Multistage Adaptive Robust Optimization for the Unit Commitment Problem

Álvaro Lorca, X. Andy Sun
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA

Eugene Litvinov, Tongxin Zheng
ISO New England, Holyoke, MA 01040 USA

September 2015

The growing uncertainty associated with the increasing penetration of wind and solar power generation has presented new challenges to the operation of large-scale electric power systems. Motivated by these challenges, we present a multistage adaptive robust optimization model for the most critical daily operational problem of power systems, namely the unit commitment (UC) problem, in the situation where nodal net electricity loads are uncertain. The proposed multistage robust UC model takes into account the time causality of the hourly unfolding of uncertainty in the power system operation process, which we show to be relevant when ramping capacities are limited and net loads present significant variability. To deal with large-scale systems, we explore the idea of simplified affine policies and develop a solution method based on constraint generation. Extensive computational experiments on the IEEE 118-bus test case and a real-world power system with 2736 buses demonstrate that the proposed algorithm is effective in handling large-scale power systems and that the proposed multistage robust UC model can significantly outperform the deterministic UC and existing two-stage robust UC models in both operational cost and system reliability.

Key words: Electric energy systems, multistage robust optimization, affine policies, constraint generation.

---

1. Introduction

Operating large-scale electric power systems is a challenging task that requires adequate decision tools and methodologies for hedging against uncertain factors such as wind and solar power generation, water inflows for hydroplants, electricity demand, transmission line and generator contingencies, and demand response (see e.g. Gómez-Expósito et al. (2008), Conejo et al. (2010), Xie et al. (2011)). The unit commitment (UC) problem consists in finding an on/off commitment schedule and generation dispatch levels for generating units in each hour of the next day, in such a way that the total production cost is minimized while electricity demand is met and various physical constraints of generators and the transmission network are satisfied. This is the most critical daily operational problem for large-scale power systems, and it is a difficult optimization problem due to its large scale and discrete nature. It becomes more complicated when wind power and other renewable generation resources are available in large quantities and present significant uncertainty in their availability. How to deal with increasing uncertainty in power systems has been identified by the electricity industry as an urgent challenge (see Hobbs et al. (2001), Keyhani et al. (2009), Conejo et al. (2010), Xie et al. (2011)).

Stochastic programming is an important approach that has been applied to managing uncertainties in the UC problem (e.g., see Ozturk et al. (2004), Wu et al. (2007), Wang et al. (2008), Ruiz et al. (2009a), Ruiz et al. (2009b), Tuohy et al. (2009), Constantinescu et al. (2011),Wang et al. (2012), Papavasiliou and Oren (2013)). These models offer a notable advancement over deterministic methods. However, they also present important computational challenges when dealing with large-scale power systems. In particular, stochastic programming models require identifying appropriate probability distributions for uncertain parameters such as load and renewable energy generation, which might be difficult; it is also difficult to construct scenario trees that represent high-dimensional stochastic processes; and large scenario trees often lead to computational difficulties. See the work by Heitsch and Römisch (2011) for an important example on scenario tree generation methods.

Robust optimization is an alternative paradigm for optimization under uncertainty, which has received wide attention and has been applied in several engineering disciplines (e.g., see Ben-Tal et al. (2009a), Bertsimas et al. (2011)). Instead of using probability distributions for uncertain parameters, robust optimization models assume that uncertain parameters are realized as elements of a deterministic *uncertainty set*. Given an uncertainty set, the problem consists of finding a solution that is feasible for any realization of the uncertain parameters in this set and also minimizes the worst-case cost. This approach is of particular interest when accurate probability distributions are not available or when uncertain parameters present high dimensionality. Further, the conservativeness of robust solutions can be controlled by the choice of uncertainty sets.

Several robust optimization formulations for the UC problem have been recently proposed. A robust formulation for the contingency constrained UC problem is proposed in Street et al. (2011). Various adaptive robust UC models dealing with demand and renewable generation uncertainty are studied in Jiang et al. (2012), Zhao and Zeng (2012), Bertsimas et al. (2013), and Wang et al. (2013). More specifically, Jiang et al. (2012) present a robust UC formulation including pumped-storage hydro units under wind uncertainty. Zhao and Zeng (2012) present a formulation with demand response under wind uncertainty. Bertsimas et al. (2013) present a security constrained robust UC formulation with system reserve requirements under nodal net injection uncertainty, including extensive computational experiments on a real-world power system. Wang et al. (2013) present a contingency constrained UC model under uncertain generator and transmission line contingencies. Zhao and Guan (2013) present a hybrid approach that combines stochastic and robust optimization by weighing expected cost and worst-case cost.

An essential feature of all the above stochastic and robust UC models is that they are *two-stage* models, where the first-stage decision is the on/off commitment decision made in the day-ahead electricity market, while the second-stage decision is the real-time dispatch decision for the entire scheduling horizon. The work in Zhao et al. (2013) presents a three-stage robust UC model, which has UC decisions in the first stage and dispatch decisions in the second stage, and then has uncertain demand response after dispatch decisions. This decision-making structure is converted to a two-stage robust model. The crucial assumption of two-stage models is that the second-stage decision is made with the full knowledge of realized uncertain parameters over the entire scheduling horizon. However, in reality, power systems are operated sequentially, where generation dispatch at each hour can only depend on the information of realized uncertain parameters up to that hour. In other words, dispatch decisions are *non-anticipative*. Two-stage stochastic and robust UC models ignore this.

In this paper, we demonstrate the importance of considering non-anticipativity constraints in power system operations and present a *multistage* adaptive robust optimization model for the UC problem, where the commitment decisions are selected here-and-now as done in the day-ahead electricity market, and the dispatch decision for each hour of the next day is the wait-and-see decision respecting non-anticipativity for the sequential revelation of uncertainty. We also address the computational challenge presented by the multistage robust UC model. To make it computationally tractable, we consider approximation schemes with decision rules, in particular, we use *affine policies* for the dispatch decisions, where generators' dispatch levels are affine functions of uncertain load.

The affinely adjustable robust optimization approach has attracted considerable attention since the seminal paper of Ben-Tal et al. (2004). Most of the existing works focus on studying multistage convex optimization problems with relatively simple and well-structured constraints, such as multi-period inventory problems studied in Bertsimas et al. (2010), Goh and Sim (2011), Hadjiyiannis et al. (2011) or multistage stochastic linear programs in Kuhn et al. (2011). These models can be transformed to deterministic counterparts through duality theory and solved by existing algorithms for convex programs, see e.g. Ben-Tal et al. (2009b), Kuhn et al. (2011). Another direction of research is to extend affine policies to more general decision rules, such as in Chen et al. (2008), Chen and Zhang (2009), and Georghiou et al. (2013).

Previous applications of affine policies for power system operations were carried out by Warrington et al. (2012, 2013), who considered a stochastic optimization model for the economic dispatch problem with storage, where the UC binary decision is assumed to be fixed. These works have been recently extended to incorporate binary UC decisions (Warrington et al. 2014). Another application of affine policies for power system operations was developed by Jabr (2013) who considered the dispatch of automatic generation control units under uncertain renewable energy outputs, with fixed UC decisions. Our work is done independently from these works. And the crucial differences of our approach with respect to these references include the proposal of a multistage robust UC model with simplified affine policies, the analysis of the relationship between the multistage and two-stage robust UC models, and the development of an algorithm based on constraint generation that allows the efficient solution of large-scale instances of the problem under a high-dimensional uncertainty set. An interesting analysis comparing two-stage and multistage robust formulations is presented by Minoux (2014), for an economic dispatch problem with one bus and one generator. However, no details on the multistage model are provided. The author also introduces the idea of state-space representable uncertainty sets, which can be used for modeling temporal dependencies in the uncertain parameters.

The proposed multistage robust UC problem in this paper presents several challenges that make existing methodologies not directly applicable. In particular, the multistage robust UC model is a large-scale mixed-integer optimization problem involving a large number of complicated constraints. Due to the mixed-integer decisions, convex optimization based modeling and solution methods cannot be applied. Furthermore, due to its very large scale, applying even the basic affine policies in a straightforward way is not computationally viable and the duality-based approach leads to reformulations with exceedingly large dimensions. To deal with these challenges, we propose new solution ideas. More specifically, instead of using more general decision rules, we descend the complexity ladder and use *simplified affine policies* through properly aggregating uncertain parameters in the dependency structure of the affine policy. The resulting multistage UC formulation has a reduced dimensionality and a structure that we can exploit. We design a solution method based on constraint generation and employ several algorithmic improvements that make the multistage robust UC problem efficiently solvable even for large-scale instances.

We conduct a thorough computational study with extensive numerical experiments on the performance of the proposed algorithm, the quality of simplified affine policies, their worst-case and average-case performance, and comparison with existing deterministic and two-stage robust UC models. The computational results show that the proposed algorithm can effectively solve the multistage robust UC model within a time frame reasonable for the day-ahead operation of large-scale power systems. The performance of the proposed multistage robust UC model demonstrates its ability to significantly reduce operational costs and at the same time improve system reliability, as we show in computational experiments where we compare this approach with the existing deterministic and two-stage robust UC models.

The contributions of the paper can be summarized as follows.

1. This paper presents a multistage adaptive robust optimization model for the UC problem under significant uncertainty in nodal net loads, respecting non-anticipativity in the dispatch process. We have also proposed a new robust dispatch model utilizing the affine policy obtained from the multistage robust UC model.

2. This paper discusses the solution concept of simplified affine policies in multistage robust optimization and demonstrates its effectiveness in power system operations.

3. This paper proposes an efficient solution algorithm based on constraint generation with various algorithmic improvements for solving large-scale multistage robust UC models with affine policy. Several aspects of the algorithm are also applicable to general large-scale robust optimization problems with mixed-integer variables.

4. This paper provides an extensive computational study of the proposed multistage robust UC model on medium and large-scale power systems. Comparison with existing deterministic and two-stage robust UC models demonstrates the potential of the proposed approach in reducing operational cost, increasing system reliability, and managing system flexibility.

The remainder of the paper proceeds as follows. Section 2 presents the deterministic and two-stage robust UC models and discusses their limitations. Section 3 proposes the multistage robust UC model and introduces the concept of simplified affine policies. Section 4 presents a traditional method based on duality and a constraint generation framework for solving robust optimization problems. Section 5 proposes several algorithmic improvements. Section 6 presents a multifaceted computational study of the performance of the proposed approach. Section 7 concludes the paper. All the proofs, unless given in the main body of the paper, are collected in the Electronic Companion.

2. Non-Causal UC models and Their Limitations

In this section, we discuss the deterministic UC and the recently developed two-stage robust UC models. We shall call the two-stage robust UC model a *non-causal* UC model, because the decisions in this model depend on future information of uncertainty and thus do not respect non-anticipativity in the physical process of dispatching generators. We show important issues with non-causal UC formulations. It serves as the motivation for the development of multistage robust UC models.

2.1. Deterministic Unit Commitment

Consider the deterministic UC model below.

$$\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{p}} \quad \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g}(G_i x_i^t + S_i u_i^t) + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} C_i p_i^t \tag{1a}$$

$$
\begin{aligned}
\text{s.t.}\quad & x_i^t,\, u_i^t,\, v_i^t \in \{0,1\} && \forall\, i \in \mathcal{N}_g,\, t \in \mathcal{T} && \text{(1b)}\\
& x_i^t - x_i^{t-1} = u_i^t - v_i^t && \forall\, i \in \mathcal{N}_g,\, t \in \mathcal{T} && \text{(1c)}\\
& \sum_{\tau=t}^{t+UT_i-1} x_i^\tau \geq UT_i u_i^t && \forall\, i \in \mathcal{N}_g,\, t \in \{1,2,\ldots,T-UT_i+1\} && \text{(1d)}\\
& \sum_{\tau=t}^{t+DT_i-1} (1 - x_i^\tau) \geq DT_i v_i^t && \forall\, i \in \mathcal{N}_g,\, t \in \{1,2,\ldots,T-DT_i+1\} && \text{(1e)}\\
& \sum_{\tau=t}^{T} (x_i^\tau - u_i^t) \geq 0 && \forall\, i \in \mathcal{N}_g,\, t \in \{T-UT_i+1,\ldots,T\} && \text{(1f)}\\
& \sum_{\tau=t}^{T} (1 - x_i^\tau - v_i^t) \geq 0 && \forall\, i \in \mathcal{N}_g,\, t \in \{T-DT_i+1,\ldots,T\} && \text{(1g)}\\
& p_i^{min} x_i^t \leq p_i^t \leq p_i^{max} x_i^t && \forall\, i \in \mathcal{N}_g,\, t \in \mathcal{T} && \text{(1h)}\\
& -RD_i x_i^t - SD_i v_i^t \leq p_i^t - p_i^{t-1} \leq RU_i x_i^{t-1} + SU_i u_i^t && \forall\, i \in \mathcal{N}_g,\, t \in \mathcal{T} && \text{(1i)}\\
& -f_l^{max} \leq \boldsymbol{\alpha}_l^T (\boldsymbol{B}^p \boldsymbol{p}^t - \boldsymbol{B}^d \boldsymbol{d}^t) \leq f_l^{max} && \forall\, t \in \mathcal{T},\, l \in \mathcal{N}_l && \text{(1j)}\\
& \sum_{i \in \mathcal{N}_g} p_i^t = \sum_{j \in \mathcal{N}_d} d_j^t && \forall\, t \in \mathcal{T}, && \text{(1k)}
\end{aligned}
$$

where $\mathcal{N}_g, \mathcal{N}_d, \mathcal{N}_l, \mathcal{T}$ denote the sets of generators, nodes with net load, transmission lines, and time periods, respectively, and N_g, N_d, N_l, T are their cardinalities; x_i^t, u_i^t, v_i^t and p_i^t are respectively the on/off, start-up, shut-down, and the generation dispatch level decisions of generator i at time t; G_i, S_i, C_i are the no-load cost, start-up cost, and variable cost of generator i; DT_i and UT_i are the minimum-down and minimum-up times of generator i; p_i^{min} and p_i^{max} are the minimum and maximum generation levels of generator i; RD_i and RU_i are the ramp-down and ramp-up rates of generator i, and SD_i and SU_i are the ramp rates when generator i shuts down and turns on; $\boldsymbol{B}^p$ and $\boldsymbol{B}^d$ are the incidence matrices for generators and loads; $\boldsymbol{\alpha}_l$ and f_l^{max} are the generation shift factor and the flow limit for line l, respectively; d_j^t is the net load at node j and time t. In this paper, nodal net load is defined as the nodal demand minus the total renewable generation such as wind and solar power connected to the same node, which is an uncertain quantity due to the uncertainty of wind and solar power generation. The objective (1a) consists of minimizing the sum of commitment costs (including no-load and start-up costs) and dispatch costs (assumed to be linear here but can be replaced with a piecewise linear function without changing the linearity of the problem). Eq. (1c) corresponds to start-up and shut-down constraints. Eq. (1d)-(1g) corresponds to minimum up and down time constraints. Constraints (1h) enforce minimum and maximum generation capacity limits when generators are on, and no generation when they are off. Constraints (1i) enforce ramping up and down limits. Constraints (1j) enforce transmission line limits. Constraints (1k) enforce energy balance at a system level. The model can also be extended to include reserve decisions and related constraints, which are omitted here for simplicity. The formulation in (1d)-(1g) follows Ostrowski et al. (2012).

