

Strengthened and Faster Linear Approximation to Joint Chance Constraints with Wasserstein Ambiguity

Yihong Zhou

Department of Engineering Science, University of Oxford, U.K., yihong.zhou@eng.ox.ac.uk,

Yuxin Xia

Department of Engineering Science, University of Oxford, U.K., yuxin.xia@eng.ox.ac.uk,

Hanbin Yang

Business School, Hong Kong University of Science and Technology, Hong Kong, China, hanbinyang@ust.hk,

Thomas Morstyn

Department of Engineering Science, University of Oxford, U.K., thomas.morstyn@eng.ox.ac.uk,

Many real-world decision-making problems in energy systems, transportation, and finance have uncertain parameters in constraints. Wasserstein distributionally robust joint chance constraints (WDRJCC) offer a promising solution by explicitly guaranteeing the probability of the simultaneous satisfaction of multiple constraints. However, WDRJCC are computationally demanding, and practical applications often require more tractable approaches, especially for large-scale and complex problems such as power system unit commitment problems and multilevel problems with chance constraints in lower levels. To address this, this paper proposes a novel convex inner-approximation for WDRJCC with right-hand-side uncertainties (RHS-WDRJCC). Motivated by the strengthening process that leads to a faster but still exact mixed-integer reformulation, we propose a Strengthened and Faster Linear Approximation (SFLA) by strengthening an existing convex inner-approximation. This strengthening process reduces the number of constraints and tightens the feasible region for ancillary variables, leading to significant computational speedup. We prove that the proposed SFLA does not introduce additional conservativeness and can even be less conservative compared to common approximations such as W-CVaR. We then extend the proposed SFLA to robustness maximization, a decision-making paradigm that can be more interpretable, where the risk level and the Wasserstein radius are determined by maximizing solution robustness subject to a utility degradation limit. We discuss the connection between risk minimization and radius maximization as two formulations of robustness maximization, and show the advantage of radius maximization.

In power system unit commitment, the proposed SFLA achieves up to $10\times$ and on average $3.8\times$ computational speedup compared to the strengthened and exact mixed-integer reformulation in finding comparable high-quality solutions. In a bilevel strategic bidding problem where the exact reformulation is not applicable due to non-convexity, the proposed SFLA can lead to $90\times$ speedup compared to existing convex approximation methods including W-CVaR. In robustness maximization, the proposed SFLA demonstrated over $100\times$ speedup than other convex-approximations.

Key words: Faster linear approximation, Wasserstein distributionally robust joint chance constraints, conditional value-at-risk, robustness maximization

1. Introduction

Many real-world decision-making problems in energy systems, transportation, and finance involve constraints with uncertain parameters due to variable supply, customer demand, and economic conditions (Gabrel et al. 2014). A typical solution approach is robust optimization (RO), which aims to ensure feasibility for each possible scenario; however, this often leads to overly conservative solutions. In contrast, chance-constrained programming (CCP) provides a less restrictive alternative (Prékopa 2013). CCP supports robust but economic decision-making by explicitly limiting the probability of constraint violations, allowing for a controlled degree of conservativeness relative to RO. As a result, CCP has found widespread applications in fields such as power systems, economics, finance, and water management (Geng and Xie 2019).

However, classic CCP models are based on the exact distribution of random variables, which is typically not available, and decision-makers generally only possess a historical dataset. This dataset may be insufficient to accurately infer the true distribution of random variables, which limits the out-of-sample performance of a CCP model. To hedge the ambiguity of the underlying distribution, distributionally robust chance-constrained programming (DRCCP) has been proposed to control the violation probability under the worst-case probability distribution over a so-called “ambiguity” set (Scarf et al. 1957). This ambiguity set may be defined as the set of probability distributions with the same statistical moments, such as mean and variance (Delage and Ye 2010), or defined as distributions within a certain distance from the reference distribution. Despite good tractability, moment-based ambiguity sets do not fully use the information of the datasets and may lead to over-conservativeness (Gao and Kleywegt 2023). For distance-based ambiguity sets, the Wasserstein distance is widely applied due to its superior out-of-sample performance (Mohajerin Esfahani and Kuhn 2018), and the advantage of Wasserstein-based ambiguity sets over other distance-based sets was discussed by Gao and Kleywegt (2023). A Wasserstein DRCCP model can be written as:

$$\min_{\mathbf{x} \in \mathcal{X}} c(\mathbf{x}) \quad (1a)$$

$$\text{s.t.} \quad \sup_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})] \leq \epsilon, \quad (1b)$$

where $c(\cdot)$ is the objective function, $\mathcal{X} \subset \mathbb{R}^L$ is a compact domain for the decision variables $\mathbf{x} \in \mathbb{R}^L$. Constraint (1b) ensures that the random vector $\boldsymbol{\xi} \in \mathbb{R}^K$ falls outside the decision-dependent safety set $\mathcal{S}(\mathbf{x})$ with a small probability, which is not greater than the risk level ϵ under the worst-case distribution $\mathbb{P} \in \mathcal{F}_N(\theta)$. Here, the Wasserstein ambiguity set $\mathcal{F}_N(\theta)$ is defined as a ball

with radius θ centered at the empirical distribution \mathbb{P}_N constructed by historical data $\mathcal{F}_N(\theta) := \{\mathbb{P} \mid d_W(\mathbb{P}_N, \mathbb{P}) \leq \theta\}$.

Following Chen et al. (2022) and Ho-Nguyen et al. (2021), we consider the 1-Wasserstein distance based on a general norm $\|\cdot\|$, expressed as $d_W(\mathbb{P}, \mathbb{P}') := \inf_{\Pi \in \mathcal{P}(\mathbb{P}, \mathbb{P}')} \mathbb{E}_{(\xi, \xi') \sim \Pi} [\|\xi - \xi'\|]$, where $\mathcal{P}(\mathbb{P}, \mathbb{P}')$ is a set of distributions with marginal distributions \mathbb{P} and \mathbb{P}' . This paper focuses on Wasserstein distributionally robust joint chance constraints under right-hand-side uncertainty (RHS-WDRJCC) (1b) with the following safety set:

$$\mathcal{S}(\mathbf{x}) := \{\xi \mid \mathbf{a}_p^\top \mathbf{x} \leq \mathbf{b}_p^\top \xi + d_p, p \in [P]\}, \quad (2)$$

where there are P constraints indexed by $[P] := \{1, \dots, P\}$ that need to be met jointly with high probability $1 - \epsilon$ with $\epsilon \in (0, 1)$. This joint satisfaction is desired in practical applications for its higher safety (Geng and Xie 2019, Ding et al. 2022), and the RHS uncertainty also arises in several practical problems, such as managing the thermal constraints of the power grid lines, securing the power system reserve (Wang et al. 2016, Wu et al. 2016, Yang et al. 2020), or restricting the load-generation imbalance (van Ackooij et al. 2018).

To exactly solve the problem with RHS-WDRJCC (1), Chen et al. (2022) proposed an exact mixed-integer programming (MIP) reformulation, which was further strengthened by Ho-Nguyen et al. (2021) by exploiting valid inequalities and reducing the values of big-M. Jiang and Xie (2024b) further proposed a method that combines inner and outer approximations to derive optimality cuts, thereby accelerating the exact MIP solution process. Alternative methods include ALSO-X, ALSO-X+, and ALSO-X# that (approximately) solve a CCP in an iterative way with high quality (Jiang and Xie 2022, 2024a). However, the MIP-based exact reformulation is NP-hard, and these ALSO methods may have numerical issues and a lack of stability due to their iterative schemes. Greater tractability is still desired, especially when the original deterministic problem is already complicated. We have identified two types of practical applications that necessitate more computationally efficient solution schemes.

One typical example is the power system unit commitment (UC) problem that determines the on/off statuses of generators to minimize operational cost within power network constraints, which is a large-scale multi-period MIP problem with thousands of generators and network nodes included. In one of the largest electricity markets in the world, managed by the Midcontinent Independent System Operator (MISO), the network model includes more than 45,000 buses, 1,400 generation sources, around 2,500 pricing nodes, and has a 36-hour look-ahead horizon with hourly resolution (Chen

et al. 2016, Sun et al. 2018). Additionally, UC problems may be solved repeatedly during the market-clearing process to accommodate necessary modifications to constraints or other evaluations. Moreover, the entire process must be completed within a stringent time frame of three to four hours (Chen et al. 2016). The UC complexity is further exacerbated by the increasing number of virtual bids (bids for speculation submitted by purely financial players with no obligation to own physical generation) and the increasing number of combined cycle units that require complicated operation models (Sun et al. 2018). However, the increasing penetration of fluctuating renewables increases the uncertainty level of the system, and thus it becomes more important to have chance-constrained UC to ensure reliable operation (Yang et al. 2020, van Ackooij et al. 2018).

The second example involves applying RHS-WDRJCC at lower levels of bilevel or multilevel optimization problems, where all levels need to reach the optimum. To solve these multilevel problems, Karush-Kuhn-Tucker (KKT) conditions or strong duality need to be exploited to derive solvable single-level counterparts, which require convexity in lower levels. Practical applications include strategic price-maker bidders in day-ahead energy markets (Paredes et al. 2023) and gas markets that involve a sequential clearing process (Heitsch et al. 2022). In these applications, market clearing is modeled at lower levels. Because the clearing process can occur before uncertainties are revealed (e.g., day-ahead markets), a CCP needs to be implemented at these lower levels to manage uncertainties (Beck et al. 2023). However, the MIP-based RHS-WDRJCC reformulation (Ho-Nguyen et al. 2021) is unsuitable due to its non-convexity, and the iterative ALSO methods (Jiang and Xie 2022, 2024a) are less capable of incorporating optimality conditions.

Due to the high complexity of these practical problems, it is desirable to have convex or linear approximations of RHS-WDRJCC, so as to bring a minimum extra computational effort while maintaining solution quality. Common approaches include the Bonferroni approximation (Chen et al. 2023), which can become overly conservative when events overlap significantly, and the worst-case conditional value at risk (W-CVaR) (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023), a common approximation for CCPs. Chen et al. (2023) also introduced the “best” convex inner-approximation for RHS-WDRJCC, which is equivalent to W-CVaR under certain hyperparameters. However, these approximation schemes introduce a large number of additional ancillary variables and constraints, which brings excessive computing burden, especially for the UC and multilevel problems discussed above; consequently, CCPs generally remain unscalable for industrial-scale instances (Zhao et al. 2024). Therefore, it is still preferable to further enhance the computational efficiency of RHS-WDRJCC.

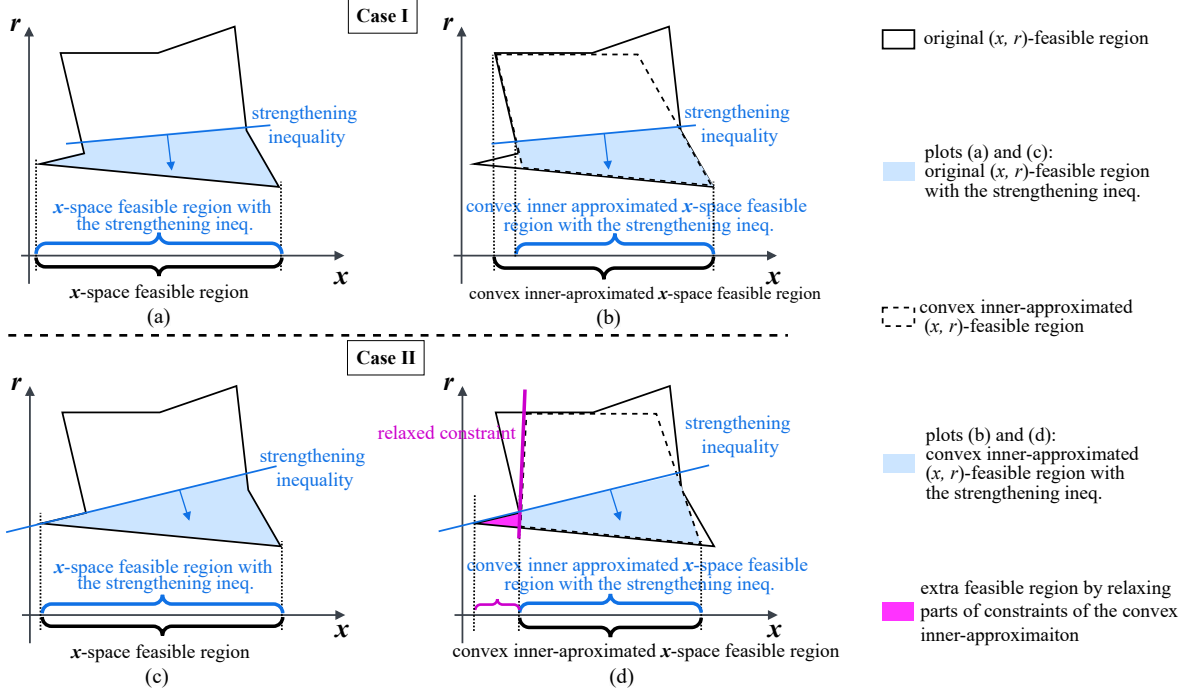


Figure 1 Two possible cases when a strengthening inequality valid for the original feasible region is applied to the convex inner-approximation.

Note. Analytical reformulations sometimes (such as for WDRJCC) introduce an ancillary variable r in addition to the original decision variable x . Case I: A strengthening inequality is valid for the original nonconvex region but leads to higher conservativeness in its inner approximation. (a) In the original nonconvex feasible region of x and r , the strengthening (valid) inequality excludes a redundant portion without reducing the feasible region of x . (b) The convex inner-approximation (dashed) has a smaller feasible region of x compared to the original region, and the strengthening inequality further reduces the x feasible region, resulting in even higher conservativeness. Case II: A strengthening inequality is valid for both the original nonconvex region and its inner approximation. (c) In the original nonconvex feasible region, the strengthening inequality excludes a redundant portion without reducing the feasible region of x . (d) The convex inner-approximation (dashed) has a smaller feasible region of x compared to the original region, and the strengthening inequality does not further reduce the x feasible region. Furthermore, by relaxing certain constraints of the convex inner-approximation, it is possible to include an extra non-empty feasible region (depicted in purple) and thus expand the feasible region of x (less conservativeness). This extra feasible region is small relative to the excluded region; therefore, the overall search space is still reduced, preserving the computational efficiency provided by the strengthening inequality.

