On data-driven robust optimization with multiple uncertainty subsets: Unified uncertainty set representation and mitigating conservatism

Yun Li [a],*, Neil Yorke-Smith [b], Tamas Keviczky [a]

[a] *Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands*
[b] *STAR Lab, Delft University of Technology, Delft, The Netherlands*

ARTICLE INFO

ABSTRACT

Constructing uncertainty sets as unions of multiple subsets has emerged as an effective approach for creating compact and flexible uncertainty representations in data-driven robust optimization (RO). This paper focuses on two separate research questions. The first concerns the computational challenge in applying these uncertainty sets in RO-based predictive control. To address this, a monolithic mixed-integer representation of the uncertainty set is proposed to uniformly describe the union of multiple subsets, enabling the computation of the worst-case uncertainty scenario across all subsets within a single mixed-integer linear programming (MILP) problem. The second research question focuses on mitigating the conservatism of conventional RO formulations by leveraging the structure of the uncertainty set. To achieve this, a novel objective function is proposed to exploit the uncertainty set structure and integrate the existing RO and distributionally robust optimization (DRO) formulations, yielding less conservative solutions than conventional RO formulations, while avoiding the high-dimensional continuous uncertainty distributions and the high computational burden typically associated with existing DRO formulations. Given the proposed formulations, numerically efficient computation methods based on *column-and-constraint generation* (CCG) are also developed. Extensive simulations across three case studies are performed to demonstrate the effectiveness of the proposed schemes.

1. Introduction

Decision-making under uncertainty has gained increasing attention in both scientific research and engineering applications. One of the solutions for handling uncertainties in decisions is robust optimization (RO). By representing uncertainties via uncertainty sets, RO can ensure constraint satisfaction for all possible uncertainties without incurring an unaffordable computational burden and requiring the complete information of uncertainty distributions [1–3]. Despite the popularity and effectiveness of RO formulations, it is well-known that the optimal solution of RO problems can be conservative owing to two main factors.

The first factor contributing to the conservatism of the robust optimal solution is how the uncertainty set is represented. In recent years, with the availability of abundant measurement data and machine learning techniques, data-driven approaches have been developed to construct compact uncertainty sets by uncovering and exploiting the latent features of the uncertainty data, see [4–9] and references therein. Among these data-driven approaches, one innovative methodology is constructing the uncertainty set as a union of multiple subsets, see [6, 7,10] and references therein. By adopting such a methodology, the data-driven uncertainty set is able to adapt to irregular uncertainty

distributions and ensure compactness, which is beneficial for reducing the conservatism of the corresponding robust optimal solutions.

It should be noted that, while the methodology of representing uncertainty sets as unions of multiple subsets has been proven to be effective in several applications, such as power system operation [11] and building climate control [12], this uncertainty set representation together with the existing *column-and-constraint generation* (CCG)-based algorithm could result in computationally demanding optimization problems when applied in robust predictive control settings. This computational tractability in practice is generally neglected in the existing works.

One widely adopted algorithmic solution for solving two-stage RO problems is the so-called CCG algorithm [13,14]. When applying the existing CCG algorithm to solve the RO problem with multiple uncertainty subsets, each uncertainty subset entails solving one optimization problem to compute the worst-case uncertainty in each iteration of the CCG algorithm [6]. As elaborated in Section 3, the total number of uncertainty subsets in RO-based predictive control problems could increase exponentially with the length of the prediction horizon, which

leads to computationally demanding or even intractable problems and limits the applicability of the existing approaches.

The second major factor influencing the conservatism of robust optimization solutions lies in how the design objective is formulated. In the classical RO framework, the optimization problem is designed to minimize the cost function while ensuring constraint satisfaction under the worst-case realization of the uncertainty. While this approach provides strong guarantees, it often leads to overly conservative solutions, especially when the worst-case scenario is highly improbable [1,2,6,15]. To alleviate the conservatism of the optimal solution, alternative formulations have been explored. Stochastic optimization (SO) focuses on minimizing the expected performance with respect to a known probability distribution of the uncertainty, thereby potentially achieving better average-case performance [16–18]. However, SO relies on accurate knowledge of the true distribution of uncertainties, which is rarely available in practice. Further, it is practically possible that distribution shifts between the empirical distribution and the true distribution might exist to degrade the out-of-sample performance. Distributionally robust optimization (DRO) offers a middle ground by optimizing the worst-case expected performance over an ambiguity set of probability distributions centered around a nominal or empirical distribution [15,19–23]. This enables DRO to hedge against distributional misspecification while avoiding some of the excessive conservatism inherent in worst-case RO.

Despite the effectiveness of these existing design options, they either require approximating the probability distribution of uncertainties, which could be high-dimensional and is non-trivial to obtain in practice, or entail solving a computationally demanding optimization problem. For example, in the moment-based DRO formulation [19,22], a semi-definite programming problem needs to be solved at each iteration of the corresponding CCG algorithm; in scenario-based SO [17] and Wasserstein-based DRO formulation [20], the consideration of a large number of uncertainty scenarios leads to large-scale optimization problems and increased computational burden. Moreover, most existing designs of conservatism-reduced objective functions do not exploit the structure of uncertainty sets, especially the promising option of representing the uncertainty set as a union of multiple subsets in data-driven RO settings. In [24,25], data-driven DRO formulations are proposed with the combination of multiple-source uncertainty modeling, where data-driven ambiguity sets are developed for each uncertainty subset separately. While these approaches exploit the latent property in empirical uncertainty samples and reduce dependence on prior knowledge to construct an ambiguity set, they still inherit the computational limitations in corresponding conventional DRO formulations, as the underlying problem structure remains unchanged.

By considering that the uncertainty set consists of multiple subsets, this paper investigates two separate research questions: (1) *how to properly represent the uncertainty set to deal with the computational challenge when applying it in RO-based predictive control problems*; and (2) *how to exploit the uncertainty set structure to design new objective function for mitigating the conservatism of the robust optimal solution*. The main contributions of this paper are summarized as follows:

- A monolithic representation of the uncertainty set consisting of multiple subsets is proposed for RO-based predictive control problems. Given the proposed representation, a computationally efficient CCG algorithm is developed such that only a single MILP problem needs to be solved to compute the worst-case uncertainty scenario across all uncertainty subsets, regardless of the length of the prediction horizon. In comparison, the direct extension of the conventional solutions, e.g., the methods developed in [6,14], to predictive control problems entails solving a set of optimization problems whose number grows exponentially with the prediction horizon.

- Aiming at reducing the conservatism for optimizing the worst-case performance in conventional RO formulations, a novel objective function exploiting the structure of the uncertainty set, together with a CCG-based computation method, is proposed. By combining conventional RO and DRO formulations, the proposed objective function not only achieves less conservative solutions than conventional RO formulations but also avoids approximating the continuous and possibly high-dimensional uncertainty distribution and incurring high computational burden as with the existing SO and DRO formulations, especially in the case of fully adaptive recourse decision and large numbers of uncertainty samples.

- Numerical experiments of three case studies are performed to extensively illustrate the issues of the existing approaches with the uncertainty set consisting of multiple subsets and demonstrate the effectiveness of the proposed schemes.

The remainder of this paper is organized as follows. Section 2 briefly describes the preliminaries and research gaps. Section 3 clarifies the issues of the existing approach and proposes a new uncertainty set representation with a CCG-based computation method. In Section 4, a novel objective function exploiting the structure of the uncertainty set is proposed to mitigate the conservatism of robust optimal solutions. Given the proposed objective function, a CCG-based computation method is also derived. Section 5 presents numerical experiment results. Section 6 concludes this paper.

Notation: boldface capital letters denote matrices, and boldface lowercase letters denote vectors. Sets are represented by calligraphic capital letters. $\mathbf{e}$ denotes all-ones vectors with appropriate dimensions. $\otimes$ is the Cartesian product. $\mathbb{E}[\cdot]$ indicates the expectation of a given random variable. The operator $\mathrm{col}(x_1, \ldots, x_n)$ represents the stacked vector $[x_1^{\mathrm{T}}, \ldots, x_n^{\mathrm{T}}]$.

2. Preliminaries

This section introduces some preliminaries of the two-stage linear RO problem to be investigated, and briefly discusses the research gaps when considering an uncertainty set consisting of multiple subsets.

The following two-stage linear RO problem is considered in this work

$$\min_{\mathbf{x}} \; \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\mathbf{v}\in\mathcal{V}} \min_{\mathbf{y}\in\mathcal{Y}(\mathbf{x},\mathbf{v})} \mathbf{b}^{\mathrm{T}}\mathbf{y} \right\} \tag{1a}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} \le \mathbf{q}, \tag{1b}$$

$$\mathcal{Y}(\mathbf{x}, \mathbf{v}) := \{\mathbf{y} \mid \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \le \mathbf{h}\}, \tag{1c}$$

where $\mathbf{x} \in \mathbb{R}^p$ are the first-stage decision variables (interchangeably called *here-and-now* decision variables), which can contain both continuous and binary elements, $\mathbf{y} \in \mathbb{R}^q$ are the second-stage decision variables (interchangeably called *wait-and-see* decision variables or *recourse* decision variables), which are assumed to be continuous, $\mathbf{v} \in \mathbb{R}^m$ denotes the uncertainties with $\mathcal{V}$ as the corresponding uncertainty set, $\mathcal{Y}(\mathbf{x}, \mathbf{v})$ is the admissible set of $\mathbf{y}$, $(\mathbf{c}, \mathbf{b}, \mathbf{q}, \mathbf{h})$ and $(\mathbf{A}, \mathbf{T}, \mathbf{W}, \mathbf{M})$ are parameter vectors and matrices, respectively, with appropriate dimensions. Without loss of generality, in the remaining parts of this paper, the above two-stage RO problem is assumed to be relatively complete recourse, i.e., the optimization problem is feasible for any possible $\mathbf{x}$ and $\mathbf{v}$ [14]. For problems that do not satisfy this assumption, some other strategic solutions discussed in [13] are also applicable in our proposed design.

