-A.4), the conservatism of DB-DRUC is dependent on the index of ambiguity. Therefore, when the index of ambiguity is equal to zero, the DB-DRUC are less conservative than those of MB-DRUC. As the index of ambiguity approaches infinity, the DB-DRUC is more conservative than the MB-DRUC. Compared with MB-DRUC, one advantage of DB-DRUC is that its conservatism can be easily adjusted.

C. Modified IEEE 118-Bus Test System

This section conducts experiments on a modified IEEE 118-bus test system, with details provided in [26] and [42]. The hypothetical historical data of relative wind prediction errors are the same as those in Section V-A. The ambiguity sets are constructed using nonparametric methods and constraints (43) are added to the MPs. The results are shown in Table III. The convergence tolerance of the iterative algorithm is 10^{-3}. We use the Parallel Computing Toolbox in MATLAB to perform parallel computations. The number of parallel workers in MATLAB is set to 1, 2 or 4. Fig. 4 shows the curves of calculating time with different number of parallel workers.

When the index of ambiguity, number of workers and number of iterations are held constant, the calculation time increases approximately linearly with the number of scenarios.

When the number of scenarios is large (such as 80 and 100), most of the calculation time is used to solve the SL-SPs and SL-FPs. In this case we can reduce the calculation time by about 40% by using two parallel workers or by about 65% using four parallel workers.

The solution is only guaranteed to be feasible for the sampled scenarios, so it may be infeasible for other scenarios. Hence, 100 scenarios may be insufficient in practice. We test our algorithm further by conducting the following experiments. Table IV shows that all cases converge.

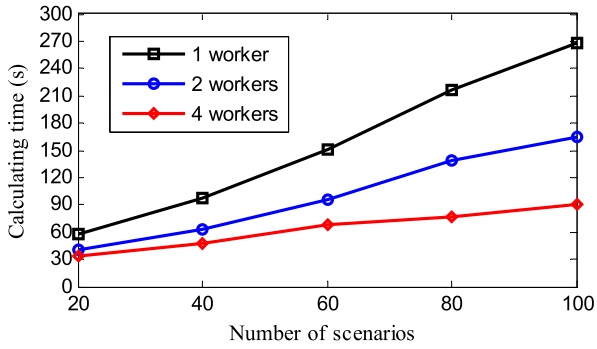

Fig. 4. Calculating time with different number of parallel workers.

TABLE IV
OPTIMAL VALUES OF RDB-DRUC WITH DIFFERENT NUMBER OF SCENARIOS

Number of scenarios	Index of ambiguity	Number of workers	Optimal value	Calculating time	Number of iterations
500	0.3	4	$1,838,202	371 s	8
1000	0.3	4	$1,838,594	869 s	8
1500	0.3	4	$1,838,879	1208 s	8

TABLE V
OPTIMAL VALUES OF RDB-DRUC WITH DIFFERENT CONFIDENCE LEVEL

Number of scenarios	Confidence level	Index of ambiguity	Optimal value	Calculating time	Number of iterations
1000	0.99	0.1845	¥2,700,302	66 min	23
1000	0.95	0.1791	¥2,699,225	62 min	23
1000	0.90	0.1763	¥2,698,581	71 min	25

D. Large-Scale Practical System

The 319-bus system is a practical system located in Jilin province of Northeast China. The data for this system are provided in [43]. The values for the expected electrical load and wind energy generation are based on real data. There are 65 units and 34 wind farms in the test system. For simplicity, we assume that the relative prediction error is the same for all wind farms. The hypothetical historical data of relative wind prediction errors are the same as those in Section V-A. Nonparametric ambiguity sets are constructed with confidence levels of 99%, 95% and 90%. Constraints (43) are added to the MPs. There are more than 30,000 constraints per SL-SP. The results are shown in Table V. The convergence tolerance of the iterative algorithm is 2×10^{-3}. Four MATLAB workers are used in the parallel computations.

