On Tractability, Complexity, and Mixed-Integer Convex Programming Representability of Distributionally Favorable Optimization

Nan Jiang · **Weijun Xie**

Abstract Distributionally Favorable Optimization (DFO) is an important framework for decision-making under uncertainty, with applications across fields such as reinforcement learning, online learning, robust statistics, chance-constrained programming, and two-stage stochastic optimization without relatively complete recourse. In contrast to the traditional Distributionally Robust Optimization (DRO) paradigm, DFO presents a unique challenge– the application of the inner infimum operator often fails to retain the convexity. In light of this challenge, we study the tractability and complexity of DFO. We establish sufficient and necessary conditions for determining when DFO problems are tractable or intractable. Despite the typical nonconvex nature of DFO problems, our findings show that they are mixed-integer convex programming representable (MICP-R), thereby enabling solutions via standard optimization solvers. Finally, we numerically validate the efficacy of our MICP-R formulations.

Keywords. Distributionally Favorable Optimization; Tractability; Complexity; Mixed-Integer Convex Programming Representability

1 Introduction

We study Distributionally Favorable Optimization (DFO) that admits the following form:

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right], \tag{1}$$

where set $\mathcal{X} \subseteq \mathbb{R}^n$ is deterministic, set $\mathcal{P}$ denotes an ambiguity set with $\mathcal{P} \subseteq \{\mathbb{P} \colon \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{U}\} = 1\}$ and support $\mathcal{U} \subseteq \mathbb{R}^m$ (also called "uncertainty set" throughout this paper), and function $Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \colon \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}$. Note that if only support information $\mathcal{U}$ is available (i.e., $\mathcal{P} = \{\mathbb{P} \colon \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{U}\} = 1\}$), then DFO (1) simplifies to a singular DFO (sDFO), i.e.,

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \inf_{\boldsymbol{\xi} \in \mathcal{U}} Q(\boldsymbol{x}, \boldsymbol{\xi}). \tag{2}$$

The special cases of DFO (1) and their variants have been applied to various fields, including reinforcement learning [3, 5, 38], image reconstruction [19], Bayesian optimization [28–30], among others. Recent

First Author: Nan Jiang
Affiliation: Georgia Institute of Technology, Atlanta, GA
E-mail: nanjiang@gatech.edu

Corresponding Author: Weijun Xie
Affiliation: Georgia Institute of Technology, Atlanta, GA
E-mail: wxie@gatech.edu

advancements in [22] demonstrate that DFO (1) can recover many robust statistics and machine learning problems.

It is worth noting that as an opposite counterpart of DFO (1), the Distributionally Robust Optimization (DRO) of the form

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right], \tag{3}$$

has recently successfully addressed many decision-making problems under uncertainty to achieve better out-of-sample performance guarantees (see more discussions in [14, 27, 35, 39]). A natural way to bridge DFO (1) and DRO (3) is through the so-called Hurwicz criterion, proposed in [4, 20], which can be used to effectively balance the tradeoff between DFO (1) and DRO (3). The Hurwicz model has been studied in many decision-making problems (see, e.g., [10, 11, 23, 33]). For example, [33] presented empirical evidence supporting the use of the Hurwicz criterion and showcased its potential predictive capability in path selection and traffic equilibrium. [11] provided the analysis on using the Hurwicz model to reduce over-conservatism and achieve better out-of-sample performance in the newsvendor model. Specifically, for a given level of optimism $\bar{\lambda} \in [0, 1]$, the Hurwicz model considers

$$\min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \bar{\lambda} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] + (1 - \bar{\lambda}) \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] \right\}. \tag{Hurwicz Model}$$

Moreover, the discrepancy between DRO (3) and DFO (1) can be used to upper bound the regret. That is, in the face of the ambiguity set $\mathcal{P}$, the decision-maker chooses a solution $\boldsymbol{x}$ to minimize the distributionally robust regret, which is

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \left\{ E_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] - \min_{\boldsymbol{y} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{y}, \tilde{\boldsymbol{\xi}}) \right] \right\}. \tag{Regret}$$

The regret is upper bounded by the following value

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] - \min_{\boldsymbol{x} \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right],$$

which represents the possible range of objective function values. The difference between DRO (3) and DFO (1) can also be used to quantify the level of uncertainty and to bound the value of the stochastic program (see, e.g., [11, 12]).