2.2. Two-stage Adaptive Robust Unit Commitment

To deal with uncertainties in the nodal net electricity loads, the following two-stage adaptive robust UC model has been proposed (e.g. see Jiang et al. (2012), Zhao and Zeng (2012), Bertsimas et al. (2013)):

$$\min_{\boldsymbol{x}\in X} \left\{ F(\boldsymbol{x}) + \max_{\boldsymbol{d}\in\mathcal{D}} \min_{\boldsymbol{p}\in\Omega(\boldsymbol{x},\boldsymbol{d})} c(\boldsymbol{p}) \right\}, \tag{2}$$

where $\boldsymbol{x}$ denotes all the commitment related binary variables (x_i^t, u_i^t, v_i^t in the deterministic UC model (1)), $\boldsymbol{d}$ is the vector of net load at all nodes and all time periods, $\boldsymbol{p}$ is the vector of dispatch variables, set X is the feasible region of the commitment decisions defined by Eq. (1b)-(1g), $\mathcal{D}$ is the uncertainty set of net loads, $\Omega(\boldsymbol{x},\boldsymbol{d})$ is the feasible region of the dispatch variables parameterized by the commitment decisions and realized net load, as defined in Eq. (1h)-(1k), and

$$F(\boldsymbol{x}) = \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} \left(G_i x_i^t + S_i u_i^t\right), \quad c(\boldsymbol{p}) = \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} C_i p_i^t.$$

In this paper, we use the following budget uncertainty set:

$$\mathcal{D}^t = \left\{ \boldsymbol{d}^t = (d_1^t,\ldots,d_{N_d}^t): \sum_{j\in\mathcal{N}_d} \frac{|d_j^t - \overline{d}_j^t|}{\hat{d}_j^t} \le \Gamma\sqrt{N_d},\ d_j^t \in [\overline{d}_j^t - \Gamma\hat{d}_j^t, \overline{d}_j^t + \Gamma\hat{d}_j^t]\ \forall j\in\mathcal{N}_d \right\}. \tag{3}$$

Notice that d_j^t lies in an interval centered around the nominal value $\overline{d}_j^t$ within a deviation denoted by $\Gamma\hat{d}_j^t$. The budget constraint with budget $\Gamma\sqrt{N_d}$ controls the size of the uncertainty set, where Γ represents the conservativeness of the model. For $\Gamma = 0$ we have $\mathcal{D}^t = \{\overline{\boldsymbol{d}}^t\}$, i.e., the uncertainty set only contains the nominal net load vector and the two-stage robust UC model (2) becomes the deterministic UC model (1). As Γ increases, more net load vectors are contained in the uncertainty set. The square root $\sqrt{N_d}$ scaling is motivated by a central limit theorem type argument, where the standard deviation of the aggregated randomness scales proportionally to the square root of the number of random variables (see Bertsimas et al. (2013)). Robust constraints using uncertainty set (3) also guarantees a probabilistic feasibility condition (see Chen et al. (2007)). We define $\mathcal{D} = \prod_{t\in\mathcal{T}} \mathcal{D}^t$ as the uncertainty set for the net load trajectory $\boldsymbol{d}$ over the entire scheduling horizon. With this choice, notice the separability of $\mathcal{D}$ over time periods, i.e., the temporal independency.

As seen from the above two-stage robust UC model, the dispatch decision $\boldsymbol{p}$ in the inner minimization problem over $\Omega(\boldsymbol{x},\boldsymbol{d})$ is made with perfect knowledge of the realization of uncertain net loads $\boldsymbol{d}$ over the entire scheduling horizon. In reality, system operators only have perfect information about uncertain parameters that are realized up to the current operating time. The key questions are: What is the consequence of assuming the full knowledge of nodal net loads in the dispatch process? How to properly tackle the sequential nature of this process?

2.3. Example that Illustrates the Limitations of Non-causal UC Models

We present a simple example to illustrate that the UC solution from the two-stage robust UC model can lead to infeasibility in the real-time dispatch.

EXAMPLE 1. The system has 2 buses, A and B, and two periods, so $T=2$. Each bus has a conventional generator. The transmission line has a flow limit of 1 unit of power. The ramp rates of both generators are also 1 unit of power, i.e., $RU_A = RD_A = R_A = 1$, and $RU_B = RD_B = R_B = 1$. The initial generation levels of the two generators are at 12, i.e., $p_A^0 = p_B^0 = 12$ at $t=0$.

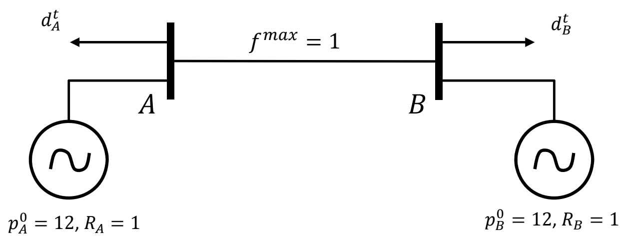

Figure 1 Simple two-bus system to illustrate the limitation of non-causal UC models.

The uncertainty sets for nodal net loads (d_A^t, d_B^t) at $t=1,2$ are given as follows:

$$\mathcal{D}^1 = \left\{(d_A^1, d_B^1) = (12,12)\right\}, \text{and } \mathcal{D}^2 = \left\{(d_A^2, d_B^2): \ d_A^2 \in [10,15], \ d_B^2 \in [10,15], \ d_A^2 + d_B^2 = 25\right\}.$$

That is, the first period loads are deterministic with power level of 12 at each bus, and the net loads in the second period are uncertain, but the total net load is known to be 25. Denote $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$.

CLAIM 1. *The two-stage robust UC model* (2) *is feasible for the system in Example 1.*

Proof: Consider $\boldsymbol{x}_{2S} = ((x_A^1, x_B^1), (x_A^2, x_B^2)) = ((1,1),(1,1))$. To prove that $\boldsymbol{x}_{2S}$ is feasible for the two-stage robust UC model, we construct a feasible dispatch policy. In particular, for any $\boldsymbol{d} = ((d_A^1, d_B^1), (d_A^2, d_B^2)) \in \mathcal{D}$, consider the following policy:

$$p_A^1(\boldsymbol{d}) = 12 + (2/5)(d_A^2 - 12.5), \ \ p_B^1(\boldsymbol{d}) = 12 - (2/5)(d_A^2 - 12.5), \tag{4a}$$

$$p_A^2(\boldsymbol{d}) = 12.5 + (3/5)(d_A^2 - 12.5), \ \ p_B^2(\boldsymbol{d}) = 12.5 - (3/5)(d_A^2 - 12.5). \tag{4b}$$

From (4), we can see that for $t=1$, $p_A^1(\boldsymbol{d}) + p_B^1(\boldsymbol{d}) = 24$ for all $\boldsymbol{d} \in \mathcal{D}$, so energy balance is respected. By the definition of the uncertainty sets, we have $p_A^1(\boldsymbol{d}), p_B^1(\boldsymbol{d}) \in [11,13]$ for all $\boldsymbol{d} \in \mathcal{D}$, so the ramping constraints from the initial states ($\boldsymbol{p}^0 = (12,12)$) are respected. Furthermore, $p_A^1(\boldsymbol{d}) - d_A^1 = p_A^1(\boldsymbol{d}) - 12 \in [-1,1]$ for all $\boldsymbol{d} \in \mathcal{D}$, so transmission constraints are respected. Similarly for $t=2$, we have $p_A^2(\boldsymbol{d}) + p_B^2(\boldsymbol{d}) = 25$ for all $\boldsymbol{d} \in \mathcal{D}$, so energy balance is satisfied. Since $d_A^2 \in [10,15]$, we can see that $p_A^2(\boldsymbol{d}) - p_A^1(\boldsymbol{d}) = 0.5 + (1/5)(d_A^2 - 12.5) \in [0,1]$ and $p_B^2(\boldsymbol{d}) - p_B^1(\boldsymbol{d}) = 0.5 - (1/5)(d_A^2 - 12.5) \in [0,1]$, hence ramping constraints are respected. Finally, $p_A^2(\boldsymbol{d}) - d_A^2 = 5 - (2/5)d_A^2 \in [-1,1]$, so transmission constraints are respected. Therefore, $\boldsymbol{p}(\boldsymbol{d})$ given in (4) satisfies all constraints in (2), thus $\boldsymbol{x}_{2S}$ is feasible for the two-stage robust UC model. □

Notice that the dispatch policy (4) is non-causal, because the dispatch decision at $t=1$ depends on the uncertainty realization at $t=2$. If this UC solution is implemented, the real-time dispatch under this UC solution can be infeasible, as shown in the following result.

Claim 2. *Let $\boldsymbol{x}_{2S}^*$ be the optimal UC solution of the two-stage robust UC model for Example 1. Under $\boldsymbol{x}_{2S}^*$, there does not exist any feasible dispatch policy that respects time causality, i.e. where $\boldsymbol{p}^1(\cdot)$ does not depend on $\boldsymbol{d}^2$.*

Proof Notice that $\boldsymbol{x}_{2S}^* = ((x_A^1, x_B^1), (x_A^2, x_B^2)) = ((1,1),(1,1))$ is the optimal solution of the two-stage robust UC for the system in Example 1, since keeping both generators online is the only candidate solution to satisfy net load.

Now consider the real-time sequential operation under this commitment decision $\boldsymbol{x}_{2S}^*$, where the causal dispatch policy at t can only depend on the information available up to t. We want to show that there does not exist any causal dispatch policy that can make the system feasible for all net load vectors in the uncertainty set. For this, we need to show that there is no $\boldsymbol{p}^1(\boldsymbol{d}^1)$ such that, for all $\boldsymbol{d}^2 \in \mathcal{D}^2$, there always exists a feasible $\boldsymbol{p}^2(\boldsymbol{d}^1, \boldsymbol{d}^2)$ at $t=2$.

Since $\boldsymbol{d}^1$ is fixed at $(12,12)$, we write $\boldsymbol{p}^1(\boldsymbol{d}^1)$ as $\boldsymbol{p}^1 = (p_A^1, p_B^1)$ for brevity. Notice that due to the energy balance constraint, we must have $p_A^1 + p_B^1 = 24$, and due to the ramping capacity and transmission constraints, we must have $p_A^1, p_B^1 \in [11,13]$. Suppose we choose $p_A^1 \leq 12$. Then take $\boldsymbol{d}^2 = (15,10)$ from the uncertainty set $\mathcal{D}^2$. Due to ramping constraints, we must have $p_A^2 \leq 13$. However, it is impossible to satisfy energy balance at location A, because the transmission limit is 1. Similarly, if we choose $p_A^1 \geq 12$, the adversary can take $\boldsymbol{d}^2 = (10,15) \in \mathcal{D}^2$, which leads to the impossibility of satisfying net load at location B. This means that no matter what $\boldsymbol{p}^1$ we choose to satisfy the constraints at $t=1$, there always exists a $\boldsymbol{d}^2 \in \mathcal{D}^2$ so that the constraints at $t=2$ cannot be satisfied. □

With this result, we see that the two-stage robust UC decision $\boldsymbol{x}_{2S}^*$ can lead to infeasibility in the real-time dispatch problem. This simple example demonstrates that when the transmission and generation ramping capability is limited, the two-stage robust UC model can make an infeasible problem appear to be feasible. When such a UC solution is implemented, the real-time operation can become infeasible under uncertain parameters realized within the uncertainty set. Also notice that, if we add expensive generators at each bus in Example 1, we can obtain a system where the multistage robust UC model produces a UC solution under which feasible real-time dispatch is guaranteed, while the optimal commitment solution from the two-stage robust UC solution again leads to infeasibility in real-time operation.

With high penetration of renewable energy resources, power systems frequently experience fast swings in net loads, which pushes the generators toward the regime of limited ramping capability, which motivates us to consider multistage robust UC models.

3. Multistage Adaptive Robust UC and Simplified Affine Policy

In this section, we first propose the multistage robust UC model and give a theoretical analysis on the relationship between the two-stage and multistage robust UC models. Then, we introduce affine dispatch policies and the concept of simplified affine policies.

3.1. Multistage Adaptive Robust UC Model

In the operation of power systems, the commitment decision $\boldsymbol{x}$ is made several hours before the observations of uncertain net loads, and then the dispatch decisions are sequentially optimized in real time with observations of realized uncertainty up to the operating hour. To faithfully model this process, the dispatch decision $\boldsymbol{p}^t$ at time t in the UC model should depend on the history of net load $\boldsymbol{d}^{[t]} \triangleq (\boldsymbol{d}^1, ..., \boldsymbol{d}^t)$. Based on this requirement, we formulate the following multistage adaptive robust UC model.

$$\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{p}(\cdot)} \left\{ \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} (G_i x_i^t + S_i u_i^t) + \max_{\boldsymbol{d}\in\mathcal{D}} \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} C_i\, p_i^t(\boldsymbol{d}^{[t]}) \right\} \tag{5a}$$

$$\text{s.t.} \quad \text{Constraints (1b)-(1g) for } (\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$$

$$p_i^{min} x_i^t \le p_i^t(\boldsymbol{d}^{[t]}) \le p_i^{max} x_i^t \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, i\in\mathcal{N}_g,\, t\in\mathcal{T} \tag{5b}$$

$$-RD_i x_i^t - SD_i v_i^t \le p_i^t(\boldsymbol{d}^{[t]}) - p_i^{t-1}(\boldsymbol{d}^{[t-1]}) \le RU_i x_i^{t-1} + SU_i u_i^t \qquad \forall \boldsymbol{d}\in\mathcal{D},\, i\in\mathcal{N}_g,\, t\in\mathcal{T} \tag{5c}$$

$$-f_l^{max} \le \boldsymbol{\alpha}_l^T (\boldsymbol{B}^p \boldsymbol{p}^t(\boldsymbol{d}^{[t]}) - \boldsymbol{B}^d \boldsymbol{d}^t) \le f_l^{max} \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, t\in\mathcal{T},\, l\in\mathcal{N}_l \tag{5d}$$

$$\sum_{i\in\mathcal{N}_g} p_i^t(\boldsymbol{d}^{[t]}) = \sum_{j\in\mathcal{N}_d} d_j^t \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, t\in\mathcal{T}. \tag{5e}$$

The crucial feature of this formulation is the expression $p_i^t(\boldsymbol{d}^{[t]})$, which makes the generation of unit i at time t a function of net load uncertainty realized up to time t, thus respecting non-anticipativity. Constraints (5b)-(5e) enforce generation limits, ramping capacities, transmission line capacities, and energy balance, for any realization of $\boldsymbol{d}\in\mathcal{D}$. Note that binary decisions $\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}$ are not adaptive, they are decided "here-and-now" before observing any uncertainty.

The multistage decision making structure of (5) can be equivalently represented in the following nested formulation, using the separability of the uncertainty set over time periods:

$$\min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X} \left\{ \boldsymbol{G}^\top \boldsymbol{x} + \boldsymbol{S}^\top \boldsymbol{u} + \max_{\boldsymbol{d}^1\in\mathcal{D}^1} \min_{\boldsymbol{p}^1\in\Omega_1(\boldsymbol{x},\boldsymbol{d}^1,\boldsymbol{p}^0)} \left\{ \boldsymbol{C}^\top \boldsymbol{p}^1 + \cdots + \max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T(\boldsymbol{x},\boldsymbol{d}^T,\boldsymbol{p}^{T-1})} \boldsymbol{C}^\top \boldsymbol{p}^T \right\} \right\}, \tag{6}$$

where $\Omega_t(\boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{p}^{t-1}) \triangleq \left\{\boldsymbol{p}^t : \text{ (1h)-(1k) are satisfied } \quad \forall\, i \in \mathcal{N}_g\right\}$. Notice that the feasible region $\Omega_t(\boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{p}^{t-1})$ of the dispatch decision at stage t depends on previous stage $t-1$'s dispatch levels $\boldsymbol{p}^{t-1}$ and stage t's realized demand $\boldsymbol{d}^t$. Due to discrete decision variables and the large scale of the formulation, numerical solution of the multistage robust UC model ((5) or (6)) presents a major computational challenge. In the following, we first make further discussion on the relation between the two-stage and multistage models, then propose approximate decision rules and tractable solution methods for solving the multistage robust UC model.