As shown in Table V, it can be concluded that our decomposition method and solution algorithm also work very well in large-scale systems with large number of scenarios. The calculation time can be reduced further by adding more parallel workers.

VI. Conclusion

In many practical situations, it is possible to derive a fitted distribution from historical data. However, moment-based DRO only use expectation, variance, and other statistical features to form the ambiguity set, and does not fully utilize this distribution knowledge. Thus, we have proposed a new DB-DRUC model where the ambiguity set is a family of distributions within a fixed distance from the fitted distribution. The reformulation methods for DRO and stochastic programming technologies were applied to reformulate the "min-max-min" DB-DRUC model into a one-level RDB-DRUC model. A two-level decomposition method was then presented, along with an iteration algorithm to solve the RDB-DRUC model, which addresses the computational difficulties under the scenario representation of uncertainties. The iteration algorithm guarantees the global convergence within finite iterations and the simulation results demonstrate that the conservatism of DB-DRUC is between those of basic RUC and SUC. Compared with MB-DRUC, one advantage of DB-DRUC is that its conservatism is dependent on the index of ambiguity and can be easily adjusted. Global optimality and finite convergence of the algorithm are also validated by numerical examples.

Appendix I

We consider a general subproblem (GSP):

$$\begin{aligned}\text{GSP}: \min_{\boldsymbol{z}\in\mathcal{Z}} \quad & f_0(\alpha,\boldsymbol{\theta}) = \alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\theta_s/\alpha}\right)+\alpha\eta \\ \text{s.t.} \quad & h_s(\theta_s,\boldsymbol{x}_s)=0, \forall s=1,2,\ldots,S \\ & g_s(\theta_s,\boldsymbol{x}_s)\le 0, \forall s=1,2,\ldots,S \end{aligned} \tag{47}$$

where vector $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_S)^T$. $h_s(\theta_s,x_s)$ is affine and $g_s(\theta_s,x_s)$ is convex. $\mathcal{Z}$ is the domain of the GSP. We can also write down the general second-level subproblem (GSL-SP) and general second level master problem (GSL-MP):

$$\begin{aligned}\text{GSL-SP}: \tilde{\theta}_s = \min_{\theta_s,\boldsymbol{x}_s} \quad & f_s(\theta_s)=\theta_s \\ \text{s.t.} \quad & h_s(\theta_s,\boldsymbol{x}_s)=0, \\ & g_s(\theta_s,\boldsymbol{x}_s)\le 0 \end{aligned} \tag{48}$$

$$\text{GSL-MP}: \min_{\alpha\ge 0} \quad f_0(\alpha,\tilde{\boldsymbol{\theta}}) = \alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}\right)+\alpha\eta. \tag{49}$$

The KKT conditions of the GSP are shown in (50)–(54). Symbols marked with a hat ($\hat{}$) are also used to denote the Lagrange multipliers and optimal variables of GSP.

$$\nabla_\alpha f_0(\hat{\alpha},\hat{\boldsymbol{\theta}})=0, \tag{50}$$

$$\begin{aligned}&\rho_s e^{\hat{\theta}_s/\hat{\alpha}}\Big/\sum_{s'=1}^{S}\rho_{s'}e^{\hat{\theta}_{s'}/\hat{\alpha}}+\hat{\boldsymbol{\mu}}_s^T\nabla_{\theta_s}h_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s) \\ &+\hat{\boldsymbol{\lambda}}_s^T\nabla_{\theta_s}g_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)=0, \forall s=1,2,\ldots,S,\end{aligned} \tag{51}$$

$$\hat{\boldsymbol{\mu}}_s^T\nabla_{\boldsymbol{x}_s}h_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)+\hat{\boldsymbol{\lambda}}_s^T\nabla_{\boldsymbol{x}_s}g_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)=0, \forall s=1,2,\ldots,S, \tag{52}$$

$$h_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)=0, g_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)\le 0, \forall s=1,2,\ldots,S, \tag{53}$$

$$\hat{\boldsymbol{\lambda}}_s\ge 0, \hat{\boldsymbol{\lambda}}_s.*g_s(\hat{\theta}_s,\hat{\boldsymbol{x}}_s)=0, \forall s=1,2,\ldots,S. \tag{54}$$