While DFO (1) can be applied to many domains, a thorough complexity analysis of DFO (1) remains absent from existing works. This paper aims to bridge this gap by exploring the computational complexities of DFO (1) in depth. Specifically, when comparing DFO (1) with DRO (3), we see that the inner supremum operator in DRO (3) maintains the convexity, but the inner infimum in DFO (1) typically undermines this convexity (see, e.g., [7]). To address this challenge, we revisit and expand upon the mixed-integer convex programming representability concept, adapting it to DFO (1). Our results allow standard optimization solvers to solve DFO (1) efficiently. We expect that the mixed-integer convex programming representability results presented in this paper can be applied to decision-dependent uncertainty (see the discussions in [16, 31]). Throughout this paper, we make the following assumptions of set $\mathcal{X}$:

Assumption 1 *Set $\mathcal{X}$ is compact and has a non-empty relative interior.*

Assumption 1 is standard in the stochastic optimization literature (see, e.g., [36]). It is worth mentioning that Assumption 1 is useful to prove that DFO may not be mixed-integer convex programming representable. We also note that the result in this paper can be directly extended to mixed-integer compact convex set $\mathcal{X}$, which can be written as a finite union of compact convex sets with a non-empty relative interior.

1.1 Relevant Literature

Recent studies on DFO address outliers and uncertainties of decision-making problems (see, e.g., [17, 22, 32, 34, 38]). Notably, [34] studied two-stage stochastic optimization without relatively complete recourse from the DFO perspective, using the Rockafellian relaxation for perturbation analysis to tackle model uncertainty. [32] applied DFO to manage noises and outliers in machine learning, while [17] investigated the out-of-sample performance and sensitivity analysis of DFO solutions, especially those involving regularization. For a comprehensive understanding of DFO, readers are directed to the discussions in [22, 34] and the references cited therein.

Despite its vital importance and wide applications, DFO (1) often faces a significant challenge: the absence of convexity, which renders it more complex to solve compared to its DRO counterpart. For example, [32] investigated intractable machine learning problems by employing a nonconvex regularizer based on DFO. This paper is quite different from [22]. The latter work provided a comprehensive introduction to the DFO framework to illustrate its suitability for decision-making, such as showing how DFO (1) can recover many robust statistics. However, [22] did not study the computational tractability of DFO (1). In fact, to date, the computational tractability of DFO has not been extensively explored in the literature. To fill this gap, our paper studies the tractability and complexity of DFO (1). Recognizing that strong conditions are required for the tractability of DFO (1), we study the conditions that allow for its representation as a mixed-integer convex program. Our results build upon recent findings by [24] in mixed-integer convex programming representability, focusing on special cases of DFO that can be precisely formulated as mixed-integer convex programs. Leveraging this notion of mixed-integer convex programming representability advances our understanding of DFO and opens avenues for more efficient computational approaches.

1.2 Contributions

This paper complements the literature by providing the tractability and complexity analysis for DFO (1). We prove that solving DFO (1) is NP-hard in general, and we provide conditions under which DFO (1) can be tractable, i.e., when the function $Q(x, \xi)$ is convex or concave piecewise affine in x in sDFO (2) under mild conditions. We generalize the mixed-integer convex programming (MICP) representable (MICP-R) notion, introduced in [24]. Particularly, we show the sufficient conditions under which sDFO (2) and DFO (1) can be MICP-R. Table 1 summarizes our main MICP-R results. We numerically demonstrate the value of MICP-R for DFO (1) and find that the MICP-R formulation can dramatically improve the running time, even for small-scale instances. The MICP-R results can be directly applied to the Hurwicz model, which can effectively reduce over-conservatism and achieve better out-of-sample performance. This is also the first-known complexity result of the Hurwicz model.