One promising solution approach is to utilize valid inequalities, which are additional constraints introduced to an optimization problem to tighten the formulation without excluding any feasible solutions from the original problem's feasible region. Motivated by the work of Ho-Nguyen et al. (2021), where valid inequalities were exploited to strengthen the exact MIP reformulation so as to speed up computation, this paper shows that the strengthening process can also be migrated to a convex inner-approximation that admits W-CVaR equivalence. This migration reduces the number of constraints in the original convex approximation and tightens the feasible region for ancillary variables, resulting in significant computational speedups. Although the migration may

appear direct, there is a subtle source of conservativeness: migrating a strengthening process can introduce additional conservativeness when applied to a convex inner approximation, as shown in Case I of Figure 1. However, and fortunately, we prove that the proposed SFLA will always lead to Case II in Figure 1, which does not introduce extra conservativeness and can even produce a less conservative inner-approximation. The computational speedup and reduced conservativeness of the proposed SFLA address the practical needs of industry-scale problems, such as UC and bilevel strategic bidding. In summary, the main contributions of this paper include:

- (1) We proposed a Strengthened and Faster Linear Approximation (SFLA) to RHS-WDRJCC. By exploiting valid inequalities for a convex inner-approximation that is equivalent to W-CVaR (Chen et al. 2023) under specific hyperparameters, the number of constraints is reduced, and the feasible region of ancillary variables is tightened, which leads to significant computational speedup. Even with the tightening, we demonstrate that the proposed SFLA does not introduce additional conservativeness and can even provide a less conservative inner approximation. In addition, we offer theoretical guidance for setting the SFLA hyperparameter and analyze when SFLA becomes equivalent to exact reformulation or less conservative (or equivalent) to existing convex approximations such as W-CVaR.
- (2) We extend the proposed SFLA to robustness maximization, a decision-making paradigm that can be more interpretable to set the risk level and the Wasserstein radius. We show that the proposed SFLA applies to both risk minimization and radius maximization (the two formulations for robustness maximization), offering computational speedups without introducing extra conservativeness. We further show the equivalence of the feasible regions between risk minimization and radius maximization, and demonstrate that radius maximization can be both more computationally efficient and more robust.
- (3) We demonstrated the superiority of the proposed SFLA through extensive numerical studies on two important real-world large-scale chance-constrained optimization problems. First, in the UC problem, the proposed SFLA achieves up to $10\times$ and on average $3.8\times$ computational speedup compared to the existing strengthened and exact MIP reformulation (Ho-Nguyen et al. 2021) in finding comparable high-quality solutions. We also demonstrate that the proposed SFLA achieves approximation quality comparable to the exact reformulation, except for minor inferiority only occurring under the combination of a high risk level, a large number of historical data, and a small ambiguity set radius. Second, in the bilevel strategic bidding problem where the exact formulation is not applicable due to non-convexity, the proposed SFLA demonstrates $90\times$

speedup in computation time compared to convex inner-approximations such as W-CVaR. In robustness maximization, the proposed SFLA demonstrated over 100× computational speedup compared to other convex approximations with the same solution quality.

This paper is organized as follows. Section 2 lists the existing exact reformulation and convex inner-approximations. Section 3 introduces the proposed SFLA and demonstrates its theoretical properties that make it particularly useful for real-world large-scale problems. Section 4 extends SFLA to a more interpretable robustness-maximization framework. Section 5 carries out numerical studies and Section 6 concludes this paper.

2. Problem Formulation

This paper focuses on RHS-WDRJCC (1b), with the safety set described in Eq. (2). In what follows, we assume $\epsilon \in (0, 1)$ and $\theta > 0$, which are conditions necessary for the subsequent exact reformulation on which this work is based (Chen et al. 2022). This section summarizes the current exact reformulation and approximation methods for RHS-WDRJCC.

2.1. Exact Reformulation

Suppose we have collected N independent and identically distributed (i.i.d.) samples $\{\xi_i\}_{i \in [N]}$ for the random vector ξ , where the index set is defined as $[N] := \{1, \dots, N\}$. Given a decision $\mathbf{x} \in \mathcal{X}$ and a sample ξ_i , we define the distance from ξ_i to the complement of $\mathcal{S}(\mathbf{x})$ as:

$$\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \inf_{\xi' \in \mathbb{R}^K} \{\|\xi_i - \xi'\| \mid \xi' \notin \mathcal{S}(\mathbf{x})\}. \quad (3)$$

Building on the results given by Chen et al. (2022) and the safety set defined in (2), we can reformulate $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))$ to the following analytical expression:

$$\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \left(\min_{p \in [P]} \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ = \min_{p \in [P]} \left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+, \quad (4)$$

where $\|\cdot\|_*$ is the dual norm and $(\cdot)^+$ extracts the non-negative component of the argument, setting negative values to zero. By introducing ancillary variables $s \in \mathbb{R}$ and $\mathbf{r} \in \mathbb{R}^N$, Chen et al. (2022) showed that RHS-WDRJCC (1b) can be expressed as the following distance-based formulation:

$$s \geq 0, \mathbf{r} \geq \mathbf{0}, \quad (5a)$$

$$\epsilon N s - \sum_{i=1}^N r_i \geq \theta N, \quad (5b)$$

$$\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ \geq s - r_i, \quad \forall i \in [N], p \in [P]. \quad (5c)$$

While this distance-based formulation cannot be directly processed by commercial solvers such as Gurobi (Gurobi Optimization, LLC 2024), it can be transformed into an equivalent MIP formulation:

This MIP reformulation, based on Chen et al. (2022), is originally designed for open safety sets, but Ho-Nguyen et al. (2021) demonstrated that it applies to both open and closed sets.

Despite being solvable by advanced MIP solvers, the MIP reformulation remains NP-hard (Luedtke et al. 2010) and its complexity becomes particularly problematic when the original deterministic problem is already complicated or requires convexity for tractable reformulation such as in multilevel problems. Therefore, we focus on convex inner-approximation (for safety) of the \mathbf{x} -feasible region $\mathcal{X}_{\text{Exact}} := \{\mathbf{x} \in \mathcal{X} \mid \exists \mathbf{s}, \mathbf{r} : (5a)–(5c)\}$ defined by the RHS-WDRJCC in model (1) for the decision variable \mathbf{x} , because this is the region that affects the objective value $c(\mathbf{x})$ in (1a).

2.2. Convex Inner-Approximations

In this section, we introduce three representative convex inner-approximation methods for RHS-WDRJCC. The first is the Bonferroni approximation, which is a straightforward and computationally efficient inner approximation that replaces a joint chance constraint (JCC) with individual chance constraints at reduced individual risk levels ϵ_p . According to Chen et al. (2023), the Bonferroni approximation of RHS-WDRJCC (1b) can be formulated as:

$$\mathbf{a}_p^\top \mathbf{x} \leq d_p - \sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\epsilon_p}(-\mathbf{b}_p^\top \boldsymbol{\xi}), \quad \forall p \in [P], \quad (6)$$

where the preset individual risk levels ϵ_p must satisfy $\sum_{p \in P} \epsilon_p \leq \epsilon$. Here, $\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\epsilon_p}(-\mathbf{b}_p^\top \boldsymbol{\xi})$ is the optimal value of a bilinear optimization problem (Chen et al. 2023) for a specific value of ϵ_p . The Bonferroni approximation (6) does not add extra variables or constraints to the original deterministic problem, resulting in minimal computational overhead. A common practice is to set $\epsilon_p = \epsilon/P$ (Geng and Xie 2019), but this often leads to excessive conservativeness, especially when P is large or when the safety constraints are correlated (Chen et al. 2007). However, even with optimized ϵ_p , the approach can still be overly conservative in some cases, and finding optimal values for ϵ_p is generally intractable (Chen et al. 2023). As will be shown in our case studies, the Bonferroni approximation often leads to infeasibility due to this conservativeness. Consequently, we exclude this approach from theoretical comparisons in the following discussion.

An alternative and widely applied inner-approximation is W-CVaR (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023), which can be formulated as the following set of linear constraints:

$$\boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\beta} \in \mathbb{R}, \tau \in \mathbb{R}, \quad (7a)$$

$$\tau + \frac{1}{\epsilon} \left(\theta\beta + \frac{1}{N} \sum_{i \in [N]} \alpha_i \right) \leq 0, \quad (7b)$$

$$\alpha_i \geq w_p \left(\mathbf{a}_p^\top \mathbf{x} - \mathbf{b}_p^\top \boldsymbol{\xi}_i - d_p \right) - \tau, \quad \forall i \in [N], p \in [P], \quad (7c)$$

$$\beta \geq w_p \|\mathbf{b}_p\|_*, \quad \forall p \in [P], \quad (7d)$$

where $\mathbf{w} := [w_p]_{p \in [P]}$ subject to $\mathbf{w} \in \Delta_{++} := \{\mathbf{w} \in (0, 1)^P \mid \sum_{p \in [P]} w_p = 1\}$ is a tunable hyperparameter that affects the performance of the W-CVaR approximation by prioritizing specific constraints in $\mathcal{S}(\mathbf{x})$ (Chen et al. 2023, Ordoudis et al. 2021). Similarly, we define its \mathbf{x} -feasible region as $\mathcal{X}_{\text{WCVaR}}(\mathbf{w}) := \{\mathbf{x} \in \mathcal{X} \mid \exists \alpha, \beta, \tau : (7a)–(7d)\}$.

Another convex approximation (Chen et al. 2023) linearizes constraint (5c) by overcoming the non-convexity caused by the distance function (4). Specifically, it replaces the original distance function $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$ with a conservative approximation $\widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$ given by:

$$\widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) := \kappa_i \left(\min_{p \in [P]} \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right), \quad (8)$$

where $\kappa_i \in [0, 1]$ are pre-selected parameters. Since $\widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) \leq \text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$, the replacement yields the following linear inner-approximation (LA) of the constraints in (5):

$$s \geq 0, \mathbf{r} \geq \mathbf{0}, \quad (9a)$$

$$\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N, \quad (9b)$$

$$\kappa_i \left(\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \geq s - r_i, \quad \forall i \in [N], p \in [P]. \quad (9c)$$

Chen et al. (2023) showed that LA (9) represents a family of the “best” inner-approximation in the $(\mathbf{x}, s, \mathbf{r})$ -feasible region of (5) and LA (9) becomes exact by optimizing $\boldsymbol{\kappa} := [\kappa_i]_{i \in [N]}$. Furthermore, as demonstrated by Chen et al. (2023), LA (9) with $\boldsymbol{\kappa} = \mathbf{1}$ is equivalent to W-CVaR for a specific setting of \mathbf{w} , and becomes exact under certain conditions. These conditions are also extended to our proposed approximation as will be illustrated in Corollaries 3 and 5. The \mathbf{x} -feasible region of the LA approximation can be expressed as: $\mathcal{X}_{\text{LA}}(\boldsymbol{\kappa}) = \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r} : (9a)–(9c)\}$.

3. Strengthened and Faster Linear Approximation

Our proposed approximation starts from the strengthening process of the exact MIP reformulation. As discussed, the RHS-WDRJCC reformulation (5) can be exactly reformulated as a set of MIP

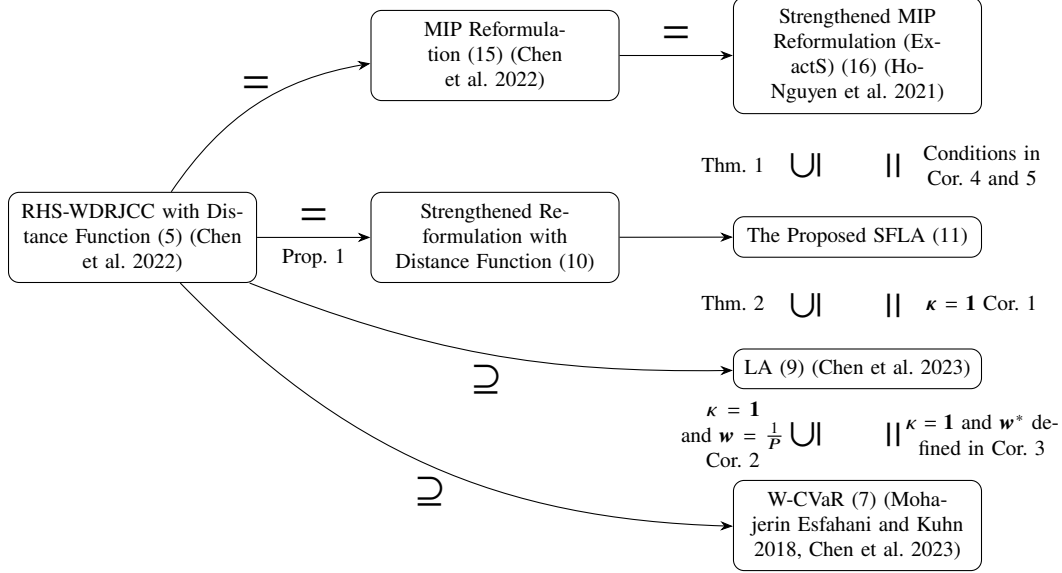


Figure 2 The derivation flow and the formulation comparisons of the x -feasible region defined by different reformulation of RHS-WDRJCC.

constraints in (15). Ho-Nguyen et al. (2021) showed that the exact MIP reformulation (15) can be further strengthened to another exact MIP reformulation (denoted by ExactS), by replacing $NP - \lfloor \epsilon N \rfloor P$ constraints with P valid inequalities. For convenience, the formulation of ExactS is reproduced as Eq. (16) in Appendix A. The x -feasible region of ExactS remains equivalent to the exact MIP set (15), but the number of constraints and the complexity of the whole $(\mathbf{x}, s, \mathbf{r})$ -feasible region are reduced. In this section, we will show that this strengthening process can be migrated to the RHS-WDRJCC (5) expressed by the distance function (Proposition 1). Then, combining with the derivation process for LA, we arrive at our proposed approximation, termed as *Strengthened and Faster Linear Approximation* (SFLA). By combining ExactS and LA, rather than compromising the x -feasible region as illustrated in Case I of Figure 1, we prove that the proposed SFLA is always the Case II: the proposed SFLA will not introduce additional conservativeness and can even be less conservative than LA proposed by Chen et al. (2023) (Theorem 2). The proof of all the theoretical results in this section is provided in Appendix B.

3.1. Derivation and Formulation

Let $k := \lfloor \epsilon N \rfloor$ and assume an ordering $\mathbf{b}_p^\top \boldsymbol{\xi}_1 \leq \dots \leq \mathbf{b}_p^\top \boldsymbol{\xi}_N$ without loss of generality. We further denote by q_p the $(k + 1)$ -th smallest value as $q_p := \mathbf{b}_p^\top \boldsymbol{\xi}_{k+1}$. Then we introduce the index set $[N]_p$ with only $k = \lfloor \epsilon N \rfloor$ elements as $[N]_p := \{i \in [N] \mid \mathbf{b}_p^\top \boldsymbol{\xi}_i < q_p\}$ for $p \in [P]$. We propose the following proposition which demonstrates that the strengthening process for the exact MIP

reformulation ExactS (15) in Ho-Nguyen et al. (2021) can be migrated to the exact RHS-WDRJCC reformulation (5) expressed by the distance function $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$.

PROPOSITION 1. *The feasible set $\mathcal{X}_{\text{Exact}}$ defined through constraints (5) can be equivalently defined by keeping only $\lfloor \epsilon N \rfloor P$ constraints in (5c) as (10a) and replacing the remaining $NP - \lfloor \epsilon N \rfloor P$ constraints in (5c) involving the distance function with P valid inequalities (10b):*

$$\left(\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ \geq s - r_i, \quad \forall i \in [N]_p, p \in [P], \quad (10a)$$

$$\frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s, \quad \forall p \in [P]. \quad (10b)$$

In other words, $\mathcal{X}_{\text{Exact}} = \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r} : (5a), (5b), (10a), (10b)\} = \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r} : (5a)–(5c)\}$.