To solve the above two-stage RO problem, one commonly adopted approach is the so-called *column-and-constraint generation* (CCG) algorithm. In CCG algorithm, the optimal solution of the RO problem is computed by iteratively solving a master problem and several subproblems (the number of the subproblems to be solved is dependent on the number of uncertainty subsets) to update the lower bound and

upper bound of the objective function until a predefined optimality gap is achieved [6,14]. In this paper, unless mentioned specifically, the existing approach for solving the RO problem (1) refers to the CCG algorithm.

For reducing the conservatism of the robust optimal solutions, including but not limited to the two-stage RO problem in (1), data-driven approaches exploiting historical uncertainty data and machine learning techniques have been extensively explored for constructing compact uncertainty sets. Among these data-driven approaches, one promising methodology is representing the uncertainty set as a union of multiple subsets, such as polytopes and ellipsoids, which has been shown to be effective in improving the compactness and flexibility of the uncertainty sets in handling irregular and complex uncertainty distributions. An uncertainty set $\mathcal{V}$ represented as a union of multiple subsets can be expressed as

$$\mathcal{V} := \bigcup_{k=1}^{K} \mathcal{V}_k \tag{2}$$

where $\mathcal{V}_k$ denotes the kth uncertainty subset with K as the total number of subsets. The uncertainty subsets can be any basic sets, such as box, ellipsoid, and polytope.

While the existing works have demonstrated the effectiveness of formulating an uncertainty set as a union of multiple subsets as in (2), there are two research gaps accompanying such uncertainty set representation to be filled.

- *RO-based Predictive Control With Multiple Uncertainty Subsets*: The RO formulation (1) has been applied in predictive control problems due to its effectiveness in ensuring robust constraint satisfaction [5]. However, when employing the data-driven uncertainty set representation (2), the number of uncertainty subsets could grow exponentially with the prediction horizon. This exponential growth leads to a rapid increase in the number of subproblems solved in each iteration of the existing CCG-based algorithms, making the optimization problem computationally demanding or even intractable. Therefore, developing computationally efficient solutions to address this challenge is crucial for enhancing the practicality of the data-driven RO design with the uncertainty set representation (2) in robust predictive control.

- *Conservatism-Reduced Objective Function With Multiple Uncertainty Subsets*: The conventional RO formulation could suffer from conservatism due to the optimization of worst-case performance, as shown in (1). Given the existing RO formulations with multiple uncertainty subsets, the structure of the uncertainty set is not fully considered in designing the objective function. Consequently, how to properly exploit the property that the uncertainty set consists of multiple subsets to design novel RO formulations is beneficial to reduce the conservatism of the robust optimal solution.

The above research gaps, together with our proposed solutions, will be further elaborated in detail in the subsequent sections.

3. RO-based predictive control with multiple uncertainty subsets

This section details the research gap when applying the uncertainty set structure (2) in robust predictive control, and proposes a novel mixed-integer representation to uniformly describe the unified uncertainty set such that the computational efficiency of the CCG algorithm for solving the RO problem is remarkably improved.

3.1. Problem formulation

Theoretically, even with the uncertainty set represented as in (2), existing data-driven methods for constructing uncertainty sets and the solutions for solving (1) can be directly extended to robust predictive control problems. However, practical challenges arise when applying these methods, which have been overlooked in existing works.

For simplicity of illustration, consider the following deterministic formulation of a predictive control problem

$$\min_{\mathbf{u}_t} \sum_{t=1}^{N} l(\mathbf{s}_t, \mathbf{u}_t) \tag{3a}$$

$$\text{s.t. } \mathbf{s}_{t+1} = \boldsymbol{\Phi}\mathbf{s}_t + \boldsymbol{\Gamma}\mathbf{u}_t + \mathbf{v}_t \tag{3b}$$

$$\mathbf{s}_t \in S, \mathbf{u}_t \in \mathcal{U}, \ \mathbf{v}_t \in \mathcal{V}_t \tag{3c}$$

$$t = 1, \dots, N \tag{3d}$$

where $\mathbf{s}_t$ denotes the system states vector, $\mathbf{u}_t$ denotes the control input vector, $\mathbf{v}_t$ the uncertainties, $l(\mathbf{s}_t, \mathbf{u}_t)$ the stage cost function, S and $\mathcal{U}$ are feasible sets of system states and control inputs, respectively, $\mathcal{V}_t$ is the uncertainty set of $\mathbf{v}_t$, subscript t denotes the tth time step, and N the length of the prediction horizon. In predictive control settings, it is common that the uncertainties $\mathbf{v}_t$ ($t = 1, \dots, N$) are assumed to be independent and identically distributed (I.I.D.) [26,27].

After reformulating the predictive control problem (3) into the corresponding RO formulation (1), the uncertainty vector $\mathbf{v}$ is the stacked uncertainty sequences $\mathbf{v}_t$ within the prediction horizon, i.e., $\mathbf{v} := \text{col}(\mathbf{v}_1, \dots, \mathbf{v}_N)$. In the existing data-driven RO framework, the uncertainty set $\mathcal{V}$ is directly constructed for the stacked uncertainty vector $\mathbf{v}$. While this approach is still theoretically feasible, it will incur some issues. On one hand, for a fixed number of uncertainty samples $\mathbf{v}_t$, considering the stacked uncertainty $\mathbf{v}$ reduces the size of the training dataset (shrinking it to $\frac{1}{N}$ of the original dataset for $\mathbf{v}_t$). This reduction may result in insufficient data and degrade the performance of data-driven approaches for constructing uncertainty sets. On the other hand, directly modeling the uncertainty set for the stacked uncertainty $\mathbf{v}$ increases the dimensionality of uncertainty samples, particularly for long prediction horizon N. As shown in existing studies, high-dimensional uncertainties pose challenges in implementing data-driven methods, including difficulties in hyper-parameter tuning and performance validation.

For example, in building climate predictive control, the prediction horizon could span 12 h with a sampling period of 30 min, and a typical uncertainty is the prediction error of ambient temperature. Given one year of historical uncertainty data of $\mathbf{v}_t$, the number of uncertainty samples for $\mathbf{v}_t$ is 365*24*2. However, when considering the stacked uncertainty samples for $\mathbf{v} = \text{col}(\mathbf{v}_1, \dots, \mathbf{v}_N)$ with $N = 24$, the training set size reduces to $365 * 2$. Consequently, the dataset for $\mathbf{v}$ may be too small to ensure the effectiveness of data-driven approaches in constructing uncertainty sets.

Under the I.I.D. assumption of $\mathbf{v}_t$, one solution is to construct a data-driven uncertainty set for $\mathbf{v}_t$, denoted as $\mathcal{V}_t$, and then extend the per-step uncertainty set $\mathcal{V}_t$ to construct the uncertainty set of $\mathbf{v}$. Assuming that the uncertainty set for $\mathbf{v}_t$ consists of K subsets and denoting the kth uncertainty subset as $\mathcal{V}_{t,k}$, the uncertainty set for $\mathbf{v} = \text{col}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, denoted as $\mathcal{V}$, can be expressed as

$$\mathcal{V} = \underbrace{\mathcal{V}_t \otimes \cdots \otimes \mathcal{V}_t}_{N \text{ times}}, \quad \mathcal{V}_t = \bigcup_{k=1}^{K} \mathcal{V}_{t,k} \tag{4}$$

where N is the length of the prediction horizon considered in the predictive control problem (3).

Remark 1. It can be seen from (4) that the number of subsets constituting $\mathcal{V}$ is K^N, which grows exponentially with the prediction horizon N. By adopting the existing CCG algorithm to solve the RO problem (1) with the uncertainty set (4), each uncertainty subset requires solving a bilinear or mixed-integer linear optimization problem to compute the worst-case uncertainty in every iteration of the algorithm [6]. Consequently, handling K^N subsets translates to solving K^N optimization problems per iteration. In case of large values of K

and N, this approach could be computationally demanding and practically inapplicable. Therefore, the computational challenges arising from representing the uncertainty set as a union of multiple subsets in (4) highlight the need for an effective formulation of the uncertainty set for $\mathbf{v}$.

3.2. Unified uncertainty set representation and computational method design

This section aims to propose a novel representation to uniformly represent the uncertainty set of $\mathbf{v}$ based on the uncertainty subsets of $\mathbf{v}_t$ and derive a computationally efficient solution for solving the corresponding RO problem.

Assumption 1. The uncertainty set of $\mathbf{v}_t$, denoted as $\mathcal{V}_t$, consists of K subsets $\mathcal{V}_{t,k}$ with each uncertainty subset as a nonempty and bounded polytope defined as $\mathcal{V}_{t,k} := \{\mathbf{v}_t | \mathbf{D}_k \mathbf{v}_t \le \mathbf{d}_k\}$.

On the basis of Assumption 1, the following monolithic mixed-integer formulation of the uncertainty set $\mathcal{V}$ is derived

$$\mathcal{V} := \left\{ \text{col}(\mathbf{v}_1, \ldots, \mathbf{v}_N) \middle| \sum_{k=1}^{K} \delta_{t,k} \mathbf{D}_k \mathbf{v}_t \le \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \right.$$
$$\left. \delta_{t,k} \in \mathbb{B}, \sum_{k=1}^{K} \delta_{t,k} = 1, \ t = 1, \ldots, N \right\} \quad (5)$$

where $\delta_{t,k} \in \{0,1\}$ are auxiliary variables. For the formulation (5), given any subset of $\mathcal{V}$, a set of feasible variables $\{\delta_{t,k}^* \mid t = 1, \ldots, N, k = 1, \ldots, K\}$ can be found to uniquely represent a specific uncertainty subset. Conversely, any feasible $\{\delta_{t,k}^* \mid t = 1, \ldots, N, k = 1, \ldots, K\}$ will also define an admissible uncertainty subset of $\mathcal{V}$. Instead of explicitly representing the uncertainty set $\mathcal{V}$ as a union of multiple subsets as in (4), the proposed mixed-integer representation (5) gives a monolithic implicit representation of $\mathcal{V}$ that is feasible for any size of uncertainty subsets K and prediction horizon N.

