Robust Scheduling with Improved Uncertainty Sets via Purchase of Distributed Predictive Information

Rui Xie, Pierre Pinson, *Fellow, IEEE,* Yue Chen, *Member, IEEE*

***Abstract*—Robust scheduling is an essential way to cope with uncertainty. However, the rising unpredictability in net demand of distributed prosumers and the lack of relevant data make it difficult for the operator to forecast the uncertainty well. This leads to inaccurate, or even infeasible, robust scheduling strategies. In this paper, a novel two-stage robust scheduling model is developed, which enables the operator to purchase predictive information from distributed prosumers to enhance scheduling efficiency. An improved uncertainty set with a smaller variation range is developed by combining the forecasts from the operator and prosumers. Since the improved uncertainty set is influenced by the first-stage information purchase related decisions, the proposed model eventually becomes a case of robust optimization with decision-dependent uncertainty (DDU). An adaptive column-and-constraint generation (C&CG) algorithm is developed to solve the problem within a finite number of iterations. The potential failures of traditional algorithms in detecting feasibility, guaranteeing convergence, and reaching optimal strategies under DDU are successfully circumvented by the proposed algorithm. Case studies demonstrate the effectiveness, necessity, and scalability of the proposed method.**

***Index Terms*—decision-dependent uncertainty, robust scheduling, distributed information, adaptive C&CG**

I. Introduction

With the proliferation of distributed energy resources (DERs), such as rooftop PV panels and electric vehicles, traditional passive consumers are now turning into responsive prosumers [1]. The net demand of a prosumer may be highly uncertain due to both fluctuating distributed renewable generation and volatile consumption behaviors. This inherent uncertainty is exerting great challenges on power system operations, such as higher risks of power imbalance and inadequate ramping capacities [2]. Two-stage robust scheduling is an essential way to deal with uncertainty and has been applied for virtual power plants [3], smart buildings [4], and charging stations [5], among others.

In a robust scheduling problem, the operator minimizes the worst-case operation cost considering the variation of the random factor within an uncertainty set. Hence, the choice of an uncertainty set will influence the feasibility and optimality of the obtained robust scheduling strategy. A vast literature has been devoted to improving the uncertainty set. The parameters of an uncertainty set were set to achieve the best trade-off between security and conservativeness [6]. A method for building polyhedral uncertainty sets based on the theory of coherent risk measures was introduced [7]. In addition to better selection of parameters for the uncertainty set, enhancing the prediction accuracy is another important way. An improved wind forecast framework making use of the spatio-temporal correlation in wind speed was developed and used to build a more accurate uncertainty set [8]. A dynamic uncertainty set was proposed in [9] to characterize the impact of uncertainty realization in the last period on the current uncertainty set. However, in the aforementioned studies, the uncertainty set is estimated by the operator through processing, structuring, and organizing their own data. This has been proved to be effective for traditional power systems, where uncertainties mainly stem from large-scale renewable generators, for which the operator has enough data to have a fair representation of uncertainties involved. However, in future power systems, the uncertain net demand of prosumers will become an non-negligible part, whose data are stored locally.

Some recent work has proposed data markets to tackle this challenge. For example, regression markets were developed to aggregate local data for energy forecasting with proper incentives based on cooperative game theory [10], [11]. A data sharing mechanism was designed for electricity retailers to improve their profits in the wholesale market [12]. Due to the high communication and computational burdens and the risk of private data leakage, sharing data to perform a central prediction may not always be a good way. An alternative approach is to build an information market for aggregating distributed predictions [13], [14]. This paper chooses to focus on the latter approach, i.e., to help the operator improve the uncertainty set by purchasing distributed predictions.

Despite the efforts mentioned in the above, the use of distributed predictive information in power system scheduling is an open question that remains to be addressed. This paper makes an initial attempt by allowing the operator to make use of (and actually purchase) distributed predictive information for better robust scheduling. The proposed model turns out to be a case of robust optimization (RO) with decision-dependent uncertainty (DDU). The decision-dependent feature makes the traditional algorithms, such as Benders decomposition [15] and column-and-constraint generation (C&CG) [16], inapplicable. A reformulation method was proposed to solve RO with DDU [17] and was extended to a more general uncertainty set [18]. These two studies focused on static robust models, and the adaptive two-stage robust model is even more challenging. Modified Benders decomposition [3], [19], improved C&CG [20], and multi-parametric programming [21] methods were established to provide an exact solution. However, they might be computational expensive with an increasing number of

R. Xie and Y. Chen are with the Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, Hong Kong SAR. (email: ruixie@cuhk.edu.hk; yuechen@mae.cuhk.edu.hk)

P. Pinson is with the Dyson School of Design Engineering, Imperial College London, UK. (email: p.pinson@imperial.ac.uk)

 (Corresponding author: Y. Chen)

resources. An efficient solution algorithm that is suitable for the DDU set studied in this paper is necessary.

Our core objective is to investigate how to exploit distributed predictive information to build an improved uncertainty set and how to apply it in robust scheduling. An adaptive C&CG algorithm is developed to solve the problem. Our contributions are two-fold:

(1) **Robust Scheduling Model**. We build an improved uncertainty set using conditional expectations and variances, which are derived by combining forecasts from the operator and distributed agents (prosumers). Based on this, a two-stage robust scheduling model with DDU is proposed. Distinct from prior robust models that employed a pre-determined uncertainty set and focused on energy decisions only, the proposed model incorporates the operator's decision on purchasing distributed information to get a more accurate uncertainty set to enhance their scheduling efficiency.

(2) **Solution Algorithm**. An adaptive C&CG algorithm is developed to solve the proposed model. As mentioned above, traditional RO techniques may fail to detect feasibility and guarantee optimality when dealing with DDU. The proposed algorithm is proved to converge to the optimal solution within a finite number of iterations. Case studies demonstrate the necessity and effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. Section II builds an improved uncertainty set based on the distributed predictions. A two-stage robust scheduling model with DDU and its solution algorithm are developed in Section III and IV, respectively. Numerical experiments are carried out in Section V. Section VI concludes the paper.

II. Improved Uncertainty Set Construction

The two-stage RO has wide potential applications, such as in economic dispatch [22] and unit commitment [23]. It can be generally formulated as

$$\min_{x} \left\{ f(x) + \max_{u\in\mathcal{U}} \min_{y\in\mathcal{Y}(x,u)} g(y) \right\}, \quad \text{s.t. } x \in \mathcal{X} \cap \mathcal{X}_R, \tag{1}$$

with

$$\mathcal{X}_R = \{x \mid \mathcal{Y}(x,u) \neq \emptyset, \forall u \in \mathcal{U}\}, \tag{2}$$

where x and y are the first- and second-stage decision variables, respectively. In parallel, u represents the potential uncertainty realization varying within an uncertainty set $\mathcal{U}$. $\mathcal{X}$ and $\mathcal{X}_R$ are the feasible and robust feasible sets of x, respectively. According to (2), a first-stage decision x is robust feasible if and only if for any realization of uncertainty u, we can find a feasible second-stage solution $y \in \mathcal{Y}(x,u)$.

Conventionally, the uncertainty set $\mathcal{U}$ is estimated by the operator using their own data and determined in advance. However, in future power systems, the uncertainty may largely come from distributed agents, e.g., distributed renewable generators and responsive demands. Due to privacy concerns, the operator may lack sufficient data and the expertise to estimate $\mathcal{U}$. A more practical way is to let the operator purchase predictive information from distributed agents for better estimation.

To this end, in the following, we first investigate how, if provided with distributed predictions, the operator can improve their forecasts and obtain an improved uncertainty set.

A. Improved Forecasts

Suppose there are I agents indexed by $i \in \mathcal{I} = \{1,2,...,I\}$. The uncertainty of agent $i \in \mathcal{I}$ can be represented as a random variable U_i in $\mathbb{R}^+$ with an unknown distribution, and u_i is its realization. In this paper, we focus on the polyhedral uncertainty set that is widely used in RO. The two key parameters to determine such a set are the expectation and variance of the uncertain factor. Suppose the operator can estimate the expectation and variance of U_i using historical data, denoted by $\bar{u}_i := \mathbb{E}[U_i]$ and $\sigma_{U_i}^2 := \text{var}(U_i)$, respectively. As mentioned above, these two estimates may be inaccurate due to the limited data of the operator. To predict more precisely, the operator can buy distributed predictive information from the agents. Suppose agent i's forecast of U_i is U_i^{pre}, then we have $U_i = U_i^{pre} + \varepsilon_i$, where ε_i is a random noise. The prediction U_i^{pre} is also a random variable and let u_i^{pre} be its realization. Denote by $\sigma_{\varepsilon_i}^2 := \text{var}(\varepsilon_i)$ the variance of ε_i. The higher the agent's prediction accuracy, the smaller $\sigma_{\varepsilon_i}^2$ is. Throughout the paper, we assume that:

A1: $\{\varepsilon_i, \forall i \in \mathcal{I}\}$ are independent. Each ε_i has zero expectation, i.e., $\mathbb{E}[\varepsilon_i] = 0$, and ε_i is independent of U_i. In particular, $\text{cov}(U_i, \varepsilon_i) = 0$.

First, let us see how the operator can improve their estimation of U_i with the help of distributed predictive information u_i^{pre}. We propose to use the conditional probability $\mathcal{P}(U_i | U_i^{pre} = u_i^{pre})$ as an approximation of the actual probability $\mathcal{P}(U_i)$ of U_i. Then the uncertainty set can be constructed based on the conditional expectation $\mathbb{E}[U_i | U_i^{pre} = u_i^{pre}]$ and variance $\text{var}(U_i | U_i^{pre} = u_i^{pre})$. Generally, these two parameters can be complicated nonlinear functions of u_i^{pre}. For simplicity, in this paper, we adopt the *best linear predictor* of U_i, which is $U_i^e := \alpha_i + \beta_i U_i^{pre}$, so that the squared error expectation is minimized as follows,

$$\min_{\alpha_i, \beta_i} \mathbb{E}\left[\left(U_i - (\alpha_i + \beta_i U_i^{pre}) \right)^2 \right], \tag{3}$$

where α_i and β_i are parameters to be determined. Denote the realization of U_i^e by u_i^e.