The KKT conditions of the GSL-SP are shown in (55)–(58). Symbols marked with a tilde ($\sim$) are also used to denote the Lagrange multipliers and optimal variables of GSL-MP and GSL-SP.

$$1+\tilde{\boldsymbol{\mu}}_s^T\nabla_{\theta_s}h_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)+\tilde{\boldsymbol{\lambda}}_s^T\nabla_{\theta_s}g_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)=0, \tag{55}$$

$$\tilde{\boldsymbol{\mu}}_s^T\nabla_{\boldsymbol{x}_s}h_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)+\tilde{\boldsymbol{\lambda}}_s^T\nabla_{\boldsymbol{x}_s}g_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)=0, \tag{56}$$

$$h_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)=0, g_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)\le 0, \tag{57}$$

$$\tilde{\boldsymbol{\lambda}}_s\ge 0, \tilde{\boldsymbol{\lambda}}_s.*g_s(\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s)=0. \tag{58}$$

The KKT condition of the GSL-MP is expressed by

$$\nabla_\alpha f_0(\tilde{\alpha},\tilde{\boldsymbol{\theta}})=0. \tag{59}$$

Comparing (50)–(54) with (55)–(59), we find that differences only exist between (51) and (55). If $\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s,\tilde{\boldsymbol{\mu}}_s$ and $\tilde{\boldsymbol{\lambda}}_s$ are the solution of (55)–(59),$\tilde{\theta}_s,\tilde{\boldsymbol{x}}_s,\tilde{o}_s\tilde{\boldsymbol{\mu}}_s$ and $\tilde{o}_s\tilde{\boldsymbol{\lambda}}_s$ are the solution of (50)–(54), where $\tilde{o}_s$ is expressed by

$$\tilde{o}_s=\rho_s e^{\tilde{\theta}_s/\tilde{\alpha}}\Big/\sum_{s'=1}^{S}\rho_{s'}e^{\tilde{\theta}_{s'}/\tilde{\alpha}}, \forall s=1,2,\ldots,S. \tag{60}$$

Thus, the Lagrange multipliers and optimal solution of the GSP can be obtained by

$$\begin{aligned}&\hat{\alpha}=\tilde{\alpha},\hat{\theta}_s=\tilde{\theta}_s,\hat{\boldsymbol{x}}_s=\tilde{\boldsymbol{x}}_s,\forall s=1,2,\ldots,S \\ &\hat{\boldsymbol{\mu}}_s=\tilde{\boldsymbol{\mu}}_s\rho_s e^{\tilde{\theta}_s/\tilde{\alpha}}\Big/\sum_{s'=1}^{S}\rho_{s'}e^{\tilde{\theta}_{s'}/\tilde{\alpha}}, \forall s=1,2,\ldots,S \\ &\hat{\boldsymbol{\lambda}}_s=\tilde{\boldsymbol{\lambda}}_s\rho_s e^{\tilde{\theta}_s/\tilde{\alpha}}\Big/\sum_{s'=1}^{S}\rho_{s'}e^{\tilde{\theta}_{s'}/\tilde{\alpha}}, \forall s=1,2,\ldots,S.\end{aligned} \tag{61}$$

Appendix II

For notational convenience, the subscript 'k' is omitted below. Based on (36) and (40), the support functions in (33) can be reformulated as

$$\begin{aligned}\xi=&\inf_{\alpha\ge 0,\boldsymbol{\theta}}\left(\alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\theta_s/\alpha}\right)+\alpha\eta+\sum_{s=1}^{S}\hat{\mu}_s\theta_s\right) \\ &+\inf_{\boldsymbol{x}_s}\sum_{s=1}^{S}\left(-\hat{\mu}_s\boldsymbol{q}_s^T+\hat{\boldsymbol{\lambda}}_s^T\boldsymbol{T}\right)x_s+\sum_{s=1}^{S}\hat{\boldsymbol{\lambda}}_s^T\left(\boldsymbol{W}\boldsymbol{y}-\boldsymbol{h}_s\right)\end{aligned} \tag{62}$$