3.2. Discussion on Two-Stage and Multistage Robust UC Models

The key difference between the two-stage and multistage models is that the multistage robust UC provides a *causal* policy $\boldsymbol{p}^t(\boldsymbol{d}^{[t]})$, which only relies on information observed up to the respective time period when the dispatch decision is made. Clearly, the two-stage robust UC model lacks this property. It turns out that, when the system is not constrained by its ramping capability, i.e., the generators have enough ramping capacity to follow the rapidly varying wind, the two-stage robust UC model is equivalent to the multistage model.

Proposition 1. *Consider the two-stage robust UC* (2) *and the multistage robust UC* (5)*, where the uncertainty set is given by Eq.* (3)*. If ramping constraints* (5c) *are neglected, the two-stage robust UC* (2) *and the multistage robust UC* (5) *have the same optimal objective value and a same optimal UC solution.*

The proof follows from the fact that the dispatch component of both problems can be separated into T non-coupled problems when there are no ramping constraints. Please see the Electronic Companion for details.

Proposition 1 suggests that the multistage robust UC model is important precisely when the system's ramping capability is a limited resource, which is the case for power systems with a high penetration of uncertain wind and solar power generation.

3.3. Affine Multistage Robust UC

To computationally solve the proposed multistage robust UC model (5), we propose to consider approximation schemes using linear decision rules. In particular, to make the problem tractable, we restrict the dispatch decision $\boldsymbol{p}^t(\cdot)$ to have the form of an affine function as

$$p_i^t(\boldsymbol{d}^{[t]}) = w_i^t + \sum_{j \in \mathcal{N}_d} \sum_{s \in [1:t]} W_{itjs} d_j^s, \tag{7}$$

where $[1:t] \triangleq \{1,...,t\}$ and (w_i^t, W_{itjs}) are the coefficients of the affine policy. It is important to notice that the affine policy (7) automatically respects non-anticipativity. Using this affine dispatch policy, the multistage robust UC model has the following form

$$\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},z,\boldsymbol{w},\boldsymbol{W}} \quad \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g}(G_i x_i^t + S_i u_i^t) + z \tag{8a}$$

$$\text{s.t.} \quad \text{Constraints (1b)-(1g) for } (\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$$

$$\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} C_i\left(w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s\right) \le z \qquad \forall\, \boldsymbol{d}\in\mathcal{D} \tag{8b}$$

$$p_i^{min} x_i^t \le w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s \le p_i^{max} x_i^t \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, i\in\mathcal{N}_g,\, t\in\mathcal{T} \tag{8c}$$

$$\left(w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s\right) - \left(w_i^{t-1} + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t-1]} W_{i,t-1,js} d_j^s\right) \ge -RD_i x_i^t - SD_i v_i^t \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, i\in\mathcal{N}_g,\, t\in\mathcal{T} \tag{8d}$$

$$\left(w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s\right) - \left(w_i^{t-1} + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t-1]} W_{i,t-1,js} d_j^s\right) \le RU_i x_i^{t-1} + SU_i u_i^t \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, i\in\mathcal{N}_g,\, t\in\mathcal{T} \tag{8e}$$

$$-f_l^{max} \le \sum_m \sum_{i\in\mathcal{N}_g} \alpha_{lm} B_{mi}^p \left(w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s\right) - \sum_m\sum_{j\in\mathcal{N}_d} \alpha_{lm} B_{mj}^d d_j^t \;\le\; f_l^{max} \qquad \forall\, \boldsymbol{d}\in\mathcal{D},\, t\in\mathcal{T},\, l\in\mathcal{N}_l \tag{8f}$$

$$\sum_{i\in\mathcal{N}_g}\left(w_i^t + \sum_{j\in\mathcal{N}_d}\sum_{s\in[1:t]} W_{itjs} d_j^s\right) = \sum_{j\in\mathcal{N}_d} d_j^t \qquad \forall \boldsymbol{d}\in\mathcal{D}, t\in\mathcal{T}. \tag{8g}$$

We have created variable z to denote the worst-case dispatch cost in constraint (8b). Constraints (8c)-(8g) correspond to (5b)-(5e), obtained by replacing $p_i^t(\boldsymbol{d}^{[t]})$ with the affine policy (7). Note that constraints (8b)-(8g) are robust constraints that should hold for all $\boldsymbol{d}\in\mathcal{D}$. We call (8) the affine multistage robust UC model.

3.4. Simplified Affine Policies

In the affine policy (7), the dispatch decision $p_i^t(\boldsymbol{d}^{[t]})$ of generator i at time t depends on the entire history of realized net load at every node and every time period up to t. This full affine dependency requires defining a large number of W_{itjs} variables, which can quickly lead to scalability issues in large-scale power systems.

To make the affine multistage robust UC model (8) a practical decision tool for the operation of large-scale power systems, we introduce further restrictions on the affine policy form. In particular, we consider affine policies with *simplified* structures by limiting the degrees of freedom in W_{itjs}. There are several ways to do this: we can restrict $\boldsymbol{p}^t(\cdot)$ to only depend on the most recently revealed information at time t, rather than on the whole history; we can partition time periods into peak-load, medium-load, and low-load periods and assume affine policies in each period have the same form; we can also partition the transmission network into zones and make generators' dispatch policies depend on the aggregated load in each zone. We use the following two very simple policies:

$$p_i^t(\boldsymbol{d}^{[t]}) = w_i^t + W_i \sum_{j\in\mathcal{N}_d} d_j^t \qquad \forall i \in \mathcal{N}_g, t \in \mathcal{T} \tag{9}$$

$$p_i^t(\boldsymbol{d}^{[t]}) = w_i^t + W_{it} \sum_{j\in\mathcal{N}_d} d_j^t \qquad \forall i \in \mathcal{N}_g, t \in \mathcal{T}. \tag{10}$$

We call (9) the W_i-policy, where the coefficients W_i of the affine policy only depend on generators but not on time, and the dispatch level of each generator at time t depends on the total load in the system at time t. Eq. (10) presents a finer policy, which we call the W_{it}-policy, where the coefficients W_{it} of the affine policy can change over time. Surprisingly, it will be shown that these two very simplified affine policies are already quite powerful and produce close-to-optimal performance for the multistage robust UC model. We also want to remark that a static policy, i.e. $p_i^t(\boldsymbol{d}^{[t]}) = w_i^t$, is the simplest (and trivial) form of an affine policy, however, with this choice it becomes impossible to satisfy energy balance equality over all net load vectors in the uncertainty set. This shows that the simpler but not the simplest affine policies work and the non-trivial affine dependence in the dispatch policy is very important.

4. Basic Algorithmic Framework

In this section, we discuss basic solution methods for the affine multistage robust UC problem (8). We first discuss the traditional approach using duality theory and point out its limitation in solving large-scale robust optimization problems. Then, we present a constraint generation framework as the basis for further algorithmic improvements developed in this paper. Then, we close this section with some discussion.

4.1. Duality-Based Approach

The robust constraints in (8b)-(8f) have the following structure:

$$\boldsymbol{c}(\boldsymbol{W})^\top \boldsymbol{d} \le h(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},z) \quad \forall \boldsymbol{d} \in \mathcal{D}, \tag{11}$$

where $\boldsymbol{c}(\boldsymbol{W})$ and $h(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},z)$ are affine functions of the respective decision variables. To simplify notations, we write (11) as $\boldsymbol{c}^\top \boldsymbol{d} \le h$ for all $\boldsymbol{d} \in \mathcal{D}$. This robust constraint can be reformulated by using linear programming duality. In particular, (11) is equivalent to $\max_{\boldsymbol{d}\in\mathcal{D}} \boldsymbol{c}^\top \boldsymbol{d} \le h$. Since the uncertainty set $\mathcal{D}$ is a polytope, the maximization problem always attains a finite optimum, therefore, the maximization problem can be replaced by the dual minimization problem. Thus, (11) is equivalent to $\min_{\boldsymbol{\pi}\in\Pi(\boldsymbol{c})} \boldsymbol{b}^\top \boldsymbol{\pi} \le h$, where $\boldsymbol{b}$ comes from the definition of $\mathcal{D}$ and $\Pi(\boldsymbol{c})$ is a polyhedron that depends on $\boldsymbol{c}$. With this, (11) is further equivalent to the existence of $\boldsymbol{\pi} \in \Pi(\boldsymbol{c})$ such that $\boldsymbol{b}^\top \boldsymbol{\pi} \le h$. In this way, (11) is reformulated as a finite number of linear constraints involving dual variables. This duality-based approach is general and widely used in reformulating robust constraints, see the book by Ben-Tal et al. (2009a). For our problem, the deterministic counterpart of (11) with uncertainty set (3) is given below.

PROPOSITION 2. *The robust constraint* $\boldsymbol{c}^\top \boldsymbol{d} = \sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{N}_d} c_j^t d_j^t \le h \ \forall \boldsymbol{d} \in \mathcal{D}$*, where* $\mathcal{D}$ *is given by* (3)*, is equivalent to the existence of a vector of dual variables* $\boldsymbol{\pi}$ *that satisfies the following linear constraints:*

$$\sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{N}_d} \left[\overline{d}_j^t \pi_{jt}^1 - \underline{d}_j^t \pi_{jt}^2 + (\Gamma \hat{d}_j^t - \overline{d}_j^t)\pi_{jt}^3 + (\Gamma \hat{d}_j^t + \underline{d}_j^t)\pi_{jt}^4 \right] + \sum_{t\in\mathcal{T}} \Gamma \sqrt{N_d} \pi_t^5 \le h \tag{12a}$$

$$\pi_{jt}^1 - \pi_{jt}^2 - \pi_{jt}^3 + \pi_{jt}^4 = c_j^t \quad \forall j \in \mathcal{N}_d, t \in \mathcal{T} \tag{12b}$$

$$-\hat{d}_j^t \pi_{jt}^1 - \hat{d}_j^t \pi_{jt}^2 + \pi_t^5 = 0 \quad \forall j \in \mathcal{N}_d, t \in \mathcal{T} \tag{12c}$$

$$\pi_{jt}^1, \pi_{jt}^2, \pi_{jt}^3, \pi_{jt}^4, \pi_t^5 \ge 0 \qquad \forall j \in \mathcal{N}_d, t \in \mathcal{T}. \tag{12d}$$

Each robust constraint in (8b)-(8f) can be replaced by a set of equivalent deterministic constraints defined in (12a)-(12d) for the corresponding $\boldsymbol{c}$ and h. Notice that we need to introduce a respective vector $\boldsymbol{\pi}$ of dual variables for each of these robust constraints. The size of the resulting MIP reformulation is very large. In the affine multistage robust UC (8), there are $1 + 2T(2N_g + N_l)$ robust constraints, each requiring a vector $\boldsymbol{\pi}$ of dimension up to $(4N_d + 1)T$ if the W_{itjs}-policy is used or $(4N_d + 1)$ if the W_{it}-policy is used. Table 1 shows the number of variables required in the respective MIPs. Even with the W_i-policy or W_{it}-policy, the resulting MIP is too large to solve for moderate-sized power systems. For example, for a 2736-bus test case considered in Section 6, using the W_{it}-policy under this method would lead to more than 250 million $\boldsymbol{\pi}$-variables. We need a solution method that is more scalable.

Table 1 Number of variables in the MIPs obtained directly using the duality-based approach.

Policy structure	W_i	W_{it}	W_{itjs}
Binary variables	$3N_gT$	$3N_gT$	$3N_gT$
$(\boldsymbol{w}, \boldsymbol{W})$-variables	$N_gT + N_g$	$2N_gT$	$N_gT + N_dN_gT(T+1)/2$
$\boldsymbol{\pi}$-variables	$8N_d(2N_g + N_l)T$	$8N_d(2N_g + N_l)T$	$4N_d(2N_g + N_l)T(T+1)$

4.2. Constraint Generation

Since the constraints in the affine multistage robust UC model have the form of (11), where the left-hand side is a linear function in $\boldsymbol{d}$ and the uncertainty set $\mathcal{D}$ is a polytope, each robust constraint is equivalent to an enumeration of the finitely many extreme points of the uncertainty set, in the following form:

$$\boldsymbol{c}^\top \boldsymbol{d} \leq h \qquad \forall\, \boldsymbol{d} \in \operatorname{ext}(\mathcal{D}), \tag{13}$$

where $\operatorname{ext}(\mathcal{D}) = \{d_1^*, ..., d_N^*\}$ is the set of extreme points of $\mathcal{D}$ (see Ben-Tal et al. (2009b)). This applies to every robust inequality in the affine multistage robust UC model. Furthermore, the energy balance equality constraints in the affine multistage robust UC model can be reformulated using the full-dimensionality of the uncertainty sets (equivalently, the existence of an interior point).

Proposition 3. *For a full-dimensional uncertainty set $\mathcal{D}$, the robust energy balance equation* (8g) *of the W_i-policy and W_{it}-policy is equivalent to the following equalities*

$$W_i\text{-}Policy: \quad \sum_{i\in\mathcal{N}_g} w_i^t = 0, \quad \sum_{i\in\mathcal{N}_g} W_i = 1 \qquad \forall t \in \mathcal{T} \tag{14}$$

$$W_{it}\text{-}Policy: \quad \sum_{i\in\mathcal{N}_g} w_i^t = 0, \quad \sum_{i\in\mathcal{N}_g} W_{it} = 1 \quad \forall t \in \mathcal{T}. \tag{15}$$

With the above observations, we can reformulate the multistage affine robust UC model (8) in the following compact form:

$$\min_{\boldsymbol{y}\in Y} \quad f(\boldsymbol{y}) \tag{16a}$$

$$\text{s.t.} \quad g_k(\boldsymbol{y}, \boldsymbol{d}) \leq 0 \qquad \forall \boldsymbol{d} \in \operatorname{ext}(\mathcal{D}), \quad \forall k \in \{1, \dots, K\}, \tag{16b}$$

where $\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{W}, z)$ includes all decision variables in (8), the objective $f(\boldsymbol{y})$ represents (8a), and the set Y in (16a) is defined by (1b)-(1g) and (14) or (15) according to the policy structure used. Constraints (16b) represent (8b)-(8f), where $g_k(\boldsymbol{y}, \boldsymbol{d})$ is a bilinear function in $\boldsymbol{y}$ and $\boldsymbol{d}$, and $K = 1 + 2T(2N_g + N_l)$ in (16b) is the total number of robust constraints.

This reformulation suggests a constraint generation framework. It starts with an initial set of extreme points for each robust constraint, and at each iteration, finds the worst-case scenario $\boldsymbol{d}$ for each robust constraint that achieves the highest constraint violation and adds it to the master problem, which is defined as

$$(MP) \quad \begin{aligned} \min_{\boldsymbol{y}\in Y} \quad & f(\boldsymbol{y}) \\ \text{s.t.} \quad & g_k(\boldsymbol{y}, \boldsymbol{d}) \leq 0 \quad \forall\, \boldsymbol{d} \in D_k, \quad \forall\, k \in \{1, ..., K\}, \end{aligned} \tag{17}$$

where $D_k \subseteq \operatorname{ext}(\mathcal{D})$ is the list of extreme points that are identified from the constraint generation procedure for each robust constraint k in (16b). The constraint generation framework is outlined in Algorithm 1.