Lemma 1. When A1 holds, the two parameters of the best linear predictor $U_i^e = \alpha_i + \beta_i U_i^{pre}$ are

$$\beta_i = \frac{\sigma_{U_i}^2}{\sigma_{U_i}^2 + \sigma_{\varepsilon_i}^2}, \ \alpha_i = (1 - \beta_i)\bar{u}_i. \tag{4}$$

Moreover, $\mathbb{E}[U_i - U_i^e] = 0$ and $\text{cov}(U_i - U_i^e, U_i^{pre}) = 0$.

The proof of Lemma 1 can be found in Appendix A. When the error is very small ($\sigma_{\varepsilon_i}^2 \to 0$), we have $\alpha_i = 0, \beta_i = 1$, and thus $u_i^e = u_i^{pre}$. It means that the distributed predictive information u_i^{pre} is accurate so we just use it. When the error is very high ($\sigma_{\varepsilon_i}^2 \to \infty$), we have $\alpha_i = \bar{u}_i, \beta_i = 0$, and thus

$u_i^e = \bar{u}_i$. It means that the distributed predictive information is so inaccurate that we cannot get a better estimation than the original one $\bar{u}_i$.

Based on the best linear predictor, the conditional variance $\mathrm{var}(U_i|U_i^{pre} = u_i^{pre})$ can be calculated by Lemma 2. Let $\eta_i := U_i - U_i^e$. Note that by Lemma 1, we have $\mathrm{cov}(\eta_i, U_i^{pre}) = 0$. So it is reasonable to make a stronger assumption:

A2: η_i and U_i^{pre} are independent random variables.

Lemma 2. When A1 and A2 hold, we have

$$\mathrm{var}(U_i|U_i^{pre} = u_i^{pre}) = (1-\beta_i)^2\sigma_{U_i}^2 + \beta_i^2\sigma_{\varepsilon_i}^2 \tag{5}$$

and $\mathrm{var}(U_i|U_i^{pre} = u_i^{pre}) \le \sigma_{U_i}^2$.

The proof of Lemma 2 can be found in Appendix B. From Lemma 1, we have observed that the more accurate the distributed predictive information (i.e., the smaller the $\sigma_{\varepsilon_i}^2$), the larger the β_i (weight on u_i^{pre}). In an extreme case when $\sigma_{\varepsilon_i}^2$ is zero, we have $u_i^e = u_i^{pre}$ and $\mathrm{var}(U_i|U_i^{pre} = u_i^{pre}) = 0$. This indicates that with distributed predictive information, the operator can know the exact value of u_i, and so there is no uncertainty. On the contrary, when $\sigma_{\varepsilon_i}^2 \to \infty$, we have $u_i^e = \bar{u}_i$, and $\mathrm{var}(U_i|U_i^{pre} = u_i^{pre}) = \sigma_{U_i}^2$, meaning that the distributed prediction is so inaccurate that the operator still uses the same estimation as if there were no distributed information. From the above analysis, we find that $\beta_i \in [0,1]$ can be a good indicator of the prediction accuracy of distributed information. We borrow similar concepts from economics and define the prediction accuracy as follows.

Definition 1. (Prediction accuracy [24]) The parameter β_i in (4) can be formally defined as the prediction accuracy, i.e.,

$$\tau_i := \frac{\sigma_{U_i}^2}{\sigma_{U_i}^2 + \sigma_{\varepsilon_i}^2}. \tag{6}$$

The accuracy of each agent's prediction to the operator is influenced both by the agent's forecasting technology and the incentive paid by the operator.

B. Improved Uncertainty Set

With the above improved forecasts, the operator can then construct an improved uncertainty set. Suppose there are T periods indexed by $t \in \mathcal{T} = \{1,...,T\}$, then we have the best linear predictor u_{it}^e and variance $\mathrm{var}(U_{it}|U_{it}^{pre} = u_{it}^{pre})$ for all $i \in \mathcal{I}$ and $t \in \mathcal{T}$. We adopt a polyhedral uncertainty set with the following form.

$$\begin{aligned}\mathcal{U}(\tau) = \ & \{u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}: \\ & u_{it}^0 - u_{it}^h \le u_{it} \le u_{it}^0 + u_{it}^h, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \\ & \sum_i \frac{|u_{it} - u_{it}^0|}{u_{it}^h} \le \Gamma_S, \ \forall t \in \mathcal{T} \\ & \sum_t \frac{|u_{it} - u_{it}^0|}{u_{it}^h} \le \Gamma_T, \ \forall i \in \mathcal{I}\},\end{aligned} \tag{7}$$

where Γ_S and Γ_T are the uncertainty budgets to restrain the spatial and temporal deviations from the forecast u_{it}^0. An illustration of the original and improved uncertainty sets is given in Fig. 1. Denote $v_{it} := |u_{it} - u_{it}^0|/u_{it}^h, \forall i, \forall t$, so $v_{it} = |\eta_{it}/u_{it}^h|$ when $u_{it}^0 = u_{it}^e$. For simplicity, we assume that:

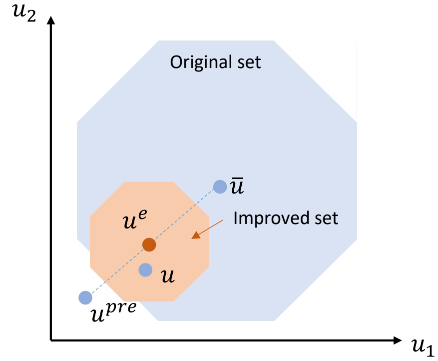

Fig. 1. The original uncertainty set (blue region) with $u_i^0 = \bar{u}$ and the improved uncertainty set (orange region) with $u_i^0 = u_i^e$.

A3: $\{\eta_{it}/\mathrm{var}[\eta_{it}], \forall i, \forall t\}$ are independent and identically distributed (i.i.d.).

The parameters u_{it}^0, $u_{it}^h, \forall i, \forall t$ and Γ_S, Γ_T are determined according to Lemma 3.

Lemma 3. When A1-A3 hold, if the parameters of the uncertainty set (7) are chosen as

$$u_{it}^0 = \ u_{it}^e = (1-\tau_{it})\bar{u}_{it} + \tau_{it}u_{it}^{pre}, \ \forall i, \forall t, \tag{8a}$$

$$\begin{aligned}u_{it}^h = & \sqrt{\mathrm{var}[U_{it}|U_{it}^{pre} = u_{it}^{pre}]/(1-\delta)} \\ = & \sqrt{\left[(1-\tau_{it})^2\sigma_{U_i}^2 + \tau_{it}^2\sigma_{\varepsilon_i}^2\right]/(1-\delta)}, \ \forall i, \forall t,\end{aligned} \tag{8b}$$

$$\Gamma_S = \sqrt{\frac{I(1-\delta)(1+I-I\xi)}{1-\xi}}, \tag{8c}$$

$$\Gamma_T = \sqrt{\frac{T(1-\delta)(1+T-T\xi)}{1-\xi}}. \tag{8d}$$

Then, we can ensure that $\mathcal{P}(v_{it} \ge 1) \le 1-\delta, \forall i, \forall t$, $\mathcal{P}\left(\sum_i v_{it} \ge \Gamma_S\right) \le 1-\xi, \forall t$ and $\mathcal{P}\left(\sum_t v_{it} \ge \Gamma_T\right) \le 1-\xi, \forall i$.

The proof of Lemma 3 can be found in Appendix C. The improved uncertainty set (7) depends on the prediction accuracy $\tau_{it}, \forall t$ which is influenced by the payment of operator to agent i for buying the agent's prediction (denoted by $C_i, \forall i$). The higher the payment, the more accurate the distributed information. The relationship between the operator's payment and the prediction accuracy will be given in the next section. The improved uncertainty set is impacted by the payment $C_i, \forall i$ determined in the first-stage, and thus, is decision-dependent.

III. Two-Stage Robust Scheduling With an Improved Uncertainty Set

In this section, we first introduce the general model of a two-stage robust scheduling problem considering the purchase and utilization of distributed predictive information. Then an example of virtual power plant is given.

A. General Model

To derive the relationship between operator's payment C_i and the distributed prediction accuracy τ_{it}, first, we quantify the prediction cost for distributed agents as

$$h_i(\tau_i) = \sum_{t=1}^{T} \frac{m}{\sigma^2_{U_{it}}} \frac{\tau_{it}}{1-\tau_{it}}, \tag{9}$$

where m is a given cost parameter. For notation conciseness, we assume that $\tau_{i1} = \tau_{i2} = ... = \tau_{iT} =: \tau_i$ and $\sigma^2_{U_{i1}} = \sigma^2_{U_{i2}} = ... = \sigma^2_{U_{iT}} =: \sigma^2_{U_i}$. But it is worth noting that the proposed model and algorithm can also be applied to the cases with heterogeneous $\tau_{it}, \forall t$ and $\sigma^2_{U_{it}}, \forall t$. Let $\hat{m} = Tm$, then

$$h_i(\tau_i) = \frac{\hat{m}}{\sigma^2_{U_i}} \frac{\tau_i}{1-\tau_i}. \tag{10}$$

The function $h_i(\tau_i)$ grows with an increasing τ_i, meaning that a higher accuracy will cost more. To ensure that the agents are willing to provide their information, the payment should be able to cover their prediction cost, i.e.,

$$C_i \geq h_i(\tau_i) = \frac{\hat{m}}{\sigma^2_{U_i}} \frac{\tau_i}{1-\tau_i}, \forall i \in \mathcal{I}. \tag{11}$$

Definition 2. (Value of distributed information) The operator's payment C_i to agent i can be formally defined as the value of distributed information from agent i.