$$\bar{\xi}=\inf_{\boldsymbol{x}_s}\bar{\boldsymbol{\lambda}}_s^T\boldsymbol{T}x_s+\bar{\boldsymbol{\lambda}}_s^T\left(\boldsymbol{W}\boldsymbol{y}-\boldsymbol{h}_s\right). \tag{63}$$

By comparing the SP (32) to the GSP (47), (50)–(52) are reformulated, respectively, as

$$\log\left(\sum_{s=1}^{S}\rho_s e^{\hat{\theta}_s/\hat{\alpha}}\right)-\frac{\sum_{s=1}^{S}\rho_s\hat{\theta}_s e^{\hat{\theta}_s/\hat{\alpha}}}{\hat{\alpha}\sum_{s=1}^{S}\rho_s e^{\hat{\theta}_s/\hat{\alpha}}}+\eta=0 \quad (64)$$

$$\rho_s e^{\hat{\theta}_s/\hat{\alpha}}\Big/\sum_{s'=1}^{S}\rho_{s'}e^{\hat{\theta}_{s'}/\hat{\alpha}}+\hat{\mu}_s=0,\forall s=1,2,\ldots,S \quad (65)$$

$$-\hat{\mu}_s\boldsymbol{q}_s^T+\hat{\boldsymbol{\lambda}}_s^T\boldsymbol{T}=0,\forall s=1,2,\ldots,S. \quad (66)$$

Equations (64) and (65) are also the optimal conditions of

$$\inf_{\alpha\geq 0,\boldsymbol{\theta}}\left(\alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\theta_s/\alpha}\right)+\alpha\eta+\sum\nolimits_{s=1}^{S}\hat{\mu}_s\theta_s\right). \quad (67)$$

Thus $(\hat{\alpha},\hat{\boldsymbol{\theta}})$ are the optimal solutions of (67). Substituting (64)–(66) into (62) we obtain

$$\xi=\sum_{s=1}^{S}\hat{\boldsymbol{\lambda}}_s^T\left(\boldsymbol{W}\boldsymbol{y}-\boldsymbol{h}_s\right). \quad (68)$$

Similarly, (63) can be reformulated as

$$\bar{\xi}=\bar{\boldsymbol{\lambda}}_s^T\left(\boldsymbol{W}\boldsymbol{y}-\boldsymbol{h}_s\right). \quad (69)$$

Appendix III

The objective function of the SL-MP (35) contains exponential terms. This means that numerical issues are likely to arise when solving the SL-MP using Newton's method. Below we will introduce the method to reduce the value of exponential terms.

$f_0(\alpha,\tilde{\boldsymbol{\theta}})$ is used to denote the objective function of the SL-MP. In addition to $f_0(\alpha,\tilde{\boldsymbol{\theta}})$, we need to calculate the partial derivatives $\partial f_0(\alpha,\tilde{\boldsymbol{\theta}})/\partial\alpha$ and $\partial^2 f_0(\alpha,\tilde{\boldsymbol{\theta}})/\partial\alpha^2$. The expressions for $f_0(\alpha,\tilde{\boldsymbol{\theta}})$, $\partial f_0(\alpha,\tilde{\boldsymbol{\theta}})/\partial\alpha$ and $\partial^2 f_0(\alpha,\tilde{\boldsymbol{\theta}})/\partial\alpha^2$ are shown, respectively, in

$$f_0(\alpha,\tilde{\boldsymbol{\theta}})=\alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}\right)+\alpha\eta \quad (70)$$

$$\frac{\partial f_0(\alpha,\tilde{\boldsymbol{\theta}})}{\partial\alpha}=\log\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}\right)-\frac{\sum_{s=1}^{S}\rho_s\tilde{\theta}_s e^{\tilde{\theta}_s/\alpha}}{\alpha\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}}+\eta \quad (71)$$