Theorem 1. *Considering the uncertainty set (5) and decomposing the parameter matrix $\mathbf{M}$ into N column blocks $\{\mathbf{M}_t\}_{t=1}^{N}$ with appropriate dimensions, the RO problem in (1) can be reformulated as in (6), which can be solved via Algorithm 1 within finite iterations.*

$$\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \max_{\delta_{t,k}, \mathbf{v}_t} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \quad (6a)$$

$$s.t. \ \mathbf{Ax} \le \mathbf{q}, \quad (6b)$$

$$\mathbf{Tx} + \mathbf{Wy} + \sum_{t=1}^{N} \mathbf{M}_t \mathbf{v}_t \le \mathbf{h}, \quad (6c)$$

$$\sum_{k=1}^{K} \delta_{t,k} \mathbf{D}_k \mathbf{v}_t \le \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \quad (6d)$$

$$\delta_{t,k} \in \mathbb{B}, \ k = 1, \ldots, K, \ t = 1, \ldots, N, \quad (6e)$$

$$\sum_{k=1}^{K} \delta_{t,k} = 1, \ t = 1, \ldots, N. \quad (6f)$$

Proof. the proof of this theorem is presented in Appendix. $\square$

The master problem **MP** and the subproblem **SP**1 entailed in Algorithm 1 are defined as:

$$\mathbf{MP} : \quad \min_{\mathbf{x}, \eta, \mathbf{y}_i} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \quad (7a)$$

$$\text{s.t.} \ \mathbf{b}^{\mathrm{T}}\mathbf{y}_i \le \eta, \quad (7b)$$

$$\mathbf{Ax} \le \mathbf{q}, \quad (7c)$$

$$\mathbf{Tx} + \mathbf{Wy}_i + \sum_{t=1}^{N} \mathbf{M}_t \mathbf{v}_{t,i} \le \mathbf{h}, \quad (7d)$$

$$i = 1, 2, \ldots, r. \quad (7e)$$

$$\mathbf{SP}1 : \quad \max_{\delta_{t,k}, \mathbf{v}_t} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \quad (8a)$$

$$\text{s.t.} \ \mathbf{Tx} + \mathbf{Wy} + \sum_{t=1}^{N} \mathbf{M}_t \mathbf{v}_t \le \mathbf{h}, \quad (8b)$$

$$\sum_{k=1}^{K} \delta_{t,k} \mathbf{D}_k \mathbf{v}_t \le \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \quad (8c)$$

$$\sum_{k=1}^{K} \delta_{t,k} = 1, \ \delta_{t,k} \in \mathbb{B}, \quad (8d)$$

$$k = 1, \ldots, K, \ t = 1, \ldots, N. \quad (8e)$$

Algorithm 1 *column-and-constraint generation* algorithm for solving (6).

Input: suboptimality gap ϵ
Output: the optimal decision variable $\mathbf{x}$ and objective function value $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$
1: Set $LB = -\infty, UB = \infty, r = 0$
2: **while** $|UB - LB| > \epsilon$ **do**
3:     Solve **MP** in (7) to derive solutions $(\mathbf{x}^*, \eta^*)$ and update $LB = \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$
4:     Solve **SP**1 in (8) or **SP**2 in (9) to derive solutions $\{\mathbf{v}_t^*, \mathbf{y}^*\}$, and update UB as

$$UB = \min\left\{ UB, \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mathbf{b}^{\mathrm{T}}\mathbf{y}^* \right\}$$

5:     Create decision variables $\mathbf{y}_r$, set parameters $\mathbf{v}_{t,r} = \mathbf{v}_t^*$ ($t = 1, \cdots, N$), and add the following constraints to **MP** in (7)

$$\begin{cases} \mathbf{b}^{\mathrm{T}}\mathbf{y}_r \le \eta, \\ \mathbf{Tx} + \mathbf{Wy}_r + \sum_t \mathbf{M}_t \mathbf{v}_{t,r}^* \le \mathbf{h}. \end{cases}$$

6:     $r \leftarrow r + 1$
7: **end while**
8: **Return**: $\mathbf{x}^*$ and $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

The subproblem (8) contains bilinear terms $\delta_{t,k}\mathbf{v}_t$ due to the auxiliary binary decision variable $\delta_{t,k}$, and hence is non-convex and might be intractable to some solvers. These bilinear terms can be further reformulated as mixed-integer linear constraints via big-M formulation. The reformulated subproblem **SP**2 is given in (9).

$$\mathbf{SP}2 : \quad \max_{\delta_{t,k}, \mathbf{v}_t, \mathbf{w}_{t,k}} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \quad (9a)$$

$$\text{s.t.} \ \mathbf{Tx} + \mathbf{Wy} + \sum_{t=1}^{N} \mathbf{M}_t \mathbf{v}_t \le \mathbf{h}, \quad (9b)$$

$$\sum_{k=1}^{K} \mathbf{D}_k \mathbf{w}_{t,k} \le \sum_{k=1}^{K} \delta_{t,k} \mathbf{d}_k, \quad (9c)$$

$$- \Delta(1 - \delta_{t,k})\mathbf{e} \le \mathbf{v}_t - \mathbf{w}_{t,k} \le \Delta(1 - \delta_{t,k})\mathbf{e}, \quad (9d)$$

$$- \Delta\delta_{t,k}\mathbf{e} \le \mathbf{w}_{t,k} \le \Delta\delta_{t,k}\mathbf{e}, \quad (9e)$$

$$k = 1, \ldots, K, \ t = 1, 2, \ldots, N \quad (9f)$$

where $\Delta > 0$ is a sufficiently large constant, and $\mathbf{e}$ is an all-ones vector with appropriate dimension.

Remark 2. It is worth noting that the proposed formulation in (6) is equivalent to the conventional RO formulation (1) with the uncertainty set (4), because the mixed-integer representation of the uncertainty set (5) exactly matches (4) that unifies K^N subsets. The number of possible parameter combinations $\{\delta_{t,k} \mid t = 1, \ldots, N, k = 1, \ldots, K\}$ in (5) is equal to the number of uncertainty subsets in (4), i.e., K^N. In addition, each feasible parameter combination $\{\delta_{t,k} \mid t = 1, \ldots, N, k = 1, \ldots, K\}$ uniquely specifies an uncertainty subset for $\mathbf{v}$ via (5).

Remark 3. In contrast to the existing solution for solving the RO problem (1) with uncertainty set (4), in which K^N numbers of subproblems need to be solved in each algorithm iteration, the reformulated uncertainty set (5) enables a monolithic description of the uncertainty set and ensures that only a single subproblem (**SP**1 or **SP**2), whose decision variables and constraints increase linearly with K and N, to be solved in Algorithm 1 for computing the worst-case uncertainty. This can remarkably improve the computational efficiency of the proposed approach in the case of large values of K and/or N. Additionally, the proposed approach enables the construction of data-driven uncertainty sets for $\mathbf{v}_t$ instead of the stacked uncertainty $\mathbf{v}$ without incurring a high computational burden for solving the corresponding RO problem (1). In comparison with $\mathbf{v}$, $\mathbf{v}_t$ has a lower dimension and more data samples, which will further boost the performance of the data-driven approaches for constructing uncertainty sets. For example, a lower uncertainty dimension makes it easier to tune the parameters/hyper-parameters and evaluate the performance of the data-driven approaches. A larger dataset can also prevent the data-driven approaches from suffering performance degradation, such as overfitting, underfitting, poor model calibration, etc., due to insufficient training data. It is worth highlighting that, while the proposed MILP formulation is NP-hard in theory, it can often be efficiently solved in practice with modern off-the-shelf solvers given the developments in mixed-integer optimization [28].

4. Mitigating conservatism via objective function design

This section proposes an alternative objective function for RO problems by leveraging the structure of uncertainty sets composed of multiple subsets. Additionally, an algorithmic solution is developed to efficiently solve the corresponding RO problem.

4.1. Problem formulation

Based on the structure of the uncertainty set – a union of multiple subsets – this section presents a new formulation of the objective function for RO problems to reduce the conservatism of the optimal solution.

The conventional RO formulation in (1) could lead to conservative solutions due to the objective of worst-case performance optimization. Many existing works have been done to propose new objective functions to reduce the conservatism of the optimal solution. Two prominent alternative objective functions are given as follows.

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\mathrm{T}\mathbf{x} + Q(\mathbf{v}, \mathbf{x}) \right\} \tag{10}$$

where two options of $Q(\mathbf{v}, \mathbf{x})$ are

Option 1: $Q(\mathbf{v}, \mathbf{x}) := \mathbb{E}_{f(\mathbf{v})} \left[\min_{\mathbf{y}} \mathbf{b}^\mathrm{T}\mathbf{y} \right]$, $\tag{11a}$

Option 2: $Q(\mathbf{v}, \mathbf{x}) := \max_{f(\mathbf{v}) \in \mathcal{P}} \mathbb{E}_{f(\mathbf{v})} \left[\min_{\mathbf{y}} \mathbf{b}^\mathrm{T}\mathbf{y} \right]$ $\tag{11b}$

with $f(\mathbf{v})$ as the probability distribution of the uncertainty $\mathbf{v}$, and $\mathcal{P}$ denotes the ambiguity set defining the admissible set of $f(\mathbf{v})$.

The objective function in (11a) seeks to optimize the expected performance to the probability distribution of the uncertainty $\mathbf{v}$, which is the standard setting in stochastic optimization. This approach can reduce conservatism compared with worst-case formulations, but it relies on explicit and accurate knowledge of the underlying uncertainty distribution. Moreover, closed-form or tractable solutions are generally available only for specific distributional families (e.g., Gaussian) or problem structures (e.g., linear decision rules). When samples of the uncertainty are available, sample-based (randomized) methods are often used to approximate the expectation in (11a), resulting in a large-scale deterministic optimization problem whose size grows with the number of samples, which can be computationally demanding. In addition, distribution shifts between the empirical and the true distributions

may exist so that the resulting solution of (11a) may exhibit poor out-of-sample performance.