The total payment for buying distributed predictive information is also a cost of the operator in the first stage. Therefore, the two-stage RO model considering purchase and utilization of distributed information can be formulated as

$$\begin{aligned} &\min_{x} \left\{ f(x) + \sum_{i \in \mathcal{I}} C_i + \max_{u \in \mathcal{U}(\tau)} \min_{y \in \mathcal{Y}(x,u)} g(y) \right\}, \qquad (12)\\ &\text{s.t. } x \in \mathcal{X} \cap \tilde{\mathcal{X}}_R, \\ &\quad (11), \forall i \in \mathcal{I}, \\ &\quad 0 \leq \tau_i \leq 1, \forall i \in \mathcal{I}, \end{aligned}$$

where

$$\tilde{\mathcal{X}}_R = \{x \mid \mathcal{Y}(x,u) \neq \emptyset, \forall u \in \mathcal{U}(\tau)\}. \tag{13}$$

Since the focus of this paper is the operator's decision-making in two-stage robust scheduling, the detailed information market design between the operator and distributed agents will be left for future study. The model (12) is an RO with DDU. The traditional algorithms for solving an RO with decision-independent uncertainty (DIU), such as the Benders decomposition and the C&CG, cannot be directly applied since it may fail to converge or lead to suboptimal solutions. In Section IV, an adaptive C&CG algorithm is developed with proof of convergence. Before that, we first use the energy scheduling problem in a virtual power plant as an example to illustrate how the proposed model (12) looks like in applications.

B. Example: Virtual Power Plant Scheduling

Suppose there are J controllable generators indexed by $j \in \mathcal{J} = \{1,...,J\}$ and I prosumers indexed by $i \in \mathcal{I} = \{1,...,I\}$ in a virtual power plant. The prosumers can have loads and/or renewable generators, whose net demand $u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}$ are uncertain. In the first-stage (pre-dispatch stage), the VPP operator will decide on the reference output and reserve capacity of the controllable generators, the buying/selling energy from/to the main grid, and the payment for buying distributed information. In the second-stage stage (re-dispatch stage), knowing the exact net demand, the VPP operator can adjust the power output of controllable generators within their reserve capacity or buy/sell from/to the main grid to maintain power balance. According to the above analysis, the VPP's robust scheduling problem can be formulated as

$$\begin{aligned} \min_{\substack{p,r,C,\tau\\E^{bf},E^{sf}}} \quad & \underbrace{\sum_t \left(\sum_j (\rho_j p_{jt} + \gamma_j r_{jt}) + (\lambda^{bf}_t E^{bf}_t - \lambda^{sf}_t E^{sf}_t) \right)}_{f(x)} + \sum_i C_i \\ & + \max_{u \in \mathcal{U}(\tau)} \min_{\substack{\Delta p, E^{bs}, E^{ss}\\ \in \mathcal{Y}(p,r,u)}} \underbrace{\sum_t \left(\sum_j (\rho_j \Delta p_{jt}) + \lambda^{bs}_t E^{bs}_t - \lambda^{ss}_t E^{ss}_t \right)}_{g(y)}, \end{aligned} \tag{14a}$$

$$\text{s.t. } (p, r, E^{bf}, E^{sf}) \in \mathcal{X} \cap \tilde{\mathcal{X}}_R, \tag{14b}$$

$$C_i \geq \frac{m}{\sigma^2_{U_i}} \frac{\tau_i}{1-\tau_i}, \; 0 \leq \tau_i \leq 1, \; \forall i, \tag{14c}$$

where

$$\mathcal{X} = \Big\{ (p, r, E^{bf}, E^{sf}) \mid$$

$$\sum_j p_{jt} + E^{bf}_t - E^{sf}_t = \sum_i u^e_{it}, \forall t, \tag{15a}$$

$$0 \leq r_{jt} \leq \mathcal{R}^{max}_j, \forall j, \forall t, \tag{15b}$$

$$P^{min}_j + r_{jt} \leq p_{jt} \leq P^{max}_j - r_{jt}, \forall j, \forall t, \tag{15c}$$

$$\begin{aligned} &-\mathcal{R}^{max}_j \leq (p_{jt} + r_{jt}) \\ &-(p_{j(t-1)} - r_{j(t-1)}) \leq \mathcal{R}^{max}_j, \forall j, \forall t = 2, ..., T, \end{aligned} \tag{15d}$$

$$\begin{aligned} &-\mathcal{R}^{max}_j \leq (p_{jt} - r_{jt}) \\ &-(p_{j(t-1)} + r_{j(t-1)}) \leq \mathcal{R}^{max}_j, \forall j, \forall t = 2, ..., T, \end{aligned} \tag{15e}$$

$$0 \leq E^{bf}_t, E^{sf}_t \leq E^{max}, \forall t, \tag{15f}$$

$$-F_l \leq \sum_j \pi_{jl} p_{jt} + \pi_{gl}(E^{bf}_t - E^{sf}_t) - \sum_i \pi_{il} u^e_{it} \leq F_l, \forall l, \forall t \Big\}, \tag{15g}$$

and

$$\mathcal{Y}(p, r, u) = \Big\{ (\Delta p, E^{bs}, E^{ss}) \mid$$

$$-r_{jt} \leq \Delta p_{jt} \leq r_{jt}, \forall j, \forall t, \tag{16a}$$

$$\sum_j (p_{jt} + \Delta p_{jt}) + E^{bf}_t - E^{sf}_t + E^{bs}_t - E^{ss}_t = \sum_i u_{it}, \forall t, \tag{16b}$$

$$0 \leq E^{bs}_t, E^{ss}_t \leq E^{max}, \forall t, \tag{16c}$$

$$\begin{aligned} &-F_l \leq \sum_j \pi_{jl}(p_{jt} + \Delta p_{jt}) - \sum_i \pi_{il} u_{it} \\ &+ \pi_{gl}(E^{bf}_t - E^{sf}_t + E^{bs}_t - E^{ss}_t) \leq F_l, \forall l, \forall t \Big\}. \end{aligned} \tag{16d}$$

The first-stage decision variable x consists of the reference output $\{p_{jt}, \forall j, \forall t\}$ and reserve capacity $\{r_{jt}, \forall j, \forall t\}$ of controllable generators, the buying energy E_t^{bf} (selling energy E_t^{sf}) from (to) the main grid, and the payment $C_i, \forall i$ for buying information from the prosumers. The second-stage decision variable y includes the power output adjustment $\{\Delta p_{jt}, \forall j, \forall t\}$ of controllable units and the buying energy E_t^{bs} (selling energy E_t^{ss}) from (to) the main grid. ρ_j and γ_j are the power output and reserve cost coefficients, respectively; λ_t^{bf} and λ_t^{sf} are the day-ahead buying and selling electricity prices, respectively, while λ_t^{bs} and λ_t^{ss} are those for real-time. The objective function (14a) minimizes the total cost under the worst-case scenario, including the generator operation cost, electricity bills, and information payment in the first-stage and the adjustment cost and electricity bills in the second-stage.

Constraints (15a) and (16b) are the power balance conditions. The reserve capacity should not exceed the ramping limit $\mathcal{R}_j^{max}$ as in (15b). The upper/lower power limits of controllable generators considering reserve requirements are given in (15c), where P_j^{min}/P_j^{max} are the minimum/maximum power output. (15d)-(15e) ensure the satisfaction of ramping limits when offering reserves. The information payment should be able to cover the prediction cost as in (14c). Constraints (15f) and (16c) limit the buying/selling energy from/to the main grid in day-ahead and real-time stages, respectively, where E^{max} is the upper bound. The network capacity limits are imposed in (15g) and (16d); F_l is the power flow limit of line l and $\pi_{il}, \pi_{jl}, \pi_{gl}$ are the line flow distribution factors. Constraints (16a) ensure that the power adjustment be within the reserve capacity. In Section V, we use this example for illustrating the effectiveness and scalability of the proposed method.

IV. Solution Algorithm

In this section, an adaptive C&CG algorithm is developed to solve the RO with DDU. Notice that the re-dispatch problem (16) is a linear program, $g(y)$ and $\mathcal{Y}(x,u)$ can be expressed by

$$g(y) = c^\top y, \tag{17}$$

$$\mathcal{Y}(x,u) = \{y \in \mathbb{R}^{n_y} \mid Ax + By + Du \le q\}. \tag{18}$$

A. Second-Stage Problem Transformation

Given the first-stage decision $x \in \mathcal{X}$ and $\tau_i \in [0,1], \forall i$, the second-stage problem is a bilevel optimization:

$$S(x,\tau) = \max_{u \in \mathcal{U}(\tau)} \min_{y \in \mathcal{Y}(x,u)} c^\top y, \tag{19}$$

which is equivalent to the sub-problem (**SP**) (20) by converting the inner "min" problem to its KKT condition.

$$\begin{aligned} \mathbf{SP}: \max_{u \in \mathcal{U}(\tau), y, \pi} \;& c^\top y, && (20a)\\ \text{s.t. } & Ax + By + Du \le q, && (20b)\\ & B^\top \pi = c, && (20c)\\ & 0 \le -\pi \perp -(Ax + By + Du) + q \ge 0. && (20d)\end{aligned}$$

The complementary slackness condition (20d) can be linearized by the Big-M method [25].

Furthermore, for a given first-stage decision (x, τ), the problem (19) may be infeasible. Remember that we need to ensure x is robust feasible ($x \in \tilde{\mathcal{X}}_R$), to this end, we construct the following relaxed problem for checking feasibility.

$$\begin{aligned} F(x,\tau) = & \max_{u \in \mathcal{U}(\tau)} \min_{y,s} 1^\top s, && (21a)\\ & \text{s.t. } Ax + By + Du - s \le q, s \ge 0. && (21b)\end{aligned}$$

Obviously, the relaxed problem (21) is always feasible. Moreover, we have the original problem (19) is feasible if and only if s is an all zero vector at optimum. Similarly, the relaxed problem (21) is equivalent to

$$\begin{aligned} \mathbf{FC}: \max_{\substack{u \in \mathcal{U}(\tau)\\ y,s,\pi,\mu}} \;& 1^\top s, && (22a)\\ \text{s.t. } & Ax + By + Du - s \le q, s \ge 0, && (22b)\\ & B^\top \pi = 0, && (22c)\\ & -\pi + \mu = 1, && (22d)\\ & 0 \le -\pi \perp -(Ax + By + Du - s) + q \ge 0, && (22e)\\ & 0 \le \mu \perp s \ge 0. && (22f)\end{aligned}$$

We call (22) the feasibility-check (**FC**) problem.