$$\frac{\partial^2 f_0(\alpha,\tilde{\boldsymbol{\theta}})}{\partial\alpha^2}=\frac{1}{\alpha^3}\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}\sum_{s=1}^{S}\rho_s\tilde{\theta}_s^2 e^{\tilde{\theta}_s/\alpha}-\left(\sum_{s=1}^{S}\rho_s\tilde{\theta}_s e^{\tilde{\theta}_s/\alpha}\right)^2\right)\Big/\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha}\right)^2. \quad (72)$$

We reduce the value of the exponential term by introducing an intermediate variable ϑ as follows:

$$\vartheta=\frac{1}{S}\sum_{s=1}^{S}\frac{\tilde{\theta}_s}{\alpha}, \quad (73)$$

where $\alpha\vartheta$ is equal to the average cost of all the scenarios. $f_0(\alpha,\tilde{\boldsymbol{\theta}})$ can be reformulated as follows:

$$f_0(\alpha,\tilde{\boldsymbol{\theta}})=\alpha\log\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}\right)+\alpha\vartheta+\alpha\eta. \quad (74)$$

In most practical systems, it can be believed that the uncertainty in wind power does not have a significant effect on the total cost of generation. In the other words, the total generation cost of each scenario $\tilde{\theta}_s$ is close to the average cost of all the scenarios $\alpha\vartheta$ and the following formula holds:

$$e^{\tilde{\theta}_s/\alpha-\vartheta}<<e^{\tilde{\theta}_s/\alpha},\forall s=1,2,\ldots,S. \quad (75)$$

Thus (74), compared with (70), can effectively limit the value of the exponential term. Similarly, (71) and (72) can be reformulated as follows:

$$\frac{\partial f_0(\alpha,\tilde{\boldsymbol{\theta}})}{\partial\alpha}=\log\left(\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}\right)+\vartheta+\eta-\sum_{s=1}^{S}\rho_s\tilde{\theta}_s e^{\tilde{\theta}_s/\alpha-\vartheta}\Big/\left(\alpha\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}\right) \quad (76)$$

$$\frac{\partial^2 f_0(\alpha,\tilde{\boldsymbol{\theta}})}{\partial\alpha^2}=\frac{1}{\alpha^3}\left(\frac{\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}}{\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}}\frac{\sum_{s=1}^{S}\rho_s\tilde{\theta}_s^2 e^{\tilde{\theta}_s/\alpha-\vartheta}}{\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}}-\left(\frac{\sum_{s=1}^{S}\rho_s\tilde{\theta}_s e^{\tilde{\theta}_s/\alpha-\vartheta}}{\sum_{s=1}^{S}\rho_s e^{\tilde{\theta}_s/\alpha-\vartheta}}\right)^2\right) \quad (77)$$

In (77), we further limit the issues with the numerical calculations by exchanging the order of the multiplications and divisions.

References

[1] J. Li *et al.*, *2014 China Wind Power Review and Outlook, Chinese Renewable Energy Ind. Assoc. (CREIA)*, Beijing, China, 2014.

[2] G. W. E. Council, "Global wind statistics 2012," *Global Wind Rep.*, Brussels, Belgium, 2013. [Online]. Available: http://www.gwec.net/wp-content/uploads/2013/02/GWEC-PRstats-2012_english.pdf

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[4] R. Jiang, J. Wang, M. Zhang, and Y. Guan, "Two-stage minimax regret robust unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2271–2282, Aug. 2013.

[5] L. Zhao and B. Zeng, "Robust unit commitment problem with demand response and wind energy," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–8.

[6] D. Bertsimas, E. Litvinov, X. 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.

[7] C. Zhao, J. Wang, J.-P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Trans. Power. Syst.*, vol. 28, no. 3, pp. 2708–2717, Aug. 2013.