To address these issues, distributionally robust optimization (DRO), as in (11b), accounts for distributional ambiguity by assuming that the true distribution $f(\mathbf{v})$ is unknown but lies within an ambiguity set $\mathcal{P}$, typically defined around a nominal or empirical distribution. DRO then optimizes the worst-case expected performance over all distributions in $\mathcal{P}$. This approach improves robustness against sampling errors and model misspecification. Nevertheless, achieving both high-probability constraint satisfaction and low conservatism in DRO often requires a sufficiently large and representative set of samples to construct the ambiguity set. Furthermore, as noted in [29], when the decision rule is fully adaptive to the uncertainty, which is the case considered in this work, and the sample size is large, DRO formulations can suffer from the curse of dimensionality, leading to high computational cost.

Motivated by the objective functions in (11b) and the property that the uncertainty set consists of multiple subsets, we propose a new objective function

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\mathrm{T}\mathbf{x} + \max_{\mathbf{p} \in \mathcal{P}} \mathbb{E}_{\mathbf{p}} \left[\max_{\mathbf{v} \in \mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^\mathrm{T}\mathbf{y} \right] \right\} \tag{12}$$

where $\mathbf{p} = [\mathbf{p}_1, \mathbf{p}_2 \cdots, \mathbf{p}_K]^\mathrm{T}$ is the probability vector with $\mathbf{p}_k := \mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$, $\mathcal{P}$ is the ambiguity set defining the admissible set of $\mathbf{p}$.

Remark 4. In contrast to the conventional RO formulation in (1), where the worst-case performance among all possible uncertainty scenarios are optimized, the proposed formulation (12) optimizes the distributionally robust solution w.r.to the expected worst-case performance $\max_{\mathbf{v} \in \mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^\mathrm{T}\mathbf{y}$ over all uncertainty subsets. In (12), the worst-case performance in each uncertainty subset $\mathcal{V}_k$ is weighted based on the probability $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$, and hence a less conservative solution is expected with our proposed formulation than the conventional RO formulation.

Remark 5. In contrast to the SO and DRO formulations in (11), which entail the information of the joint probability distribution $f(\mathbf{v})$ of the continuous uncertainty $\mathbf{v}$, our proposed formulation (12) only considers the discrete probability distribution $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$. Compared with $f(\mathbf{v})$, $\mathbb{P}(\mathbf{v} \in \mathcal{V}_k)$ is much easier to estimate with high accuracy since $f(\mathbf{v})$ is the continuous probability distribution of random variable $\mathbf{v}$ that might be high dimensional. In contrast, the probability vector $\mathbf{p}$ is the distribution of the discrete random variable $\mathbb{I}(\mathbf{v} \in \mathcal{V}_k)$, where $\mathbb{I}(\cdot)$ is the indicator function. An approximation of $\mathbf{p}$ can be readily computed by counting the frequency of $\mathbf{v} \in \mathcal{V}_k$ for all uncertainty data samples, and is a multinomial distribution when uncertainty subsets $\{\mathcal{V}_k, \ k = 1, \ldots, K\}$ are disjoint. Furthermore, the uncertainty is still described via an uncertainty set in (12), and uncertainty samples are not explicitly considered, which are beneficial for maintaining the computational efficiency as in conventional RO formulation. On the contrary, the resulting optimization problems for the existing SO and DRO formulations in the format of (11), e.g., the scenario-based SO formulation [17] and Wasserstein-based DRO formulation [20], need to consider all uncertainty samples and are generally computationally demanding in the presence of large numbers of uncertainty scenarios.

In this work, the ambiguity set $\mathcal{P}$ is defined based on Kullback–Leibler (KL) divergence as

$$\mathcal{P} := \left\{ \mathbf{p} \mid \mathbf{p} \geq 0, \ \mathbf{e}^\mathrm{T}\mathbf{p} = 1, \ KL(\bar{\mathbf{p}}, \mathbf{p}) \leq \rho \right\}, \tag{13}$$

where $KL(\bar{\mathbf{p}}, \mathbf{p}) := \sum_k \bar{\mathbf{p}}_k \log(\frac{\bar{\mathbf{p}}_k}{\mathbf{p}_k})$ is the KL divergence function measuring the similarity of two probability vectors $\bar{\mathbf{p}}$ and $\mathbf{p}$ with $\bar{\mathbf{p}}$ as the approximated probability vector extracted from historical data, $\rho \geq 0$ is a user-defined parameter determining the size of the ambiguity set.

Considering the objective function (12) as well as the ambiguity set definition (13) and applying epigraphical reformulation, it gives to the following optimization problem

$$\min_{\mathbf{x},\eta} \mathbf{c}^\mathrm{T}\mathbf{x} + \eta \tag{14a}$$

$$\text{s.t.} \max_{\mathbf{p} \in \mathcal{P}} \left\{ \mathbb{E}_{\mathbf{p}} \left[\max_{\mathbf{v} \in \mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}} \mathbf{y} \right] \right\} \leq \eta, \tag{14b}$$

$$\mathbf{p} \geq 0, \ \mathbf{e}^{\mathrm{T}} \mathbf{p} = 1, \ KL(\bar{\mathbf{p}}, \mathbf{p}) \leq \rho, \tag{14c}$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{q}, \tag{14d}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \leq \mathbf{h}. \tag{14e}$$

Remark 6. For the proposed formulation in (14), the conservatism of the optimal solution depends on both the size of the ambiguity set $\mathcal{P}$, determined by the parameter ρ, and the number of uncertainty subsets K. Specifically, conservatism can be tuned in two ways: by adjusting K when constructing the data-driven uncertainty set, and by varying ρ when designing the ambiguity set. In general, a larger ρ yields more conservative solutions, whereas a larger K reduces conservatism. In extreme cases, for a fixed K, a sufficiently large $\mathcal{P}$ leads to the worst-case optimization, as in the RO formulation (1), since the ambiguity set then includes the probability vector $\mathbf{p}$ that places unit mass on the subset containing the worst-case sample. Similarly, when the uncertainty set has no subset (i.e., $K = 1$), the formulation reduces to the worst-case optimization regardless of the chosen ρ. In practice, ρ can be tuned through cross-validation or out-of-sample testing, by evaluating different ambiguity set sizes and selecting the one that best balances robustness and performance.

4.2. Computational method design

The proposed formulation (14) contains quadruple-level optimization problems and is computationally intractable to numerical solvers. This section will present a computational method based on the CCG algorithm to efficiently solve the optimization problem (14).

Assumption 2. For the robust optimization problem (14), the uncertainty set $\mathcal{V}$ is a union of K nonempty and bounded subsets $\mathcal{V}_k$ defined as $\mathcal{V}_k := \{\mathbf{v} \mid \mathbf{D}_k \mathbf{v} \leq \mathbf{d}_k\}$.

Theorem 2. *Assuming Assumption 2 holds, the optimization problem (14) can be solved via Algorithm 2 by iteratively solving a master problem $\mathbf{MP}_{DRO}$ in (15) and K subproblems $\mathbf{SP}_{DRO}^k$ in (16) within finite number of iterations.*

$$\mathbf{MP}_{DRO}: \quad \min_{\mathbf{x}, \eta, \nu, \mu, \mathbf{y}_{k,i}} \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta \tag{15a}$$

$$\text{s.t.} \ \nu \sum_{k=1}^{K} \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}} \mathbf{y}_{k,i} - \mu}{\nu} - 1 \right) + \mu + \rho\nu \leq \eta, \tag{15b}$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{q}, \tag{15c}$$

$$\mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_{k,i} + \mathbf{M}\mathbf{v}_{k,i}^* \leq \mathbf{h}, \tag{15d}$$

$$\nu \geq 0, \tag{15e}$$

$$i = 1, \dots, r, \ k = 1, \dots, K. \tag{15f}$$

$$\mathbf{SP}_{DRO}^k: \quad \max_{\mathbf{v}} \min_{\mathbf{y}} \mathbf{b}^{\mathrm{T}}\mathbf{y} \tag{16a}$$

$$s.t. \ \mathbf{T}\mathbf{x}^* + \mathbf{W}\mathbf{y} + \mathbf{M}\mathbf{v} \leq \mathbf{h}, \tag{16b}$$

$$\mathbf{D}_k \mathbf{v} \leq \mathbf{d}_k. \tag{16c}$$

Proof. The proof is provided in Appendix. □

Remark 7. It should be noted that the master problem $\mathbf{MP}_{DRO}$ (15) is nonlinear due to the constraint (15b) even if the original constraints and objective function in (14) are linear. Besides, it can be seen from Algorithm 2 that the number of these nonlinear constraints will increase

with the iteration of the algorithm, which could degrade the computational efficiency of our proposed scheme. One solution to alleviate this issue is applying the following reformulation for constraint (15b):

$$\mu + \rho\nu + \nu \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\phi_k - \mu}{\nu} - 1 \right) \leq \eta, \tag{17a}$$

$$\mathbf{b}^{\mathrm{T}} \mathbf{y}_{k,i} \leq \phi_k, \ i = 1, \dots, r, \ k = 1, \dots, K, \tag{17b}$$

where ϕ_k are auxiliary decision variables. The above reformulation is valid because the left-hand side of (15b) is strictly increasing in $\mathbf{b}^{\mathrm{T}} \mathbf{y}_{k,i}$. By replacing (15b) via (17), only linear constraints are added in each iteration, and the number of nonlinear constraints (17a) is fixed regardless of the algorithm iteration, which is beneficial in improving the computational efficiency of Algorithm 2.

Algorithm 2 *column-and-constraint generation* algorithm for solving (14).