Given a candidate first-stage decision, we first solve the **FC** problem to check whether $x \in \tilde{\mathcal{X}}_R$. If not, a feasibility cut will be returned; otherwise, we continue to solve the **SP** problem to identify an optimality cut.

Lemma 4. Suppose u^* is the optimal solution of **SP** or **FC**, then u^* can be reached at a vertex of $\mathcal{U}(\tau)$.

The proof of Lemma 4 is similar to that in [26]. To find the vertex set of $\mathcal{U}(\tau)$, first $\mathcal{U}(\tau)$ can be rewritten as

$$\begin{aligned} \mathcal{U}(\tau) = \; & \{u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ & u_{it} = u_{it}^e(\tau_i) + u_i^h(\tau_i)\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \\ & \{\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}\} \in \Phi\},\end{aligned}$$

where

$$\begin{aligned} \Phi := \; & \{\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ & -1 \le \phi_{it} \le 1, \phi_{it} \le \psi_{it}, -\phi_{it} \le \psi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \\ & \sum_i \psi_{it} \le \Gamma_S, \forall t \in \mathcal{T}, \sum_t \psi_{it} \le \Gamma_T, \forall i \in \mathcal{I}\}.\end{aligned}$$

In fact, ϕ_{it} represents $(u_{it} - u_{it}^e)/u_{it}^h$ and ψ_{it} represents $|u_{it} - u_{it}^e|/u_{it}^h$. Φ is a polyhedron. Denote the vertex set of Φ by $V(\Phi)$, then the vertex set of $\mathcal{U}(\tau)$ can be represented as

$$\begin{aligned} V(\mathcal{U}(\tau)) = \; & \{u_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T} \mid \\ & u_{it} = u_{it}^e(\tau_i) + u_i^h(\tau_i)\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \\ & \{\phi_{it}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}\} \in V(\Phi)\}.\end{aligned}$$

The traditional algorithms return the worst-case scenario $\{u_{it}^*, \forall i, \forall t\}$ directly to the master problem to generate a feasibility/optimality cut. However, when facing the RO with DDU, a previously selected scenario may no longer be a vertex of the new uncertainty set when the first-stage decision changes. This is the root of possible failure of traditional algorithms. To tackle this problem, instead of returning the scenario $u_{it}^*, \forall i, \forall t$ directly, we return the following active constraints:

$$u_{it} = u_{it}^e(\tau_i) + u_i^h(\tau_i)\phi_{it}^*, \forall i, \forall t. \tag{23}$$

Here, ϕ_{it}^* is the value of ϕ_{it} corresponding to the scenario u_{it}^*; u_{it}, $u_{it}^e(\tau_i)$, and $u_i^h(\tau_i)$ are all variables. An illustrative example is given in Fig. 2 to show how (23) works.

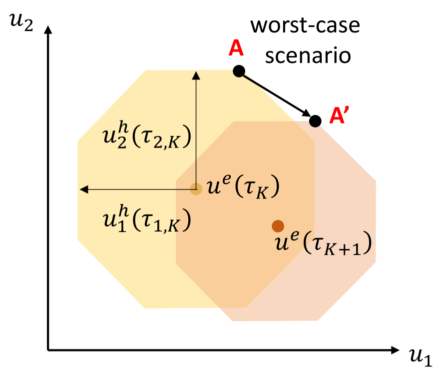

Fig. 2. Illustration of (23). The yellow and orange regions are the uncertainty sets in the K and $K+1$ iterations, respectively. Point A is the worst-case scenario in the K iteration. When u^e and u^h changes with τ, (23) moves point A to point A′, which is a vertex of the new uncertainty set.

B. Adaptive C&CG Algorithm

After introducing the second-stage problem transformation and the active constraints returned, the master problem (**MP**) can be formulated as

$$\textbf{MP}: \min_{x,C,\tau,\zeta,y^k,u^k} \; f(x) + \sum_i C_i + \zeta, \tag{24a}$$

$$\text{s.t. } x \in \mathcal{X}, \tau_i \in [0,1], \forall i \in \mathcal{I}, \tag{24b}$$

$$(11), \forall i \in \mathcal{I}, \tag{24c}$$

$$\zeta \geq c^\top y^k, \forall k \in [K], \tag{24d}$$

$$Ax + By^k + Du^k \leq q, \forall k \in [K], \tag{24e}$$

$$u^k = u^e(\tau) + u^h(\tau)\phi_k, \forall k \in [K], \tag{24f}$$

where symbols with superscript k are variables while symbols with subscript k are given parameters. $[K]$ represents all positive integers not exceeding K.

The overall procedure is given in Algorithm 1.

The adaptive C&CG algorithm is different from the traditional C&CG algorithm [16] because it returns the active constraints instead of the worst-case scenarios $\{u_{it}^*, \forall i, \forall t\}$ to the master problem to generate new cuts. To be specific, for the traditional C&CG algorithm, the step 6 in Algorithm 1 is replaced by: "6: Create variables y^{K+1} and add the following constraints to **MP** (24):

$$\zeta \geq c^\top y^{K+1},$$
$$Ax + By^{K+1} + Cu_{K+1}^* \leq q.$$

Update $K = K+1$ and go to Step 2." Note that u_{K+1}^* is a constant vector obtained by solving **FC** (22) or **SP** (20).

Algorithm 1: Adaptive C&CG Algorithm

1: **Initiation**: Error tolerance $\varepsilon > 0$; $K = 0$; $UB_K = +\infty$.

2: **Solve the Master Problem**
Solve the **MP** (24). Derive the optimal solution $(x_{K+1}^*, C_{K+1}^* \tau_{K+1}^*, \zeta_{K+1}^*, y^{1*}, ..., y^{K*}, u^{1*}, ..., u^{K*})$ and update $LB_{K+1} = f(x_{K+1}^*) + \sum_i C_{i,K+1}^* + \zeta_{K+1}^*$.

3: **Solve the Feasibility-check Problem**
Solve the **FC** (22) with $(x_{K+1}^*, \tau_{K+1}^*)$.
Let $(u_{K+1}^*, \phi_{K+1}^*, \pi_{K+1}^*, \mu_{K+1}^*, y_{K+1}^*, s_{K+1}^*)$ be the optimal solution. If $1^\top s_{K+1}^* > 0$, let $UB_{K+1} = UB_K$ and go to Step 6. Otherwise, go to Step 4.

4: **Solve the Sub-problem**
Solve the **SP** (20) with $(x_{K+1}^*, \tau_{K+1}^*)$. Denote the optimal solution by $(u_{K+1}^*, \phi_{K+1}^*, \pi_{K+1}^*, y_{K+1}^*)$. Let

$$UB_{K+1} = f(x_{K+1}^*) + \sum_i C_{i,K+1}^* + c^\top y_{K+1}^*$$

5: If $|UB_{K+1} - LB_{K+1}| \leq \varepsilon$, terminate and output $(x_{K+1}^*, \tau_{K+1}^*)$. Otherwise, go to Step 6.

6: Create variables (y^{K+1}, u^{K+1}) and add the following constraints to **MP** (24):

$$\zeta \geq c^\top y^{K+1},$$
$$Ax + By^{K+1} + Cu^{K+1} \leq q,$$
$$u^{K+1} = u^e(\tau) + u^h(\tau)\phi_{K+1}^*.$$

Update $K = K+1$ and go to Step 2.

Theorem 1. Let $n_U := |V(\Phi)|$ be the number of extreme points of Φ, which is also the maximum number of vertex of $\mathcal{U}(\tau)$. The adaptive C&CG Algorithm generates the optimal solution to problem (12) within $\mathcal{O}(n_U)$ iterations.

The proof of Theorem 1 can be found in Appendix D. While the traditional algorithms fail to guarantee finite-step convergence and optimality of the obtained strategy, the proposed adaptive C&CG algorithm can overcome these limitations.

C. Transformation and Linearization

The master problem (24) is highly nonlinear due to the term $u^e(\tau)$, $u^h(\tau)$, and the information payment constraint (11). In the following, we show how to turn it into a solvable form.

First, it is easy to prove that at the robust optimum, we have $C_i = h_i(\tau_i)$. Otherwise, if $C_i > h_i(\tau_i)$, we can always reduce C_i a little bit without changing the value of other variables, so that all constraints are still satisfied but the objective value decreases. This contradicts the definition of the robust optimum. Therefore, we can eliminate constraint (11) and replace $\sum_i C_i$ in the objective function with $\sum_i h_i(\tau_i)$.

Second, if we let $u_i^h, \forall i$ be the decision variables and use them to represent $\tau_i, \forall i$, then the prediction cost $\sum_i h_i(\tau_i)$ can be represented by

$$\sum_i h_i(\tau_i) = \sum_i \hat{m} \left(\frac{1}{(1-\delta)(u_i^h)^2} - \frac{1}{\sigma_{U_i}^2} \right) = \sum_i \tilde{h}_i(u_i^h), \tag{25}$$

where $\tilde{h}_i(u_i^h) = h_i(\tau_i)$ but u_i^h is regarded as the decision variable in $\tilde{h}_i(u_i^h)$. In fact, $\tilde{h}_i(u_i^h)$ is a convex function. Similarly, $u^e(\tau)$ can be represented by u^h, which is

$$u_{it}^e(u_i^h) = \frac{(1-\delta)(u_i^h)^2}{\sigma_{\bar{U}_i}^2}\bar{u}_{it} + \left(1 - \frac{(1-\delta)(u_i^h)^2}{\sigma_{\bar{U}_i}^2}\right) u_{it}^{pre}, \forall i, \forall t. \tag{26}$$

After the above transformations, we only need to deal with the nonlinear (nonconvex) term (26). To this end, we introduce a new variable $\tilde{u}_{it}^e, \forall i, \forall t$, use it to replace $u^e(\tau)$ in (24f), and add the following penalty function to the objective:

$$\mathcal{H}(\tilde{u}^e, u^h) = L \sum_i \sum_t \left(\tilde{u}_{it}^e - u_{it}^e(u_i^h)\right)^2, \tag{27}$$

where L is a large constant.