Algorithm 1 Constraint generation

```
Start with some initial D_k for all k ∈ {1, ..., K}
repeat
    y' ← optimal solution of the Master Problem (17).
    for all k ∈ {1, ..., K} do
        d_k ← argmax_{d∈𝒟} g_k(y', d)
        If g_k(y', d_k) > 0 let D_k ← D_k ∪ {d_k}
    end for
until g_k(y', d_k) ≤ 0 for all k ∈ {1, ..., K}
output: y' is an optimal solution to (16)
```

PROPOSITION 4. *The constraint generation algorithm presented in Algorithm 1 for solving the affine multistage robust UC problem* (8) *with uncertainty sets defined in* (3) *converges to the global optimum or reports infeasibility in a finite number of steps.*

Proof: The finite convergence follows from the fact that the uncertainty sets in (3) are bounded polyhedrons with a finite number of extreme points. □

4.3. Discussion

The constraint generation framework of Algorithm 1 can be also viewed as an embodiment of the cutting-plane method. A similar framework has been used in solving infinitely constrained optimization problems (see e.g. Blankenship and Falk (1976)). Fischetti and Monaci (2012) studied the computational performance of a similar cutting-plane algorithm for solving static robust integer and linear programs with uncertainty in the constraint coefficients and using budgeted uncertainty sets. They find that the cutting-plane algorithm is more efficient than the duality-based approach for solving uncertain linear programs, and is less efficient when the problem involves integer decisions. Recently, Bertsimas et al. (2014) extended this comparison to different types of uncertainty sets including ellipsoidal uncertainty sets and explored different algorithmic strategies. Similar conclusions are reached. Bertsimas and Georghiou (2014) presented a solution method based on constraint generation for adaptive optimization problems with a specific type of decision rule that can handle adaptive integer variables. Despite these interesting works, the computational study of solving large-scale multistage robust optimization problems with mixed-integer decisions still seems to be at an early stage.

In comparison to the above mentioned works, the affine multistage robust UC problem has some special characteristics such as high dimensionality in the numbers of continuous and integer variables and constraints, and also specific structures that can be exploited. Another point worth making is that the constraint generation framework of Algorithm 1 is equivalent to applying a Benders decomposition procedure to the MIP obtained with the duality-based approach, i.e., the feasibility cut generated through Benders decomposition on the dual system is equivalent to the primal constraint generated by constraint generation. As we will show in our experiments, the duality-based approach fails to solve even moderate sized problems with the simplest affine policy structure due to the scalability issue. Constraint generation, or the cutting-plane method, becomes necessary to handle this problem. However, a direct implementation of Algorithm 1 also has limited success. Exploiting the special structure of the affine multistage robust UC model is crucial to devise an efficient constraint generation method.

5. Algorithmic Improvements

The constraint generation framework summarized in Algorithm 1 is still not efficient enough to handle large-scale problems. However, it does provide a basis for further algorithmic improvements, which proves to be critical in making the large-scale affine multistage robust UC model efficiently solvable. In the following, we develop an efficient procedure for the separation problem, an effective initialization for the master problem, a method to reduce the number of MIPs solved in the algorithm, and formulations to fully exploit the special structure of the W_i-policy and W_{it}-policy. We would also like to remark that the constraint generation framework with the proposed algorithmic improvements are not restricted to solving the robust UC problem, but can be applied to solve general multistage robust optimization problems with affine policies.

5.1. Efficient Separation Procedure

The separation procedure in the constraint generation algorithm involves solving the problem

$$\max_{\boldsymbol{d}\in\mathcal{D}} g_k(\boldsymbol{y},\boldsymbol{d}) \tag{18}$$

for each robust constraint k in (8), in each iteration of the master problem. Thus, it is important to solve it as fast as possible. We can exploit two special structures of (18). First, as discussed above, $g_k(\boldsymbol{y},\boldsymbol{d})$ is a linear function in $\boldsymbol{d}$ for any fixed $\boldsymbol{y}$. Second, the structure of the budgeted uncertainty set (3) allows us to solve the separation problem (18) by a simple sorting procedure, as we show below.

Proposition 5. *Consider the separation problem* $\max_{\boldsymbol{d}\in\mathcal{D}} \boldsymbol{c}^\top \boldsymbol{d}$*, where the uncertainty set* $\mathcal{D}$ *is defined in* (3)*. An optimal solution for this problem is given by* $(d_j^s)^* = \overline{d}_j^s + \Gamma \hat{d}_j^s (u_j^s)^*$ *for each* $s \in \mathcal{T}$*, where* $(u_j^s)^*$ *is obtained by the following procedure: let* $\{|c^s_{\sigma(j)}\, \hat{d}^s_{\sigma(j)}|\}_{j\in\mathcal{N}_d}$ *be a non-increasing* *ordering of* $\{|c_j^s \hat{d}_j^s|\}_{j\in\mathcal{N}_d}$*, where* $\sigma(\cdot)$ *determines the indices of the non-increasing order, and* $(u_j^s)^*$ *is given as follows:*

$$(u^s_{\sigma(j)})^* = \begin{cases} \text{sign}(c^s_{\sigma(j)}) & \textit{if } \sigma(j) \leq \lfloor\sqrt{N_d}\rfloor, \\ (\sqrt{N_d} - \lfloor\sqrt{N_d}\rfloor)\cdot \text{sign}(c^s_{\sigma(j)}) & \textit{if } \sigma(j) = \lfloor\sqrt{N_d}\rfloor + 1, \\ 0 & \textit{if } \sigma(j) \geq \lfloor\sqrt{N_d}\rfloor + 2, \end{cases}$$

where $\text{sign}(x) = 1$ *if* $x \geq 0$ *and* -1 *otherwise.*

5.2. Initialization with Specific Uncertainty Scenarios

The constraint generation approach consists of iteratively finding extreme points of the uncertainty sets for each robust constraint until all robust constraints are satisfied. If there are extreme points that we believe to be strong candidates for being violated at some point in the constraint generation procedure, it would be useful to add them in the beginning.

Consider the vector $\boldsymbol{d}_{max}$ that achieves the maximum total net load in each time period, i.e., each component of $\boldsymbol{d}_{max}$ is defined as

$$\boldsymbol{d}^t_{max} \in \text{argmax}_{\boldsymbol{d}\in\mathcal{D}} \sum_{j\in\mathcal{N}_d} d^t_j \quad \forall\, t \in \mathcal{T}. \tag{19}$$

This net load vector is clearly an important scenario in the uncertainty set for determining the worst-case dispatch costs. Thus, to speed up the constraint generation algorithm, we add $\boldsymbol{d}_{max}$ to D_k in the worst-case dispatch cost constraint (8b).

Similarly, we consider the minimum total net load $\boldsymbol{d}_{min} \in \mathcal{D}$ for (8b), which is defined as

$$\boldsymbol{d}^t_{min} \in \text{argmin}_{\boldsymbol{d}\in\mathcal{D}} \sum_{j\in\mathcal{N}_d} d^t_j \quad \forall\, t \in \mathcal{T}. \tag{20}$$

We can also add $\boldsymbol{d}_{min}$ and $\boldsymbol{d}_{max}$ to D_k for every k representing the generation upper and lower bound constraints (8c).

For robust ramping constraints (8d) and (8e), consider the following scenarios for each t:

$$\boldsymbol{d}_{minmax}(t) = \left(\boldsymbol{d}^1_{min}, ..., \boldsymbol{d}^{t-1}_{min}, \boldsymbol{d}^t_{max}, ..., \boldsymbol{d}^T_{max}\right), \tag{21}$$

$$\boldsymbol{d}_{maxmin}(t) = \left(\boldsymbol{d}^1_{max}, ..., \boldsymbol{d}^{t-1}_{max}, \boldsymbol{d}^t_{min}, ..., \boldsymbol{d}^T_{min}\right), \tag{22}$$

which are the net loads with the largest up or down variations at period t. At initialization, we add $\boldsymbol{d}_{min}$, $\boldsymbol{d}_{max}$, $\boldsymbol{d}_{minmax}(t)$, $\boldsymbol{d}_{maxmin}(t)$ to D_k for every k representing the ramping constraints (8d)-(8e) at time t.

5.3. Complete Characterization for the W_{it}-Policy

The initialization technique in Section 5.2 is applicable to any affine policy. However, it has a very important consequence for the W_{it}-policy. Essentially, the robust constraints for generation limits and ramping can be completely characterized by a few uncertainty scenarios identified above, when using the W_{it}-policy. The computational benefit is huge.

Recall that the W_{it}-policy is described as $p_i^t(\boldsymbol{d}) = w_i^t + W_{it} \sum_{j \in \mathcal{N}_d} d_j^t$. When using the W_{it}-policy structure or any simpler policy such as the W_i-policy, generation output constraints (8c) and ramping constraints (8d)-(8e) are exactly equivalent to only considering the respective $\boldsymbol{d}$'s identified in (19)-(22), as we show below.

PROPOSITION 6. *Under the W_{it}-policy or any simpler policy, and using the uncertainty set in* (3)*, the following statements hold:*

(i) The robust constraints on generation limits (8c) *are equivalent to the ones with the uncertainty set $\mathcal{D}$ replaced by the finite set $\{\boldsymbol{d}_{min}^t, \boldsymbol{d}_{max}^t\}$, where $\boldsymbol{d}_{min}$ and $\boldsymbol{d}_{max}$ are defined in* (20) *and* (19)*, respectively.*

(ii) The robust constraints on ramping limits (8d)-(8e) *at time t are equivalent to the ones with the uncertainty set $\mathcal{D}$ replaced by the finite set $\{\boldsymbol{d}_{min}, \boldsymbol{d}_{max}, \boldsymbol{d}_{minmax}(t), \boldsymbol{d}_{maxmin}(t)\}$, where $\boldsymbol{d}_{minmax}(t)$ and $\boldsymbol{d}_{maxmin}(t)$ are defined in* (21) *and* (22)*, respectively.*

In the proof for ramping constraints we use the fact that the uncertainty set, given in (3), is separable over time periods.

This result implies that if we use the W_{it}-policy or any simpler policy such as the W_i-policy, the robust constraints corresponding to generation output limits and ramping capacities can be pre-computed before starting the constraint generation process. The only robust constraints left to deal with using constraint generation are the worst-case dispatch cost constraint (8b) and the transmission constraints (8f). This saves a tremendous amount of time checking feasibility and generating violated constraints. The overall convergence time of the constraint generation algorithm is significantly reduced.

5.4. Generating Multiple Cuts to the Master Problem

The difficulty in solving the affine multistage robust UC lies in finding all the necessary uncertainty scenarios $\boldsymbol{d}$'s for each robust constraint. This can lead to the undesired situation of solving the master problem (17) many times, which itself is a MIP with a large number of constraints. To strengthen the master problem, we employ a procedure that generates constraints by keeping all the binary variables fixed in the master problem. This can be helpful in reducing the number of MIP problems solved in the overall algorithm.

Algorithm 2 Generating multiple cuts for a fixed $\boldsymbol{x}'$

1: **input**: x', $\{D_k\}_{k=1}^K$
2: **repeat**
3: $\boldsymbol{y}' = (\boldsymbol{x}', \boldsymbol{u}', \boldsymbol{v}', \boldsymbol{w}', \boldsymbol{W}', z') \leftarrow$ optimal solution of the master problem (17) with $\boldsymbol{x} = \boldsymbol{x}'$ fixed
4: **for all** $k \in \{1, ..., K\}$ **do**
5: $\boldsymbol{d}_k \leftarrow \text{argmax}_{\boldsymbol{d} \in \mathcal{D}}\, g_k(\boldsymbol{y}', \boldsymbol{d})$
6: If $g_k(\boldsymbol{y}', \boldsymbol{d}_k) > 0$ let $D_k \leftarrow D_k \cup \{\boldsymbol{d}_k\}$
7: **end for**
8: **until** $g_k(\boldsymbol{y}', \boldsymbol{d}_k) \leq 0$ for all $k \in \{1, ..., K\}$
9: **output**: $\{D_k\}_{k=1}^K$

In particular, fix the commitment vector at the current solution $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})$ of the master problem, then the master problem becomes a linear program (LP) in the dispatch policy variables $(\boldsymbol{w}, \boldsymbol{W})$. Apply constraint generation to the resulting problem, starting from the current set of uncertainty scenarios $\boldsymbol{d}$'s until all the violated scenarios are identified for each robust constraint. This procedure is presented in Algorithm 2.

Furthermore, this technique can also be applied at the initialization phase of the overall constraint generation method. In particular, we can solve a static robust UC, which we define as a simplification of (8a)-(8g) by forcing $\boldsymbol{W} = \boldsymbol{0}$ and replacing robust energy balance constraints (8g) by enforcing it only for maximum total net load in the uncertainty set. This problem is very fast to solve and provides a good starting point for $\boldsymbol{x}$.

The concept of generating several cuts in each iteration of a constraint generation framework has been studied before with different formats. For example, Birge and Louveaux (1988) extended the L-shaped method for two-stage stochastic linear programs to a multicut version where each subproblem can induce a different cut. We make use of this idea in our algorithm, where each subproblem corresponds to checking the feasibility of a robust constraint. However, the enhancement presented here is different in that we proceed with the constraint generation algorithm solving an LP master problem with fixed binary variables, as many times as needed, inducing the fast generation of many “useful cuts” before solving each MIP master problem where binary variables are allowed to change. Another relevant idea that could be explored to enhance the algorithm is the concept of on-demand accuracy; see the work by de Oliveira and Sagastizábal (2014) and references therein. For example, some of the subproblems could be solved partially, and as the method develops the quality of their solutions could be increased as needed, potentially making the overall algorithm faster.

5.5. Algorithm Summary

The overall constraint generation algorithm with the above proposed algorithmic improvements is summarized in Algorithm 3. The initialization consists of finding $\boldsymbol{d}$'s described in Sections 5.2 and 5.3, and solving the static robust UC described in Section 5.4. Then the algorithm solves the master problem, and updates in each iteration the lists $\{D_k\}_{k=1}^K$ using each commitment solution found as described in Section 5.4.

Algorithm 3 Proposed solution method

$D_k \leftarrow \emptyset \quad \forall k = 1, 2, \ldots, K$
Add $\boldsymbol{d}$ from (19) to the D_k representing (8b)
Add $\boldsymbol{d}$'s from (19)-(20) to all D_k's representing (8c)
Add respective $\boldsymbol{d}$'s from (19)-(22) to all D_k's representing (8d)-(8e)
$\boldsymbol{x}' \leftarrow$ optimal solution of static robust UC
repeat
    Update $\{D_k\}_{k=1}^K$ using Algorithm 2 for $\boldsymbol{x}'$
    $\boldsymbol{y}' = (\boldsymbol{x}', \boldsymbol{u}', \boldsymbol{v}', \boldsymbol{w}', \boldsymbol{W}', z') \leftarrow$ optimal solution of (17)
    for all $k \in \{1, ..., K\}$ **do**
        $\boldsymbol{d}_k \leftarrow \text{argmax}_{\boldsymbol{d} \in \mathcal{D}} \, g_k(\boldsymbol{y}', \boldsymbol{d})$
        If $g_k(\boldsymbol{y}', \boldsymbol{d}_k) > 0$ let $D_k \leftarrow D_k \cup \{\boldsymbol{d}_k\}$
    end for
until $g_k(\boldsymbol{y}', \boldsymbol{d}_k) \leq 0$ for all $k \in \{1, ..., K\}$
output: $\boldsymbol{y}' = (\boldsymbol{x}', \boldsymbol{u}', \boldsymbol{v}', \boldsymbol{w}', \boldsymbol{W}', z')$ is an optimal solution for (16)

For simplicity, in our description of this algorithm we ignore the case where the master problem (17) reports infeasibility at some point. If such event ever occurs, the algorithm stops and reports infeasibility of the affine multistage robust UC problem under the affine policy used. Also, notice that checking for violated robust constraints can be parallelized, because it consists of solving K separate problems with the procedure described in Section 5.1.