The structure of the constraints (10) motivates SFLA, which linearizes the non-convex constraints in (10a) using the approximate distance function $\widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$ (8) as was done for LA, while preserving the strengthened formulation in (10). Therefore, SFLA can be expressed as follows:

$$s \geq 0, \mathbf{r} \geq \mathbf{0}, \quad (11a)$$

$$\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N, \quad (11b)$$

$$\kappa_i \left(\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \geq s - r_i, \quad \forall i \in [N]_p, p \in [P], \quad (11c)$$

$$\frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s, \quad \forall p \in [P], \quad (11d)$$

where we still have the preset hyperparameters $\kappa_i \in [0, 1]$ for all $i \in [N]$. The \mathbf{x} -feasible region of the proposed SFLA can be expressed as $\mathcal{X}_{\text{SFLA}}(\boldsymbol{\kappa}) := \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r} : (11a)–(11d)\}$.

3.2. Key Properties and Practical Relevance

By integrating the strengthening process used in ExactS (Ho-Nguyen et al. 2021) with the approximate distance function in LA (Chen et al. 2023), SFLA is not only more computationally efficient but also less (or equally) conservative than existing convex approximation approaches, making it well-suited for practical large-scale settings.

3.2.1. Computational Tractability Many real-world decision-making problems prioritize computational tractability over strict optimality due to limited time budgets (while still requiring good solution quality). As discussed in Section 1, this is the case for applications such as unit commitment and multilevel optimization with chance-constrained lower levels, where even the deterministic

models are already computationally challenging, and convexity is important for tractable reformulation. In these settings, the value of approximation of chance constraints hinges not only on its probabilistic guarantees but also on whether it can be embedded into a large optimization pipeline and solved reliably at scale.

The proposed SFLA is therefore of high practical importance for two reasons. First, compared with the exact MIP reformulation ExactS, SFLA *convexifies* and *linearizes* the WDRJCC (1b) through an approximate distance function. This convexity is crucial for maintaining reformulation tractability, scalability, and numerical robustness, especially when SFLA is applied to the lower-level JCC of multi-level problems or repeatedly solved within a decomposition and rolling-horizon procedure.

Second, compared with common convex approximation benchmarks such as LA (9) and W-CVaR (7) for WDRJCC, SFLA admits a *more compact* representation. Due to the migration of the strengthening process, the valid inequality (11d) cuts off redundant space of ancillary variables and thus brings computational speedup. In addition, this valid inequality enables replacing the original $[N]$ index set with $[N]_p$ in constraint (11c). Due to this replacement, the proposed SFLA only has $\lfloor \epsilon N \rfloor P + P + 1$ constraints (except for basic domain constraints such as non-negativity), while WCVaR has $NP + P + 1$ constraints and LA has $NP + 1$ constraints. For decision-making problems that require high reliability, ϵ will be set small, which reduces the number of constraints significantly. This will, in turn, reduce the memory footprint and is important for large-scale problems.

3.2.2. Safety and No Extra Conservativeness Although computational tractability is prioritized, a good (although not necessarily optimal) solution quality is still important. Also, as an approximation, real-world applications that demand high reliability often prefer inner-approximation over outer-approximation due to safety. This safety property is preserved for the proposed SFLA:

THEOREM 1. *The set $\mathcal{X}_{SFLA}(\boldsymbol{\kappa})$ is a subset of \mathcal{X}_{Exact} , i.e., $\mathcal{X}_{SFLA}(\boldsymbol{\kappa}) \subseteq \mathcal{X}_{Exact}$.*

Regarding solution quality, we show that although being more computationally tractable than the exact MIP reformulation and the two common convex approximations, including LA and W-CVaR, the proposed SFLA will not bring extra conservativeness than LA and W-CVaR, and can even be exact for certain conditions.

First, we show that SFLA does not introduce additional conservativeness compared to LA:

THEOREM 2. *The set $\mathcal{X}_{SFLA}(\boldsymbol{\kappa})$ is a superset of $\mathcal{X}_{LA}(\boldsymbol{\kappa})$, i.e., $\mathcal{X}_{LA}(\boldsymbol{\kappa}) \subseteq \mathcal{X}_{SFLA}(\boldsymbol{\kappa})$.*

Moreover, under a natural choice of the parameter, the two formulations coincide:

COROLLARY 1. We have $\mathcal{X}_{SFLA}(\mathbf{1}) = \mathcal{X}_{LA}(\mathbf{1})$.

Second, SFLA with $\boldsymbol{\kappa} = \mathbf{1}$ dominates W-CVaR with $\boldsymbol{w} = \mathbf{1}/P$, which is the standard W-CVaR in Chen et al. (2023) and Jiang and Xie (2024a):

COROLLARY 2. We have $\mathcal{X}_{WCVaR}(\mathbf{1}/P) \subseteq \mathcal{X}_{SFLA}(\mathbf{1})$.

In addition, they coincide when W-CVaR has tuned hyperparameters:

COROLLARY 3. We have $\mathcal{X}_{SFLA}(\mathbf{1}) = \mathcal{X}_{WCVaR}(\boldsymbol{w}^*)$, where $\boldsymbol{w}^* = [w_p^*]_{p \in [P]}$ and $w_p^* = \frac{\|\boldsymbol{b}_p\|_*^{-1}}{\sum_{l \in [P]} \|\boldsymbol{b}_l\|_*^{-1}}$.

Finally, SFLA can recover the exact feasible set by allowing $\boldsymbol{\kappa}$ to be decision-dependent:

COROLLARY 4. Let $\boldsymbol{\kappa} \in [0, 1]^N$ be extra decision variables and define the updated SFLA set as $\mathcal{X}_{SFLA}^* = \{\boldsymbol{x} \in \mathcal{X} \mid \exists \boldsymbol{\kappa} \in [0, 1]^N, \boldsymbol{s}, \boldsymbol{r} : (11a)–(11d)\}$. Then we have $\mathcal{X}_{SFLA}^* = \mathcal{X}_{Exact}$.

This exactness primarily serves as a theoretical benchmark for understanding the tightness of SFLA since treating $\boldsymbol{\kappa}$ as extra decision variables renders the problem nonconvex and bilinear. Nevertheless, even with fixed $\boldsymbol{\kappa} = \mathbf{1}$, exactness can still arise:

COROLLARY 5. The equality $\mathcal{X}_{SFLA}(\mathbf{1}) = \mathcal{X}_{Exact}$ holds under either of the following conditions: (i) We have $\boldsymbol{\xi}_i \in \mathcal{S}(\boldsymbol{x})$ for all $i \in [N]$ and $\boldsymbol{x} \in \mathcal{X}_{Exact}$; (ii) We have $\epsilon \leq 1/N$.

The second condition in Corollary 5 shows that the proposed SFLA becomes exact when the number of data N is small or the risk level ϵ is low. This also implies that a smaller N or a smaller ϵ can lead to better approximation quality for the proposed SFLA. This implication is particularly important, as DRCCP is often used in cases with limited historical data due to its strong out-of-sample performance. Additionally, safety-critical applications, such as in power systems, generally demand a low risk level. Thus, the proposed SFLA can achieve high approximation quality across a range of practical applications.

Figure 2 summaries the set relationships between the proposed SFLA and other reformulations. These many theoretical properties regarding conservativeness can hold for the proposed SFLA because SFLA combines the derivation process of the both exact reformulation (Ho-Nguyen et al. 2021) and LA (Chen et al. 2023). A detailed proof can be found in Appendix B.

Finally, we provide a simple prior guideline for setting the hyperparameters. The following monotonicity property implies that $\boldsymbol{\kappa} = \mathbf{1}$ yields the largest (least conservative) inner approximation within this family, and hence serves as a natural default choice when no additional tuning information is available:

COROLLARY 6. We have $\mathcal{X}_{SFLA}(\mathbf{1}) \supseteq \mathcal{X}_{SFLA}(\boldsymbol{\kappa}\mathbf{1})$ for all $\boldsymbol{\kappa} \in [0, 1]$.

4. Robustness Maximization

In WDRJCC, the risk level ϵ and the Wasserstein radius θ do admit clear interpretations in terms of risk and ambiguity. However, for decision-makers that demand high reliability but have other backgrounds (e.g., power system operators), it can be easier and more interpretable to first specify a bearable cost as a constraint that bounds the original objective, and then minimize the risk level ϵ or maximize the Wasserstein radius θ to align with their high reliability principle. This paper refers to this alternative decision-making paradigm as *robustness maximization*. In this section, we illustrate how the proposed SFLA can be applied within this paradigm and clarify the link between minimizing the risk level ϵ and maximizing the Wasserstein radius θ .

Note that our robustness maximization framework can be related to the concept of *robust satisficing* (Schwartz et al. 2011), in which a decision-maker specifies an acceptable utility target and then seeks to maximize the robustness of achieving it. A direct formulation of this concept treats the risk level ϵ as an additional decision variable and minimizes ϵ in the objective (Zhao et al. 2023, Online Supplement D). Long et al. (2023) further formalize this concept by proposing a framework in which the *adjusted* bearable cost is required to hold for *all* distributions, with the adjustment determined by their distance from a reference distribution.

4.1. Risk Minimization

In this formulation, the objective is to minimize the risk level ϵ :

$$\min_{\mathbf{x} \in \mathcal{X}, 0 \leq \epsilon \leq \epsilon_{\max}} \epsilon \quad (12a)$$

$$\text{s.t.} \quad \sup_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[\xi \notin \mathcal{S}(\mathbf{x})] \leq \epsilon, \quad (12b)$$

$$c(\mathbf{x}) \leq \tau, \quad (12c)$$

where $\epsilon_{\max} \in [0, 1]$ is an upper bound on the risk level ϵ , determined by decision maker's risk tolerance. Then constraint (12c) ensures that the cost $c(\mathbf{x})$ does not exceed the cost level τ . Moreover, choosing a smaller ϵ_{\max} can also tighten the problem to speedup the computation. Note that, because our objective is to maximize robustness (i.e., to minimize ϵ), the bearable cost constraint (12c) already regulates the optimization. Therefore, there is no need to explicitly penalize ϵ as the case for Zhang (2025)—where the risk level is intended to increase.

The Exact reformulation (5), WCVaR (7), and LA (9) can be easily extended to this setting, by taking $\epsilon \in [0, \epsilon_{\max}]$ as an extra decision variable and making corresponding constraints (5b), (9b)

and (7b) bilinear and nonconvex. Take LA as an example, the \mathbf{x} -feasible region of the extended LA is denoted as $\mathcal{X}'_{LA}(\boldsymbol{\kappa}) := \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r}, \epsilon : (9a)-(9c), \epsilon \in [0, \epsilon_{\max}]\}$.

The risk minimization formulation is not directly applicable to the proposed SFLA in (11) and ExactS in (16), because the valid inequalities and the index set $[N]_p$ for these two sets depend on ϵ (recall the ranking-related definition where $k := \lfloor \epsilon N \rfloor$), which, if formulated, leads to complicated non-convex combinatorial formulations. However, given ϵ_{\max} , we propose to extend SFLA that works for the risk minimization problem, by replacing constraints (11c) and (11d) with the following:

$$\kappa_i \left(\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \geq s - r_i, \quad \forall i \in [N]'_p, p \in [P], \quad (13a)$$

$$\frac{q'_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s, \quad \forall p \in [P]. \quad (13b)$$

where we define $q'_p := \mathbf{b}_p^\top \boldsymbol{\xi}_{k'+1}$ and $[N]'_p := \{i \in [N] \mid \mathbf{b}_p^\top \boldsymbol{\xi}_i < q'_p\}$, with $k' := \lfloor \epsilon_{\max} N \rfloor$ and the same ordering assumption $\mathbf{b}_p^\top \boldsymbol{\xi}_1 \leq \dots \leq \mathbf{b}_p^\top \boldsymbol{\xi}_N$ without loss of generality. In this way, we eliminate the dependency on ϵ . We further denotes its \mathbf{x} -feasible region as $\mathcal{X}'_{SFLA}(\boldsymbol{\kappa}) := \{\mathbf{x} \in \mathcal{X} \mid \exists s, \mathbf{r}, \epsilon : (11a), (11b), (13a), (13b), \epsilon \in [0, \epsilon_{\max}]\}$. We have the following result:

COROLLARY 7. $\mathcal{X}'_{LA}(\boldsymbol{\kappa}) \subseteq \mathcal{X}'_{SFLA}(\boldsymbol{\kappa})$.

The proof can follow the same process as the proof for Theorem 2, noticing the fact that $q'_p \geq q_p$ and $[N]_p \subseteq [N]'_p$, such that the derivation process that holds for constraints (11c) and (11d) in $\mathcal{X}_{SFLA}(\boldsymbol{\kappa})$ will also hold for (13a) and (13b) in $\mathcal{X}'_{SFLA}(\boldsymbol{\kappa})$.

4.2. Radius Maximization

To maximise robustness, one may alternatively maximize the radius of the ambiguity set, subject to the same tolerable cost constraint (14c):

$$\max_{\mathbf{x} \in \mathcal{X}, \theta_{\min} \leq \theta} \theta \quad (14a)$$

$$\text{s.t.} \quad \sup_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})] \leq \epsilon, \quad (14b)$$

$$c(\mathbf{x}) \leq \tau. \quad (14c)$$

This formulation is closely related to the risk minimization formulation (12). Specifically, let $\mathcal{X}_{\text{risk}}(\theta)$ denote the \mathbf{x} -feasible region of (12) for a fixed θ , and let $\mathcal{X}_{\text{rad}}(\epsilon)$ denote the \mathbf{x} -feasible region of (14) for a fixed ϵ . Then we have:

THEOREM 3. *Given $\theta_{\min} \geq 0$ and $\epsilon_{\max} \in [0, 1]$, we have $\mathcal{X}_{\text{risk}}(\theta_{\min}) = \mathcal{X}_{\text{rad}}(\epsilon_{\max})$.*

The proof only focuses on the constraint $\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N$, which appears in most WDRJCC reformulations such as the proposed SFLA, the exact reformulation, ExactS, LA, and WCVaR. We can show that any feasible solution $(\mathbf{x}, s, \mathbf{r})$ to one formulation is also feasible to the other under appropriately chosen values of θ and ϵ . Therefore, Theorem 3 is not restricted to the exact reformulation but extends naturally to SFLA, LA, WCVaR, and ExactS.

Theorem 3 has several implications. On the one hand, although treating $0 \leq \epsilon \leq \epsilon_{\max}$ as an extra decision variable will make the problem bilinear and non-convex, the \mathbf{x} -feasible space (or the $(\mathbf{x}, s, \mathbf{r})$ -feasible space for reformulations such as SFLA) is still convex and linear (because treating θ as an extra decision variable is still linear and convex). This may promote a broader range of optimization techniques for solving problems with ϵ being decision variables, for example, still minimizing the original cost function but plus a penalty cost for the increased risk (Zhang 2025), or possibly enlightening the methods for tuning the individual risk levels ϵ_p in the Bonferroni approximation (this is more complicated due to the coupling $\sum_{p \in [P]} \epsilon_p \leq \epsilon$) (Ding et al. 2022).

On the other hand, Theorem 3 implies that the risk minimization model (12) and the radius maximization model (14) share the same x -feasible region. Therefore, comparing the two formulations reduces to evaluating which objective—minimizing ϵ or maximizing θ —is a more *effective* criterion for selecting a *robust* solution. First, radius maximization is generally more computationally tractable, as it preserves the linearity and convexity of the constraint $\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N$. This allows reformulation techniques such as ExactS, the proposed SFLA, LA, and WCVaR to be directly applied by treating θ as a one-dimensional decision variable. Second, our numerical results (see Appendix E) show that radius maximization tends to achieve slightly higher out-of-sample robustness than risk minimization. Therefore, for robustness maximization, we advocate using radius maximization over risk minimization.