Input: suboptimality gap ϵ

Output: optimal decision variable $\mathbf{x}^*$ and objective function value $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

1: Set $LB = -\infty$, $UB = \infty$, $r = 0$
2: **while** $|UB - LB| > \epsilon$ **do**
3:     Solve $\mathbf{MP}_{DRO}$ (15) to obtain solutions $(\mathbf{x}^*, \eta^*, \mu^*, \nu^*)$ and update $LB = \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$
4:     Solve K subproblems $\mathbf{SP}_{DRO}^k$ (16) to obtain solutions $\{\mathbf{v}_k^*, \mathbf{y}_k^*\}_{k=1}^K$, and update UB as

$$UB = \min\Bigg\{ UB, \ \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* +$$

$$\nu^* \sum_{k=1}^{K} \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_k^* - \mu^*}{\nu^*} - 1 \right) \Bigg\}$$

5:     Create decision variables $\{\mathbf{y}_{k,r}\}_{k=1}^K$, set parameters $\mathbf{v}_{k,r}^* = \mathbf{v}_k^*$ ($k = 1, \cdots, K$), and add the following constraints to $\mathbf{MP}_{DRO}$ in (15)

$$\begin{cases} \mu + \rho\nu + \nu \sum_{k=1}^{K} \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,r} - \mu}{\nu} - 1 \right) \leq \eta, \\ \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y}_{k,r} + \mathbf{M}\mathbf{v}_{k,r}^* \leq \mathbf{h}, \ k = 1, \cdots, K. \end{cases}$$

6:     $r \leftarrow r + 1$
7: **end while**
8: **Return**: $\mathbf{x}^*$ and $\mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \eta^*$

Remark 8. It is worth noting that, while the constraints (15b) and (17a) are nonlinear, they are convex, and $\mathbf{MP}_{DRO}$ becomes a convex optimization when the decision variable $\mathbf{x}$ is continuous. The convexity of the constraints (15b) and (17a) is proved in Appendix. Several off-the-shelf solvers, such as Gurobi, Ipopt and MadNLP, can deal with this type of nonlinearity. The subproblem $\mathbf{SP}_{DRO}$ (16) is a bi-level linear optimization problem that is not numerically tractable to solvers. However, by applying strong duality or KKT-based reformulations and big-M approach, the inner-level optimization problem in (16) can be eliminated, and this bilevel optimization problem can be reformulated as a mixed-integer linear programming problem, see [13,14] for more details.

Remark 9. Compared to conventional RO formulations, both the proposed formulation (14) and existing DRO formulations, e.g., Wasserstein-based DRO formulation [20], result in increased computational burden, though for different reasons. In our proposed formulation, this computational burden arises primarily from the nonlinear constraint (15b), particularly when the first-stage decision variables $\mathbf{x}$ include binary components, making the master problem $\mathbf{MP}_{DRO}$ (15) a mixed-integer nonlinear optimization problem. As for the existing DRO formulation, we take the Wasserstein-based DRO [20] as an example. When considering the fully adaptive recourse decision rule

and large numbers of uncertainty samples, while the resulting optimization problem still retains the linearity of the original deterministic formulation, the increased computational burden is mainly caused by solving numerous subproblems in every iteration of the CCG algorithm since each uncertainty scenario will incur an optimization problem, see Algorithm 1 in the supplementary material [30]. The curse of dimensionality issue of the Wasserstein-based DRO formulation is also mentioned in [29].

5. Simulation results

This section presents three case studies to illustrate the effectiveness of the proposed schemes in this paper. *Case Study 1* considers robust predictive control of building climate to demonstrate the effectiveness of the proposed approach in Section 3. *Case Study 2* and *Case Study 3* showcase the approach designed in Section 4 with robust location transportation planning and chemical process network planning problems, respectively.

All simulations are implemented on an Intel Xeon W-2223 CPU at 3.6 GHz with 16 GB RAM. Optimization problems are modeled via Python package `gurobipy` and solved via `Gurobi 11.0` [31]. The values of the parameters used in our case studies are provided in the supplementary material [30].

5.1. Case Study 1: Robust predictive control of building climate

This case study considers robust predictive control of building climate. Building systems suffer from weather uncertainties, such as prediction errors of ambient temperature, solar irradiation, etc. Properly considering these uncertainties can improve indoor climate comfort.

A building climate predictive control problem in its deterministic form can be formulated as [5]

$$
\begin{aligned}
\min_{\mathbf{u}_t} \quad & \sum_{t=1}^{N} l(\mathbf{s}_t, \mathbf{u}_t) \\
\text{s.t.} \quad & \mathbf{s}_{t+1} = \boldsymbol{\Phi}\mathbf{s}_t + \boldsymbol{\Gamma}_u\mathbf{u}_t + \boldsymbol{\Gamma}_w\mathbf{w}_t + \boldsymbol{\Gamma}_v\mathbf{v}_t, \\
& \underline{\mathbf{s}}_t \leq \mathbf{s}_t \leq \bar{\mathbf{s}}_t, \quad \underline{\mathbf{u}}_t \leq \mathbf{u}_t \leq \bar{\mathbf{u}}_t, \\
& \forall \mathbf{v}_t \in \mathcal{V}_t, \ t = 1, \ldots, N
\end{aligned}
$$

where $l(\cdot, \cdot)$ is the stage cost function, $\mathbf{s}_t$ is the system states consisting of indoor temperature, roof temperature, wall temperature and floor temperature; $\mathbf{u}_t$ denotes the heating power, $\mathbf{w}_t$ is the predicted ambient conditions, $\mathbf{v}_t$ is the prediction error of ambient temperature, $\mathcal{V}_t$ denotes the uncertainty set, N is the length of prediction horizon, $\underline{\mathbf{s}}_t/\underline{\mathbf{u}}_t$ and $\bar{\mathbf{s}}_t/\bar{\mathbf{u}}_t$ are lower bound and upper bound of system states/control inputs, respectively. The stage cost function is defined as $l(\mathbf{s}_t, \mathbf{u}_t) := \mathbf{u}_t$ to minimize energy usage, system state constraints are defined to keep the indoor temperature above 21°C during 7:00–18:00 and above 15°C during the remaining hours. Heating power constraints are $0 \leq \mathbf{u}_t \leq 150$. The values of system matrices $(\boldsymbol{\Phi}, \boldsymbol{\Gamma}_u, \boldsymbol{\Gamma}_w, \boldsymbol{\Gamma}_v)$ are adopted from [4].

In the simulation, uncertainties $\mathbf{v}_t$ $(t = 1, \ldots, N)$ are assumed to be I.I.D., and the uncertainty set $\mathcal{V}_t$ for $\mathbf{v}_t$ is a union of 2 subsets. As a result, the total number of subsets for the stacked uncertainty $\mathbf{v} = [\mathbf{v}_1^{\mathrm{T}}, \ldots, \mathbf{v}_N^{\mathrm{T}}]^{\mathrm{T}}$ is 2^N. Two schemes are considered in this case study for solving the building climate control problem. One is the conventional RO formulation with the explicit description of each uncertainty subset (4) and is solved via the CCG-based algorithm in [6]. Another one is our proposed formulation (6) and is solved via Algorithm 1. To fully demonstrate the computational efficiency of the proposed scheme, different values of the prediction horizon N are tested. Simulation results are shown in Fig. 1. It can be seen that, with the increase of the prediction horizon N, the computational time of applying the conventional approach increases exponentially and is much larger than that of our proposed formulation since 2^N numbers of subproblems
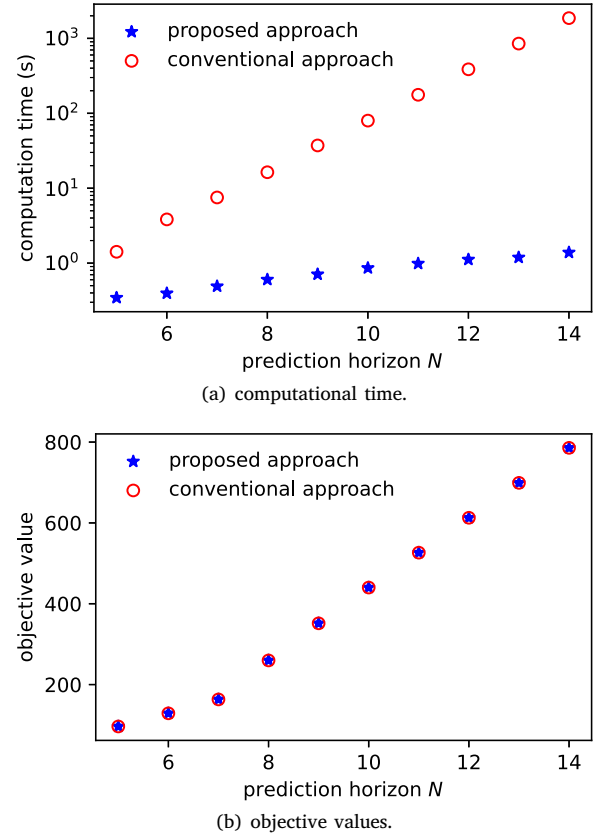

(a) computational time.

(b) objective values.

Fig. 1. Simulation results of *Case Study 1* with the proposed approach and the conventional approach.

have to be solved in each algorithm iteration. In contrast, the computation time with our proposed formulation (6) and Algorithm 1 only increases linearly with the prediction horizon N. Besides, Fig. 1(b) shows that both approaches give the same objective function, which indicates that the proposed formulation does not sacrifice optimality while remarkably improving the computational efficiency.

5.2. Case Study 2: Robust location transportation planning

In this subsection, the effectiveness of the proposed formulation (14) and the corresponding Algorithm 2 is validated via a robust location transportation planning problem, which is also considered as a benchmark problem in [14]. The deterministic formulation of this problem is

$$
\begin{aligned}
\min \ & 400x_1 + 414x_2 + 326x_3 + 18x_4 + 25x_5 + 20x_6 \\
& + 22y_{11} + 33y_{12} + 24y_{13} + 33y_{21} + 23y_{22} + 30y_{23} \\
& + 20y_{31} + 25y_{32} + 27y_{33} \tag{18a}
\end{aligned}
$$

$$
\text{s.t.} \ x_{i+3} \leq 800x_i, \quad i = 1, 2, 3, \tag{18b}
$$

$$
\sum_j y_{ij} \leq x_{i+3}, \quad \forall i = 1, 2, 3, \tag{18c}
$$

$$
\sum_i y_{ij} \geq d_j + 40 * v_j \tag{18d}
$$

$$
x_i \in \mathbb{B}, \quad x_{i+3} \leq 0, \quad i = 1, 2, 3, \tag{18e}
$$

$$
y_{ij} \geq 0, \quad \forall i = 1, 2, 3, \quad j = 1, 2, 3 \tag{18f}
$$

where binary first-stage decision variables (x_1, x_2, x_3) determine the location of the facilities; continuous first-stage decision variables $(x_4, x_5,$

Table 1
Computational results of different RO formulations for *Case Study 2*.