6. Computational Experiments

We conduct extensive computational experiments on the IEEE 118-bus and the 2736-bus Polish systems (c.f. Zimmerman et al. (2011)). The major aspects of these instances are summarized in Table 2. In all cases, the UC problems involve a planning horizon of $T = 24$ hours. Uncertain net loads are located at every node with electricity demand. The uncertainty sets are given by (3), where we choose $\hat{d}_j^t = 0.1\overline{d}_j^t$ with various budget levels Γ, unless stated otherwise. All the experiments have been implemented using Python 2.7 in a PC laptop with an Intel Core i5 at 2.4 GHz and 4GB memory with CPLEX 12.5 as MIP and LP solver.

Table 2 Summary of test cases used

Buses	118	2736
Units	54	289
Loads	99	2011
Lines	186	100
Total generation capacity (MW)	7106	28880
Min total nominal net load (MW)	3327	10851
Max total nominal net load (MW)	4931	18075

Section 6.1 demonstrates the computational efficiency of the proposed algorithm. Section 6.2 shows that the simplified affine policies, as an approximation to the fully-adaptive policy, achieve close-to-optimal performance. Section 6.3 studies the impact of the UC solutions on the real-time dispatch operation from a worst-case perspective. In particular, it compares the worst-case performance of the real-time dispatch problem based on the UC solutions obtained from the two-stage robust UC model against those obtained from the affine multistage robust UC model. Section 6.4 studies the average performance of the affine multistage robust UC model in a rolling horizon simulation framework, and compares it with the deterministic and two-stage robust UC models.

6.1. Computational Performance of the Proposed Algorithm

In this section, we demonstrate the efficiency of the proposed solution methods for solving the affine multistage robust UC model in (8) with the W_{it}-policy structure. We show the efficiency enhancement achieved by individual algorithmic improvement techniques as well as the ultimate improvement achieved by their combination, and compare them with the two traditional solution methods, namely the duality-based approach (DBA) introduced in Section 4.1 and the basic constraint generation (CG) algorithm discussed in Section 4.2.

More specifically, we show the performance of the proposed algorithmic improvements in the following order. (a) The algorithm based on basic CG and Algorithm 2, which generates Multiple Cuts (MC) in each iteration for a fixed commitment solution (see Section 5.4). We denote this procedure as "CG + MC". (b) The algorithm based on basic CG and the method that exploits the Problem Structure (PS) of the W_{it}-policy (see Section 5.3). We denote this procedure as "CG + PS". (c) The combination of (a) and (b), denoted as "CG + MC + PS". (d) The combination of (a)(b)(c) along with the generation of an Initial Scenario (IS) of specific $\boldsymbol{d}$ for the worst-case dispatch cost constraint (see Section 5.2). This is the final solution algorithm summarized in Section 5.5. We denote it as "CG + MC + PS + IS".

All of the above four algorithms are implemented to solve the multistage robust UC model (8) with the W_{it}-policy on the 118-bus system. Table 3 shows the solution time (in seconds) of all these methods on the 118-bus system with different values of budget Γ for the uncertainty sets in (3).

The stopping criterion of 0.1% optimality gap is used for solving each MIP problem. A time limit of 15,000 seconds is imposed on each algorithm. "M" and "T" in Table 3 stand for out-of-memory and out-of-time limits, respectively.

Table 3 Solution time (seconds) of various algorithms for solving affine multistage robust UC under the W_{it}-policy for the 118-bus system.

Method	$\Gamma = 0.25$	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$
DBA	M	M	M	M	M	M
CG	T	T	T	T	T	T
CG + MC	6,807	8,475	5,639	3,488	10295	6,965
CG + PS	563	80	961	1,011	1183	1,227
CG + MC + PS	175	67	77	78	161	218
CG + MC + PS + IS	66	64	47	63	155	178

Notice that DBA and the basic CG are not efficient in solving the simple W_{it}-policy for the 118-bus system — either running out of memory or time limits. Applying the techniques of fixing the UC solution to find $\boldsymbol{d}$'s (CG + MC) or exploiting the policy structure (CG + PS) leads to a substantial improvement in solution times, especially when the special structure of the W_{it}-policy is exploited (CG + PS). When the two techniques are combined (i.e., CG + MC + PS), the solution times are reduced to within 218 seconds (less than 4 minutes) for all sizes of tested uncertainty sets, and even faster for problems with small uncertainty sets (around 1 minute). Running time is further reduced by initializing the algorithm with one more valid $\boldsymbol{d}$ for the worst-case dispatch cost constraint (CG + MC + PS + IS).

At this point, let us try to understand why the algorithm with "CG + MC + PS + IS" is significantly more efficient than DBA. For the medium-sized 118-bus system with the W_{it}-policy, the MIP obtained by directly applying DBA requires the creation of approximately 70 million dual variables and 35 million associated constraints for the explicit representation of all the robust constraints of the original formulation. If the special structure of the W_{it}-policy is exploited (which can also be combined with DBA) these numbers are reduced to approximately 3.5 million dual variables and 1.8 million associated constraints, which still runs into memory issues. In contrast, the algorithm with "CG + MC + PS + IS" requires initially creating about 15,000 constraints for policy structure exploitation, but does not require the creation of dual variables. Furthermore, if we take $\Gamma = 4$, a total number of 263 constraints are generated in the master problem along the algorithm, where a total of 5 MIPs and 7 LPs are solved (while these numbers are even smaller for the other Γ's tested). This significantly saves the computation time comparing to a naive CG method.

Algorithm "CG + MC + PS + IS", thus identified as the most efficient algorithm among the six tested methods, is applied to the 2736-bus Polish system. Table 4 presents the solution times of this algorithm for solving the multistage robust UC model with the W_{it}-policy structure, for different values of Γ. For the 2736-bus system, an optimality gap of 1% is used for all MIP problems solved in the algorithm. In Table 4, "inf" indicates that the algorithm detects the problem being infeasible, which is caused by the large size of the uncertainty set. The solution time variations for the 2736-bus system are explained by the variability in the time taken for the MIPs to be solved.

Table 4 Solution time using "CG + MC + PS + IS" algorithm for both systems studied under the W_{it}-policy.

System	$\Gamma = 0.25$	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$
118-bus	66s	64s	47s	63s	155s	178s
2736-bus	3.6h	3.2h	2.3h	2.0h	2.4h	0.4h (inf)

For the 2736-bus Polish system, when $\Gamma = 1$, a total number of 6 MIPs and 5 LPs are solved, and 727 constraints are generated by the proposed constraint generation algorithm. Similar numbers are obtained for the other values of Γ tested. In comparison, if DBA is used with exploitation of the special structure of the W_{it}-policy, the MIP obtained would require approximately 39 million dual variables and 19 million associated constraints for the explicit representation of all the robust constraints of the original formulation. Furthermore, without exploiting the structure of the W_{it}-policy, more than 250 million dual variables would be required in DBA.

From Table 4 we can see that the proposed algorithm can efficiently solve the real-world 2736-bus system within a time framework reasonable for the day-ahead operation. Considering the complexity of the multistage robust UC model and the simple computation resources (a moderate personal computer) that our experiments rely on, these computational experiments show that the affine multistage robust UC model and the proposed algorithms are very promising for practical applications in large-scale power system operations.

6.2. Optimality Gap for Simplified Affine Policies

The affine multistage robust UC model proposed in (8) is an approximation scheme to the original fully-adaptive multistage robust UC model (5). The UC solution and the affine dispatch policy thus obtained are feasible, but may not be optimal for the fully-adaptive model. In this section, we study the approximation quality of the simplified affine policies. As will be shown, affine policies with the very simple W_i-policy or W_{it}-policy perform surprisingly well as approximate solutions to the fully adaptive problem. This is a particularly encouraging result for the large-scale 2736-bus system.

6.2.1. Bounding the Approximation Quality of Affine Policies The two-stage robust UC formulation (2) is a relaxation of the fully adaptive multistage robust UC model (5) by ignoring non-anticipativity on dispatch decisions. Thus, the optimal objective value of the two-stage robust UC problem, denoted as v^*_{2S}, provides a *lower* bound to the optimal objective value of the fully adaptive multistage robust UC, denoted as v^*_{MS}. However, obtaining a globally optimal solution of the two-stage robust UC problem for large-scale power systems is still computationally challenging (e.g. see Bertsimas et al. (2013)). To reduce computation time, we employ the heuristic used by Lorca and Sun (2015), which generates a *lower* bound to v^*_{2S}, denoted as $\underline{v}_{2S}$. Furthermore, since the affine policy is an approximation to the fully adaptive policy, its optimal objective value, denoted as v^*_{AFF}, provides an *upper* bound to the optimal objective value of the fully adaptive multistage robust UC. Because the MIP solver is terminated within a certain accuracy (e.g. with a 0.1% MIP gap), the solution at termination gives a further upper bound to v^*_{AFF}, denoted as $\overline{v}_{AFF}$. In summary, we have the following relations between objective values of different solutions: $\underline{v}_{2S} \le v^*_{2S} \le v^*_{MS} \le v^*_{AFF} \le \overline{v}_{AFF}$. Then, the optimality gap between v^*_{AFF} and v^*_{MS}, i.e., $(v^*_{AFF} - v^*_{MS})/v^*_{MS}$, is upper bounded as

$$0 \le \frac{v^*_{AFF} - v^*_{MS}}{v^*_{MS}} \le \frac{\overline{v}_{AFF} - \underline{v}_{2S}}{\underline{v}_{2S}} \triangleq \text{Guaranteed Optimality Gap.}$$

We call the upper bound to the optimality gap the *guaranteed optimality gap* of the affine multistage robust UC problem.

6.2.2. Computational Results for Guaranteed Optimality Gap Table 5 presents the guaranteed optimality gaps of two simple affine policy structures for the 118-bus and 2736-bus systems with different values of the uncertainty set size parameter Γ.

Table 5 Guaranteed opt. gap under different policy structures ("inf" indicates infeasibility).

118-bus system

Policy	$\Gamma = 0.25$	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 1.5$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$
W_i	0.04%	0.02%	0.04%	0.08%	0.10%	0.26%	0.67%
W_{it}	0.04%	0.02%	0.03%	0.07%	0.07%	0.17%	0.35%

2736-bus system

Policy	$\Gamma = 0.25$	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 1.5$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$
W_i	0.09%	0.22%	0.42%	0.55%	1.05%	inf	inf
W_{it}	0.07%	0.11%	0.25%	0.35%	0.53%	0.94%	inf

From these results, we offer the following observations.

1. For each test system, the W_{it}-policy achieves a better guaranteed optimality gap than the W_i policy, especially for larger uncertainty sets. For example, for the 2736-bus system with $\Gamma = 2$, the guaranteed optimality gap is improved from 1.05% of the W_i-policy to 0.53% by the W_{it}-policy. For smaller uncertainty sets, the W_i-policy has a more comparable performance to the W_{it}-policy.
2. The simple W_{it}-policy achieves surprisingly good performance in both test cases. The guaranteed optimality gap is at most 0.94% for all sizes of uncertainty sets in both test systems. Due to its strong performance and computational tractability, we will use the W_{it}-policy in all the following experiments.

6.3. Worst-Case Performance Analysis

As discussed in Section 2.3, the proposed multistage robust UC formulation is motivated by a critical issue of the two-stage robust UC model, namely that it ignores non-anticipativity in the dispatch process for the sequential revelation of uncertain net loads, and thus may not be prepared in real-time operations for all realizations of net loads within the uncertainty set. Indeed, Claims 1 and 2 in Section 2.3 show that this is possible based on a simple two-bus example. This section will further study this issue on the 118-bus and the 2736-bus systems. In particular, we want to estimate "how much" infeasibility can be caused in the real-time dispatch under the commitment solutions of the two-stage robust UC model. For this purpose, the two-stage model is solved for different sizes of the uncertainty sets, then the obtained UC solutions are fed into the affine multistage robust model (8). That is, the UC decision in (8) is fixed at the two-stage UC solution, and the remaining affine multistage robust dispatch problem is solved. The dispatch model is properly augmented with penalty variables in the energy balance and transmission constraints, so that the degree of infeasibility can be quantified by the amount of penalty costs incurred (see Section EC.2 in the Electronic Companion for details on the penalty variables). In this way, we can compare the worst-case operational costs (including penalty costs) of the real-time dispatch under the two-stage robust UC solutions against those obtained under the affine multistage robust UC solutions. It is important to carry out this type of worst-case performance study of the real-time dispatch under different UC solutions, because power system operations require extremely high reliability. Infeasibility in real-time operation has to be resolved by starting expensive fast-start units or shedding load, both of which bear significant economic consequences.

Table 6 presents the results. "Total Cost" is the worst-case dispatch cost plus penalty cost of the affine multistage robust dispatch model under a specific UC solution. "Penalty" is the total penalty cost associated with constraint violations in the dispatch model, where \$5000/MW is used as the unit penalty cost. "Rel Diff" is the relative difference between the total costs obtained by the multistage and two-stage UC solutions. We can make the following observations.

1. The multistage UC solutions do not cause any infeasibility in real-time operation for $\Gamma \leq 3$, whereas even though the two-stage UC model is feasible in itself for both 118-bus and 2736-bus systems, its UC solutions cause infeasibility to the multistage robust dispatch and incur quite significant penalties in the real-time operation.

2. The penalty costs and the total costs of the two-stage UC solutions increase as the size of the uncertainty set grows. For the 118-bus system, the two-stage model has 62.87% more total cost than the multistage model at $\Gamma = 3$, and the penalty cost is over \$1.2M. For the 2736-bus system, the two-stage UC model incurs 25.70% more total cost than the multistage model at $\Gamma = 3$, and the absolute amount of penalty cost exceeds \$2.7M.

These results further demonstrate the importance of non-anticipative constraints and the multistage robust UC model in power system operations.

Table 6 Worst-case cost (US\$) of multistage robust dispatch under the two-stage and multistage UC solutions. Multistage models use the W_{it}-policy.

118-bus system	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 1.5$	$\Gamma = 2$	$\Gamma = 3$
Affine multistage UC solutions					
Total Cost	1,696,304	1,725,470	1,755,398	1,784,543	1,845,218
Penalty	0	0	0	0	0
Two-stage UC solutions					
Total Cost	1,696,456	1,749,766	1,797,503	1,897,212	3,005,290
Penalty	0	52,501	55,268	196,101	1,229,300
Rel Diff	0.01%	1.41%	2.40%	6.31%	62.87%

2736-bus system	$\Gamma = 0.5$	$\Gamma = 1$	$\Gamma = 1.5$	$\Gamma = 2$	$\Gamma = 3$
Affine multistage UC solutions					
Total Cost	9,445,069	9,596,788	9,746,685	9,905,527	10,234,459
Penalty	0	0	0	0	0
Two-stage UC solutions					
Total Cost	9,505,651	9,745,889	10,183,433	10,975,403	12,864,719
Penalty	96,313	224,952	591,661	1,165,324	2,703,522
Rel Diff	0.64%	1.55%	4.49%	10.80%	25.70%

We have the following further discussion. In the above experiment, the same uncertainty set sizes Γ are used in the two-stage and multistage UC models. It is also interesting to test if the two-stage UC solution would perform better in the multistage dispatch if the two-stage UC model uses a larger value of Γ than the Γ later used in the multistage dispatch problem. In this way, a larger Γ might "compensate" the two-stage UC solution for its lack of non-anticipativity. For this purpose, we feed the two-stage UC solutions obtained using $\Gamma = 3$ to the multistage dispatch problem with

$\Gamma = 0.5, 1, 1.5, 2$. For the 2736-bus system, the worst-case costs of the total UC costs thus obtained are respectively 1.35%, 1.04%, 0.84% and 0.66% higher than those obtained by directly solving the affine multistage robust UC problem, which indeed are better than the performance reported in Table 6. From this we can see that if the two-stage robust UC is solved under a conservative "over-robustness" request, better solutions can be obtained than using smaller Γ's, however the performance is still not as cost-effective as those obtained by directly solving the multistage robust UC problem. Similar results are obtained for the 118-bus system.