The above bi-variate function can be linearized by a convex combination approach [27]. In particular, the following constraints are added and the term $\sum_i \tilde{h}_i(u_i^h) + \mathcal{H}(\tilde{u}^e, u^h)$ in the objective function is replaced by χ:

$$\tilde{u}_{it}^e = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} \kappa_{n_1 n_2}^{it} \tilde{u}_{it,n_2}^e, \forall i, \forall t, \tag{28a}$$

$$u_i^h = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} \kappa_{n_1 n_2}^{it} u_{i,n_1}^h, \forall i, \forall t, \tag{28b}$$

$$\chi = \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} \kappa_{n_1 n_2}^{it} \chi_{n_1 n_2}^{it}, \tag{28c}$$

$$\kappa_{n_1 n_2}^{it} \geq 0, \forall n_1, \forall n_2, \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} \kappa_{n_1 n_2}^{it} = 1, \forall i, \forall t, \tag{28d}$$

where u_{i,n_1}^h and $\tilde{u}_{it,n_2}^e$ are sample points; $\chi_{n_1 n_2}^{it} = \tilde{h}_i(u_{i,n_1}^h)/T + L(\tilde{u}_{it,n_2}^e - u_{it}^e(u_{i,n_1}^h))^2$. The $\kappa_{n_1 n_2}^{it}$ are superposition coefficients. The sample points are selected as follows: Evenly pick sample points of $\tau_i \in (0,1)$ and then the samples of $\sigma_{\varepsilon_i}^2$, u_{it}^e, and u_i^h can be calculated according to (6), (8a) and (8b).

Up to now, the master problem **MP** has been turned into a linear programming problem that can be solved efficiently.

V. Case Studies

We first use a simple 5-bus system to verify the proposed algorithm and reveal some interesting phenomena; then, larger systems are tested to show the scalability of the proposed method. Detailed data can be found in [28]. The adaptive C&CG algorithm is implemented in MATLAB with GUROBI 9.5. All the simulations are conducted on a laptop with Intel i7-12700H processor and 16 GB RAM.

A. Benchmark

A 5-bus system with 3 controllable generators and 5 prosumers is tested. Some parameters are shown in Table I. The electricity buying price from the grid is \$35/MWh from 9am to 9pm and \$25/MWh in other periods, while the selling price to the grid is \$1/MWh lower than the buying price. The operator's original forecasts $\bar{u}_i$ and the actual value u_i of prosumer i's net demand are depicted in Fig. 3. Suppose in every period $\sigma_U^2 = [8000, 2000, 4000, 9000, 1000]$ MW^2.

TABLE I
PARAMETERS

Parameter	Value	Parameter	Value
T	24	m	2×10^4 \$·MW^2
E^{max}	210 MW	δ	0.95
ρ	[30,40,20] \$/MWh	ξ	0.95
P^{max}	[520,200,600] MW	ε	\$1
P^{min}	[156,60,180] MW	L	10^4 \$/MW^2
$\mathcal{R}^{max}$	[208,80,240] MW	N_1, N_2	30

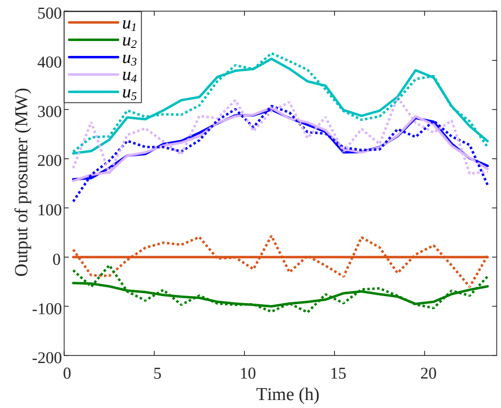

Fig. 3. Original forecasts (solid lines) and actual values (dashed lines) of prosumer net demands.

The proposed algorithm converges after 10 iterations in 244 s, which is acceptable for day-ahead scheduling. At the optimal point, we have $\mathcal{H}(\tilde{u}^e, u^h) < \$0.1$, meaning that the linearization approximation is highly accurate with $\tilde{u}_{it}^e$ nearly equals $u_{it}^e(u_i^h)$ for all i. The operation cost of controllable generator is $\$5.64\times10^5$, the revenue from the main grid is $\$0.62\times10^5$, while the prediction payments C_i for 5 prosumers are $[7.99, 7.52, 8.87, 7.10, 0.00]\times10^3$ (\$), respectively. Hence, the total cost under the worst-case scenario is $\$5.33\times10^5$. We also test the performance of the obtained robust strategy when dealing with the actual prosumer net demands. The obtained strategy is still feasible while the total cost is a bit lower, i.e., $\$4.31\times10^5$. The reference, worst-case, and actual power outputs of controllable generators are shown in Fig. 4.

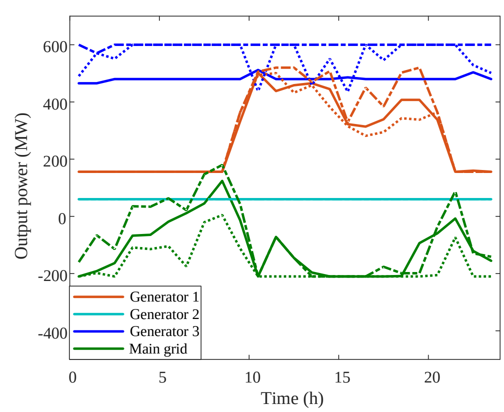

Fig. 4. The reference (solid lines), worst-case (dash-dotted lines), and actual (dashed lines) power outputs of controllable generators and power purchased from the main grid.

To visualize the impact of distributed predictive information on forecasting, the original (green area) and improved (blue area) uncertainty sets of prosumers 3 and 4 are shown in Fig. 5. Both sets have a confidence probability $\delta = 0.95$ in each period. The centers of the original uncertainty sets are the original forecasts $\bar{u}_{it}, \forall i, \forall t$, respectively. With the help of distributed information, the centers become the best linear predictors $u^e_{it}, \forall i, \forall t$, which are closer to the actual uncertainty realizations. Both uncertainty sets contain the actual prosumer net demand, but the improved set is much narrower so the operator is facing less uncertainty. The original forecasts of prosumer 3 and 4 are the same, but the uncertainty variance of prosumer 4 is larger, so the width of prosumer 4's original uncertainty set is larger ($u^h_3 = 134$ MW) than that of prosumer 3 ($u^h_4 = 89$ MW). The widths of their improved sets are similar with $u^h_3 = 31$ MW and $u^h_4 = 35$ MW. Moreover, Fig. 6 shows how the uncertainty sets narrow as the prediction payments increase. Note that Fig. 6 is a semi-log plot and the prediction payment grows very quickly when the width of uncertainty set is small. This is because the marginal prediction cost increases with a higher accuracy. In other words, it costs more to improve the accuracy of an already quite accurate prediction. The optimal payments and the corresponding widths $2u^{h*}_i, \forall i$ are also marked in Fig. 6. We can find that the values of $u^{h*}_i, \forall i$ are similar, which is due to the equal incremental principle, i.e., at the optimum $\partial C_i / \partial u^h_i, \forall i$ are equal, and (25). The subtle difference between $u^{h*}_i, \forall i$ is because of the linearization approximation errors.

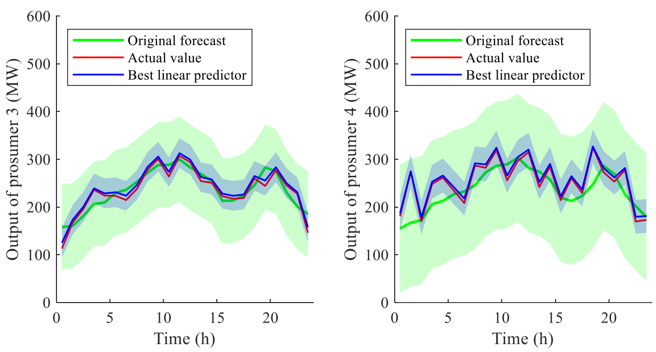

Fig. 5. Original and improved uncertainty sets of prosumers 3 and 4.

B. Sensitivity Analysis

In the following, we investigate the impacts of three different factors: the prosumer's prediction cost coefficient m, the probability parameters δ, ξ of the uncertainty set and the variance of the uncertain factor σ^2_U.

1) Impact of Prediction Cost Coefficient: First, we test how the strategy of operator changes with a rising prosumer prediction cost by changing m from 0 to 5×10^5 \$·MW^2. The total costs, operation costs, and the prediction payments under different m are shown in Fig. 7. The change of prediction accuracy τ and the width of the improved uncertainty set are given in Fig. 8. We can find that when $m = 0$, the prosumers' prediction payments are zero since the operator

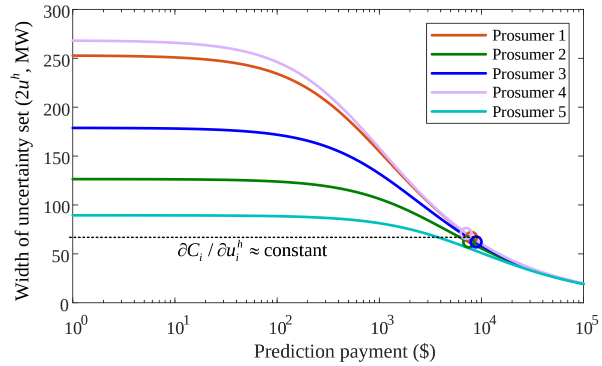

Fig. 6. Width of uncertainty sets under different prediction payments; circles represent the results at the optimum (the optimal payment for prosumer 5 is zero, which is outside of this figure).

can pay nothing to know the exact value of u, and thus, there is no uncertainty ($\tau_i = 1, \forall i$). When m is very large, e.g., 5×10^5 \$·MW^2, the prosumers' prediction costs are extremely high, so the operator cannot afford to purchase predictions from the prosumers. Therefore, as shown in Fig. 8, the final prediction accuracy τ_i is zero for each prosumer i and the uncertainty sets have the largest width. As m grows, from Fig. 7, the operation cost and the total cost are always less than the cost of the traditional model (1) without distributed predictive information. The lower the m, the higher the operation cost reduction, which shows the potential of our model. According to Definition 2, the prediction payment C_i can be interpreted as the value of distributed information. This value is influenced by prosumer's prediction cost coefficient m and the system parameters. From Fig. 7, the value of distributed information of all prosumers follow a similar trend (first increases and then declines) and the peak value of the prosumer with a larger uncertainty variance $\sigma^2_{U_i}$ is higher. This indicates that distributed predictive information will play an increasingly important role in future power systems with higher uncertainties. In Fig. 8, as m increases, the prediction accuracy τ at the optimum declines while the widths of the uncertainty sets increase. When m is extremely large, both of them reach the same value as if there were no distributed predictive information.