5. Numerical Experiments

This section presents numerical case studies to demonstrate the reduction of computational complexity and also the approximation quality of the proposed SFLA. For all norm calculations in the RHS-WDRJCC formulations, we adopt the L2 norm which has the property of being self-dual. We set $\kappa = \mathbf{1}$ for the proposed SFLA and LA, $w_p = 1/P$ for W-CVaR following Ordoudis et al. (2021), and $\epsilon_p = \epsilon/P$ for Bonferroni approximation. To reach robust conclusions, we implement multiple random runs for each parameter setting as detailed in Appendices C.1 and D.1. The code, the mathematical formulations, and detailed case study settings for our UC and bilevel problems are available at our GitHub repository (Zhou et al. 2026).

5.1. Unit Commitment Problem

The first case study examines a chance-constrained UC problem. UC represents a typical real-world large-scale problem, and its primary objective is to determine the generator on/off statuses for the next scheduling period with a minimal operation cost (Chen et al. 2016). UC is typically formulated as an MIP problem, where each generator requires at least T binary variables to represent on/off statuses over the next T time steps. With thousands of generators (e.g., 1,400 in MISO), non-linear generator cost functions (Xu et al. 2017), and complex network constraints, UC is challenging to solve even in a deterministic case. However, the integration of renewables introduces additional uncertainty into UC, making chance constraints essential to ensure reliability (Wang et al. 2016, Wu et al. 2016, Yang et al. 2020). Given UC’s inherent complexity, minimizing the added computational burden of JCCs is critical.

We compare the proposed SFLA with benchmarks including the exact and strengthened MIP formulation (ExactS) (Ho-Nguyen et al. 2021) in Eq. (16), LA (Chen et al. 2023) in Eq. (9), W-CVaR (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023) in Eq. (7), Bonferroni approximation (Chen et al. 2023) in Eq. (6), and ALSO-X# (Jiang and Xie 2024a).

5.1.1. Evaluation Metrics The UC problem is primarily evaluated using the following metrics:

- 1) *TimeF* (s): This metric records the time (in seconds) to obtain the first “comparable high-quality” solution, defined as the solution with an objective within the `MIPGap` of the final objective achieved by the proposed SFLA (whether optimal or the best found within the `TimeLimit`). If SFLA does not yield a feasible solution within the `TimeLimit`, the *TimeF* of the proposed SFLA is set to the `TimeLimit`, and the *TimeF* for other benchmarks is set to the time when the first feasible solution is found (if any, otherwise set to the `TimeLimit`). Because MIP admits early stop, this metric offers a practical perspective in cases where the solver identifies a solution close to optimal early but requires extra time for verifying optimality.
- 2) *Time* (s): This measures the total time (in seconds) spent to solve the problem to optimality (within `MIPGap`), and is set to the `TimeLimit` if no optimal solution is found within the `TimeLimit`. This metric compliments the previous *TimeF* by providing verified optimality.
- 3) *Obj. Diff.* (%): This is the percentage difference between the optimal value (minimum UC cost) of a benchmark method and that of the SFLA, relative to the optimal value of the SFLA. A positive value indicates that the benchmark achieves a lower cost (better optimality) than the SFLA. We only keep certain random runs where the benchmark method and the proposed SFLA are both solved to optimality.

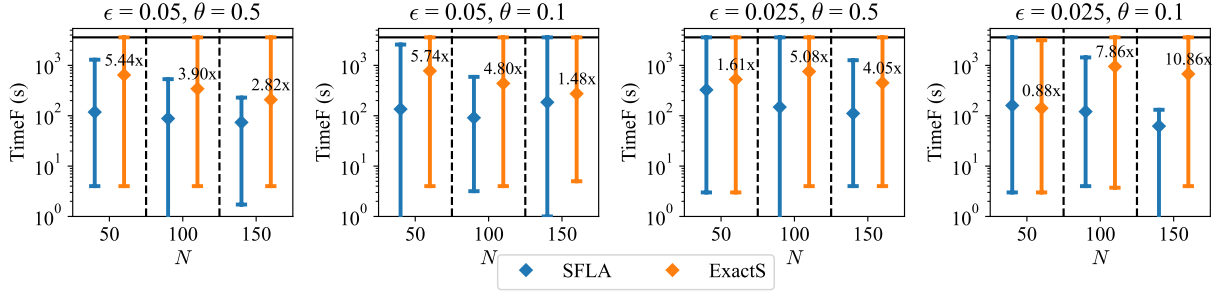


Figure 3 Computation time to obtain the first comparable high-quality solution for the proposed SFLA and the benchmark ExactS for the UC problem.

Note. Dots represent the mean value of the 150 random runs, with error bars indicating the 95% percentile interval (from 2.5th to 97.5th). The black horizontal line represents the 3600s TimeLimit. The optimization horizon is set to $T = 24$. Numbers ending with “x” represent the speedup in the average computing time of the proposed SFLA compared to ExactS.

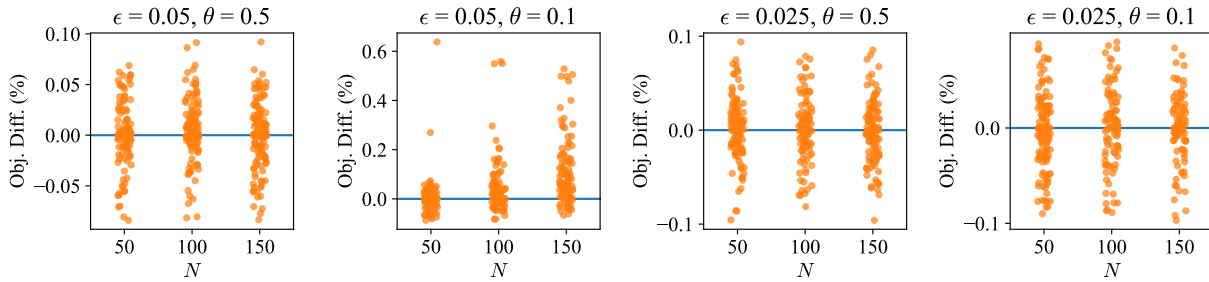


Figure 4 Performance comparison of the optimality for the proposed SFLA and the benchmark ExactS for the UC problem.

Note. Each round dot represents a result of one of the 150 random runs. The blue horizontal line indicates a zero difference of the optimal value of a benchmark method compared to the proposed SFLA. Values higher than this horizontal line indicate that the benchmark achieves a lower cost (better optimality) than the proposed SFLA. The optimization horizon is set to $T = 24$.

4) *Reli. (%)*: This is the out-of-sample joint satisfaction rate of the RHS-WDRJCC. This metric is used to justify the reasonability of the selection of Wasserstein radius θ , and is calculated on 5000 held-out samples.

This section prioritizes *TimeF* over the total computation time *Time* due to comparison with ExactS. MIP can always employ early stopping, and ExactS can achieve better optimality compared to inner approximations (SFLA, LA, W-CVaR, Bonferroni, and ALSO-X#). Therefore, a higher *Time* for ExactS does not necessarily suggest its disadvantage.

5.1.2. Results for Unit Commitment Our main content specifically focuses on the comparison between the proposed SFLA and ExactS. Comparisons with LA, W-CVaR, Bonferroni, and ALSO-X# are excluded here due to their significant disadvantages relative to the proposed SFLA as demonstrated in Appendix C.2. Also, our numerical results suggest that both the proposed SFLA and ExactS exhibit variability. To reach a more robust conclusion, we test a range of parameter settings and implement 150 random runs for each setting. Figure 3 visualizes the distribution of the

TimeF metric. As can be seen, except for a single parameter setting ($\epsilon = 0.025, \theta = 0.1, N = 50$) where the proposed SFLA shows a minor underperformance (12% slower), the SFLA demonstrates significant computational speedup in *TimeF* over ExactS: the computational speedup reaches $10\times$ for a parameter setting ($\epsilon = 0.025, \theta = 0.1, N = 150$), and the speedup is $3.8\times$ for the average *TimeF* of the 12 parameter settings.

Figure 4 compares the quality of the optimal solution of the proposed SFLA with that of ExactS. In most random runs and most parameter settings, the *Obj. Diff.* is within the preset MIPGap, demonstrating comparable optimal solution quality between the proposed SFLA and ExactS. Slight underperformance (less than 0.65%) for certain random runs appears only under the combination of a high risk ($\epsilon = 0.05$), a small radius ($\theta = 0.1$), and a large number of samples ($N \geq 100$), which is consistent with Corollary 5 and the related discussions. However, among these underperformed parameter settings, the average underperformance is still below 0.107%. In addition, real-world applications such as those found in power systems have a small tolerance for risk, making the proposed SFLA an effective approximation method.

Our Appendix C.2 further compares the proposed SFLA and ExactS in different problem scales (different T). It is observed that the proposed SFLA preserves computational efficiency in more complex cases without sacrificing optimality, making it particularly advantageous for practical high-complexity applications such as UC.

5.2. Bilevel Storage Strategic Bidding Problem

This section explores a bilevel optimization problem that allows a price-maker energy storage operator to maximize its profit by participating in a joint day-ahead energy and reserve market. The market participation process can be summarized as follows: Participants submit their bids and offers for each energy and reserve product for specific future delivery times. Bids represent the provision of services to the power system, such as generating power or supplying reserve capacity, while offers involve procuring services, such as purchasing energy for storage charging. In line with most established power markets, we model each bid or offer as a price-quantity pair that is submitted to the market for clearing, which then determines a market-clearing price (marginal price) and cleared quantities bounded by the submitted bid-offer quantities. Upon settlement, market participants receive payment or incur charges at the market-clearing price for their respective cleared quantities, with a binding obligation to fulfill these quantities.

An energy storage operator transitions from a price-taker to a price-maker when its storage power capacity is large enough to influence market outcomes. To maximize profit, the storage

operator determines the optimal bidding or offering strategy by solving a bilevel strategic bidding problem, where the lower level models the market-clearing process (Nasrolahpour et al. 2018, Tómasson et al. 2020, Dimitriadis et al. 2022, Wang et al. 2017). With increasing uncertainty due to renewable generation, the market clearing process can incorporate chance constraints (RHS-WDRJCC), such as those in the UC problem (see Section 5.1). Therefore, to accurately model the market-clearing process, the bilevel problem needs to incorporate an RHS-WDRJCC in the lower level. Consequently, convex approximations to RHS-WDRJCC become important for transforming the bilevel problem into a solvable single-level counterpart via KKT conditions or strong duality. Note that chance constraints at lower levels frequently arise in bilevel and multilevel optimization problems, where uncertainty is revealed after lower-level decisions are made (Beck et al. 2023). Another example is the European multilevel gas market clearing problem (Heitsch et al. 2022).

This section compares the proposed SFLA with typical convex approximation benchmarks, including LA (Chen et al. 2023) in Eq. (9), W-CVaR (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023) in Eq. (7), and Bonferroni approximation (Chen et al. 2023) in Eq. (6). For all the RHS-WDRJCC approximations, the corresponding bilevel problems are reformulated as single-level mixed-integer nonlinear non-convex programming problems, where the lower-level problem is replaced by the corresponding KKT optimality conditions, which can be solved using the Gurobi solver. ExactS is not included because bilevel problem solving requires convexity at the lower level (ALSO-X# is also not applicable due to its iterative nature). The bilevel formulation and also the corresponding single-level reformulation using KKT can be found in our online GitHub repository.

5.2.1. Results for Bilevel Strategic Bidding Figure 5 visualizes the distributions of the evaluation metrics for the proposed SFLA and three benchmarks, i.e., LA, W-CVaR, and Bonferroni. As can be seen from the first two columns of Figure 5, SFLA demonstrates significant speedup in both *Time* and *TimeF*. In particular, SFLA achieves a speedup of up to 90× in solving to optimality (*Time*) compared to LA and W-CVaR, and the speedup is up to 40× in finding the first comparable high-quality solution (*TimeF*) relative to LA and 180× relative to W-CVaR. Although the Bonferroni approximation achieves the fastest computation, its over-conservativeness leads to infeasibility in the majority of parameter settings. It should be noted that the speedup rates compared to LA and W-CVaR decrease for a large T , especially for $\epsilon = 0.05$; this is because LA and W-CVaR reach the `TimeLimit` rather than due to a reduction in the computational superiority of the proposed SFLA.

Our Appendix D.3 provides supplemental analysis on N *Infeasible* and the actual profits that can be attained by using the proposed SFLA and other benchmarks (the third and fourth columns

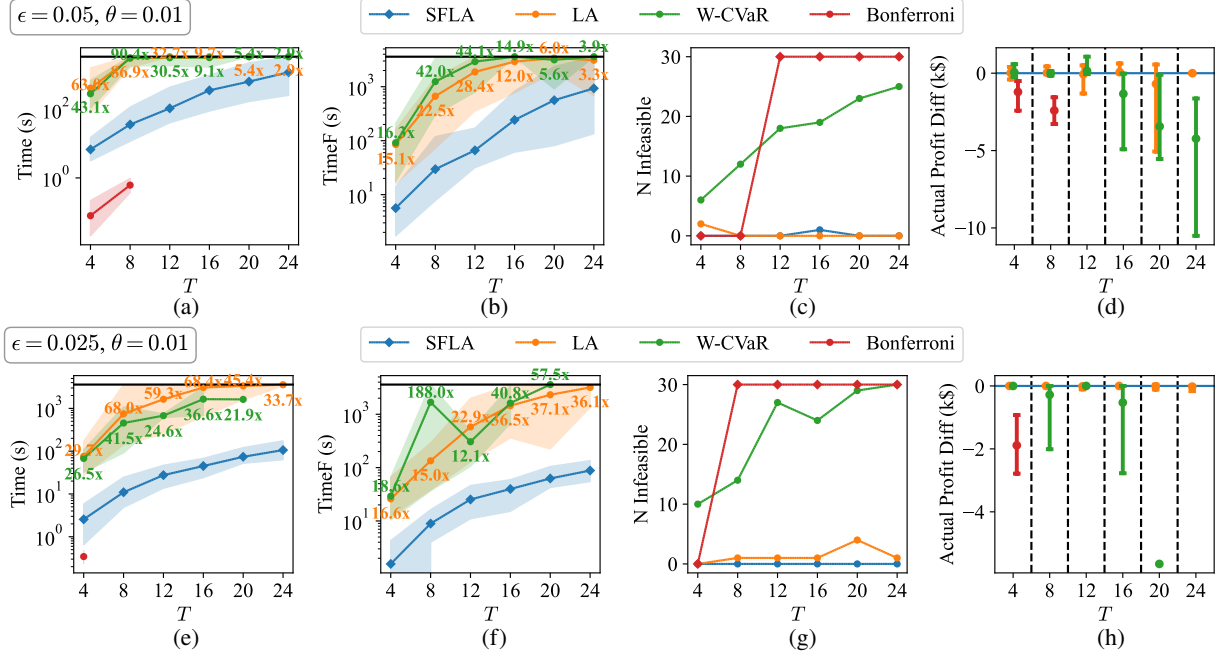


Figure 5 Performance comparison for the bilevel problem.