Table 1 Summary of MICP-R and Not MICP-R Results

	Function $Q(x, \xi)$	Uncertainty Set	Results
sDFO (2)	Concave Piecewise Affine Section 3.1	$\mathcal{U} = \{\xi : \|\xi - \xi^0\|_p \leq \theta\}$	$p \in \{1, \infty\}$: MICP-R
			$p \in (1, \infty)$: Not MICP-R
	Convex Piecewise Affine Section 3.2		$p \in \{1, \infty\}$ & Mild Conditions: MICP-R
			$p \in (1, \infty)$: Not MICP-R
		Ambiguity Set	Results
DFO (1)	Function $Q(x, \xi)$ Can be MICP-R	Type-∞ Wasserstein Section 4.1	$p \in \{1, \infty\}$ & Mild Conditions: MICP-R
			$p \in (1, \infty)$: Not MICP-R
		Finite Support Section 4.2	MICP-R

Organization. The remainder of the paper is organized as follows. Section 2 reviews and extends the MICP-R notion. Section 3 discusses the tractability analysis, complexity analysis, and MICP-R formulations for sDFO (2). Section 4 discusses the tractability analysis, complexity analysis, and MICP-R formulations for

DFO (1). Section 5 numerically demonstrates the value of MICP-R formulations for DFO (1). Section 6 concludes the paper.

Notation. The following notation is used throughout the paper. We use bold letters (e.g., $\boldsymbol{x}$, $\boldsymbol{A}$) to denote vectors and matrices and use corresponding non-bold letters to denote their components. We let $\|\cdot\|_*$ denote the dual norm of a general norm $\|\cdot\|$. We let $\boldsymbol{e}$ be the vector or matrix of all ones, and let $\boldsymbol{e}_i$ be the ith standard basis vector. Given an integer n, we let $[n] := \{1, 2, \ldots, n\}$, and use $\mathbb{R}_+^n := \{\boldsymbol{x} \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$. Given a real number t, we let $(t)_+ := \max\{t, 0\}$ and $(t)_- := \min\{t, 0\}$. Given a set I, we use $\text{ext}\{I\}$ to represent its extreme points. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector and denote its realizations by $\boldsymbol{\xi}$. Given a probability distribution $\mathbb{P}$ defined on support $\mathcal{U}$ with sigma-algebra $\mathcal{F}$ and a $\mathbb{P}$-measurable function $g(\boldsymbol{\xi})$, we use $\mathbb{P}\{A\}$ to denote $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \text{condition } A(\tilde{\boldsymbol{\xi}}) \text{ holds}\}$ when $A(\tilde{\boldsymbol{\xi}})$ is a condition on $\boldsymbol{\xi}$, and to denote $\mathbb{P}\{\tilde{\boldsymbol{\xi}} : \tilde{\boldsymbol{\xi}} \in A\}$ when $A \in \mathcal{F}$ is $\mathbb{P}$-measurable, and we let $\text{ess.sup}_{\mathbb{P}}(g(\tilde{\boldsymbol{\xi}}))$ denote the essential supremum of the random function $g(\tilde{\boldsymbol{\xi}})$. Given a set R, the characteristic function $\chi_R(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in R$, and ∞, otherwise. We let δ_ω denote for the Dirac distribution that places unit mass on the realization ω. We use $\lfloor x \rfloor$ and $\lceil x \rceil$ to denote the largest integer y satisfying $y \leq x$ and the smallest integer y satisfying $y \geq x$ for any $x \in \mathbb{R}$, respectively. Additional notations will be introduced as needed.

2 NP-hardness of DFO and the MICP-R Notion

DFO (1) can be viewed as a biconvex program, which is notoriously known to be computationally challenging. Hence, in this section, we first show that solving DFO (1) is, in general, an NP-hard problem. Nevertheless, we are able to identify tractable DFO special cases, where in this section, we formally define the tractability. In contrast to biconvex programs, mixed-integer convex programs have recently been shown to be more scalable and capable of solving many large-scale problems (see more discussions in [1, 2] and the references cited therein). Therefore, we focus on studying MICP-R reformulations of DFO (1), and this section formally defines the MICP-R notions as a preliminary of our main results.

2.1 NP-hardness and Tractability of DFO

We observe that evaluating the most favorable objective function value of DFO (1) for a given decision can be NP-hard, even under a very simple setting.