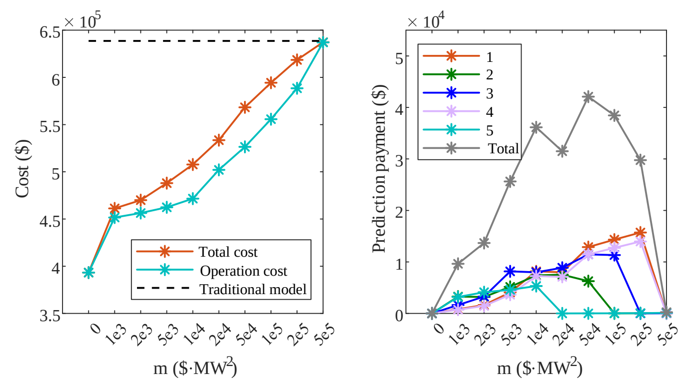

Fig. 7. Costs and prediction payments under different m.

2) Impact of Uncertainty Set Probability Parameters: We next change the probability parameters δ and ξ simultane-

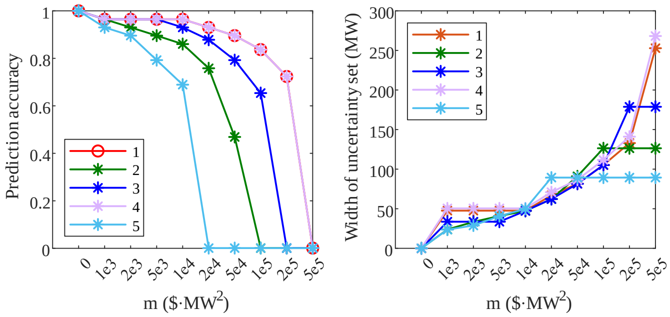

Fig. 8. Prediction accuracy and width of uncertainty set under different m.

ously, i.e., keeping $\delta = \xi$. The costs as well as the widths of uncertainty sets of prosumer 3 and 4 are shown in Fig. 9. When δ and ξ increase, the uncertainty sets expand, giving a more robust optimal scheduling strategy but also resulting in higher total and operation costs. Moreover, the total cost of the proposed model is always less than that of the traditional model without distributed information. The improved uncertainty sets are much narrower than the original uncertainty set derived by $\bar{u}$ and σ_U^2.The widths of the improved uncertainty sets of prosumers 3 and 4 are similar under different δ and ξ due to the same reason as in Section V-A.

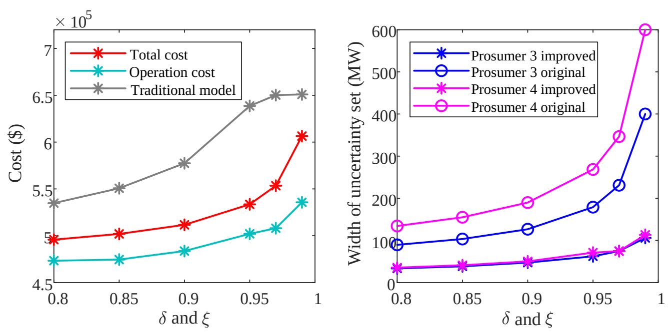

Fig. 9. Costs and widths of uncertainty sets under different δ and ξ.

3) Impact of Uncertainty Variance: We further investigate the impact of σ_U^2, which is the variance of the uncertain factor (or the operator's original estimate). To do this, we multiply σ_U^2 by a positive constant. The original forecasts and actual values are still the same as those in Fig. 3. The cost and prediction accuracy $\tau_i, \forall i$ are shown in Fig. 10. When the variance is small, the operator already has a relatively good original estimate, so they tend to pay less for distributed information. As the variance grows, the advantage of using distributed information becomes more significant. When the variance is large, at the optimum, the prediction accuracy is close to 1 for every prosumer, meaning that the operator relies on the distributed predictive information to make decisions.

C. Comparison With the Traditional C&CG Algorithm

To show the necessity of the adaptive C&CG algorithm, we compare it with the traditional C&CG algorithm [16] using the benchmark case. The iteration processes are depicted in Fig. 11. The adaptive C&CG algorithm converges to the optimal

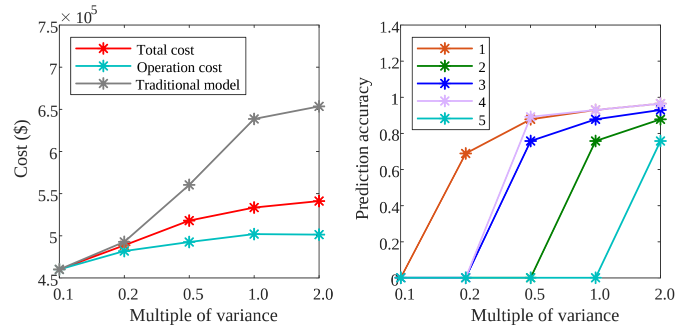

Fig. 10. Cost and prediction accuracy τ under different multiples of variance.

solution introduced in Section V-A. The traditional algorithm stops in 12 iterations with the optimal objective value equals $\$6.39 \times 10^5$, which is much higher than that of the proposed algorithm ($\$5.64 \times 10^5$). This is because a previously added worst-case scenario may lie outside of the uncertainty set when the first-stage decision changes. Therefore, the master problem is no longer a relaxation of the robust optimization, which may lead to over-conservative results. Moreover, the previous scenarios that are outside of the uncertainty set hinders the improvement using distributed information. At the optimum of the traditional C&CG, we can find that the prediction payments are zero. Given the above reasons, the proposed algorithm is necessary to deal with decision-dependent uncertainty.

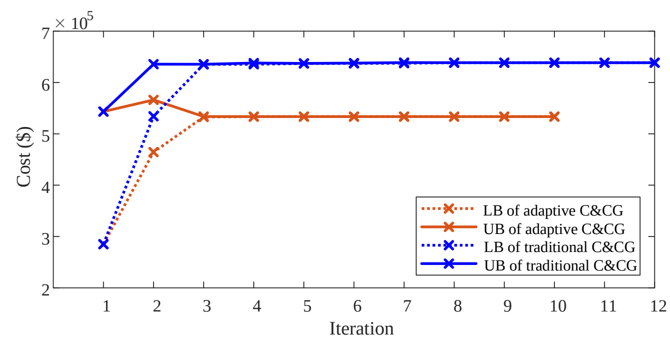

Fig. 11. Iteration processes of the proposed and traditional C&CG algorithms.

D. Out-of-Sample Test

To analyze the robustness of the obtained strategy, out-of-sample tests are conducted. To be specific, we randomly generate scenarios from a uniform and a Gaussian distribution with the same expectation u and standard deviation, respectively. We change the standard deviation via multiplying $\sqrt{\text{var}(U|U^{pre})}$ by a constant from 1.0 to 1.8 and test whether the obtained robust strategy can ensure a feasible second-stage strategy y under the selected scenarios. Ten thousand (10000) scenarios are tested for each setting. The percentage of infeasible scenarios of the proposed algorithm and the traditional C&CG are given in Table II. We can find that the proposed algorithm has a better performance in terms of feasibility thanks to the purchase and use of distributed predictive information.

TABLE II
OUT-OF-SAMPLE TEST OF THE PROPOSED ALGORITHM AND THE TRADITIONAL C&CG: INFEASIBLE RATE (%)

Multiple of standard deviation		1.0	1.2	1.4	1.6	1.8
Uniform	Proposed	0.00	0.01	0.18	0.58	1.74
	Traditional	3.52	7.22	9.95	13.84	16.39
Gaussian	Proposed	0.01	0.07	0.34	0.97	2.27
	Traditional	3.95	6.82	9.74	13.16	15.02

E. Scalability

To show the scalability of the proposed algorithm, larger systems are tested. The computational time and the number of iterations under different settings are recorded in Table III. The time needed are all less than 2 h, which is acceptable for the day-ahead scheduling.

TABLE III
NUMBER OF ITERATIONS/COMPUTATIONAL TIME UNDER DIFFERENT SETTINGS

No. of prosumers	4	8	16	24
33-bus	7 / 491 s	8 / 531 s	14 / 1549 s	14 / 2293 s
69-bus	4 / 282 s	6 / 468 s	6 / 1275 s	8 / 4080 s
123-bus	2 / 127 s	9 / 693 s	11 / 1991 s	7 / 4422 s

VI. CONCLUSION

This paper proposes a novel robust scheduling model in which the operator is able to purchase predictive information from distributed agents to obtain a more precise uncertainty set and make better decisions. The proposed model renders a two-stage RO with DDU. An adaptive C&CG algorithm is developed to solve it, which is guaranteed to converge in a finite number of iterations. Some interesting findings are:

- Compared with the traditional models without using distributed predictive information, the proposed model can help the operator greatly narrow the uncertainty set and reduce the total cost.
- The value of distributed information grows with the variance of uncertainty, indicating that it will play an increasingly important role in future power systems with more volatile renewable generation and flexible demand.
- When dealing with DDU, the proposed algorithm outperforms the traditional C&CG algorithm in terms of solution robustness and optimality.