Note. The evaluations are based on different parameter combinations of risk level (ϵ), radius (θ) parameters, and the number of time steps T . Subplots (e)–(h) replicate (a)–(d) but for varying (ϵ, θ) combinations. The black horizontal line represents the 3600s TimeLimit. Dots represent the mean values of the 30 random runs, with shaded areas representing the 95% percentile interval (from 2.5th to 97.5th). The blue horizontal line indicates a zero difference of the optimal value compared to the proposed SFLA. The *TimeF* (s) of Bonferroni approximation is not displayed because Bonferroni approximation is overly conservative such that it can easily find high *Bilevel Profit* (k\$) (the bilevel objective value) but poor *Actual Profit* (k\$) (plots 5d and 5h). Numbers ending with “x” represent the speedup in the average computing time of the proposed SFLA compared to benchmarks.

of Figure 5). Results demonstrate that the proposed SFLA is more numerically stable and attains comparable actual market profits.

5.3. Risk Minimization for Robustness Maximization

This section compares the proposed SFLA and LA extended for risk minimization. ExactS, WCVaR, and Bonferroni are not compared as their limitations have been illustrated in previous sections. The test case considered is the UC problem in Section 5.1. We first solve the chance-constrained UC with SFLA at ϵ_{\max} to obtain the base cost, and then set the bearable cost τ in Problem (12) to the base cost multiplied by $1 + \text{Cost Tol} (\%)$. Regarding the computing time, as can be seen in Figure 6a, the proposed SFLA leads to significant computational speedup over LA, and the speedup increases from $52.2\times$ to more than $400\times$ for $\epsilon_{\max} = 0.2$ as the problem scales up (a greater optimization horizon T). Moreover, when $\epsilon_{\max} = 0.1$ in Figure 6c, the speedup is up to $763\times$. This is because a smaller $\epsilon_{\max} = 0.1$ leads to a stronger strengthening. Finally, as can be seen in Figures 6b and 6d, both SFLA and LA improve the problem robustness (*Reli. Improve (%)*), which

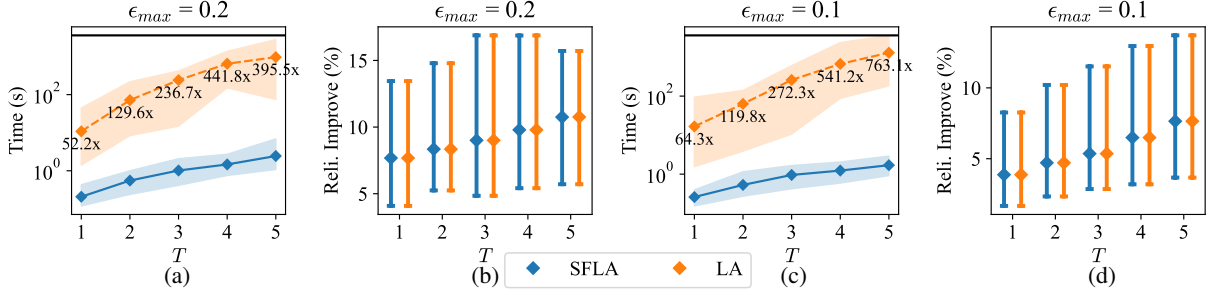


Figure 6 Performance comparison of risk minimization by the proposed SFLA and LA.

Note. The square dots represent the mean values over 30 random runs, with shaded areas and error bars indicating the 95% percentile interval. The x-axis represents the length of the optimization horizon T . Numbers ending with “x” represent the speedup in the average computing time of the proposed SFLA compared to LA. The first two plots correspond to results where the upper bound of the risk level ϵ is set to 0.2, while the remaining two use $\epsilon = 0.1$. In the original UC problem, the Wasserstein radius θ is set to 0.1, and the risk levels are set to the upper bounds mentioned above. $Cost Tol$ (%) is set to 3%.

represents the absolute improvement of the out-of-sample $Reli.$ (%) after robustness maximization). The difference in the mean values of the $Reli. Improve$ (%) is less than 0.01% (the negligible difference is attributed to the existence of multiple optimal solutions). This supports our theoretical results that the extended SFLA is still valid and brings computational speedup compared to LA without being more conservative, even when applied to risk minimization. Our Appendix E further shows the radius maximization results for robustness maximization.

6. Conclusion

This work proposes a novel Strengthened and Faster Linear Approximation (SFLA) for Wasserstein distributionally robust joint chance constraints under right-hand-side uncertainty (RHS-WDRJCC). The proposed SFLA enables more efficient and less conservative solutions to a class of real-world large-scale chance-constrained optimization problems, such as chance-constrained unit commitment (UC) and bilevel optimization for strategic bidding in electricity markets with chance-constrained economic dispatch. By applying valid inequalities to an existing linear approximation (LA), the proposed SFLA reduces the number of constraints and tightens the feasible region for the ancillary variables, resulting in a significant computational speedup. Despite this tightening, we demonstrate that the proposed SFLA can even be a less conservative inner approximation compared to LA and the worst-case conditional value-at-risk (W-CVaR) method. We also extend the proposed SFLA to robustness maximizing, a decision-making paradigm that can be more interpretable to set the risk level and the Wasserstein radius. We prove the equivalence of the feasible regions between risk minimization and radius maximization as two formulations for robustness maximization, and show that radius maximization can be both more computationally efficient and more robust.

Extensive numerical experiments demonstrate the superiority and significance of the proposed SFLA. In the UC problem, we demonstrate that the proposed SFLA achieves up to $10\times$ and on average $3.8\times$ computational speedup compared to the existing exact and strengthened MIP reformulation (ExactS) in finding comparable high-quality solutions. We also demonstrate that the proposed SFLA achieves approximation quality comparable to the exact reformulation ExactS, except for minor inferiority only under the combination of a high risk ϵ , a large number of historical data (N), and a small ambiguity set radius θ . In the bilevel strategic bidding problem where the exact formulation is not applicable due to its non-convexity, the proposed SFLA can lead to $90\times$ computational speedup than existing linear approximation methods (LA and W-CVaR). In robustness maximization, the proposed SFLA demonstrated over $100\times$ computational speedup compared to other convex approximations with the same solution quality.

It is worth noting that the applications of the proposed SFLA are not limited to the UC and bilevel energy bidding problems demonstrated in this paper. The method is also applicable to other large-scale, real-world problems that are already complex in their deterministic form and require tractable approaches to handle uncertainty. Examples include large-scale transportation problems, multilevel market-clearing problems, and multilevel infrastructure planning that involves a lower-level market being cleared before uncertainty is realized.

Historical data plays a significant role in affecting the complexity and solution quality. Future work could also explore scenario reduction techniques (Arpón et al. 2018) to further improve SFLA.

Acknowledgments

This paper was supported by the Engineering and Physical Sciences Research Council (EPSRC) (grant reference number EP/W027321/1). Yihong Zhou's and Yuxin Xia's works were also supported by the Engineering Studentship from the University of Edinburgh. Hanbin Yang's work is funded by the National Natural Science Foundation of China under Grant 72201232 and 72231008, Guangdong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence (2023B1212010001), and Shenzhen Key Laboratory of Crowd Intelligence Empowered Low-Carbon Energy Network, China (project number ZDSYS20220606100601002).

References

- Arpón S, Homem-de Mello T, Pagnoncelli B (2018) Scenario reduction for stochastic programs with conditional value-at-risk. *Mathematical Programming* 170(1):327–356.
- Beck Y, Ljubić I, Schmidt M (2023) A survey on bilevel optimization under uncertainty. *European Journal of Operational Research* 311(2):401–426.
- Chen X, Sim M, Sun P (2007) A robust optimization perspective on stochastic programming. *Operations research* 55(6):1058–1071.

- Chen Y, Casto A, Wang F, Wang Q, Wang X, Wan J (2016) Improving large scale day-ahead security constrained unit commitment performance. *IEEE Transactions on Power Systems* 31(6):4732–4743.
- Chen Z, Kuhn D, Wiesemann W (2022) Data-driven chance constrained programs over Wasserstein balls. *Operations Research* .
- Chen Z, Kuhn D, Wiesemann W (2023) On approximations of data-driven chance constrained programs over Wasserstein balls. *Operations Research Letters* 51(3):226–233.
- Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* 58(3):595–612.
- Dimitriadis CN, Tsimopoulos EG, Georgiadis MC (2022) Strategic bidding of an energy storage agent in a joint energy and reserve market under stochastic generation. *Energy* 242:123026, ISSN 0360-5442.
- Ding Y, Morstyn T, McCulloch MD (2022) Distributionally robust joint chance-constrained optimization for networked microgrids considering contingencies and renewable uncertainty. *IEEE Transactions on Smart Grid* .
- Gabrel V, Murat C, Thiele A (2014) Recent advances in robust optimization: An overview. *European journal of operational research* 235(3):471–483.
- Gao R, Kleywegt A (2023) Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research* 48(2):603–655.
- Geng X, Xie L (2019) Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization. *Annual reviews in control* 47.
- Gurobi Optimization, LLC (2024) Gurobi Optimizer Reference Manual. URL <https://www.gurobi.com>.
- Heitsch H, Henrion R, Kleinert T, Schmidt M (2022) On convex lower-level black-box constraints in bilevel optimization with an application to gas market models with chance constraints. *Journal of Global Optimization* 84(3):651–685.
- Ho-Nguyen N, Kılınç-Karzan F, Küçükyavuz S, Lee D (2021) Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity. *Mathematical Programming* .
- Jiang N, Xie W (2022) ALSO-X and ALSO-X+: Better convex approximations for chance constrained programs. *Operations Research* 70(6):3581–3600.
- Jiang N, Xie W (2024a) ALSO-X#: Better convex approximations for distributionally robust chance constrained programs. *Mathematical Programming* 1–64.
- Jiang N, Xie W (2024b) The terminator: An integration of inner and outer approximations for solving Wasserstein distributionally robust chance constrained programs via variable fixing. *INFORMS Journal on Computing* .
- Long DZ, Sim M, Zhou M (2023) Robust satisficing. *Operations Research* 71(1):61–82.
- Luedtke J, Ahmed S, Nemhauser GL (2010) An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming* 122(2):247–272.
- Mohajerin Esfahani P, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171(1):115–166.

-
- Nasrolahpour E, Kazempour J, Zareipour H, Rosehart WD (2018) A bilevel model for participation of a storage system in energy and reserve markets. *IEEE Transactions on Sustainable Energy* 9(2):582–598.
- Ordoudis C, Nguyen VA, Kuhn D, Pinson P (2021) Energy and reserve dispatch with distributionally robust joint chance constraints. *Operations Research Letters* 49(3):291–299.
- Paredes Á, Aguado JA, Essayeh C, Xia Y, Savelli I, Morstyn T (2023) Stacking revenues from flexible ders in multi-scale markets using tri-level optimization. *IEEE Transactions on Power Systems* 39(2):3949–3961.
- Prékopa A (2013) *Stochastic programming*, volume 324 (Springer Science & Business Media).
- Scarf HE, Arrow K, Karlin S (1957) *A min-max solution of an inventory problem* (Rand Corporation Santa Monica).
- Schwartz B, Ben-Haim Y, Dacso C (2011) What makes a good decision? robust satisficing as a normative standard of rational decision making. *Journal for the Theory of Social Behaviour* 41(2):209–227.
- Sun X, Luh PB, Bragin MA, Chen Y, Wan J, Wang F (2018) A novel decomposition and coordination approach for large day-ahead unit commitment with combined cycle units. *IEEE Transactions on Power Systems* 33(5):5297–5308.
- Tómasson E, Hesamzadeh MR, Wolak FA (2020) Optimal offer-bid strategy of an energy storage portfolio: A linear quasi-relaxation approach. *Applied Energy* 260:114251, ISSN 0306-2619.
- van Ackooij W, Finardi EC, Ramalho GM (2018) An exact solution method for the hydrothermal unit commitment under wind power uncertainty with joint probability constraints. *IEEE Transactions on Power Systems* 33(6):6487–6500.
- Wang Y, Dvorkin Y, Fernández-Blanco R, Xu B, Qiu T, Kirschen DS (2017) Look-ahead bidding strategy for energy storage. *IEEE Transactions on Sustainable Energy* 8(3):1106–1117.
- Wang Y, Zhao S, Zhou Z, Botterud A, Xu Y, Chen R (2016) Risk adjustable day-ahead unit commitment with wind power based on chance constrained goal programming. *IEEE Transactions on Sustainable Energy* 8(2):530–541.
- Wu Z, Zeng P, Zhang XP, Zhou Q (2016) A solution to the chance-constrained two-stage stochastic program for unit commitment with wind energy integration. *IEEE Transactions on Power Systems* 31(6):4185–4196.
- Xu T, Birchfield AB, Gegner KM, Shetye KS, Overbye TJ (2017) Application of large-scale synthetic power system models for energy economic studies.
- Yang Y, Wu W, Wang B, Li M (2020) Analytical reformulation for stochastic unit commitment considering wind power uncertainty with gaussian mixture model. *IEEE Transactions on Power Systems* 35(4):2769–2782.
- Zhang Y (2025) Integer programming approaches for distributionally robust chance constraints with adjustable risks. *Computers & Operations Research* 177:106974.
- Zhao H, Tanneau M, Hentenryck PV (2024) On the viability of stochastic economic dispatch for real-time energy market clearing. URL <https://arxiv.org/abs/2308.06386>.
- Zhao Y, Chen Z, Zhang Z (2023) Distributionally robust chance-constrained p-hub center problem. *INFORMS Journal on Computing* 35(6):1361–1382.
- Zhou Y, Xia Y, Yang H, Morstyn T (2026) Strengthened and faster linear approximation to joint chance constraints with wasserstein ambiguity. URL <http://dx.doi.org/10.1287/ijoc.2024.1073.cd>, available for download at <https://github.com/INFORMSJoC/2024.1073>.

Appendix. Supplemental Information

A. Supplemental Formulations

The MIP reformulation of the exact reformulation (5) is provided below:

$$z \in \{0, 1\}^N, s \geq 0, \mathbf{r} \geq \mathbf{0}, \quad (15a)$$

$$\epsilon N s - \sum_{i=1}^N r_i \geq \theta N, \quad (15b)$$

$$M(1 - z_i) \geq s - r_i, \quad \forall i \in [N], \quad (15c)$$

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + M z_i \geq s - r_i, \quad \forall i \in [N], p \in [P]. \quad (15d)$$

The strengthened exact MIP reformulation (ExactS), presented as Formulation (20) in (Ho-Nguyen et al. 2021), is reproduced below for convenience:

$$z, s, \mathbf{r} \text{ satisfy (15a), (15b), (15c) and } \sum_{i \in [N]} z_i \leq \lfloor \epsilon N \rfloor, \quad (16a)$$

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i + q_p}{\|\mathbf{b}_p\|_*} z_i \geq s - r_i, \quad \forall i \in [N], p \in [P], \quad (16b)$$

$$\frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s, \quad \forall p \in [P]. \quad (16c)$$

where all the notations have been defined in the main content. The \mathbf{x} -feasible region of the ExactS can be expressed as $\mathcal{X}_{\text{ExactS}} = \{\mathbf{x} \in \mathcal{X} \mid \exists z, s, \mathbf{r} : (16a)–(16c)\}$.