	Conventional RO formulation (Scheme 1)	Existing DRO formulation with Wasserstein metric (Scheme 2)	The proposed RO formulation (14)	
			With (15b) (Scheme 3)	With (17) (Scheme 4)
Optimal objective	36 632	35 238	35 482	35 482
Decision variable **x**	[1,0,1,274,0,570]	[1,0,1,324,0,520]	[1,0,1,364,0,480]	[1,0,1,364,0,480]
CPU time (s)	0.40	425.80	17.52	7.06
Iterations	2	3	2	2

x_6) denote the facility capacities; recourse decision variables y_{ij} ($i = 1, 2, 3$, $j = 1, 2, 3$) are transportations, $[d_1, d_2, d_3] = [206, 274, 220]$ are basic demands, and v_j ($j = 1, 2, 3$) are scaled demand uncertainties.

In our simulation, four schemes are considered:

- Scheme 1: the conventional RO formulation (1).
- Scheme 2: the Wasserstein-based DRO formulation [20].
- Scheme 3: the proposed formulation (14) and Algorithm 2 with the constraints (15b).
- Scheme 4: the proposed formulation (14) and Algorithm 2 with the reformulated constraints (17).

The uncertainty set $\mathcal{V}$ for $[v_1, v_2, v_3]^T$ is supposed to have 4 polyhedral subsets. For our proposed schemes (Scheme 3 & 4). The nominal probability distribution for defining the ambiguity set $\mathcal{P}$ is selected as $\bar{\mathbf{p}} = [0.5, 0.1, 0.2, 0.2]$ to represent a practical situation where uncertainty samples are unevenly distributed among different uncertainty subsets. The upper bound of KL divergence for defining the ambiguity set $\mathcal{P}$ in (13) is set as $\rho = 0.5$. For Scheme 2, since its implementation entails considering specific uncertainty scenarios, 1000 uncertainty samples in $\mathcal{V}$ are randomly generated, which is a reasonable choice for the uncertainty with dimension 3 [32,33]. Since we mainly focus on demonstrating the computational efficiency of the DRO formulation, the Wasserstein distance ϵ, which does not affect its computational efficiency, is set as $\epsilon = 1$ to give a comparable performance as with the other schemes. It should be pointed out that the Wasserstein-based DRO formulation (Scheme 2) can be expected to outperform our proposed methods with a well-calibrated radius. Nevertheless, within the Wasserstein-based DRO framework, achieving both low conservatism and strong out-of-sample performance hinges on carefully tuning the Wasserstein radius and considering a large number of representative uncertainty samples to avoid overfitting to the empirical distribution, which will lead to increased computational cost for solving the corresponding optimization problem. Numerical results for the DRO formulation and the proposed formulation with an extreme size of ambiguity set (the size of the ambiguity set is set as zero) are provided in the supplementary materials, showing that Scheme 2 would outperform our proposed scheme at the cost of increased computational burden.

Simulation results are summarized in Table 1. It can be seen that these different RO formulations derive distinct first-stage decision variables with non-trivial differences. Compared with the conventional RO formulation (Scheme 1) for optimizing the worst-case performance, the remaining formulations give less conservative solutions, i.e., smaller optimal objective values, at the price of increased computational burden. As discussed in Remark 8, the increased computational burden for the DRO formulation (Scheme 2) is due to its curse of dimensionality when considering the fully adaptive decision rule and a large number of uncertainty samples [20,29], while for our proposed schemes (Scheme 3 & 4), the increased computational burden is caused by the nonlinear constraint (15b), especially when the first-stage decision variable **x** contains integer ingredients. It can be seen that, in comparison with the DRO formulation (Scheme 2), our proposed formulations (Scheme 3 & 4) are much more computationally efficient. Furthermore, the computational time of Scheme 4 is less than that of Scheme 3, which indicates the efficacy of the proposed reformulation (17) for improving

the computational efficiency of Algorithm 2. It is worth noting that, in comparison with Scheme 2, the reduced computation time of our proposed scheme does not simply result from switching the divergence measure. Without altering the problem structure, directly replacing the Wasserstein metric with the KL divergence in Scheme 2 leads to the same issue observed in the Wasserstein-based design. Namely, as shown in Algorithm 1 in the supplementary material, both the number of subproblems and the number of constraints in the master problem are proportional to the number of uncertainty samples.

5.3. Case Study 3: Chemical process network planning

To further illustrate the viability and effectiveness of the proposed formulation (14) and Algorithm 2, a chemical process network planning (CPNP) problem is investigated. The CPNP problem is a typical engineering problem that fits into RO settings and has been considered as a benchmark problem in several existing literature, see [4,7,34].

A chemical process consists of raw materials, intermediate chemicals, final products, and multiple interconnected processes. The design objective of CPNP is to maximize the net present value (NPV) of the entire network while respecting system constraints for all possible uncertainties. Our case study considers a chemical process network consisting of 8 processes and 7 chemicals. The network sketch is shown in Fig. 2, where chemicals are denoted as red circles ($A, B, \ldots, G$), processes are blue rectangles ($1, 2, \ldots, 8$), and process flows are indicated as arrows. Among all chemicals, (A, E) are raw materials, and (D, G) are products.

The RO problem for CPNP within our proposed design framework (14) can be formulated as

$$\max_{QE_{i,t}, Y_{i,t}} \sum_i \sum_t (-\alpha_{i,t} QE_{i,t} - \beta_{i,t} Y_{i,t}) + $$

$$\min_{\mathbf{p} \in \mathcal{P}} \mathbb{E}_{\mathbf{p}} \left[\min_{\mathbf{v} \in \mathcal{V}_k} \max_{\substack{P_{j,t}, QE_{i,t} \\ S_{j,t}, W_{i,t}}} \left(-\sum_i \sum_t \gamma_{i,t} W_{i,t} \right. \right.$$

$$\left. \left. - \sum_j \sum_t \varphi_{j,t} P_{j,t} + \sum_j \sum_t \tau_{j,t} S_{j,t} \right) \right] \tag{19a}$$

$$\text{s.t. } qe_{i,t}^L \cdot Y_{i,t} \leq QE_{i,t} \leq qe_{i,t}^U \cdot Y_{i,t}, \ \forall i \in \mathcal{I}, \ \forall t \in \mathcal{T} \tag{19b}$$

$$Q_{i,t} = Q_{i,t-1} + QE_{i,t}, \ \forall i \in \mathcal{I}, \ \forall t \in \mathcal{T} \tag{19c}$$

$$\sum_t Y_{i,t} \leq ce_i, \ \forall i \in \mathcal{I} \tag{19d}$$

$$\sum_i \alpha_{i,t} \cdot QE_{i,t} + \beta_{i,t} Y_{i,t} \leq cb_t, \ \forall t \in \mathcal{T} \tag{19e}$$

$$W_{i,t} \leq Q_{i,t}, \ \forall i \in \mathcal{I}, \ \forall t \in \mathcal{T} \tag{19f}$$

$$P_{j,t} - \sum_i \kappa_{i,j} W_{i,t} - S_{j,t} = 0, \ \forall j \in \mathcal{J}, \ \forall t \in \mathcal{T} \tag{19g}$$

$$P_{j,t} \leq su_{j,t}, \ S_{j,t} \leq du_{j,t}, \forall j \in \mathcal{J}, \ \forall t \in \mathcal{T} \tag{19h}$$

$$QE_{i,t}, Q_{i,t}, P_{j,t}, W_{i,t}, S_{j,t} \geq 0, Y_{i,t} \in \{0, 1\},$$

$$\forall i \in \mathcal{I}, \ \forall j \in \mathcal{J}, \ \forall t \in \mathcal{T}, \tag{19i}$$

$$\mathbf{v} = \{du_{i,t}, su_{i,t}\} \in \mathcal{V}, \ \forall i \in \mathcal{I}, \ \forall t \in \mathcal{T} \tag{19j}$$

Table 2
Notations in (19).

Decision variables			
$QE_{i,t}$	Amount of capacity expansion	$Y_{i,t}$	Binary decision of capacity expansion
$P_{j,t}$	Chemical purchase amount	$S_{j,t}$	Chemical sale amount
$W_{i,t}$	Operating level	$Q_{i,t}$	Total process capacity
$su_{i,t}$	Market supply limit	$du_{i,t}$	Market demand limit
Parameters			
$\alpha_{i,t}$	Variable investment cost	$\beta_{i,t}$	Fixed investment cost
$\gamma_{j,t}$	Operating cost	$\varphi_{j,t}$	Purchase cost
$\tau_{j,t}$	Sale price	$qe_{i,t}^L$	Capacity expansion lower bound
$qe_{i,t}^U$	Capacity expansion upper bound	ce_i	Expansion number limit
cb_t	Expansion cost budget	$\kappa_{i,j}$	Mass balance coefficient
i	Index of the ith process	j	Index of the jth chemical
t	Index of the tth time period		

Table 3
Computational results of different RO formulations for *Case Study 3*.

	Conventional RO formulation (Scheme 1)	Existing DRO formulation with Wasserstein metric (Scheme 2)	The proposed RO formulation (14)	
			With (15b) (Scheme 3)	With (17) (Scheme 4)
Max. NPV ($MM)	221	284	288	288
CPU time (s)	17.63	8758.07	195.86	183.17
Iterations	2	5	2	2
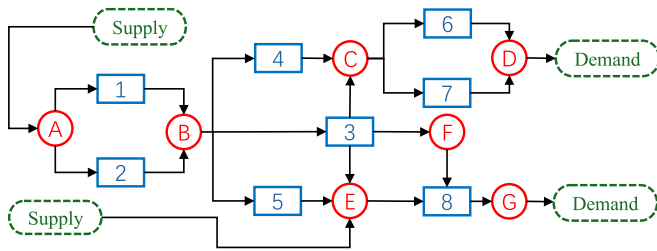
Fig. 2. The chemical process network for *Case Study 3*.