Future research directions include a detailed information market design between the operator and distributed agents, a more efficient solution algorithm, etc.

REFERENCES

[1] Y. Parag and B. K. Sovacool, “Electricity market design for the prosumer era,” *Nature energy*, vol. 1, no. 4, pp. 1–6, 2016.

[2] H. A. Rahman, M. S. Majid, A. R. Jordehi, G. C. Kim, M. Y. Hassan, and S. O. Fadhl, “Operation and control strategies of integrated distributed energy resources: A review,” *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1412–1420, 2015.

[3] Y. Zhang, F. Liu, Z. Wang, Y. Su, W. Wang, and S. Feng, “Robust scheduling of virtual power plant under exogenous and endogenous uncertainties,” *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1311–1325, 2021.

[4] C. Wang, B. Jiao, L. Guo, Z. Tian, J. Niu, and S. Li, “Robust scheduling of building energy system under uncertainty,” *Applied energy*, vol. 167, pp. 366–376, 2016.

[5] B. Kandpal, P. Pareek, and A. Verma, “A robust day-ahead scheduling strategy for ev charging stations in unbalanced distribution grid,” *Energy*, vol. 249, p. 123737, 2022.

[6] D. Bertsimas and M. Sim, “The price of robustness,” *Operations research*, vol. 52, no. 1, pp. 35–53, 2004.

[7] D. Bertsimas and D. B. Brown, “Constructing uncertainty sets for robust linear optimization,” *Operations research*, vol. 57, no. 6, pp. 1483–1495, 2009.

[8] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, “Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 511–520, 2013.

[9] A. Lorca and X. A. Sun, “Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind,” *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1702–1713, 2014.

[10] P. Pinson, L. Han, and J. Kazempour, “Regression markets and application to energy forecasting,” *TOP*, pp. 1–41, 2022.

[11] L. Han, P. Pinson, and J. Kazempour, “Trading data for wind power forecasting: A regression market with lasso regularization,” *Electric Power Systems Research*, vol. 212, p. 108442, 2022.

[12] B. Wang, Q. Guo, and Y. Yu, “Mechanism design for data sharing: An electricity retail perspective,” *Applied Energy*, vol. 314, p. 118871, 2022.

[13] A. A. Raja, P. Pinson, J. Kazempour, and S. Grammatico, “A market for trading forecasts: A wagering mechanism,” *arXiv preprint arXiv:2205.02668*, 2022.

[14] Y. Chen, T. Li, C. Zhao, and W. Wei, “Decentralized provision of renewable predictions within a virtual power plant,” *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2652–2662, 2020.

[15] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, “Adaptive robust optimization for the security constrained unit commitment problem,” *IEEE transactions on power systems*, vol. 28, no. 1, pp. 52–63, 2012.

[16] B. Zeng and L. Zhao, “Solving two-stage robust optimization problems using a column-and-constraint generation method,” *Operations Research Letters*, vol. 41, no. 5, pp. 457–461, 2013.

[17] O. Nohadani and K. Sharma, “Optimization under decision-dependent uncertainty,” *SIAM Journal on Optimization*, vol. 28, no. 2, pp. 1773–1795, 2018.

[18] N. H. Lappas and C. E. Gounaris, “Robust optimization for decision-making under endogenous uncertainty,” *Computers & Chemical Engineering*, vol. 111, pp. 252–266, 2018.

[19] Y. Zhang, F. Liu, Z. Wang, Y. Chen, S. Feng, Q. Wu, and Y. Hou, “On nash–stackelberg–nash games under decision-dependent uncertainties: Model and equilibrium,” *Automatica*, vol. 142, p. 110401, 2022.

[20] Y. Chen and W. Wei, “Robust generation dispatch with strategic renewable power curtailment and decision-dependent uncertainty,” *arXiv preprint arXiv:2203.16251*, 2022.

[21] S. Avraamidou and E. N. Pistikopoulos, “Adjustable robust optimization through multi-parametric programming,” *Optimization Letters*, vol. 14, no. 4, pp. 873–887, 2020.

[22] W. Wei, F. Liu, J. Wang, L. Chen, S. Mei, and T. Yuan, “Robust environmental-economic dispatch incorporating wind power generation and carbon capture plants,” *Applied Energy*, vol. 183, pp. 674–684, 2016.

[23] R. Jiang, M. Zhang, G. Li, and Y. Guan, “Two-stage network constrained robust unit commitment problem,” *European Journal of Operational Research*, vol. 234, no. 3, pp. 751–762, 2014.

[24] X. Vives, *Information and learning in markets: the impact of market microstructure*. Princeton University Press, 2010.

[25] S. Pineda and J. M. Morales, “Solving linear bilevel problems using big-ms: not all that glitters is gold,” *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2469–2471, 2019.

[26] H. Konno, “A cutting plane algorithm for solving bilinear programs,” *Mathematical Programming*, vol. 11, no. 1, pp. 14–27, 1976.

[27] L. Wu, “A tighter piecewise linear approximation of quadratic cost curves for unit commitment problems,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2581–2583, 2011.

[28] R. Xie, “Information valuation,” https://github.com/xieruijx/Information-Valuation, 2022.

[29] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press, 2001.

APPENDIX A
PROOF OF LEMMA 1

Set a function

$$g(\alpha_i,\beta_i) := \mathbb{E}\left[\left(U_i-(\alpha_i+\beta_i U_i^{pre})\right)^2\right] = \alpha_i^2+\mathbb{E}[(U_i^{pre})^2]\beta_i^2+2\mathbb{E}[U_i^{pre}]\alpha_i\beta_i - 2\mathbb{E}[U_i]\alpha_i-2\mathbb{E}[U_iU_i^{pre}]\beta_i+\mathbb{E}[U_i^2]. \quad (A.1)$$

Note that ε_i is independent of U_i, then

$$\begin{aligned}
&\mathbb{E}[U_i^{pre}]=\mathbb{E}[U_i-\varepsilon_i]=\mathbb{E}[U_i]=\overline{u}_i,\\
&\mathbb{E}[U_i^2]=\mathrm{var}[U_i]+(\mathbb{E}[U_i])^2=\sigma_{U_i}^2+\overline{u}_i^2,\\
&\mathbb{E}[U_i\varepsilon_i]=\mathbb{E}[U_i]\mathbb{E}[\varepsilon_i]=0,\\
&\mathbb{E}[U_iU_i^{pre}]=\mathbb{E}[U_i^2]-\mathbb{E}[U_i\varepsilon_i]=\mathbb{E}[U_i^2]=\sigma_{U_i}^2+\overline{u}_i^2,\\
&\mathbb{E}[(U_i^{pre})^2]=\mathbb{E}[U_i^2]-2\mathbb{E}[U_i\varepsilon_i]+\mathbb{E}[\varepsilon_i^2]=\sigma_{U_i}^2+\overline{u}_i^2+\sigma_{\varepsilon_i}^2.
\end{aligned}$$

Since α_i and β_i minimizes $g(\alpha_i,\beta_i)$, we have

$$\frac{\partial g(\alpha_i,\beta_i)}{\partial\alpha_i}=0,\frac{\partial g(\alpha_i,\beta_i)}{\partial\beta_i}=0, \quad (A.2)$$

whose solution is

$$\beta_i=\frac{\sigma_{U_i}^2}{\sigma_{U_i}^2+\sigma_{\varepsilon_i}^2},\ \alpha_i=(1-\beta_i)\overline{u}_i. \quad (A.3)$$

Then

$$\mathbb{E}[U_i-U_i^e]=\overline{u}_i-\alpha_i-\beta_i\overline{u}_i=0,$$

and

$$\begin{aligned}
&\mathrm{cov}(U_i-U_i^e,U_i^{pre})\\
=\ &\mathbb{E}[(U_i-U_i^e)U_i^{pre}]-\mathbb{E}[U_i-U_i^e]\mathbb{E}[U_i^{pre}]\\
=\ &\mathbb{E}[(U_i-(\alpha_i+\beta_iU_i^{pre}))U_i^{pre}]\\
=\ &\mathbb{E}[U_iU_i^{pre}]-\alpha_i\mathbb{E}[U_i^{pre}]-\beta_i\mathbb{E}[(U_i^{pre})^2]=0.
\end{aligned}$$

This completes the proof.

APPENDIX B
PROOF OF LEMMA 2

Since $U_i^e=\alpha_i+\beta_iU_i^{pre}$ is a function of U_i^{pre}, for any random variable X with finite second moment we have $\mathbb{E}[XU_i^e|U_i^{pre}]=U_i^e\mathbb{E}[X|U_i^{pre}]$ [29, p. 348]. Therefore,

$$\begin{aligned}
&\mathrm{var}[\eta_i|U_i^{pre}]\\
=\ &\mathbb{E}[(U_i-U_i^e)^2|U_i^{pre}]-(\mathbb{E}[U_i-U_i^e|U_i^{pre}])^2\\
=\ &(\mathbb{E}[U_i^2|U_i^{pre}]-2U_i^e\mathbb{E}[U_i|U_i^{pre}]+(U_i^e)^2)\\
&-((\mathbb{E}[U_i|U_i^{pre}])^2-2U_i^e\mathbb{E}[U_i|U_i^{pre}]+(U_i^e)^2)\\
=\ &\mathbb{E}[U_i^2|U_i^{pre}]-(\mathbb{E}[U_i|U_i^{pre}])^2\\
=\ &\mathrm{var}[U_i|U_i^{pre}]. \qquad (B.1)
\end{aligned}$$

Moreover, we have η_i and U_i^{pre} are independent random variables. Therefore,

$$\begin{aligned}
\mathrm{var}[U_i|U_i^{pre}]=\ &\mathrm{var}[\eta_i|U_i^{pre}]=\mathrm{var}[\eta_i]\\
=\ &\mathrm{var}[U_i-(1-\beta_i)\overline{u}_i-\beta_iU_i^{pre}]\\
=\ &\mathrm{var}[(1-\beta_i)U_i+\beta_i\varepsilon_i]\\
=\ &(1-\beta_i)^2\sigma_{U_i}^2+\beta_i^2\sigma_{\varepsilon_i}^2. \qquad (B.2)
\end{aligned}$$

Moreover,

$$\mathrm{var}[U_i|U_i^{pre}]=\sigma_{U_i}^2\frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_i}^2+\sigma_{U_i}^2}\le\sigma_{U_i}^2. \quad (B.3)$$

This completes the proof.