6.4. Average Performance of UC Models in Real-Time Dispatch

In the previous section, we have conducted a worst-case analysis to compare the two-stage and multistage robust UC models. In this section, we study the average performance of different UC solutions and their impact on real-time dispatch. We develop a rolling-horizon simulation platform to mimic the real time operation of the power system, where information about uncertain net load is revealed sequentially as time moves forward. On this platform, we conduct Monte-Carlo simulations of different economic dispatch (ED) models that are suitable for the associated UC solution concepts. In particular, we propose a new robust ED model that exploits the affine policy obtained from the multistage robust UC model. For the two-stage robust UC and the deterministic UC solutions, we use a multi-period ("look-ahead") deterministic ED model in simulation, which has started to be adopted in some ISO markets (the most prevalent in practice is still the single-period ED model) ((FERC 2014, Table 4), Navid and Rosenwald (2012)).

6.4.1. Efficient Robust Dispatch Model Exploiting Affine Policy. The proposed robust ED model is motivated by the following considerations. First, solving the affine multistage robust UC model not only produces a UC solution, but also provides an affine policy that could be exploited in the ED process. Second, any ED model needs to be solved fast within a few minutes in real-time operation.

With these considerations, we propose a new robust dispatch model in (23), which we call the *policy-enforcement* robust ED model. At each time t, the dispatch decision $\boldsymbol{p}^t$ is the first-stage decision, which satisfies all the dispatch constraints $\Omega_t(\boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{p}^{t-1})$ in the current period and will be implemented "right now" at time t. Furthermore, the policy-enforcement robust ED model also considers the next period's dispatch decision $\boldsymbol{p}^{t+1}$ and assumes that it takes the form of the affine policy with coefficients $(w_i^{t+1}, \boldsymbol{W}_i^{t+1})$ of time $t+1$ obtained from the day-ahead affine multistage robust UC model.

$$\min_{\boldsymbol{p}^t} \quad \sum_{i\in\mathcal{N}_g} C_i p_i^t \tag{23a}$$

$$\text{s.t.} \quad \boldsymbol{p}^t \in \Omega_t(\boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{p}^{t-1}) \tag{23b}$$

$$w_i^{t+1} + \boldsymbol{W}_i^{t+1}\boldsymbol{d}^{t+1} - p_i^t \geq -RD_i x_i^{t+1} - SD_i v_i^{t+1} \quad \forall \boldsymbol{d}^{t+1} \in \mathcal{D}^{t+1} \tag{23c}$$

$$w_i^{t+1} + \boldsymbol{W}_i^{t+1}\boldsymbol{d}^{t+1} - p_i^t \leq RU_i x_i^t + SU_i u_i^{t+1} \quad \forall \boldsymbol{d}^{t+1} \in \mathcal{D}^{t+1}. \tag{23d}$$

Here, $\Omega_t(\boldsymbol{x}, \boldsymbol{d}^t, \boldsymbol{p}^{t-1})$ includes all the dispatch related constraints in the deterministic UC model (1) at time t, with the observed values of the current period's net load vector $\boldsymbol{d}^t$ and the previous period's dispatch level $\boldsymbol{p}^{t-1}$. Constraints (23d) and (23c) enforce ramping limits between $\boldsymbol{p}^t$ and $\boldsymbol{p}^{t+1}$ for any realization of nodal net loads in the uncertainty set at time $t+1$. In this way, the proposed dispatch model coordinates the ramping capabilities in the two consecutive periods and hedges against unfavorable net load realizations in future periods.

It is important to note that we can also consider a multi-period model where affine policies obtained from the multistage robust UC model for all future periods $t+1, t+2, \ldots$ are used. However, this multi-period model is exactly equivalent to the above two-period model, because the affine policies obtained from the robust UC model already satisfy all the dispatch constraints in each future period as well as the ramping constraints coupling every two consecutive periods. Also notice that the above robust ED model has almost the same size as a deterministic single-period ED, since we can use the strategy in Section 5.3 to handle robust ramping constraints.

For the deterministic and two-stage robust UC solutions, there is no affine policy readily available to exploit. Instead, we use the deterministic multi-period look-ahead ED model for their dispatch simulation, where net loads in future periods use forecast values (i.e., the nominal $\overline{d}_j^t$ values), and the ED model is obtained from the deterministic UC model (1) by fixing the commitment decision.

6.4.2. Rolling-Horizon Simulation Platform for Real-Time Dispatch. We develop a rolling horizon platform to simulate the real-time dispatch process. In particular, for each UC solution, we select an ED model according to the discussion in Section 6.4.1. At each time period t in the simulation, the selected ED model is solved with the observation of nodal net load up to time t, and the dispatch solution of time period t is implemented. Then the time horizon rolls forward and the same procedure is repeated. This simulation process is different from the existing ones in the literature such as in Bertsimas et al. (2013), Jiang et al. (2012), Zhao and Zeng (2012), where net loads over the entire scheduling horizon are revealed all at once to the dispatch model, ignoring non-anticipativity.

We consider a 24-hour horizon with an hourly step size in the simulation process. At each time t, the robust ED model in (23) considers two periods t and $t+1$, i.e., a one period look-ahead, whereas the deterministic look-ahead ED model considers 4 periods, i.e. a three periods look-ahead. The look-ahead horizon shrinks in the last three periods. Each round of the rolling-horizon simulation contains $T = 24$ consecutive runs of the ED model through the entire horizon. For each UC solution and the corresponding ED model, we carry out multiple rounds of such simulations. In the uncertainty set of the robust UC problems solved, we used $\overline{d}_j^t$ and $\hat{d}_j^t$ selected equal to the expected value and standard deviation, respectively, of net load at bus j and time t. The same set of nodal net load trajectories are used in all evaluations of different UC solutions to generate a fair comparison. Penalty variables are incorporated to deal with violations of energy balance and transmission, all of which have a unit penalty cost of \$5000/MWh. Due to space restriction, we only show the results for the 2736-bus system.

6.4.3. Results for the 2736-bus system with temporally independent nodal net loads. In the experiments presented in this subsection, the nodal net load corresponds to demand sampled from a normal distribution with a standard deviation equal to 10% of its expected value (i.e., $\hat{d}_j^t = 0.10\bar{d}_j^t$) and independent across time periods. In order for the uncertainty set (3) to contain the entire ℓ_∞-ball of $\Pi_{j\in\mathcal{N}_d}[\bar{d}_j^t - \hat{d}_j^t, \bar{d}_j^t + \hat{d}_j^t]$, the budget parameter Γ has to be $\sqrt{N_d} = \sqrt{2011} = 44.8$. This corresponds to an extremely conservative robust solution. We want to choose a Γ value that is small enough to still guarantee a robust enough performance. The following experiments use $\Gamma \leq 3.0$, which corresponds to significantly smaller uncertainty sets and less conservative solutions. In each experiment, 1000 rounds of simulation are conducted.

Table 7 presents the simulation performance of the multistage robust UC solution with the W_{it}-policy and the corresponding policy-enforcement robust ED model, the two-stage robust UC solution with the deterministic look-ahead ED model, and a deterministic reserve approach, for the 2736-bus system. We compare the average total costs over the 24-hour horizon ("Cost Avg"), their standard deviation ("Cost Std"), the average penalty costs ("Penalty Avg"), and the average frequency of penalty occurrence ("Penalty Freq Avg"). We also study the performance of the deterministic UC model with adjusted reserve and look-ahead ED in the rolling-horizon simulation, which resembles the current operational practice. The reserve adjustment follows the rule used in Bertsimas et al. (2013) with various reserve levels tested.

From these results we can see that the multistage robust UC model achieves the best average total cost at $\Gamma = 0.5$, which is a 0.46% ((9319396 − 9362379)/9362379) reduction from the best average cost of the two-stage robust UC model achieved at $\Gamma = 3$, and a 0.95% reduction from that of the deterministic UC with reserve adjusted at 20%. Further comparing these three columns, we can see that the multistage robust UC solution achieves a significant improvement on system reliability, with a cost standard deviation reduced by 64.97% from the two-stage solution and

Table 7 Simulation performance of the different models for the 2736-bus system with temporally independent demand.

Affine multistage robust UC with policy-enforcement robust ED						
Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	9,397,528	9,319,396	9,342,754	9,360,359	9,379,464	9,442,858
Cost Std ($)	113,725	15,970	12,828	12,509	12,363	12,092
Penalty Cost Avg ($)	93,552	3497	727	61	5	0
Penalty Freq Avg	10.00%	1.47%	0.40%	0.01%	0.00%	0.00%

Two-stage robust UC with look-ahead ED						
Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	9,398,109	9,456,599	9,408,732	9,383,569	9,407,290	9,362,379
Cost Std ($)	93,470	195,774	173,884	144,698	162,469	45,584
Penalty Cost Avg ($)	80,127	152,637	98,113	66,801	82,864	6,103
Penalty Freq Avg	9.93%	12.26%	7.80%	5.11%	5.57%	0.37%

Deterministic UC with reserve and look-ahead ED						
Reserve	2.5%	5%	10%	15%	20%	30%
Cost Avg ($)	9,556,549	9,575,446	9,424,678	9,561,024	9,408,173	9,411,741
Cost Std ($)	261,464	288,777	121,122	196,354	92,268	69,050
Penalty Cost Avg ($)	254,627	271,672	119,127	248,658	83,938	51,907
Penalty Freq Avg	15.93%	13.37%	14.31%	18.16%	10.03%	7.22%

82.69% from the deterministic UC with reserve. Moreover, the penalty cost of the multistage robust UC solution is reduced by 42.70% and 98.43% from the two-stage robust UC and deterministic UC solutions, respectively. The penalty cost can be reduced to zero by a larger value of Γ in the multistage model, whereas both the two-stage robust UC and deterministic UC do not achieve zero penalty for all tested budget and reserve levels.

6.4.4. Results for the 2736-bus system with persistent demand. Here we present simulation results for the 2736-bus system where nodal net corresponds to demand sampled from a persistent model (Hamilton 1994), which exhibits some simple temporal correlation. In particular, we sample the trajectory of demand at bus j from the following autoregressive model:

$$\widetilde{d}_j^t = \mu_j^t + \sigma_j^t Z_j^t \quad \forall t \in \mathcal{T}$$
$$Z_j^t = \phi Z_j^{t-1} + \epsilon_j^t \quad \forall t \in \mathcal{T},$$

where the ϵ_j^t's are sampled independently from a normal distribution with an expected value of 0 and a standard deviation of $\sigma^\epsilon = \sqrt{1-\phi^2}$, and Z_j^0 is sampled from a normal distribution with an expected value of 0 and a standard deviation of 1. The value of σ^ϵ is selected so that the standard deviation of Z_j^t is always 1. Further, $\sigma_j^t = 0.1\mu_j^t$. Given this, $\widetilde{d}_j^t$ has an expected value of μ_j^t and a standard deviation of $0.1\mu_j^t$, and the correlation between $\widetilde{d}_j^t$ and $\widetilde{d}_j^{t-1}$ is ϕ. In the uncertainty set, we take $\overline{d}_j^t = \mu_j^t$ and $\hat{d}_j^t = \sigma_j^t = 0.1\mu_j^t$. The setting in Section 6.4.3 corresponds to the case $\phi = 0$. In this section we use $\phi = 0.9$ denoting a strong temporal correlation in demand.

Table 8 presents the simulation performance results for the different UC approaches studied. We can observe that the best average cost is still achieved at $\Gamma = 0.5$ for the multistage robust UC, at $\Gamma = 3$ for the two-stage robust UC, and at 20% reserve for the deterministic UC. For these cases, the multistage robust UC presents a 0.39% and 0.84% reduction in average cost with respect to the two-stage robust UC and the deterministic UC, respectively. We can also observe that with a proper choice for Γ the multistage robust UC can completely eliminate the penalty cost, while the two-stage robust UC and deterministic UC could not completely eliminate penalty under the tested values of Γ and reserve.

Table 8 Simulation performance of the different models for the 2736-bus system with persistent demand.

Affine multistage robust UC with policy-enforcement robust ED						
Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	9,395,199	9,320,462	9,343,907	9,361,336	9,380,499	9,443,846
Cost Std ($)	180,122	51,226	42,344	41,229	41,295	40,387
Penalty Cost Avg ($)	90,292	3,464	835	8	0	0
Penalty Freq Avg	10.11%	1.23%	0.43%	0.01%	0.00%	0.00%
Two-stage robust UC with look-ahead ED						
Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	9,390,163	9,409,159	9,380,783	9,386,101	9,362,994	9,357,111
Cost Std ($)	118,856	234,627	187,981	164,032	130,666	43,809
Penalty Cost Avg ($)	71,238	10,3861	69,188	60,483	45,171	5
Penalty Freq Avg	9.35%	11.68%	7.05%	5.30%	5.23%	0.01%
Deterministic UC with reserve and look-ahead ED						
Reserve	2.5%	5%	10%	15%	20%	30%
Cost Avg ($)	9,525,854	9,565,603	9,415,143	9,515,113	9,398,984	9,409,452
Cost Std ($)	372,593	369,799	163,575	271,478	131,586	108,410
Penalty Cost Avg ($)	222,972	261,016	108,863	201,491	73,849	48,801
Penalty Freq Avg	15.05%	13.08%	13.58%	16.45%	8.83%	6.11%

6.4.5. Results for the 2736-bus system using temporally correlated wind data. In the above simulations we sampled demand from specific distributions, first assuming temporal independence and then incorporating temporal correlations through an autoregressive model. However, in power systems with a high penetration of wind power, most of the uncertainty in net loads stems from the intermittency of wind power outputs, and we would like to study the performance of our approach under more realistic net load trajectories for such power systems. For this purpose, we added 140 wind farms to the 2736-bus system, with each wind farm located at a different load bus, making use of one year of wind power output data from 140 locations of NREL's Western Wind Integration Dataset (NREL 2012, Potter et al. 2008). The average hourly total wind power output is 1051MW, which corresponds to 7.37% of average hourly total demand. Each of the $N = 365$ days of wind power output data corresponds to one simulation, and for each of these 24-hour trajectories we also consider demand generated as described in Section 6.4.2, namely from a normal distribution at each bus with a standard deviation $\sigma_{demand,tj}$ corresponding to 10% of the expected value at the respective bus j and time t. Then, we use linear regression to estimate daily and semi-daily seasonality pattern and subtract it from the wind power data to estimate the standard deviation $\sigma_{wind,tj}$ of the errors of wind farm j and time t. When building the uncertainty set for robust UC, we select $\overline{d}_j^t$ as the nominal value for net load at bus j and time t, calculated as the expected demand minus the expected wind power output, and $\hat{d}_j^t$ as the standard deviation of net load, calculated as $\sqrt{\sigma_{demand,tj}^2 + \sigma_{wind,tj}^2}$, assuming independence between demand and wind power output.