Fig. 3. Error bar of the expected NPV with different ambiguity set size ρ.

where $\mathcal{I} = \{1, \ldots, I\}$, $\mathcal{J} = \{1, \ldots, J\}$, and $\mathcal{T} = \{1, \ldots, T\}$ with I, J and T as the total numbers of processes, chemicals and planning periods, respectively. All notations in the above equations are explained in Table 2. The objective function (19a) maximizes the NPV consisting of investment cost, operation cost, purchase cost of raw chemicals, and sale profit; constraint (19b) specifies the upper and lower bounds of capacity expansion for all processes and time periods; constraint (19c) updates the total available capacity of each process; constraint (19d) limits the largest process expansion times; constraint (19e) ensures the process expansion costs are within available budgets; inequality (19f) limits the production level of each process within its total capacity; equality (19g) models the mass balance of all chemicals; constraints (19h) ensures that the amounts of purchased and sold chemicals are limited by the available market supply and demand, respectively; constraints (19i) indicates all non-negative continuous decision variables and binary decision variables. For more detailed explanations of chemical process networks, please refer to [7,34].

The uncertainty variables considered in our design are the market supply limit and demand limit ($su_{i,t}, du_{i,t}$) for all raw materials and products over the planning period. Expansion decisions ($QE_{i,t}, Y_{i,t}$) are first-stage decision variables, and other remaining variables ($P_{j,t}, S_{j,t}, Q_{i,t}, W_{i,t}$) are recourse decision variables. Our case study considers a 5−year planning period with each planning period as 1 year. The uncertainties ($su_{j,t}, du_{j,t}$) are assumed to be independent and reside in 4 uncertainty sets.

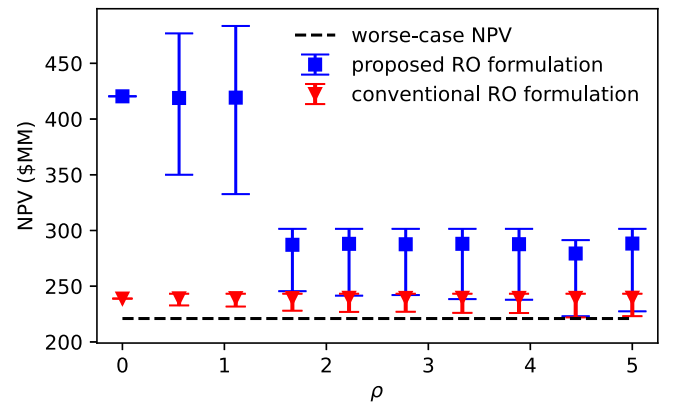As in *Case Study 2*, we consider the same four schemes. For the DRO formulation in Scheme 2, 1000 uncertainty scenarios are randomly generated, and the Wasserstein distance ε defining the size of the distribution ambiguity set is $\varepsilon = 300$ to give a comparable objective value with the other schemes. It should be noted that the number of uncertainty samples considered is far smaller than needed for a proper approximation of the uncertainty probability distribution, given the 20-dimensional uncertainty space [32,33]. However, as shown in our simulation results, even this limited number of samples can be computationally demanding for solving the corresponding DRO problem. For our proposed schemes (Scheme 3 & 4), the parameters defining the ambiguity set $\mathcal{P}$ are $\bar{\mathbf{p}} = [0.5, 0.1, 0.2, 0.2]$ and $\rho = 0.5$.

Simulation results are shown in Table 3, from which it can be concluded that consistent results are obtained as in *Case Study 2*. Compared with the conventional RO formulation (Scheme 1), the remaining formulations (Scheme 2, 3 & 4) give less conservative solutions, i.e., a larger value of Max. NPV at the price of increased computational burden (longer CPU time). Compared with the Wasserstein-based DRO formulation (Scheme 2), our proposed formulations (Scheme 3 & 4) are less computationally demanding. In addition, compared with the original constraints (15b), the reformulation in (17) is effective in improving the computational efficiency of Algorithm 2.

Furthermore, different values of the ambiguity set size ρ are tested to analyze its influence on the conservatism of the optimal solution for

our proposed formulation. In the simulation, for each ρ, 5000 random trials are implemented to sample the probability distribution $\mathbf{p}$ that resides in the corresponding ambiguity set. For all admissible samples of $\mathbf{p}$, the corresponding performance metric, which is defined as the expected NPV: $\mathbf{c}^{\mathsf{T}}\mathbf{x}^* + \sum_k \mathbf{p}_k \cdot \mathbf{b}^{\mathsf{T}}\mathbf{y}_k^*$ with $\mathbf{y}_k^*$ as the optimal recourse variable w.r.t. the worst-case uncertainty in the kth uncertainty subset, is computed. Fig. 3 depicts the error bar (mean value and envelope) of the expected NPV among all feasible samples of $\mathbf{p}$, and the worst-case NPV for all possible uncertainties. It can be seen that, compared with the conventional RO formulation, our proposed approach consistently leads to less conservative solutions for all ρ. Namely, the proposed approach gives a higher mean, minimal, and maximal NPV than the conventional RO formulation. Further, it can be seen from Fig. 3 that, with the increase of ρ (the size of the ambiguity set), the minimal NPV of the proposed formulation (14) converges to the worse NPV computed via the conventional RO formulation, which is consistent with our statement in Remark 6.

6. Conclusions

This paper investigates the RO problems where uncertainty sets comprise multiple subsets, focusing on two separate questions: (1) *how to address the computational challenge posed by exponentially increasing uncertainty subsets in RO-based predictive control?* and (2) *how to mitigate the conservatism of the robust optimal solutions by leveraging the structure of the uncertainty set?*

To address the first question, we propose a monolithic mixed-integer representation of the uncertainty set. Unlike conventional formulations requiring a separate optimization problem for each subset, our method only solves a single mixed-integer optimization problem to compute the worst-case uncertainty scenario over all subsets. This method is particularly advantageous for RO-based predictive control, where the number of uncertainty subsets could increase exponentially with the prediction horizon.

For the second question, we formulate an innovative objective function exploiting the structure of the multi-subset uncertainty set by combining the existing RO and DRO formulations. The proposed formulation achieves less conservative solutions than conventional RO formulations while showing more computational efficiency than conventional DRO formulations. Besides, a CCG-based algorithm is developed to solve the resulting optimization problem efficiently.

Numerical experiments related to the above research questions are conducted to extensively demonstrate the effectiveness of the proposed schemes.

CRediT authorship contribution statement

Yun Li: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Neil Yorke-Smith:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Tamas Keviczky:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix

Proof of Theorem 1

Proof. The main difference between the proposed Algorithm 1 with the conventional CCG algorithm in [6,14] is the definition of the subproblem (8). With the mixed-integer representation of the uncertainty set $\mathcal{V}$, each set of feasible binary variables $\{\delta_{t,k}^* \mid t = 1, \ldots, N, k = 1, \ldots, K\}$ uniquely defines a subset of $\mathcal{V}$, and vice versa. As a result, by solving the subproblem **SP**1 (8) or **SP**2 (9), the worst-case uncertainty scenario $\mathbf{v}^* = [\mathbf{v}_1^{*\mathsf{T}}, \ldots, \mathbf{v}_N^{*\mathsf{T}}]^{\mathsf{T}}$ among all subsets together with the corresponding uncertainty subset, defined by $\{\delta_{t,k}^* \mid t = 1, \ldots, N, k = 1, \ldots, K\}$, that contains the uncertainty scenario $\mathbf{v}^*$ will be computed. Based on Assumption 1, it can be readily concluded that there are finite vertices of the uncertainty set $\mathcal{V}$. Then, it follows a similar proof as shown in [14] that Algorithm 1 will solve the RO problem (6) within finite iterations. This completes the proof. □

Proof of Theorem 2

Proof. For the constraint (14b), it can be rewritten as

$$\max_{\mathbf{p} \in \mathcal{P}} \sum_{k=1}^{K} \mathbf{p}_k \cdot \left(\max_{\mathbf{v} \in \mathcal{V}_k} \mathbf{b}^{\mathsf{T}}\mathbf{y} \right) \tag{20}$$

Based on *Lemma 19.1* in [15], the inequality (20) is equivalent to finding μ and $\nu \geq 0$ such that

$$\mu + \rho\nu + \nu \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{C(\mathbf{x}, \mathcal{V}_k) - \mu}{\nu} - 1 \right) \leq \eta \tag{21}$$

where

$$C(\mathbf{x}, \mathcal{V}_k) := \max_{\mathbf{v} \in \mathcal{V}_k} \min_{\mathbf{y}} \mathbf{b}^{\mathsf{T}}\mathbf{y} \tag{22a}$$

$$\text{s.t. } \mathbf{Tx} + \mathbf{Wy} + \mathbf{Mv} \leq \mathbf{h} \tag{22b}$$

Assumption 2 implies that there are finite vertices of the uncertainty set $\mathcal{V}$. Besides, it can be seen that the left-hand side of (21) is a monotonously increasing function w.r.t. $C(\mathbf{x}, \mathcal{V}_k)$, which is LP w.r.t. $\mathbf{v}$. Consequently, it can be concluded that the optimal $\mathbf{v}$ are taken from the vertices of $\mathcal{V}$. By listing all finite vertices of each uncertainty subset $\mathcal{V}_k$, denoted as $S_k := \{\mathbf{v}_{k,1}, \ldots, \mathbf{v}_{k,H_k}\}$ with H_k as the number of all vertices, the RO problem (14) can be rewritten as

$$\min_{\substack{\mathbf{x}, \mathbf{y}_{k,i_k} \\ \eta, \mu, \nu}} \mathbf{c}^{\mathsf{T}}\mathbf{x} + \eta \tag{23a}$$

$$\text{s.t. } \mu + \rho\nu + \nu \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathsf{T}}\mathbf{y}_{k,i_k} - \mu}{\nu} - 1 \right) \leq \eta, \tag{23b}$$

$$\mathbf{Ax} \leq \mathbf{q}, \tag{23c}$$

$$\mathbf{Tx} + \mathbf{Wy}_{k,i_k} + \mathbf{Mv}_{k,i_k} \leq \mathbf{h}, \tag{23d}$$

$$\forall i_k \in \{1, \ldots, H_k\}, \ \forall k \in \{1, \ldots, K\} \tag{23e}$$

where $\mathbf{y}_{k,i_k}$ is the optimal recourse decision variable w.r.t. the uncertainty scenario $\mathbf{v}_{k,i_k}$ $(i = 1, \ldots, H_k)$. Namely,

$$\mathbf{y}_{k,i_k} = \arg\min_{\mathbf{y}} \mathbf{b}^{\mathsf{T}}\mathbf{y} \text{ s.t. } \mathbf{Tx} + \mathbf{Wy} + \mathbf{Mv}_{k,i_k} \leq \mathbf{h}.$$

Since there are finite extreme uncertainty scenarios, the optimization problem (23) can be computed by iteratively listing all possible extreme uncertainty scenarios as in the conventional CCG algorithm [13, 14], which leads to Algorithm 2.