[8] C. Zhao and Y. Guan, "Unified stochastic and robust unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3353–3361, Aug. 2013.

[9] A. Baringo and L. Baringo, "A stochastic adaptive robust optimization approach for the offering strategy of a virtual power plant," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3492–3504, Sep. 2017.

[10] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York, NY, USA: Springer, 2011.

[11] S. Takriti, J. R. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1497–1508, Aug. 1996.

[12] U. A. Ozturk, M. Mazumdar, and B. A. Norman, "A solution to the stochastic unit commitment problem using chance constrained programming," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1589–1598, Aug. 2004.
[13] T. Li, M. Shahidehpour, and Z. Li, "Risk-constrained bidding strategy with stochastic unit commitment," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 449–458, Feb. 2007.
[14] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, 2010.
[15] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen, "Robust solutions of optimization problems affected by uncertain probabilities," *Manage. Sci.*, vol. 59, no. 2, pp. 341–357, 2013.
[16] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Math. Program.*, vol. 158, no. 1/2, pp. 291–327, 2016.
[17] K. Postek, D. Den Hertog, and B. Melenberg, "Tractable counterparts of distributionally robust constraints on risk measures," *CentER Discuss. Paper Series*, 2014–031, 2014.
[18] Z. Hu and J. L. Hong, "Kullback–Leibler divergence constrained distributionally robust optimization," Nov. 2012. [Online]. Available: http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf
[19] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 24, no. 9, pp. 1–52, 2015.
[20] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.
[21] F. Qiu and J. Wang, "Distributionally robust congestion management with dynamic line ratings," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2198–2199, Jul. 2015.
[22] Q. Bian, H. Xin, Z. Wang, D. Gan, and K. P. Wong, "Distributionally robust solution to the reserve scheduling problem with partial information of wind power," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2822–2823, Sep. 2015.
[23] W. Wei, F. Liu, and S. Mei, "Distributionally robust co-optimization of energy and reserve dispatch," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 289–300, Jan. 2016.
[24] X. Chen, W. Wu, B. Zhang, and C. Lin, "Data-driven DG capacity assessment method for active distribution networks," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3946–3957, Sep. 2017.
[25] 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.
[26] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 39–49, Jan. 2017.
[27] C. Zhao and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 94–102, Jan. 2018s.
[28] Y. Guan and J. Wang, "Uncertainty sets for robust unit commitment," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1439–1440, May 2014.
[29] 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.
[30] C. Zhao, J. Wang, J. P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2708–2717, Aug. 2013.
[31] Y. An and B. Zeng, "Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 109–122, Jan. 2015.
[32] J. Shao, *Mathematical Statistics*. New York, NY, USA: Springer, 2003.
[33] H. Alizadeh Noughabi and N. Balakrishnan, "Tests of goodness of fit based on Phi-divergence," *J. Appl. Statist.*, vol. 43, no. 3, pp. 412–429, 2016.
[34] H. Heitsch and W. Romisch, "Scenario reduction algorithms in stochastic programming," *Comput. Optim. Appl.*, vol. 24, no. 2/3, pp. 187–206, 2003.
[35] L. Pardo, *Statistical Inference Based on Divergence Measures*. Boca Raton, FL, USA:CRC Press, 2005.
[36] Z. Li, W. Wu, and B. Zhang, "A Kullback-Leibler divergence-based distributionally robust optimization model for heat pump day-ahead operational schedule in distribution networks," *IET Gen., Transmission, Distrib.*, vol. 12, no. 13, pp. 3136–3144, Jul. 2018.
[37] A. M. Geoffrion, "Generalized benders decomposition," *J. Opt. Theory Appl.*, vol. 10, no. 4, pp. 237–260, 1972.
[38] C. A. Floudas, *Nonlinear and Mixed-Integer Optimization*. Oxford, U.K.: Oxford Univ. Press, 1995, pp. 135–136.
[39] CPLEX Homepage. 2017. [Online]. Available: http://www-01.ibm.com/softw are/commerce/optimization/cplex-optimizer
[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[41] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1319–1327, Aug. 2008.
[42] IEEE 118-bus system. 2004. [Online]. Available: motor.ece.iit.edu/data/JEAS_IEEE118.doc
[43] Test data for 319-bus system. 2017. [Online]. Available: https://drive.google.com/file/d/0B3NgXVM23lMuNDBlMFhtSjctdDg/view?usp=sharing