The simulation performance results are presented in Table 9. We can see that now the multistage robust UC model achieves the best average total cost at $\Gamma = 1$, which is a 1.14% reduction from the best average cost of the two-stage robust UC model, achieved at $\Gamma = 2$. The standard deviation of the cost is reduced by 6.07%. Furthermore, we can see that the multistage robust UC approach can completely remove penalty cost with $\Gamma = 3$, while penalties remain fairly large for the two-stage robust UC approach under all Γ's (we also tested Γ's larger than 3, confirming our statement). Comparing with the deterministic reserve-based approach, we can see that the deterministic UC becomes very ineffective at handling this level of temporally correlated uncertainty. In particular, the multistage robust UC model reduces the average cost by 24.52% from the deterministic UC, and reduces the standard deviation of the cost by 83.22%.

Finally, Table 10 presents simulation performance results under the temporally correlated wind power data described above, and with demand sampled from the persistent model in Section 6.4.4. The best average cost is still achieved at $\Gamma = 1$ for the multistage robust UC, at $\Gamma = 2$ for the two-stage robust UC, and at 30% reserve for the deterministic UC, with the multistage robust UC achieving a 1.23% and 24.52% reduction in average cost, with respect to the two-stage robust UC and deterministic UC, respectively. We can also observe that under $\Gamma = 3$ the multistage robust UC does not completely eliminate the penalty cost, but achieves a significant reduction as compared to the other models, with a penalty frequency average of 0.01% as compared to 2.07% for the two-stage robust UC under $\Gamma = 2$ and 15.03% for the deterministic UC under 30% reserve (which respectively achieve their minimum penalties). All of these experiments on a large-scale power system demonstrate that the multistage UC model together with the proposed robust ED approach can dominate the performance of the two-stage robust UC and the deterministic UC models with look-ahead ED in terms of both average cost and system reliability.

Table 9 Simulation performance of the different models for the 2736-bus system with wind power and temporally independent demand.

Affine multistage robust UC with policy-enforcement robust ED

Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	11,078,125	9,511,651	8,523,553	8,568,135	8,642,679	9,414,582
Cost Std ($)	3,726,779	2,006,668	528,697	452,158	423,938	454,504
Penalty Cost Avg ($)	2,761,688	1,163,142	122,558	69,472	24,334	0
Penalty Freq Avg	18.80%	14.21%	1.95%	0.48%	0.15%	0.00%

Two-stage robust UC with look-ahead ED

Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	10,431,937	11,368,530	8,780,404	8,904,157	8,622,251	8,944,327
Cost Std ($)	1,865,595	1,065,188	689,745	867,257	562,878	754,036
Penalty Cost Avg ($)	2,106,438	1,035,071	426,520	530,980	209,218	440,250
Penalty Freq Avg	13.29%	3.74%	7.84%	5.68%	2.15%	2.68%

Deterministic UC with reserve and look-ahead ED

Reserve	2.5%	5%	10%	15%	20%	30%
Cost Avg ($)	13,343,698	14,395,889	13,259,879	13,647,325	11,986,311	11,292,887
Cost Std ($)	5,652,913	7,032,034	5,634,975	6,025,708	4,134,035	3,150,160
Penalty Cost Avg ($)	5,064,315	6,128,731	4,979,369	5,365,449	3,686,461	2,957,919
Penalty Freq Avg	30.57%	29.99%	33.49%	32.85%	24.41%	15.59%

Table 10 Simulation performance of the different models for the 2736-bus system with wind power and persistent demand.

Affine multistage robust UC with policy-enforcement robust ED

Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	10,996,931	9,459,785	8,502,923	8,581,532	8,646,665	9,415,693
Cost Std ($)	3,665,301	2,007,317	490,457	466,999	424,801	458,865
Penalty Cost Avg ($)	2,679,299	1,110,032	101,234	81,834	27,344	218
Penalty Freq Avg	18.84%	14.44%	1.67%	0.47%	0.18%	0.01%

Two-stage robust UC with look-ahead ED

Γ	0.25	0.5	1	1.5	2	3
Cost Avg ($)	10,390,214	11,365,568	8,734,840	8,863,975	8,609,160	8,947,959
Cost Std ($)	1,831,279	1,059,427	620,301	802,441	522,881	793,447
Penalty Cost Avg ($)	2,064,045	1,032,109	380,451	490,562	195,681	443,401
Penalty Freq Avg	12.73%	3.68%	7.37%	5.19%	2.07%	2.66%

Deterministic UC with reserve and look-ahead ED

Reserve	2.5%	5%	10%	15%	20%	30%
Cost Avg ($)	13,186,705	14,272,477	13,110,030	13,617,194	11,879,817	11,248,546
Cost Std ($)	5,557,309	7,023,964	5,596,039	6,082,173	4,095,780	3,113,902
Penalty Cost Avg ($)	4,905,635	6,003,861	4,827,766	5,334,746	3,578,986	2,912,186
Penalty Freq Avg	30.45%	29.94%	33.00%	32.43%	23.61%	15.03%

7. Conclusion

This paper presents a systematic study of multistage adaptive robust optimization for the UC problem with the solution concept of simplified affine policy. Such a model can deal with significant uncertainty in electricity demand and renewable generation caused by a high level penetration of wind and solar resources. We also propose a solution framework based on constraint generation with various algorithmic improvements, which achieves efficient solution of the affine multistage robust UC in large-scale power systems when the traditional methods fail. We also propose an associated robust ED model for real-time dispatch, which exploits the solution of the affine multistage robust UC model and is quickly solvable in real-time operation. We conduct extensive computational experiments on medium and large-scale power systems to thoroughly study the performance of the proposed models and algorithms and to compare them with existing approaches. The results show that the proposed algorithms can effectively solve the multistage robust UC model with simplified affine policies within a time frame reasonable for the day-ahead operation of large-scale power systems. The computational results demonstrate the effectiveness of the multistage robust UC model in reducing operational costs and at the same time improving system reliability, compared to the existing two-stage robust UC model and a deterministic UC model with reserve. Built on this work, future research can further explore more complex affine or non-affine policy structures and respective solution algorithms, and also modeling techniques to combine uncertainty from heterogeneous sources such as wind and solar power, demand, and generation or transmission contingencies.

References

Ben-Tal, A., G. Boaz, S. Shimrit. 2009a. Robust multi-echelon multi-period inventory control. *European Journal of Operational Research* **199**(3) 922–935.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009b. *Robust optimization*. Princeton University Press.

Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovsky. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical programming, Ser. A* **99** 351–376.

Bertsimas, D., D. B. Brown, C. Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, D., I. Dunning, M. Lubin. 2014. Reformulations versus cutting planes for robust optimization. *Available at Optimization Online* .

Bertsimas, D., A. Georghiou. 2014. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Available at Optimization Online* .

Bertsimas, D., Dan A. Iancu, P. A. Parrilo. 2010. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research* **35**(2) 363–394.

Bertsimas, D., E. Litvinov, X. A. Sun, J. Zhao, T. Zheng. 2013. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems* **28**(1) 52–63.

Birge, J. R, F. V Louveaux. 1988. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research* **34**(3) 384–392.

Blankenship, J. W., J. E. Falk. 1976. Infinitely constrained optimization problems. *Journal of Optimization Theory and Applications* **19**(2) 261–281.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

Chen, X., Y. Zhang. 2009. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research* **57**(6) 1469–1482.

Conejo, A. J., M. Carrión, J. M. Morales. 2010. *Decision making under uncertainty in electricity markets*, vol. 153. Springer.

Constantinescu, E. M., V. M. Zavala, M. Rocklin, S. Lee, M. Anitescu. 2011. A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation. *IEEE Transactions on Power Systems* **26**(1) 431–441.

de Oliveira, W., C. Sagastizábal. 2014. Level bundle methods for oracles with on-demand accuracy. *Optimization Methods and Software* .

FERC. 2014. Operator-initiated commitment in rto and iso markets. *Price formation in organized wholesale electricity markets* **Docket No. AD14-14-000**. `http://www.ferc.gov/legal/staff-reports/2014/AD14-14-operator-actions.pdf`.

Fischetti, M., M. Monaci. 2012. Cutting plane versus compact formulations for uncertain (integer) linear programs. *Mathematical Programming Computation* **4**(3) 239–273.

Georghiou, A., W. Wiesemann, D. Kuhn. 2013. Generalized decision rule approximations for stochastic programming via liftings. Submitted.

Goh, J., M. Sim. 2011. Robust optimization made easy with rome. *Operations Research* **59**(4) 973–985.

Gómez-Expósito, A., A. J. Conejo, C. Cañizares. 2008. *Electric energy systems: analysis and operation*. CRC Press.

Hadjiyiannis, M. J., P. J. Goulart, D. Kuhn. 2011. An efficient method to estimate the suboptimality of affine controllers. *IEEE Transactions on Automatic Control* **56**(12) 2841–2853.

Hamilton, J. D. 1994. *Time series analysis*, vol. 2. Princeton University Press.

Heitsch, H., W. Römisch. 2011. Scenario tree generation for multi-stage stochastic programs. *Stochastic Optimization Methods in Finance and Energy*. Springer, 313–341.

Hobbs, B. F., M. H. Rothkopf, R. P. O'Neill, H. Chao (Eds.). 2001. *The Next Generation of Electric Power Unit Commitment Models*, *International Series in Operations Research & Management Science*, vol. 36. Springer.

Jabr, R. A. 2013. Adjustable robust OPF with renewable energy sources. *IEEE Transactions on Power Systems* **28**(4) 4742–4751.

Jiang, R., J. Wang, Y. Guan. 2012. Robust unit commitment with wind power and pumped storage hydro. *IEEE Transactions on Power Systems* **27**(2) 800–810.

Keyhani, A., M. N. Marwali, M. Dai. 2009. *Integration of green and renewable energy in electric power systems*. John Wiley & Sons.

Kuhn, D., W. Wiesemann, A. Georghiou. 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* **130**(1) 177–209.

Lorca, Á., X. A. Sun. 2015. Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems* **30**(4) 1702–1713.

Minoux, M. 2014. Two-stage robust optimization, state-space representable uncertainty and applications. *RAIRO-Operations Research* **48**(4) 455–475.

Navid, N., G. Rosenwald. 2012. Market solutions for managing ramp flexibility with high penetration of renewable resource. *IEEE Transactions on Sustainable Energy* **3**(4) 784–790.

NREL. 2012. Transmission grid integration — western wind dataset `http://www.nrel.gov/electricity/transmission/wind_integration_dataset.html`.

Ostrowski, J., M. F. Anjos, A. Vannelli. 2012. Tight mixed integer linear programming formulations for the unit commitment problem. *IEEE Transactions on Power Systems* **27**(1) 39.

Ozturk, U. A., M. Mazumdmar, B. A. Norman. 2004. A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Transactions on Power Systems* **19**(3) 1589–1598.

Papavasiliou, A., S. S. Oren. 2013. Multi-area stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research* **61**(3) 578–592.

Potter, Cameron W, Debra Lew, Jim McCaa, Sam Cheng, Scott Eichelberger, Eric Grimit. 2008. Creating the dataset for the western wind and solar integration study (usa). *Wind Engineering* **32**(4) 325–338.

Ruiz, P. A., C. R. Philbrick, P. W. Sauer. 2009a. Wind power day-ahead uncertainty management through stochastic unit commitment policies. *IEEE Power Systems Conference and Exposition (PSCE'09)* .

Ruiz, P. A., C. R. Philbrick, E. Zak, K. W. Cheung, P. W. Sauer. 2009b. Uncertainty management in the unit commitment problem. *IEEE Transactions on Power Systems* **24**(2) 642–651.

Street, A., F. Oliveira, J. M. Arroyo. 2011. Contingency-constrained unit commitment with $n-k$ security criterion: A robust optimization approach. *IEEE Transactions on Power Systems* **26**(3) 1581–1590.

Tuohy, A., P. Meibom, E. Denny, M. O'Malley. 2009. Unit commitment for systems with significant wind penetration. *IEEE Transactions on Power Systems* **24**(2) 592–601.

Wang, J., M. Shahidehpour, Z. Li. 2008. Security-constrained unit commitment with volatile wind power generation. *IEEE Transactions on Power Systems* **23**(3) 1319–1327.

Wang, Q., Y. Guan, J. Wang. 2012. A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. *IEEE Transactions on Power Systems* **27**(1) 206–215.

Wang, Q., J. P. Watson, Y. Guan. 2013. Two-stage robust optimization for $n-k$ contingency-constrained unit commitment. *IEEE Transactions on Power Systems* **28**(3) 2366–2375.

Warrington, J., P. J. Goulart, S. Mariéthoz, M. Morari. 2012. Robust reserve operation in power systems using affine policies. *51st IEEE Conference on Decision and Control*. IEEE, 1111–1117.

Warrington, J., P. J. Goulart, S. Mariéthoz, M. Morari. 2013. Policy-based reserves for power systems. *IEEE Transactions on Power Systems* **28**(4) 4427–4437.

Warrington, J., C. Hohl, P. J. Goulart, M. Morari. 2014. Optimal unit commitment accounting for robust affine reserve policies. *2014 American Control Conference*. ACC, 5049–5055.

Wu, L., M. Shahidehpour, T. Li. 2007. Stochastic security-constrained unit commitment. *IEEE Transactions on Power Systems* **22**(2) 800–811.

Xie, L., P. Carvalho, L. Ferreira, J. Liu, B. Krogh, N. Popli, M. Ilic. 2011. Wind integration in power systems: Operational challenges and possible solutions. *Proceedings of the IEEE* **99**(1) 214–232.

Zhao, C., Y. Guan. 2013. Unified stochastic and robust unit commitment. *IEEE Transactions on Power Systems* **28**(3) 3353–3361.

Zhao, C., J. Wang, J. P. Watson, Y. Guan. 2013. Multi-stage robust unit commitment considering wind and demand response uncertainties. *IEEE Transactions on Power Systems* **28**(3) 2708–2717.

Zhao, L., B. Zeng. 2012. Robust unit commitment problem with demand response and wind energy. *Power and Energy Society General Meeting, 2012 IEEE*. IEEE, 1–8.

Zimmerman, R. D., C. E. Murillo-Sánchez, R. J. Thomas. 2011. MATPOWER: Steady-state operation, planning and analysis tools for power systems research and education. *IEEE Transactions on Power Systems* **26**(1) 12–19.

Electronic Companion for "Multistage Adaptive Robust Optimization for the Unit Commitment Problem", by Lorca, Sun, Litvinov and Zheng

EC.1. Proofs for Sections 3, 4, 5

Proof of Proposition 1: To make references more explicit, we use $(2S)$ and (M) to denote the two-stage (2) and the multistage models (5) in this proof, respectively. The proof follows from the fact that, without ramping constraints (5c), the dispatch problems using uncertainty sets (3) in both $(2S)$ and (M) are separable over time periods. In fact, we show that, without ramping constraints, $(2S)$ and (M) are both equivalent to problem $(1P)$, defined as follows:

$$(1P) \quad \min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X} \Big\{ \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} (G_i x_i^t + S_i u_i^t) + \sum_{t\in\mathcal{T}} \max_{\boldsymbol{d}^t\in\mathcal{D}^t} \min_{\boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)} \sum_{i\in\mathcal{N}_g} C_i p_i^t \Big\},$$

where $X = \{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}):$ (1b)-(1g) are satisfied$\}$ and $\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)$ is the feasible dispatch set at time t without ramping constraints, i.e. $\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t) \triangleq \{\boldsymbol{p}^t:$ (1h), (1j), (1k) are satisfied$\}$.