In the following, we will prove that Algorithm 2 can terminate within finite iterations. Namely, Algorithm 2 will either find out all extreme uncertainty scenarios or terminate when a repeated uncertainty scenario is observed. Assuming at the rth iteration of the algorithm, the extreme uncertainty scenarios $\{\mathbf{v}_1^*, \ldots, \mathbf{v}_K^*\}$ computed by solving K

subproblems (16) are observed in a previous iteration t ($t \leq r - 1$), then $UB = LB$ and the algorithm terminates.

Suppose that the optimal decision variables by solving $\mathbf{MP}_{\mathrm{DRO}}$ (15) at the rth iteration are $(\mathbf{x}^*, \mu^*, \eta^*, \nu^*)$, which further lead to the optimal decision variables of K subproblems $\mathbf{SP}^k_{\mathrm{DRO}}$ $\{\mathbf{v}^*_1, \ldots, \mathbf{v}^*_K, \mathbf{y}^*_1, \ldots, \mathbf{y}^*_K\}$. It readily gives that

$$LB \leq UB \leq \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* +$$
$$\nu^* \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}^*_k - \mu^*}{\nu^*} - 1\right)$$

Since the worst-case uncertainty scenarios $\{\mathbf{v}^*_1, \ldots, \mathbf{v}^*_K\}$ was observed at a previous iteration, the optimal solution of the master problem $\mathbf{MP}_{\mathrm{DRO}}$ at $(r+1)$th iteration will be the same as with the rth iteration and satisfy

$$LB \geq \mathbf{c}^{\mathrm{T}}\mathbf{x}^* + \mu^* + \rho\nu^* + \nu^* \sum_k \bar{\mathbf{p}}_k \exp\left(\frac{\mathbf{b}^{\mathrm{T}}\mathbf{y}^*_k - \mu^*}{\nu^*} - 1\right)$$

As a result, it can be concluded that $LB = UB$ at the $(r+1)$th iteration, and the algorithm terminates. This completes the proof. $\square$

Proof of the convexity of constraint (15b)

Proof. For brevity of analysis, it is sufficient to prove that the nonlinear function $f(x, y) := x \exp(\frac{y}{x} - 1)$ with $x \geq 0$ is convex since the nonlinear constraint (15b) can be reformulated as

$$\mu + \rho\nu + \sum_k \bar{\mathbf{p}}_k f(\nu, g_{k,i}) \leq \eta, \tag{24a}$$

$$g_{k,i} = \mathbf{b}^{\mathrm{T}}\mathbf{y}_{k,i} - \mu. \tag{24b}$$

For the nonlinear function $f(x, y) := x \exp\left(\frac{y}{x} - 1\right)$, its Hessian matrix is

$$\nabla^2 f = \begin{bmatrix} \frac{y^2}{x^3} \exp(\frac{y}{x} - 1) & -\frac{y}{x^2} \exp(\frac{y}{x} - 1) \\ -\frac{y}{x^2} \exp(\frac{y}{x} - 1) & \frac{1}{x} \exp(\frac{y}{x} - 1) \end{bmatrix}. \tag{25}$$

It can be easily verified that $\nabla^2 f$ is positive semidefinite when $x \geq 0$. As a result, it follows from the *second-order condition* of convexity [35] that the nonlinear function $f(x, y)$ is convex, and hence constraint (15b) is convex. This completes the proof. $\square$

Data availability

Data will be made available on request.

References

[1] A. Ben-Tal, A. Nemirovski, Robust convex optimization, Math. Oper. Res. 23 (4) (1998) 769–805.

[2] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Vol. 28, Princeton University Press, 2009.

[3] L. El Ghaoui, F. Oustry, H. Lebret, Robust solutions to uncertain semidefinite programs, SIAM J. Optim. 9 (1) (1998) 33–52.

[4] C. Shang, X. Huang, F. You, Data-driven robust optimization based on kernel learning, Comput. Chem. Eng. 106 (2017) 464–479.

[5] C. Shang, F. You, A data-driven robust optimization approach to scenario-based stochastic model predictive control, J. Process Control 75 (2019) 24–39.

[6] C. Ning, F. You, Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty, AIChE J. 63 (9) (2017) 3790–3817.

[7] C. Ning, F. You, Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era, Comput. Chem. Eng. 111 (2018) 115–133.

[8] D. Bertsimas, V. Gupta, N. Kallus, Data-driven robust optimization, Math. Program. 167 (2018) 235–292.

[9] M. Cheramin, R.L.-Y. Chen, J. Cheng, A. Pinar, Data-driven robust optimization using scenario-induced uncertainty sets, 2021, arXiv preprint arXiv:2107.04977 URL: https://arxiv.org/pdf/2107.04977.

[10] Y. Li, N. Yorke-Smith, T. Keviczky, Machine learning enabled uncertainty set for data-driven robust optimization, J. Process Control 144 (2024) 103339.

[11] N. Zhao, F. You, Sustainable power systems operations under renewable energy induced disjunctive uncertainties via machine learning-based robust optimization, Renew. Sustain. Energy Rev. 161 (2022) 112428.

[12] G. Hu, F. You, Multi-zone building control with thermal comfort constraints under disjunctive uncertainty using data-driven robust model predictive control, Adv. Appl. Energy 9 (2023) 100124.

[13] L. Zhao, B. Zeng, An exact algorithm for two-stage robust optimization with mixed integer recourse problems, 2012, URL: https://optimization-online.org/wp-content/uploads/2012/01/3310.pdf.

[14] B. Zeng, L. Zhao, Solving two-stage robust optimization problems using a column-and-constraint generation method, Oper. Res. Lett. 41 (5) (2013) 457–461.

[15] D. Bertsimas, D. den Hertog, Robust and Adaptive Optimization, Dynamic Ideas LLC, Belmont, MA, 2022.

[16] J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer Science & Business Media, 2011.

[17] G.C. Calafiore, M.C. Campi, The scenario approach to robust control design, IEEE Trans. Autom. Control 51 (5) (2006) 742–753.

[18] M.C. Campi, S. Garatti, The exact feasibility of randomized solutions of uncertain convex programs, SIAM J. Optim. 19 (3) (2008) 1211–1230.

[19] G.A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann, A distributionally robust perspective on uncertainty quantification and chance constrained programming, Math. Program. 151 (2015) 35–62.

[20] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations, Math. Program. 171 (1) (2018) 115–166.

[21] D. Bertsimas, X.V. Doan, K. Natarajan, C.-P. Teo, Models for minimax stochastic linear optimization problems with risk aversion, Math. Oper. Res. 35 (3) (2010) 580–602.

[22] W. Wei, Tutorials on advanced optimization methods, 2020, arXiv preprint arXiv:2007.13545 URL: https://arxiv.org/abs/2007.13545.

[23] D. Kuhn, Shafiee, W.W. Soroosh, Distributionally robust optimization, 2024, arXiv preprint arXiv:2411.02549 URL: https://arxiv.org/pdf/2411.02549.

[24] L. Li, C. Ning, H. Qiu, W. Du, Z. Dong, Online data-stream-driven distributionally robust optimal energy management for hydrogen-based multimicrogrids, IEEE Trans. Ind. Inform. 20 (3) (2023) 4370–4384.

[25] X. Ma, C. Ning, L. Li, H. Qiu, W. Gu, Z. Dong, Bayesian nonparametric two-stage distributionally robust unit commitment optimization: From global multimodality to local trimming-wasserstein ambiguity, IEEE Trans. Power Syst. 39 (5) (2024) 6702–6715.

[26] G.C. Calafiore, L. Fagiano, Robust model predictive control via scenario optimization, IEEE Trans. Autom. Control 58 (1) (2012) 219–224.

[27] A. Mesbah, Stochastic model predictive control: An overview and perspectives for future research, IEEE Control Syst. Mag. 36 (6) (2016) 30–44.

[28] D. Bertsimas, J. Dunn, Machine Learning Under a Modern Optimization Lens, Dynamic Ideas LLC Waltham, 2019.

[29] Y. Wang, Y. Yang, L. Tang, W. Sun, B. Li, A wasserstein based two-stage distributionally robust optimization model for optimal operation of CCHP micro-grid under uncertainties, Int. J. Electr. Power Energy Syst. 119 (2020) 105941.

[30] Y. Li, N. Yorke-Smith, T. Keviczky, Supplementary material, 2024, URL: https://github.com/li-yun/data_driven_ro_supplementary.

[31] Gurobi Optimization, Gurobi Optimizer Reference Manual, 2024, URL: https://www.gurobi.com.

[32] H.J. van Heerde, Models for sales promotion effects based on store-level scanner data, 1999.

[33] D.W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, John Wiley & Sons, 2015.

[34] F. You, I.E. Grossmann, Stochastic inventory management for tactical process planning under uncertainties: MINLP models and algorithms, AIChE J. 57 (5) (2011) 1250–1277.

[35] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.