B. Proofs

Completing our proofs requires the introduction of the following lemmas:

LEMMA 1 (Lemma 1 in Ho-Nguyen et al. (2021)). *For any $\mathbf{x} \in \mathcal{X}_{\text{Exact}}$, there exists (\mathbf{r}, s) such that s is equal to the $(k+1)$ -th smallest element in $\{\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$ and constraints (5a)–(5c) are satisfied.*

LEMMA 2. $\mathcal{X}_{\text{LA}}(\boldsymbol{\kappa})$ defined through (9) is equivalent to the following set:

$$\left\{ \mathbf{x} \in \mathcal{X} \mid \frac{\theta}{\epsilon} + \mathbb{P}_N\text{-CVaR}_{1-\epsilon}(-\widehat{\text{dist}}(\boldsymbol{\xi}, \mathcal{S}(\mathbf{x}))) \leq 0 \right\}, \quad (17)$$

where $\mathbb{P}_N\text{-CVaR}_{1-\epsilon}(v(\boldsymbol{\xi})) := \min_{s'} \{s' + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, v(\boldsymbol{\xi}_i) - s'\}\}$ is the CVaR of the random variable $v(\boldsymbol{\xi})$ under the empirical probability distribution \mathbb{P}_N (Prékopa 2013).

proof (17) \implies (9): Denote $-\widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$ by $v(\boldsymbol{\xi}_i)$, assume s' is selected such that $s' + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, v(\boldsymbol{\xi}_i) - s'\}$ is minimized, and take $r_i^* = \max\{0, v(\boldsymbol{\xi}_i) - s'\}$ (so $r_i^* \geq 0$). The inequality in (17) indicates:

$$\frac{\theta}{\epsilon} + s' + \frac{1}{\epsilon N} \sum_{i=1}^N r_i^* \leq 0 \Leftrightarrow \epsilon N s' + \sum_{i=1}^N r_i^* \leq -\theta N. \quad (18)$$

Taking $s^* = -s'$, now we will prove that the tuple $(\mathbf{x}, s^*, \mathbf{r}^*)$ satisfy all the constraints in (9), namely (9a)–(9c), for any \mathbf{x} belonging to the set (17). Since $r_i^* \geq 0$ and $-\theta N \leq 0$, (18) implies $s' \leq 0$ (i.e., $s^* \geq 0$) and the following inequality $\epsilon N s - \sum_{i=1}^N r_i \geq \theta N$, which means (9a) and (9b) are satisfied. It remains to prove the satisfaction of (9c). Based on the setting of r_i^* , we have:

$$s^* - r_i^* = s^* - \max\{0, v(\boldsymbol{\xi}_i) + s^*\} = s^* - \max\{0, s^* - \widehat{\text{dist}}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))\}. \quad (19)$$

Notice that constraint (9c) is equivalent to:

$$\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) = \min_{p \in [P]} \left\{ \kappa_i \frac{\mathbf{b}_p^\top \xi + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\} \geq s - r_i \quad \forall i \in [N]. \quad (20)$$

When $\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) \geq s^*$, (9c) must be satisfied for $(\mathbf{x}, s^*, \mathbf{r}^*)$ due to $r_i^* \geq 0$. When $\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) < s^*$, we have $s^* - r_i^* = \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))$ based on (19). Therefore, (20) and equivalently (9c) is satisfied.

(17) \iff (9): For any \mathbf{x} belonging to the $\mathcal{X}_{LA}(\kappa)$, by definition there exist (\mathbf{r}, s) such that constraints (9a)–(9c), are satisfied. Among these constraints, (9c) (equivalently (20)) informs that $r_i \geq s - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))$. Combining with $r_i \geq 0$ in (9a), we further have $r_i \geq \max\{0, s - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}$. Then we have $\epsilon N s - \sum_{i \in [N]} r_i \leq \epsilon N s - \sum_{i \in [N]} \max\{0, s - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}$. Because of $\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N$ in (9b), we have $\epsilon N s - \sum_{i \in [N]} \max\{0, s - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\} \geq \theta N$, which means $\frac{\theta}{\epsilon} - s + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, s - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\} \leq 0$. Now take $s' = -s$, we further have $\frac{\theta}{\epsilon} + s' + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, -s' - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\} \leq 0$, which leads to $\frac{\theta}{\epsilon} + \min_{s'} \{s' + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, -s' + v(\xi_i)\}\} \leq 0$. Therefore, the constraint in (17) is satisfied. \blacksquare

LEMMA 3. For any $\mathbf{x} \in \mathcal{X}_{LA}(\kappa)$, there exists (\mathbf{r}, s) such that s is equal to the $(k+1)$ -th smallest value amongst $\{\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$ and the constraints that define $\mathcal{X}_{LA}(\kappa)$, namely (9a)–(9c), are satisfied.

proof Notice that the CVaR of a random variable $v(\xi)$ with $\xi \sim \mathbb{P}_N$ (defined in Lemma 2) can be equivalently expressed as the following primal and dual forms:

$$\text{Primal: } \mathbb{P}_N\text{-CVaR}_{1-\epsilon}(v(\xi)) = \min_{s' \in \mathbb{R}, \mathbf{r} \geq 0} s' + \frac{1}{\epsilon N} \sum_{i=1}^N r_i \quad (21a)$$

$$\text{s.t. } r_i \geq v(\xi_i) - s', \quad i \in [N]. \quad (21b)$$

$$\text{Dual: } \mathbb{P}_N\text{-CVaR}_{1-\epsilon}(v(\xi)) = \max_{\mathbf{0} \leq \mathbf{y} \leq \mathbf{1}} \frac{1}{\epsilon N} \sum_{i \in [N]} v(\xi_i) y_i \quad (22a)$$

$$\text{s.t. } \frac{1}{N} \sum_{i \in [N]} y_i = \epsilon. \quad (22b)$$

Without loss of generality, assume that we have an ordering $v(\xi_1) \geq \dots \geq v(\xi_N)$. Then a primal-dual optimal pair for CVaR is given by $s'^* = v(\xi_{k+1})$, $r_i^* = \max\{0, v(\xi_i) - s'^*\}$, and $y_i^* = \begin{cases} 1, & i = 1, \dots, k \\ \epsilon N - k, & i = k+1 \\ 0, & i > k+1. \end{cases}$ The validity of the above primal-dual optimal pair holds because of the satisfaction of primal and dual constraints ($\mathbf{y}^* \in [0, 1]$ and $\mathbf{r}^* \geq 0$) and the equality of the objectives of the primal and dual problems $\frac{1}{\epsilon N} \sum_{i \in [N]} v(\xi_i) y_i^* = s'^* + \frac{1}{\epsilon N} \sum_{i \in [N]} r_i^*$, which holds because of $\frac{1}{\epsilon N} \sum_{i \in [N]} v(\xi_i) y_i^* = \frac{1}{\epsilon N} (\sum_{i \in [k]} v(\xi_i) + v(\xi_{k+1})(\epsilon N - k)) = v(\xi_{k+1}) + \frac{1}{\epsilon N} (\sum_{i \in [k]} v(\xi_i) - k v(\xi_{k+1}))$ and $s'^* + \frac{1}{\epsilon N} \sum_{i \in [N]} r_i^* = v(\xi_{k+1}) + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, v(\xi_i) - v(\xi_{k+1})\} = v(\xi_{k+1}) + \frac{1}{\epsilon N} \sum_{i \in [k]} (v(\xi_i) - v(\xi_{k+1})) = v(\xi_{k+1}) + \frac{1}{\epsilon N} (\sum_{i \in [k]} v(\xi_i) - k v(\xi_{k+1}))$, where the reduction from $\sum_{i \in [N]}$ to $\sum_{i \in [k]}$ is valid because of our ordering of $v(\xi_i)$ such that $v(\xi_i) - v(\xi_{k+1}) \leq 0$ for $i \geq k+1$.

Now take $v(\xi_i) = -\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))$. Based on the CVaR interpretation of $\mathcal{X}_{LA}(\kappa)$ in (17) and the CVaR value expressed as (21a), for any $\mathbf{x} \in \mathcal{X}_{LA}(\kappa)$, we have:

$$\frac{\theta}{\epsilon} + \mathbb{P}_N\text{-CVaR}_{1-\epsilon} \left(-\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) \right) \leq 0 \iff \frac{\theta}{\epsilon} + s'^* + \frac{1}{\epsilon N} \sum_{i=1}^N r_i^* \leq 0 \iff \epsilon N s'^* + \sum_{i=1}^N r_i^* \leq -\theta N.$$

Since $r_i^* \geq 0$ and $-\theta N \leq 0$, the above result indicates $s^* := -s'^* \geq 0$ and $\epsilon N s^* - \sum_{i \in [N]} r_i^* \geq \theta N$, which means (9a) and (9b) are satisfied. It remains to prove the satisfaction of (9c). Based on the setting of r_i^* , we have $s^* - r_i^* =$

$s^* - \max\{0, v(\xi_i) + s^*\} = s^* - \max\{0, s^* - \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}$. Recall (see (20)) that constraint (9c) is equivalent to $\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) = \min_{p \in [P]} \left\{ \kappa_i \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\} \geq s - r_i \quad i \in [N]$. When $\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) \geq s^*$, (9c) must be satisfied due to $r_i^* \geq 0$. When $\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) < s^*$, we have $s^* - r_i^* = \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))$, so (9c) is satisfied as well. Therefore, we have proved that the pair $(\mathbf{x}, s^*, \mathbf{r}^*)$ satisfies constraints in (9) for any $\mathbf{x} \in \mathcal{X}_{\text{LA}}(\boldsymbol{\kappa})$, with $s^* = -s'^* = -v(\xi_{k+1})$ and $r_i^* = \max\{0, v(\xi_i) + s^*\}$. Finally, based on our ordering, $v(\xi_{k+1}) = -\widehat{\text{dist}}(\xi_{k+1}, \mathcal{S}(\mathbf{x}))$ is the $(k+1)$ -th greatest value amongst $\{v(\xi_i)\}_{i \in [N]}$, so $s^* = -v(\xi_{k+1})$ is the $(k+1)$ -th smallest value amongst $\{\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$. ■

Proof of Proposition 1

proof It suffices to show that replacing the constraint (5c) with constraints (10a)–(10b) has no impact on the \mathbf{x} -feasible region.

(10) \implies (5): Based on the definition of $[N]_p$, we have $\mathbf{b}_p^\top \xi_i \geq q_p, \forall i \in [N] \setminus [N]_p, p \in [P]$. Then, based on (10b), for any $i \in [N] \setminus [N]_p$ with $p \in [P]$, we have: $\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ \geq \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq \frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s$. As $r_i \geq 0$,

we further have: $\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ \geq s \geq s - r_i$, which, combined with (10a), implies the satisfaction of (5c).

(10) \longleftarrow (5): Based on Lemma 1, for any $\mathbf{x} \in \mathcal{X}_{\text{Exact}}$ defined though (5), there always exists (\mathbf{r}^*, s^*) such that s^* is equal to the $(k+1)$ -th smallest value amongst the set $\{\text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$ and constraints (5a)–(5c) are satisfied. Without loss of generality, we assume the following ordering: $\text{dist}(\xi_1, \mathcal{S}(\mathbf{x})) \leq \dots \leq \text{dist}(\xi_N, \mathcal{S}(\mathbf{x}))$. Then denote the $(k+1)$ -th smallest distance value as $\text{dist}^*(\mathbf{x}) := \text{dist}(\xi_{k+1}, \mathcal{S}(\mathbf{x})) = s^*$. Combined with the analytical expression of the distance function in (4), the pair $(\mathbf{x}, \mathbf{r}^*, s^*)$ satisfies the constraint (10a). It remains to show the satisfaction of (10b) given the pair $(\mathbf{x}, \mathbf{r}^*, s^*)$. Now consider two cases:

1) If $\text{dist}^*(\mathbf{x}) = 0$, then $s^* = 0$. As (5b) must be satisfied for the given (\mathbf{r}^*, s^*) , this leads to a contradiction because $s^* = 0$ implies the non-positivity of the left-hand side of (5b), but the right-hand side of (5b) is strictly positive due to the assumption $\theta > 0$. Therefore, we must have $\text{dist}^*(\mathbf{x}) > 0$.

2) If $\text{dist}^*(\mathbf{x}) > 0$, then for any $p \in [P]$, when $i \geq k+1$, we have: $0 < s^* = \text{dist}^*(\mathbf{x}) \leq \text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))$. Given $0 < \text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))$, based on (4), we have $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \min_{p \in [P]} \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$, which leads to: $s^* \leq$

$\min_{p \in [P]} \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \leq \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$. In other words, for any $p \in [P]$, there will be at least $N - k$ elements

in the set $\left\{ \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}_{i \in [N]}$ not smaller than s^* . Since q_p is defined as the $(k+1)$ -th smallest value of $\{\mathbf{b}_p^\top \xi_i\}_{i \in [N]}$,

$\frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$ is the $(k+1)$ -th smallest element in the set $\left\{ \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}_{i \in [N]}$. Hence, it follows $s^* \leq \frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$, which leads to the satisfaction of (10b). ■

Proof of Theorem 1

proof By comparing $\mathcal{X}_{\text{SFLA}}(\boldsymbol{\kappa})$ in (11) and $\mathcal{X}_{\text{Exact}}$ through its equivalent and strengthened reformulation in (10), the proof is apparent due to $\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)^+ \geq \kappa_i \left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right), \forall i \in [N]$ given $\kappa_i \in [0, 1]$. ■

Proof of Theorem 2

proof It is equivalent to prove that for any $\mathbf{x} \in \mathcal{X}_{\text{LA}}(\boldsymbol{\kappa})$, we have $\mathbf{x} \in \mathcal{X}_{\text{SFLA}}(\boldsymbol{\kappa})$, namely that constraints (11a)–(11d) can be satisfied. Based on Lemma 3, for any $\mathbf{x} \in \mathcal{X}_{\text{LA}}(\boldsymbol{\kappa})$, there always exists (\mathbf{r}^*, s^*) such that s^* is equal to the $(k+1)$ -th smallest value amongst the set $\{\widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$ and constraints (9a)–(9c) are satisfied, which indicate the satisfaction of (11a)–(11c).