APPENDIX C
PROOF OF LEMMA 3

Recall that according to Lemma 1 we have $\mathbb{E}[\eta_{it}]=\mathbb{E}[U_{it}-U_{it}^e]=0$. Combining the fact $\mathrm{var}[\eta_{it}]=\mathrm{var}[U_{it}|U_{it}^{pre}]$ shown in (B.2), we have $\mathbb{E}[\eta_{it}/u_{it}^h]=0$ and $\mathrm{var}[\eta_{it}/u_{it}^h]=1-\delta$. Then according to the Chebyshev inequality,

$$\begin{aligned}
&\mathcal{P}\left(|\eta_{it}/u_{it}^h|\ge1\right)\\
=\ &\mathcal{P}\left(|(\eta_{it}/u_{it}^h)-\mathbb{E}[\eta_{it}/u_{it}^h]|\ge\sqrt{\mathrm{var}[\eta_{it}/u_{it}^h]/(1-\delta)}\right)\\
\le\ &1/\left(\sqrt{1/(1-\delta)}\right)^2=1-\delta. \qquad (C.1)
\end{aligned}$$

Note that $v_{it}=|\eta_{it}/u_{it}^h|$. Thus, we have $\mathcal{P}(v_{it}\ge1)=\mathcal{P}(|\eta_{it}/u_{it}^h|\ge1)\le1-\delta$.

Again by the Chebyshev inequality,

$$\begin{aligned}
&\mathcal{P}\left(\sum_i v_{it}\ge\mathbb{E}[\sum_i v_{it}]+\sqrt{\mathrm{var}[\sum_i v_{it}]/(1-\xi)}\right)\\
=\ &\mathcal{P}\left(\sum_i v_{it}-\mathbb{E}[\sum_i v_{it}]\ge\sqrt{\mathrm{var}[\sum_i v_{it}]}\sqrt{1/(1-\xi)}\right)\\
\le\ &\{1/(1-\xi)\}^{-1}=1-\xi. \qquad (C.2)
\end{aligned}$$

We find an upper bound for $\mathbb{E}[\sum_i v_{it}]+\sqrt{\mathrm{var}[\sum_i v_{it}]/(1-\xi)}$. $\{\eta_{it}/\mathrm{var}[\eta_{it}],\forall i,\forall t\}$ are independent and identically distributed (i.i.d.), so $\{v_{it},\forall i,\forall t\}$ are also i.i.d. random variables. Therefore, $\mathbb{E}[\sum_i v_{it}]=I\mathbb{E}[v_{it}]$ and $\mathrm{var}[\sum_i v_{it}]=I\mathrm{var}[v_{it}]$. Moreover,

$$\begin{aligned}
\mathrm{var}[v_{it}]=\ &\mathbb{E}[v_{it}^2]-(\mathbb{E}[v_{it}])^2\\
=\ &\mathbb{E}[(\eta_{it}/u_{it}^h)^2]-(\mathbb{E}[v_{it}])^2\\
=\ &\mathrm{var}[\eta_{it}/u_{it}^h]+(\mathbb{E}[\eta_{it}/u_{it}^h])^2-(\mathbb{E}[v_{it}])^2\\
=\ &1-\delta-(\mathbb{E}[v_{it}])^2. \qquad (C.3)
\end{aligned}$$

Then

$$\begin{aligned}
&\mathbb{E}[\sum_i v_{it}]+\sqrt{\mathrm{var}[\sum_i v_{it}]/(1-\xi)}\\
=\ &I\mathbb{E}[v_{it}]+\sqrt{I(1-\delta-(\mathbb{E}[v_{it}])^2)/(1-\xi)}\ =:G(\mathbb{E}[v_{it}])
\end{aligned}$$

is a function of $\mathbb{E}[v_{it}]$ for $0\le\mathbb{E}[v_{it}]\le\sqrt{1-\delta}$. By calculating its derivative, it is easy to show that $G(\mathbb{E}[v_{it}])$ first increases and then declines, whose unique maximum value is Γ_S. Therefore, we have $\mathcal{P}(\sum_i v_{it}\ge\Gamma_S)\le1-\xi,\forall t$. Similarly, we can prove that $\mathcal{P}(\sum_t v_{it}\ge\Gamma_T)\le1-\xi,\forall i$.

APPENDIX D
PROOF OF THEOREM 1

Suppose the optimal solution of the two-stage RO model (12) is (x^*,τ^*) and the optimal objective value is

$$O^*:=\min_{\substack{x\in\mathcal{X}\cap\tilde{\mathcal{X}}_R\\ \tau_i\in[0,1],(11),\forall i\in\mathcal{I}}} f(x)+\sum_{i\in\mathcal{I}}C_i(\tau_i)+\max_{u\in\mathcal{U}(\tau)}\min_{y\in\mathcal{Y}(x,u)}g(y).$$

We start the proof of Theorem 1 by providing the following claims. For any $K \in \mathbb{Z}^+$:

(1) **Claim 1**: $LB_K \leq O^* \leq UB_K$;

(2) **Claim 2**: If the algorithm does not terminate after K iterations, then for any $K_1, K_2 \in [K]$, we have $\phi^*_{K_1} \neq \phi^*_{K_2}$.

Proofs of claims:

(1) **Claim 1**: The master problem in the K-th iteration is equivalent to

$$LB_K = \min_{\substack{x \in \mathcal{X} \cap \mathcal{X}_{K-1} \\ \tau_i \in [0,1], (11), \forall i \in \mathcal{I}}} f(x) + \sum_{i \in \mathcal{I}} C_i(\tau_i) + \max_{u \in \mathcal{U}_{K-1}(\tau)} \min_{y \in \mathcal{Y}(x,u)} g(y),$$

where

$$\mathcal{U}_{K-1}(\tau) := \{u^k = u^e(\tau) + u^h(\tau)\phi^*_k \mid k \in [K-1]\}, \quad \text{(D.1)}$$

$$\mathcal{X}_{K-1} := \{x \mid \mathcal{Y}(x,u) \neq \emptyset, \forall u \in \mathcal{U}_{K-1}(\tau)\}. \quad \text{(D.2)}$$

Since $\mathcal{U}_{k-1}(\tau) \subset \mathcal{U}(\tau)$ and $\mathcal{X}_{K-1} \supset \tilde{\mathcal{X}}_R$, we have $LB_K \leq O^*$.

Next, we prove $UB_K \geq O^*$ by induction. First of all, $UB_0 = +\infty \geq O^*$. Suppose for the sake of induction that $UB_{K-1} \geq O^*$, then if $1^\top s^*_K > 0$, we have $UB_K = UB_{K-1} \geq O^*$; otherwise, (x^*_K, τ^*_K) is robust feasible and

$$\begin{aligned} UB_K = &\ f(x^*_K) + \sum_i C^*_{i,K} + g(y^*_K) \\ = &\ f(x^*_K) + \sum_i C^*_{i,K} + \max_{u \in \mathcal{U}(\tau^*_K)} \min_{y \in \mathcal{Y}(x^*_K, u)} g(y) \\ \geq &\ O^*. \end{aligned}$$

The last inequality is due to the optimality of (x^*, τ^*).

(2) **Claim 2**: Without loss of generality, we assume that $K_1 < K_2$. If we have $\phi^*_{K_1} = \phi^*_{K_2}$, then $u^*_{K_2} = u^e(\tau^*_{K_2}) + u^h(\tau^*_{K_2})\phi^*_{K_2} = u^e(\tau^*_{K_2}) + u^h(\tau^*_{K_2})\phi^*_{K_1} \in \mathcal{U}_{K_2-1}(\tau^*_{K_2})$, so $x^*_{K_2}$ must be robust feasible. Moreover,

$$\begin{aligned} LB_{K_2} = &\ f(x^*_{K_2}) + \sum_i C^*_{i,K_2} + \max_{u \in \mathcal{U}_{K_2-1}(\tau^*_{K_2})} \min_{y \in \mathcal{Y}(x^*_{K_2}, u)} g(y) \\ \geq &\ f(x^*_{K_2}) + \sum_i C^*_{i,K_2} + \min_{y \in \mathcal{Y}(x^*_{K_2}, u^*_{K_2})} g(y) = UB_{K_2}. \quad \text{(D.3)} \end{aligned}$$

Together with $LB_{K_2} \leq UB_{K_2}$ from Claim 1, we have $LB_{K_2} = UB_{K_2}$. This contradicts to the assumption that the algorithm does not terminate after $K \geq K_2$ iterations.

Now the proof of **Theorem 1** is given below.

First, we prove that the algorithm converges in $\mathcal{O}(n_U)$ iterations. With Lemma 4, we know that the worst-case scenario u^*_k can be achieved at a vertex of $\mathcal{U}_k(\tau^*_k)$. A vertex of the set Φ corresponds to a vertex of the set $\mathcal{U}_k(\tau^*_k)$. According to Claim 2, the same vertex of Φ will not be picked up twice. Moreover, the number of vertices of Φ is n_U. Hence, the algorithm stops in $\mathcal{O}(n_U)$ iterations.

Suppose the algorithm terminates after $K \leq n_U$ iterations.

Next, we show the robust feasibility of (x^*_K, τ^*_K). Obviously, we have $x^*_K \in \mathcal{X}$ and $\tau^*_K \in [0,1]$. Moreover, $s^*_K = 0$ when the algorithm terminates, so (x^*_K, τ^*_K) is robust feasible.

Finally, we show the optimality of (x^*_K, τ^*_K). According to Claim 1, we have $LB_K \leq O^* \leq UB_K$. Together with the condition for termination $|UB_K - LB_K| \leq \varepsilon$, we have

$$|UB_K - O^*| \leq |UB_K - LB_K| \leq \varepsilon. \quad \text{(D.4)}$$

This completes the proof.