(i) First, we show that without ramping constraints, $(2S)$ is equivalent to $(1P)$. In fact, without ramping constraints, $(2S)$ can be written as

$$\min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X} \Big\{ \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} (G_i x_i^t + S_i u_i^t) + \max_{\boldsymbol{d}\in\mathcal{D}} \min_{\{\boldsymbol{p}:\, \boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)\, \forall t\in\mathcal{T}\}} \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} C_i p_i^t \Big\},$$

and we have

$$\begin{aligned} &\max_{\boldsymbol{d}\in\mathcal{D}} \min_{\{\boldsymbol{p}:\, \boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)\, \forall t\in\mathcal{T}\}} \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} C_i p_i^t \\ &= \max_{\boldsymbol{d}\in\mathcal{D}} \sum_{t\in\mathcal{T}} \min_{\boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)} \sum_{i\in\mathcal{N}_g} C_i p_i^t \\ &= \sum_{t\in\mathcal{T}} \max_{\boldsymbol{d}^t\in\mathcal{D}^t} \min_{\boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)} \sum_{i\in\mathcal{N}_g} C_i p_i^t, \end{aligned}$$

where the first equality comes from the fact that the dispatch set $\{\boldsymbol{p}: \boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)\ \forall t\in\mathcal{T}\}$ is separable over time, and the second equality comes from the separability of the uncertainty set $\mathcal{D}$ defined in (3) over time periods. Adding $\sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} (G_i x_i^t + S_i u_i^t)$ and applying $\min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X}$ at both sides of this equality yields the desired result.

(ii) Now we show that, without ramping constraints, (M) is equivalent to $(1P)$. Without ramping constraints, $\Omega_t(\boldsymbol{x},\boldsymbol{d}^t,\boldsymbol{p}^{t-1}) = \Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)$, so the nested multistage formulation (6) is equivalent to

$$(\widetilde{M}^{NR}) \quad \min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X} \Big\{ \boldsymbol{G}^\top\boldsymbol{x} + \boldsymbol{S}^\top\boldsymbol{u} + \max_{\boldsymbol{d}^1\in\mathcal{D}^1} \min_{\boldsymbol{p}^1\in\Omega_1^{NR}(\boldsymbol{x},\boldsymbol{d}^1)} \Big\{ \boldsymbol{C}^\top\boldsymbol{p}^1 + \cdots + \max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T^{NR}(\boldsymbol{x},\boldsymbol{d}^T)} \boldsymbol{C}^\top\boldsymbol{p}^T \Big\}\Big\}.$$

Consider the max-min problem at $t = T-1$ in $(\widetilde{M}^{NR})$. Since $\mathcal{D}^T$, $\Omega_T^{NR}(\boldsymbol{x}, \boldsymbol{d}^T)$, and $\boldsymbol{C}^\top \boldsymbol{p}^T$ do not depend on $\boldsymbol{p}^{T-1}$ and $\boldsymbol{d}^{T-1}$, we obtain

$$\begin{aligned}
&\max_{\boldsymbol{d}^{T-1}\in\mathcal{D}^{T-1}} \min_{\boldsymbol{p}^{T-1}\in\Omega_{T-1}^{NR}(\boldsymbol{x},\boldsymbol{d}^{T-1})} \left\{ \boldsymbol{C}^\top \boldsymbol{p}^{T-1} + \max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T^{NR}(\boldsymbol{x},\boldsymbol{d}^T)} \boldsymbol{C}^\top \boldsymbol{p}^T \right\} \\
&= \max_{\boldsymbol{d}^{T-1}\in\mathcal{D}^{T-1}} \left\{ \left(\min_{\boldsymbol{p}^{T-1}\in\Omega_{T-1}^{NR}(\boldsymbol{x},\boldsymbol{d}^{T-1})} \boldsymbol{C}^\top \boldsymbol{p}^{T-1} \right) + \left(\max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T^{NR}(\boldsymbol{x},\boldsymbol{d}^T)} \boldsymbol{C}^\top \boldsymbol{p}^T \right) \right\} \\
&= \left(\max_{\boldsymbol{d}^{T-1}\in\mathcal{D}^{T-1}} \min_{\boldsymbol{p}^{T-1}\in\Omega_{T-1}^{NR}(\boldsymbol{x},\boldsymbol{d}^{T-1})} \boldsymbol{C}^\top \boldsymbol{p}^{T-1} \right) + \left(\max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T^{NR}(\boldsymbol{x},\boldsymbol{d}^T)} \boldsymbol{C}^\top \boldsymbol{p}^T \right),
\end{aligned}$$

and this argument can be carried out backward until $t = 1$ to obtain

$$\max_{\boldsymbol{d}^1\in\mathcal{D}^1} \min_{\boldsymbol{p}^1\in\Omega_1^{NR}(\boldsymbol{x},\boldsymbol{d}^1)} \left\{ \boldsymbol{C}^\top \boldsymbol{p}^1 + \cdots + \max_{\boldsymbol{d}^T\in\mathcal{D}^T} \min_{\boldsymbol{p}^T\in\Omega_T^{NR}(\boldsymbol{x},\boldsymbol{d}^T)} \boldsymbol{C}^\top \boldsymbol{p}^T \right\} = \sum_{t\in\mathcal{T}} \max_{\boldsymbol{d}^t\in\mathcal{D}^t} \min_{\boldsymbol{p}^t\in\Omega_t^{NR}(\boldsymbol{x},\boldsymbol{d}^t)} \boldsymbol{C}^\top \boldsymbol{p}^t.$$

Adding $\sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{N}_g} (G_i x_i^t + S_i u_i^t)$ and applying $\min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in X}$ on both sides of this equality yields that $(\widetilde{M}^{NR})$ is equivalent to $(1P)$, which completes the proof. □

Proof of Proposition 2: $\boldsymbol{c}^\top \boldsymbol{d} \le h \;\forall \boldsymbol{d} \in \mathcal{D}$ is equivalent to $\max_{\boldsymbol{d}\in\mathcal{D}} \boldsymbol{c}^\top \boldsymbol{d} \le h$. Now notice that $\mathcal{D}^t$ is the projection over d^t of

$$\widetilde{\mathcal{D}}^t = \left\{ (\boldsymbol{d}^t, \boldsymbol{z}^t) : \sum_{j\in\mathcal{N}_d} z_j^t \le \Gamma\sqrt{N_d},\; \hat{d}_j^t z_j^t \ge d_j^t - \overline{d}_j^t,\; \hat{d}_j^t z_j^t \ge \overline{d}_j^t - d_j^t,\; d_j^t \in [\overline{d}_j^t - \Gamma\hat{d}_j^t, \overline{d}_j^t + \Gamma\hat{d}_j^t] \;\forall j \in \mathcal{N}_d \right\}.$$

So by defining $\widetilde{\mathcal{D}} = \prod_{t\in[1:T]} \widetilde{\mathcal{D}}^t$, we have $\max_{\boldsymbol{d}\in\mathcal{D}} \boldsymbol{c}^\top \boldsymbol{d} = \max_{(\boldsymbol{d},\boldsymbol{z})\in\widetilde{\mathcal{D}}} \boldsymbol{c}^\top \boldsymbol{d} = \min_{\boldsymbol{\pi}\in\Pi} \boldsymbol{e}^\top \boldsymbol{\pi}$, where the last equality follows from duality theory, since $\mathcal{D}$ is bounded, where $\boldsymbol{\pi} \in \Pi$ is equivalent to (12b)-(12d) and $\boldsymbol{e}^\top \boldsymbol{\pi}$ is the left hand side of (12a). Now, $\min_{\boldsymbol{\pi}\in\Pi} \boldsymbol{e}^\top \boldsymbol{\pi} \le h$ is equivalent to the existence of $\boldsymbol{\pi} \in \Pi$ such that $\boldsymbol{e}^\top \boldsymbol{\pi} \le h$ and the result follows. □

Proof of Proposition 3: Take the W_{it}-policy. The energy balance equation (8g) can be written as

$$\sum_{i\in\mathcal{N}_g} w_i^t + \sum_{j\in\mathcal{N}_d} \left(\sum_{i\in\mathcal{N}_g} W_{it} - 1 \right) d_j^t = 0 \qquad \forall \boldsymbol{d} \in \mathcal{D},\; \forall t \in \mathcal{T}.$$

Since the uncertainty set $\mathcal{D}$ is full-dimensional, which is the case for the uncertainty set in (3), the constraint that the above affine function of $\boldsymbol{d}$ is equal to zero for all $\boldsymbol{d} \in \mathcal{D}$ can hold if and only if all the coefficients of this affine function are zero, which gives (15). We can show (14) similarly. □

Proof of Proposition 5: In the separation problem, consider the change of variables given by $d_j^s = \overline{d}_j^s + \Gamma \hat{d}_j^s u_j^s$. The equivalent problem for $\boldsymbol{u}$ is

$$\begin{aligned} \max_{\boldsymbol{u}} \quad & \sum_{s\in\mathcal{T}} \sum_{j\in\mathcal{N}_d} c_j^s \hat{d}_j^s u_j^s \\ \text{s.t.} \quad & u_j^s \in [-1, 1] \quad \forall j \in \mathcal{N}_d, s\in\mathcal{T} \\ & \sum_{j\in\mathcal{N}_d} |u_j^s| \le \sqrt{N_d} \quad \forall s\in\mathcal{T}. \end{aligned}$$

This problem is separable in s and the solution of each of the problems obtained is found by ordering $|c_j^s \hat{d}_j^s|$ in j from largest to smallest, and successively assigning the highest possible values to those $|u_j^s|$ with the largest respective values of $|c_j^s \hat{d}_j^s|$, taking each of these u_j^s with the same sign of c_j^s (notice that $\hat{d}_j^s > 0$). □

Proof of Proposition 6: The proof follows from reformulating the respective robust constraints.

(i) Constraints (8c) under the W_{it}-policy can be written, for each i and t, as

$$p_i^{min} x_i^t \le w_i^t + W_{it}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) \le p_i^{max} x_i^t \quad \forall \boldsymbol{d}\in\mathcal{D},$$

which is equivalent to

$$\begin{aligned} p_i^{min} x_i^t &\le w_i^t + \min_{\boldsymbol{d}\in\mathcal{D}} W_{it}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) \\ w_i^t + \max_{\boldsymbol{d}\in\mathcal{D}} W_{it}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) &\le p_i^{max} x_i^t. \end{aligned}$$

Depending on the sign of W_{it}, the above two inequalities are equivalent to the following four constraints,

$$\begin{aligned} p_i^{min} x_i^t \le w_i^t + W_{it} \min_{\boldsymbol{d}\in\mathcal{D}}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) \le p_i^{max} x_i^t \\ p_i^{min} x_i^t \le w_i^t + W_{it} \max_{\boldsymbol{d}\in\mathcal{D}}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) \le p_i^{max} x_i^t. \end{aligned}$$

In other words, in the robust constraints (8c) $\mathcal{D}$ can be replaced by a finite uncertainty set consisting of $\{\boldsymbol{d}_{min}, \boldsymbol{d}_{max}\}$. This completes the proof for the first part. Notice that the proof of this part is independent of the structure of $\mathcal{D}$ so the conclusion of (i) is true for any convex uncertainty set.

(ii) Under the W_{it}-policy, (8e) can be written as

$$\max_{\boldsymbol{d}\in\mathcal{D}} \left\{ W_{it}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) - W_{i,t-1}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right) \right\} \le w_i^{t-1} - w_i^t + RU_i x_i^{t-1} + SU_i u_i^t. \qquad \text{(EC.1)}$$

Notice that the uncertainty set $\mathcal{D}$ defined in (3) is separable in time periods. Therefore, the left-hand side of (EC.1) is equivalent to the following problem

$$\max_{\boldsymbol{d}^t\in\mathcal{D}^t}\left\{W_{it}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right)\right\} - \min_{\boldsymbol{d}^{t-1}\in\mathcal{D}^{t-1}}\left\{W_{i,t-1}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right)\right\}. \tag{EC.2}$$

Depending on the signs of the affine coefficients W_{it} and $W_{i,t-1}$, (EC.2) is equivalent to one of the four possible combinations:

$$W_{it}\max_{\boldsymbol{d}^t\in\mathcal{D}^t}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) - W_{i,t-1}\min_{\boldsymbol{d}^{t-1}\in\mathcal{D}^{t-1}}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right) \tag{EC.3a}$$

$$W_{it}\max_{\boldsymbol{d}^t\in\mathcal{D}^t}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) - W_{i,t-1}\max_{\boldsymbol{d}^{t-1}\in\mathcal{D}^{t-1}}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right) \tag{EC.3b}$$

$$W_{it}\min_{\boldsymbol{d}^t\in\mathcal{D}^t}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) - W_{i,t-1}\min_{\boldsymbol{d}^{t-1}\in\mathcal{D}^{t-1}}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right) \tag{EC.3c}$$

$$W_{it}\min_{\boldsymbol{d}^t\in\mathcal{D}^t}\left(\sum_{j\in\mathcal{N}_d} d_j^t\right) - W_{i,t-1}\max_{\boldsymbol{d}^{t-1}\in\mathcal{D}^{t-1}}\left(\sum_{j\in\mathcal{N}_d} d_j^{t-1}\right), \tag{EC.3d}$$

where (EC.3a)-(EC.3d) correspond to the worst-case scenarios $\boldsymbol{d}_{maxmin}(t), \boldsymbol{d}_{max}, \boldsymbol{d}_{min}, \boldsymbol{d}_{minmax}(t)$, respectively. The proof is analogous for ramping down constraints (8d). $\square$

EC.2. Incorporating Penalty Variables

In order to incorporate penalty variables into the affine multistage robust UC with W_{it}-policy we replace equations (8b), (8f) and (8g) by

$$\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g} C_i\left(w_i^t + W_{it}\sum_{j\in\mathcal{N}_d} d_j^t\right) + C^{pen}\sum_{t\in\mathcal{T}}\left[\left(w_t^+ + W_t^+\sum_{j\in\mathcal{N}_d} d_j^t\right) + \left(w_t^- + W_t^-\sum_{j\in\mathcal{N}_d} d_j^t\right) + \sum_{l\in\mathcal{N}_l} w_{tl}^f\right] \le z \quad \forall \boldsymbol{d}\in\mathcal{D}$$

$$-f_l^{max} - w_{tl}^f \le \sum_m\sum_{i\in\mathcal{N}_g}\alpha_{lm}B_{mi}^p\left(w_i^t + W_{it}\sum_{j\in\mathcal{N}_d} d_j^t\right) - \sum_m\sum_{j\in\mathcal{N}_d}\alpha_{lm}B_{mj}^d d_j^t \ \le\ f_l^{max} + w_{tl}^f \quad \forall \boldsymbol{d}\in\mathcal{D},\, t\in\mathcal{T},\, l\in\mathcal{N}_l$$

$$\sum_{i\in\mathcal{N}_g}\left(w_i^t + W_{it}\sum_{j\in\mathcal{N}_d} d_j^t\right) + \left(w_t^+ + W_t^+\sum_{j\in\mathcal{N}_d} d_j^t\right) = \sum_{j\in\mathcal{N}_d} d_j^t + \left(w_t^- + W_t^-\sum_{j\in\mathcal{N}_d} d_j^t\right) \quad \forall \boldsymbol{d}\in\mathcal{D}, t\in\mathcal{T}$$

$$w_{tl}^f \ge 0 \quad \forall t\in\mathcal{T},\, l\in\mathcal{N}_l$$

$$\left(w_t^+ + W_t^+\sum_{j\in\mathcal{N}_d} d_j^t\right),\ \left(w_t^- + W_t^-\sum_{j\in\mathcal{N}_d} d_j^t\right) \ge 0 \quad \forall \boldsymbol{d}\in\mathcal{D},\, t\in\mathcal{T},$$

where C^{pen} is the unitary penalty cost, the w_{tl}^f's are penalty variables for transmission line capacity constraints, and the w_{it}^+, W_{it}^+, w_{it}^- and W_{it}^-'s are the penalty variables for under- and over-generation in the energy balance equation.