It remains to show the satisfaction of (11d) given the same (\mathbf{r}^*, s^*) . Without loss of generality, suppose that we have the following ordering $\widehat{\text{dist}}(\xi_1, \mathcal{S}(\mathbf{x})) \leq \dots \leq \widehat{\text{dist}}(\xi_N, \mathcal{S}(\mathbf{x}))$. Then denote its $(k+1)$ -th smallest element as $\widehat{\text{dist}}^*(\mathbf{x}) := \widehat{\text{dist}}(\xi_{k+1}, \mathcal{S}(\mathbf{x})) = s^*$. Based on (8), we have $\widehat{\text{dist}}^*(\mathbf{x}) = \kappa_{k+1} \left(\min_{p \in [P]} \frac{\mathbf{b}_p^\top \xi_{k+1} + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)$. Then for any $p \in [P]$, when $i \geq k+1$, we have $s^* = \widehat{\text{dist}}^*(\mathbf{x}) \leq \widehat{\text{dist}}(\xi_i, \mathcal{S}(\mathbf{x})) = \kappa_i \left(\min_{p \in [P]} \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \leq \kappa_i \left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right)$. In other words, for any $p \in [P]$, there will be at least $N-k$ elements in the set $\{\kappa_i(\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x})/\|\mathbf{b}_p\|_*\}_{i \in [N]}$ not less than s^* . Since q_p is defined as the $(k+1)$ -th smallest value of $\{\mathbf{b}_p^\top \xi_i\}_{i \in [N]}$, $(q_p + d_p - \mathbf{a}_p^\top \mathbf{x})/\|\mathbf{b}_p\|_*$ is the $(k+1)$ -th smallest element of the set $\{(\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x})/\|\mathbf{b}_p\|_*\}_{i \in [N]}$. We can then deduce that $(q_p + d_p - \mathbf{a}_p^\top \mathbf{x})/\|\mathbf{b}_p\|_*$ is not less than the $(k+1)$ -th smallest element of the set $\{\kappa_i(\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x})/\|\mathbf{b}_p\|_*\}_{i \in [N]}$ due to $\kappa_i \in [0, 1]$. Therefore, we must have $s^* \leq \frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$, which indicates the satisfaction of (11d). ■

Proof of Corollary 1

proof Based on the definition of $[N]_p$, we have $\mathbf{b}_p^\top \xi_i \geq q_p, \forall i \in [N] \setminus [N]_p, p \in [P]$. Therefore, the constraint (11d) in $\mathcal{X}_{\text{SFLA}}(\mathbf{1})$ implies that for any $i \in [N] \setminus [N]_p$ with $p \in [P]$, we have $\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq \frac{q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s$. As $r_i \geq 0$, we further have $\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq s \geq s - r_i$, which, combined with (11c) in $\mathcal{X}_{\text{SFLA}}(\mathbf{1})$, implies the satisfaction of (9c) when $\kappa = \mathbf{1}$. The above process implies $\mathcal{X}_{\text{SFLA}}(\mathbf{1}) \subseteq \mathcal{X}_{\text{LA}}(\mathbf{1})$. Combining with Theorem 2, we have $\mathcal{X}_{\text{SFLA}}(\mathbf{1}) = \mathcal{X}_{\text{LA}}(\mathbf{1})$. ■

Proof of Corollary 2

proof In the W-CVaR formulation (7), constraint (7d) specifies the lower bound of β , $\beta \geq w_p \|\mathbf{b}_p\|_*, \forall p \in [P]$. Note that at optimality we can always safely set $\beta = \max_{p \in [P]} w_p \|\mathbf{b}_p\|_*$, which allows us to eliminate β from the W-CVaR formulation (7). Next, applying the variable substitutions $r \leftarrow \frac{\alpha}{\beta}, -s \leftarrow \frac{\tau}{\beta}, \beta = \max_{p \in [P]} w_p \|\mathbf{b}_p\|_*$ and after reorganizing, we obtain

$$s \geq 0, \mathbf{r} \geq \mathbf{0}, \quad (23a)$$

$$\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N, \quad (23b)$$

$$\frac{w_p \|\mathbf{b}_p\|_*}{\max_{p \in [P]} w_p \|\mathbf{b}_p\|_*} \left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \geq s - r_i, \quad \forall i \in [N], p \in [P]. \quad (23c)$$

Comparing (23) and the LA formulation (9), we see that the only difference is in constraint (23c) and constraint (9c). Replacing w_p with $w_p = \frac{1}{P}$, constraint (23c) can be equivalently written as $\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\max_{p \in [P]} \|\mathbf{b}_p\|_*} \right) \geq s - r_i, \forall i \in [N], p \in [P]$. This is a more conservative constraint compared to constraint (9c) in LA with $\kappa = \mathbf{1}$, because of $\left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right) \geq \left(\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\max_{p \in [P]} \|\mathbf{b}_p\|_*} \right) \geq s - r_i$. Therefore, we have $\mathcal{X}_{\text{WCvAR}}(\mathbf{1}/P) \subseteq \mathcal{X}_{\text{LA}}(\mathbf{1})$, and subsequently $\mathcal{X}_{\text{WCvAR}}(\mathbf{1}/P) \subseteq \mathcal{X}_{\text{SFLA}}(\mathbf{1})$ due to Corollary 1. ■

Proof of Corollary 4

proof Proposition 9 in Chen et al. (2023) demonstrates that the LA method can achieve an optimal value equivalent to the exact reformulation of RHS-WDRJCC by optimizing κ . We can then adjust the objective function $c(\cdot)$ and the set \mathcal{X} to identify each individual feasible solution $\mathbf{x} \in \mathcal{X}_{\text{Exact}}$, thereby establishing set equivalence. This set equivalence also applies to our proposed SFLA due to Theorem 2. ■

Proof of Corollaries 3 and 5

proof The properties in Corollaries 3 and 5 hold for $\mathcal{X}_{\text{LA}}(\kappa)$ as was proved from the perspective of the optimal value in Chen et al. (2023). This perspective can be extended to the whole set following our proof of Corollary 4, which in turn holds for our proposed SFLA $\mathcal{X}_{\text{SFLA}}(\kappa)$ due to Corollary 1. In fact, the first condition in Corollary 5 can be relaxed from $\forall \mathbf{x} \in \mathcal{X}_{\text{Exact}}$ to just the optimal solution \mathbf{x}^* , as the primary interest is typically in the optimal solution rather than the entire set. ■

Proof of Corollary 6

proof The proof is similar to that for Proposition 10 in Chen et al. (2023). ■

Proof of Theorem 3

proof The exact reformulation of the WDRJCC is given by (5). We only need to focus on the constraint $\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N$, where we can show that for any feasible solution $(\mathbf{x}, s, \mathbf{r})$ of one set, the same solution is also feasible to this constraint of the other set by having different θ or ϵ values.

$\mathcal{X}_{\text{risk}}(\theta_{\min}) \subseteq \mathcal{X}_{\text{rad}}(\epsilon_{\max})$: Let $(\mathbf{x}^{\text{risk}}, s^{\text{risk}}, \mathbf{r}^{\text{risk}})$ be any feasible solution to $\mathcal{X}_{\text{risk}}(\theta_{\min})$. We need to show the existence of θ such that $\epsilon_{\max} N s^{\text{risk}} - \sum_{i \in [N]} r_i^{\text{risk}} \geq \theta N$ for some $\theta \geq \theta_{\min}$. Since $(\mathbf{x}^{\text{risk}}, s^{\text{risk}}, \mathbf{r}^{\text{risk}})$ is feasible to $\mathcal{X}_{\text{risk}}(\theta_{\min})$, we have $\epsilon N s^{\text{risk}} - \sum_{i \in [N]} r_i^{\text{risk}} \geq \theta_{\min} N$ with $0 \leq \epsilon \leq \epsilon_{\max}$. Then, it can be verified that setting $\theta = \theta_{\min}$ leads to the result due to $\epsilon_{\max} N s^{\text{risk}} - \sum_{i \in [N]} r_i^{\text{risk}} \geq \epsilon N s^{\text{risk}} - \sum_{i \in [N]} r_i^{\text{risk}} \geq \theta_{\min} N$.

$\mathcal{X}_{\text{rad}}(\epsilon_{\max}) \subseteq \mathcal{X}_{\text{risk}}(\theta_{\min})$: Denote by $(\mathbf{x}^{\text{rad}}, s^{\text{rad}}, \mathbf{r}^{\text{rad}})$ be any feasible solution to $\mathcal{X}_{\text{rad}}(\epsilon_{\max})$. We need to show the existence of θ such that $\epsilon N s^{\text{rad}} - \sum_{i \in [N]} r_i^{\text{rad}} \geq \theta_{\min} N$ for some $\theta \in [0, \epsilon_{\max}]$. Again, because $(\mathbf{x}^{\text{rad}}, s^{\text{rad}}, \mathbf{r}^{\text{rad}})$ is feasible to $\mathcal{X}_{\text{rad}}(\epsilon_{\max})$, we have $\epsilon_{\max} N s^{\text{rad}} - \sum_{i \in [N]} r_i^{\text{rad}} \geq \theta N$ with $\theta \geq \theta_{\min}$. We can then set $\epsilon = \epsilon_{\max}$ to obtain the desired inequality, because of $\epsilon_{\max} N s^{\text{rad}} - \sum_{i \in [N]} r_i^{\text{rad}} \geq \theta N \geq \theta_{\min} N$. ■

C. Supplemental Information for the Unit Commitment Problems

C.1. Case Study Settings for the Unit Commitment Problem

To ensure robust conclusions, we conduct 30 random runs for each parameter setting of ϵ , θ , N , and T , increasing the number of random runs to 150 for specific evaluations when needed. In each random run, we:

- 1) Sample a distinct set of generator cost parameters from a uniform distribution based on the range suggested by Xu et al. (2017), which impacts the objective function of the UC problem.
- 2) Sample a different set of historical samples $\{\xi_i\}_{i \in [N]}$ for the random wind forecasting error, uniformly from a total of 1000 training data samples, which affects the feasible region of the UC problem.
- 3) Select a random starting simulation hour, uniformly sampled from the 24 hours of a day, which affects the demand and generation profiles and thereby alters the feasible region of the UC problem.
- 4) Set a different seed for the optimization solver, affecting the solver behaviors such as tie-breaking rules.

C.2. Supplemental Numerical Experiments of the Unit Commitment Problem

Figure 7a compares all methods with an optimization horizon $T = 24$. As can be seen, in terms of mean values, the proposed SFLA can have 4× faster UC computation than the exact reformulation (ExactS) and even 20× faster than the two linear approximation schemes, namely LA and W-CVaR. This demonstrates the computational efficiency of the

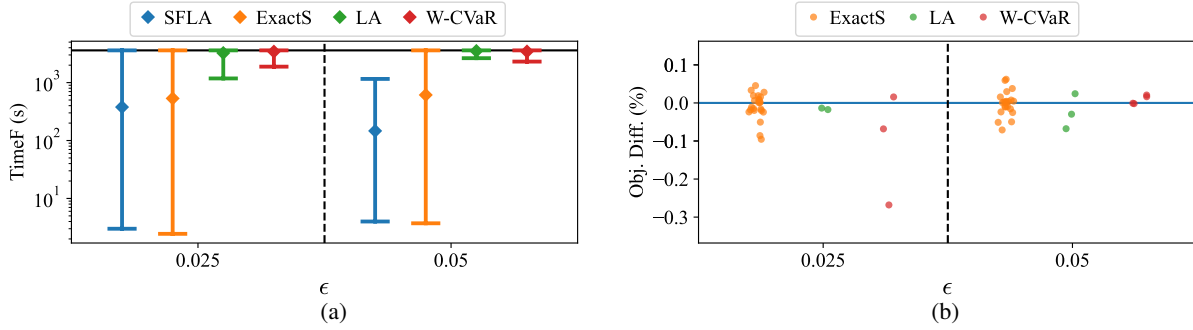


Figure 7 Performance comparison of the proposed SFLA with benchmarks for the UC problem: ExactS, LA, W-CVaR, and Bonferroni for the UC problem.

Note. The Wasserstein radius θ is set to 0.5 and the optimization horizon is set to $T = 24$. (a) Comparison of $TimeF$ (s), where dots represent the mean value of the 30 random runs, with error bars indicating the 95% percentile interval. The black horizontal line represents the one-hour $TimeLimit$. (b) Comparison of $Obj. Diff. (\%)$, where each dot represents the result for one of the 30 runs. The blue horizontal line indicates zero difference between the optimal value of a benchmark method and the proposed SFLA. Values higher than this horizontal line indicate that the benchmark achieves a lower cost (better optimality) than the proposed SFLA. Note that we only keep certain random runs where the method to be compared and the proposed method are both solved to optimum. Many runs of LA and W-CVaR fail to converge to the optimum, and thus the plot only shows a few or zero dots for these two methods. The Bonferroni approximation is infeasible due to its over-conservativeness for all runs and is therefore not displayed.

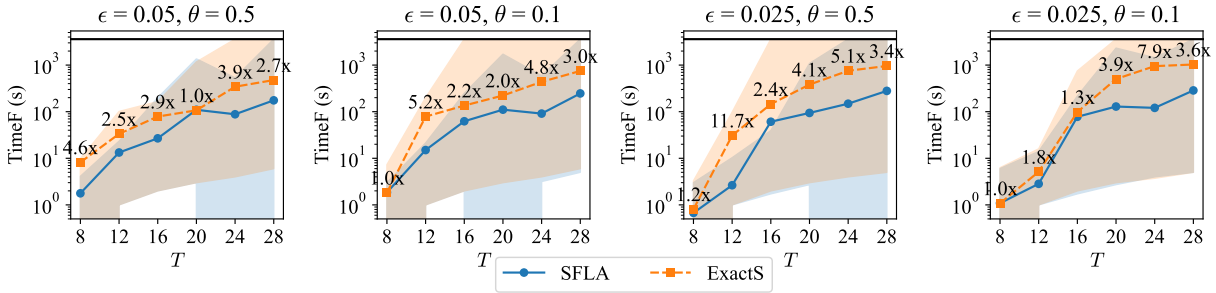


Figure 8 Computation time to obtain the first comparable high-quality solution under different optimization horizons T for the proposed SFLA and the benchmark ExactS for the UC problem.

Note. Dots represent the mean value of the 150 random runs, with shaded areas indicating the 95% percentile interval (from 2.5th to 97.5th). The black horizontal line represents the one-hour $TimeLimit$. The number of historical samples is set to $N = 100$. Numbers ending with “x” represent the speedup in the average computing time of the proposed SFLA compared to ExactS.

proposed SFLA, and also the effectiveness of the strengthening process that makes ExactS even more computationally efficient than linear approximations. Figure 7b shows that the proposed SFLA has optimality comparable to ExactS and LA, where the minor difference of less than 0.1% is attributed to the preset 0.1% $MIPGap$. We also observed a random run of W-CVaR with a 0.25% worse optimality compared to the proposed SFLA, suggesting the higher conservativeness of W-CVaR. The Bonferroni approximation is infeasible in all cases due to its over-conservatism. Overall, the observations above support both the computational gain and the approximation quality of the proposed SFLA.

Figures 8 and 9 compare the proposed SFLA and ExactS under progressively increasing optimization horizons T , a primary driver of problem complexity. Figure 8 shows that the proposed SFLA sustains its computational speedup as problem complexity increases. In contrast, the optimality differences illustrated in Figure 9 exhibit a flat or even

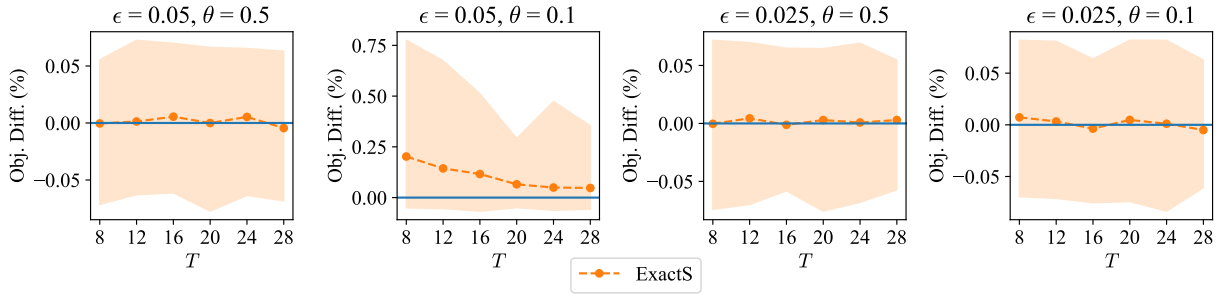


Figure 9 Comparison of optimality for different optimization horizons T for the proposed SFLA and the benchmark ExactS for the UC problem.