Yuwei Chen received the B.S. degree from Tsinghua University, Beijing, China, in 2016, where he is currently working toward the Ph.D. degree at the State Key Laboratory of Power Systems. His research interests include energy management and microgrid analysis and operation.

Qinglai Guo (SM'14) was born in Jilin City, Jilin Province, China on March 6, 1979. He received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2000 and 2005, respectively. He is currently an Associate Professor with Tsinghua University. His special fields of interest include smart grids, cyber-physical systems, and electrical power control center applications. He is a member of CIGRE C2.13 Task Force on Voltage/Var support in System Operations.

Hongbin Sun (SM'12–F'18) received the double B.S. degrees and Ph.D. degree from Tsinghua University, Beijing, China, in 1992 and 1997, respectively. He is currently Changjiang Scholar Chair Professor of Education Ministry of China, tenured Full Professor of electrical engineering, and the Director of Energy Management and Control Research Center, Tsinghua University. From September 2007 to September 2008, he was a Visiting Professor with the School of EECS, Washington State University. He is serving as a Chair of IEEE smart grid voltage control task force and IEEE Energy Internet working group. He served as the founding Chair of IEEE Conference on Energy Internet and Energy System Integration in November 2017. In recent 20 years, he led a research group in Tsinghua University to develop a commercial system-wide automatic voltage control system, which has been applied to more than 100 electrical power control centers in China as well as the control center of PJM interconnection, the largest regional power grid in USA. His research interests include energy management system, automatic voltage control, energy Internet, and energy system integration. He published more than 300 peer-reviewed papers. He was the recipient of the China National Technology Innovation Award in 2008, the National Distinguished Teacher Award in China in 2009, and the National Science Fund for Distinguished Young Scholars of China in 2010. He is a Fellow of IET.

Zhengshuo Li (S'12–M'16) received the bachelor's and the Ph.D. degrees Tsinghua University, Beijing, China, in 2011 and 2016, respectively. He is currently a Postdoctoral Fellow with the Tsinghua-Berkeley Shenzhen Institute (TBSI), Shenzhen, China. From February 2014 to May 2014, he was a Visiting Student with the Decision and Information Sciences Division, Argonne National Laboratory, Argonne, IL, USA. From January 2018 to May 2018, he is a Visiting Scholar with the University of California, Berkeley, CA, USA. His research interests include economic dispatch and security analysis of smart transmission and distribution grids, distributed storage utilization and demand response in smart grids. He was Excellent Graduates of Tsinghua University. He was the recipient of the Best Paper Award of 13th China National Doctoral Academic Annual Meeting in 2015 and 2016 IEEE PES General Meeting Best Conference Paper. He also was the recipient of the Academic Rookie Award of Department of Electrical Engineering, Tsinghua University in 2015 and the Excellent Doctoral Dissertation Award of Tsinghua University in 2016. His dissertation was selected for Springer Theses. He was also a recipient of the Best Reviewer Award for the IEEE TRANSACTIONS ON SMART GRID in 2015.

Wenchuan Wu (SM'14) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China. Currently, he is a Professor with the Department of Electrical Engineering, Tsinghua University. His research interests include energy management system, active distribution system operation and control, and EMTP-TSA hybrid real-time simulation. He is an Associate Editor of *IET Generation, Transmission & Distribution* and an Editorial Board Member of *Electric Power Components and Systems*.

Zihao Li received the B.S. degree from Tsinghua University, Beijing, China, in 2016, where he is currently working toward the Ph.D. degree at the Department of Electrical Engineering. His research interests include active distribution system and demand response.