Note. The round dots represent the mean value of the 150 random runs, with error bars indicating the 95% percentile interval. The blue horizontal line indicates a zero difference of the optimal value of a benchmark method compared to the proposed SFLA. Values higher than this horizontal line indicate that the benchmark achieves lower cost (better optimality) than the proposed SFLA. The number of historical samples is set to $N = 100$.

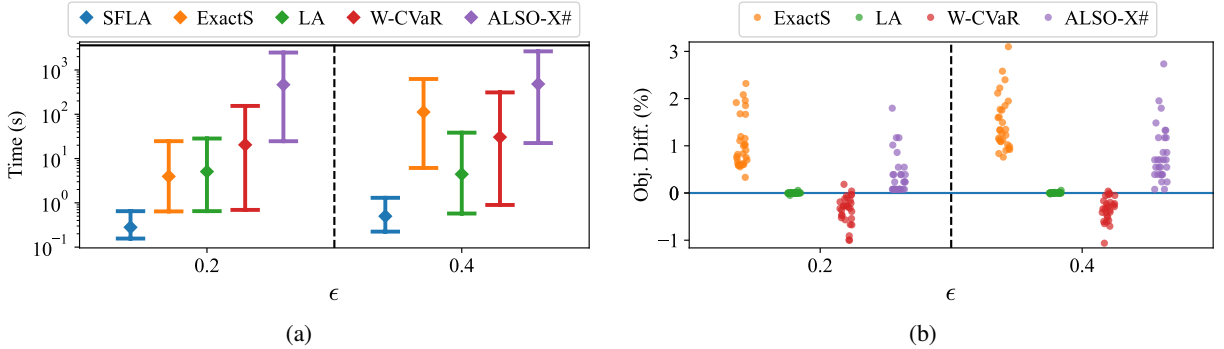


Figure 10 Performance comparison of the proposed SFLA with benchmarks for the UC problem: ExactS, LA, W-CVaR, and ALSO-X# for the UC problem.

Note. The Wasserstein radius θ is set to 0.1 and the optimization horizon is set to $T = 2$ as otherwise ALSO-X# is unsolvable. The number of generators also reduces to 10 for tractability. All graphical elements are defined as in Figure 7 and are not repeated here.

decreasing trend for $(\epsilon = 0.05, \theta = 0.1)$. These results indicate that the proposed SFLA preserves computational efficiency in more complex cases without sacrificing optimality, making it particularly advantageous for practical high-complexity applications such as UC.

We also compared against the ALSO-X# algorithm (Jiang and Xie 2024a). This algorithm interprets W-CVaR as an iterative bisection process. In each iteration, a new bound of the objective (determined by bisection) is added as a constraint, and the process terminates once the W-CVaR evaluation (a conservative approximation of WDRJCC) is satisfied. ALSO-X# modifies this procedure by replacing the conservative W-CVaR stopping criterion with the exact WDRJCC, leveraging the fact that checking WDRJCC feasibility is tractable. This substitution reduces conservativeness and can even recover exact optimality under certain conditions, since the stopping condition becomes less restrictive. However, in each iteration, an optimization problem of the same scale as W-CVaR must be solved, leading to computation times that are several times longer than those of W-CVaR, as reported in their numerical experiments (Jiang and Xie 2024a). Moreover, ALSO-X# requires initial upper and lower bounds of the objective as inputs, and the quality of these bounds can significantly affect the efficiency of the subsequent bisection process. In fact, Jiang and Xie (2024a) used

W-CVaR to calculate the upper bound. Since Figure 7 demonstrates that W-CVaR is both slower and can yield worse optimality compared to SFLA, it follows that ALSO-X# will inevitably be slower than SFLA in obtaining high-quality solutions when W-CVaR is used as the starting point.

Since ALSO-X# only requires initial bounds of the objective as inputs, it can be viewed as an add-on that further improves the solution quality of the proposed SFLA. Accordingly, in our experiments, we use the SFLA output as the initial upper bound of ALSO-X#, set the initial lower bound to 95% of this upper bound, and record only the bisection time as the computation time of ALSO-X#. As illustrated in Figure 10, ALSO-X# is still slower than the MIP-based ExactS method. The main reason is that ALSO-X# does not exploit the strengthening opportunities available in RHS-WDRJCC (a limitation similar to W-CVaR). A promising direction for future work may be the combination of SFLA with ALSO-X#. Figure 10b shows that while ALSO-X# improves solution quality, it still underperforms compared to the exact method. Notably, in this case, the exact method achieves about 3% better optimality than the proposed SFLA because we set a relatively high risk level ϵ to emphasize the improvement brought by ALSO-X#.

D. Supplemental Information for the Bilevel Bidding Problem

D.1. Case Study Settings for the Bilevel Bidding Problem

To ensure reliable conclusions, we conduct 30 random runs for each parameter setting. In each random run, we:

- 1) Sample a distinct set of generator bid prices from a uniform distribution ranging between 80% and 120% of the bid prices used by Nasrolahpour et al. (2018), which influences the objective function of the bilevel problem.
- 2) Sample a different set of historical samples $\{\xi_i\}_{i \in [N]}$ for the random wind forecasting error, uniformly from a total of 1000 training data samples, which affects the feasible region of the bilevel problem.
- 3) Select a random starting simulation hour, uniformly sampled from the 24 hours of a day, which affects the demand and generation profiles and thereby alters the feasible region of the bilevel problem.
- 4) Set a different seed for the optimization solver, affecting the solver behaviors such as tie-breaking rules.

D.2. Supplemental Evaluation Metrics

The exclusive evaluation metrics used in this supplemental case study include:

- 1) *N Infeasible*: This is the number of random runs that are infeasible. We introduce this metric because the high complexity of the bilevel problem introduces numerical issues that may make solvers incorrectly assess a feasible instance as infeasible. A large value of this metric indicates that the method is prone to numerical issues, especially for LA and W-CVaR.
- 2) *Bilevel Profit* (k\$): This represents the optimal profit of storage systems, which is the objective obtained by solving the bilevel problem.
- 3) *Actual Profit* (k\$): The market-clearing process is a single-level problem, where the market operator can use the exact MIP reformulation of RHS-WDRJCC to maximize social welfare. Therefore, the use of ExactS may cause *Bilevel Profit* to differ from *Actual Profit*. We calculate *Actual Profit* through the following steps: i) Retrieve the bids and offers from the bilevel problem solutions; ii) Adjust bid prices slightly downward and offer prices slightly upward (by $\$ 10^{-5}$), setting bid quantities to zero when the bilevel-cleared quantities are zero—these adjustments ensure the desired acceptance or rejection behaviors modeled in the bilevel problem; iii) Input these modified bids and offers into the actual market clearing problem with ExactS (MIP); iv) Calculate the actual market clearing price, which is then

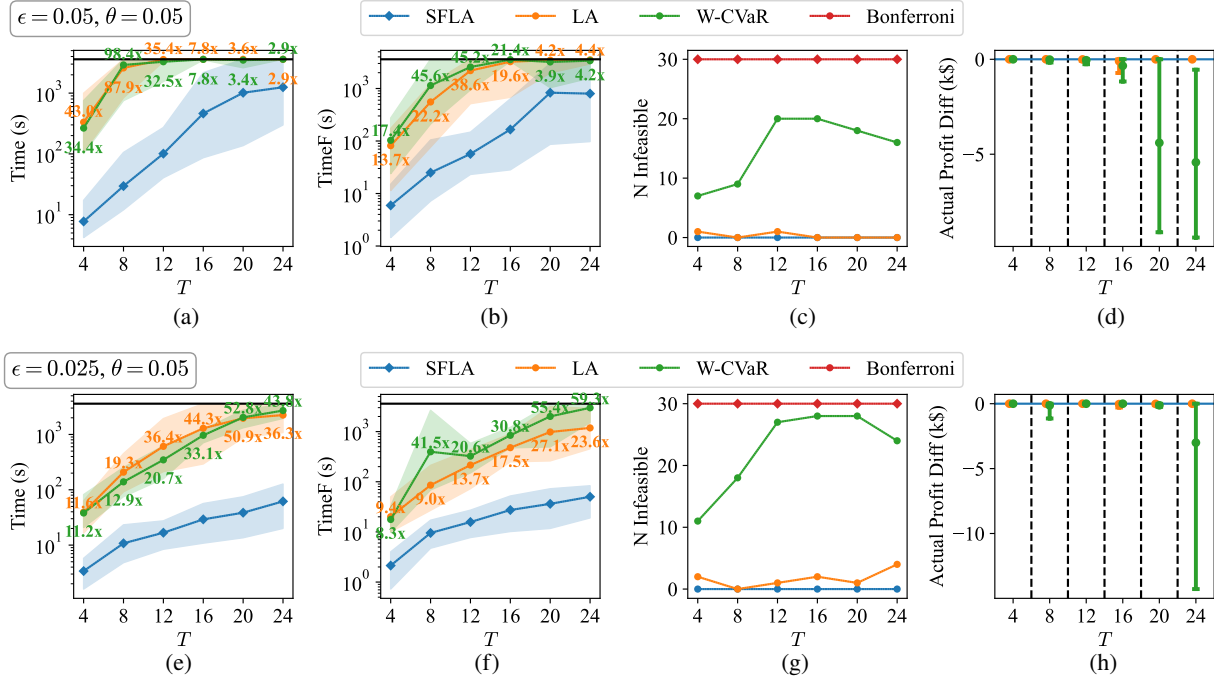


Figure 11 Supplemental performance comparison for the bilevel problem.

Note. The results supplement Figure 5 for $\theta = 0.05$.

used to compute the *Actual Profit* of the storage operator. Due to the MIP formulation in ExactS, optimal dual variables cannot be directly used to determine marginal prices. Instead, we use sensitivity analysis to calculate the clearing price, by comparing the change of the objective function of the actual market clearing after a 10^{-3} MW increase in demand for each time step.

4) *Actual Profit Diff* (k\$): The difference of *Actual Profit* between a benchmark method and that calculated by the proposed SFLA. A negative value represents a lower profit compared to SFLA. We do not use the percentage difference as is in our UC case because the profit could be close to zero, leading to exacerbated results. When calculating *Actual Profit Diff*, we consider only those random runs where both the proposed SFLA and the benchmark identify feasible solutions within the prescribed *TimeLimit*.

D.3. Supplemental Analysis for the Bilevel Bidding Problem

This section provides supplemental analysis on *N Infeasible* and the actual profits that can be attained by using the proposed SFLA and other benchmarks. As can be observed in the third column of Figures 5 and 11, W-CVaR shows a large number of infeasible runs. These infeasibilities are resolved by setting `NumericalFocus=3`, indicating numerical issues. While LA is more numerically stable, it still incurs 22 infeasible runs in total. In contrast, the proposed SFLA has only one infeasible run, demonstrating its higher numerical stability. Note that the infeasibility of the Bonferroni approximation is caused by its over-conservativeness that cannot be recovered by setting `NumericalFocus=3`.

The fourth column of Figures 5 and 11 compares the *Actual Profit* of the benchmarks with that of the proposed SFLA. Among the feasible cases, Bonferroni approximation leads to significantly lower profit than other methods because of its over-conservativeness. It is noteworthy that both LA and W-CVaR can occasionally yield either higher or lower

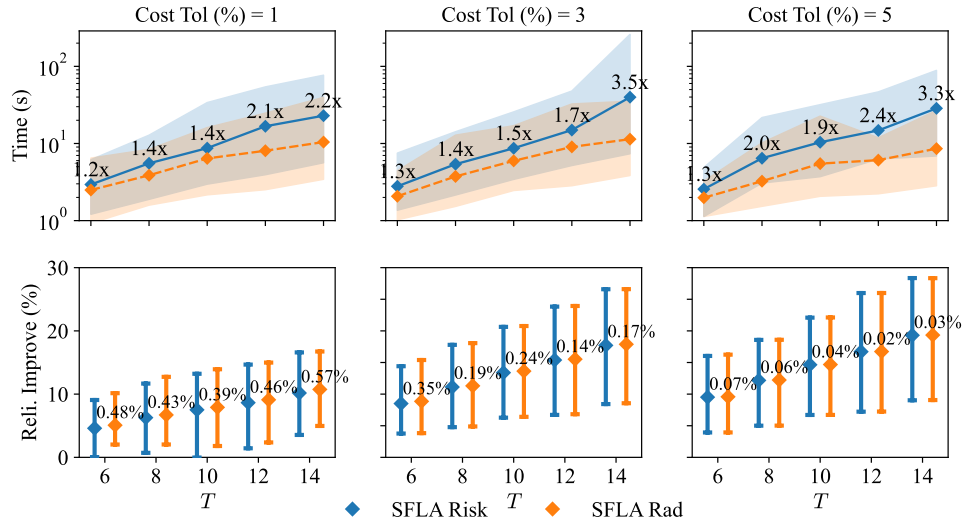


Figure 12 Performance comparison of risk minimization and radius maximization using the proposed SFLA.

Note. The square dots represent the mean values over 30 random runs, with shaded areas and error bars indicating the 95% percentile interval. T is the length of the optimization horizon. In the first row, the numbers ending with “x” are the speed-up of the average computing time of SFLA Rad compared to SFLA Risk. In the second row, the numbers represent the increased absolute amount of “Reli. Improvement” of SFLA Rad compared to SFLA Risk.

profits than the proposed SFLA. This variability arises because the market-clearing model in the bilevel problem does not precisely replicate the actual market-clearing process (cleared with exact MIP-based RHS-WDRJCC), and bilevel problems admit multiple solutions even at optimality. Thus, even when different methods yield the same *Bilevel Profit*, discrepancies can still arise in *Actual Profit*. It should be noted that W-CVaR shows significantly lower *Actual Profit*, especially for a large T , because W-CVaR fails to find a good feasible solution within the `TimeLimit`, rather than due to over-conservativeness.

Finally, among all 720 runs displayed in Figures 5 and 11, there are only three instances where *Bilevel Profit* of the proposed SFLA is slightly (less than 0.009 k\$) below that of LA and W-CVaR. This is caused by another aspect of numerical issue of the proposed SFLA that leads to an incorrect upper bound proved. Setting `NumericalFocus=3` of Gurobi effectively mitigates this issue. Compared to the numerical issue that renders the problem infeasible in LA and W-CVaR, the numerical issue associated with the proposed SFLA is minor.

E. Supplemental Results for Risk Minimization vs Radius Maximization

Figure 12 compares the proposed SFLA in risk minimization (SFLA Risk) and radius maximization (SFLA Rad). The first row shows that radius maximization is more computationally tractable and can lead to up to 3.5x computational speedup compared to risk minimization. In the second row, the result is surprising as radius maximization can even be more robust than risk minimization, and this “greater robustness” result is more significant when the cost reduction budget is low ($Cost Tol=1\%$).