Proposition 1 *Computing the inner infimum of DFO* (1), *in general, is NP-hard even when the ambiguity set* $\mathcal{P} = \{\mathbb{P} : \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{U}\} = 1\}$ *with box uncertainty set* $\mathcal{U}$ *and the recourse function* $Q(\boldsymbol{x}, \boldsymbol{\xi})$ *only involves the objective uncertainty.*

Proof Let us consider the NP-complete problem — set partition problem, which asks

> ***Set partition problem.*** *Given N nonnegative integers $w_1, w_2, \cdots, w_N$, does there exist one set partition S, such that $\sum_{i \in S} w_i = \sum_{i \in [N] \setminus S} w_i$?*

In DFO (1), let the ambiguity set $\mathcal{P} = \{\mathbb{P} : \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{U}_I\} = 1\}$ with an interval uncertainty set $\mathcal{U}_I = [-1, 1]^N$, and let

$$Q(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y} \in \mathcal{Y}} \sum_{i \in [N]} \xi^i (\boldsymbol{a}_i^\top \boldsymbol{y} - b_i),$$

where $\boldsymbol{a}_i = \boldsymbol{e}_i$ and $b_i = 0$ for each $i \in [N]$, and set $\mathcal{Y} = \{\boldsymbol{y} \in \mathbb{R}^N : -1 \leq y_i \leq 1, \forall j \in [N], \sum_{i \in [N]} w_j y_j = 0\}$. In this setting, the inner infimum of DFO (1) reduces to

$$v^* = \min_{\boldsymbol{\xi}, \boldsymbol{y}} \left\{ \sum_{i \in [N]} \xi^i y_i : -1 \leq \xi^i \leq 1, \forall i \in [N], -1 \leq y_i \leq 1, \forall i \in [N], \sum_{i \in [N]} w_i y_i = 0 \right\}. \tag{4a}$$

Above, optimizing over $\boldsymbol{\xi}$ first, problem (4a) reduces to

$$v^* = \min_{\boldsymbol{y}} \left\{ -\sum_{i \in [N]} \max(y_i, 0) + \sum_{i \in [N]} \min(y_i, 0) : -1 \le y_i \le 1, \forall i \in [N], \sum_{i \in [N]} w_i y_i = 0 \right\}. \quad (4b)$$

Then, we observe that the optimal value $v^* = -N$ in (4b) if and only if there exists an optimal solution $\boldsymbol{y}^* \in \{-1, 1\}^N$, i.e., the optimal value $v^* = -N$ in (4b) if and only if there exists a set partition such that $\sum_{i \in S} w_i = \sum_{i \in [N] \setminus S} w_i$. Since the set partition problem is NP-hard, solving problem (4b) is NP-hard. That is, computing the inner infimum of DFO (1) is NP-hard. $\qquad \square$

Proposition 1 motivates us to explore special cases under which computing the inner infimum of DFO (1) is tractable. Formally, we define tractable convex programs following the convention from work [8], as below.

Definition 1 (Tractability, theorem A.3.3 in [8]) Suppose that for any given compact set $\mathcal{X} \subseteq \mathbb{R}^n$, which has a nonempty relative interior and is contained in a Euclidean ball with radius R and is containing a Euclidean ball with radius r, then there exists an efficient algorithm to solve the favorable problem $\min_{\boldsymbol{x} \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})]$ to $\widehat{\varepsilon} > 0$ accuracy, whose running time is polynomial in $n, m, \ln(R/r), \ln(1/\widehat{\varepsilon})$, and the encoding length of $\min_{\boldsymbol{x} \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})]$.

2.2 The MICP-R Notion

As this paper aims to explore conditions under which DFO (1) can be mixed-integer convex programming (MICP) representable (MICP-R), we formally define this notion, initially introduced in the work [24], below.

Definition 2 (i) (definition 1.1 in [24]) Given $n, p, d \in \mathbb{Z}_+$, suppose that sets $\mathcal{S} \subseteq \mathbb{R}^n$ and $\mathcal{M} \subseteq \mathbb{R}^{n+p+d}$ are closed and convex. Then the tuple $(\mathcal{M}, p, d)$ induces an MICP formulation of set $\mathcal{S}$ if

$$\boldsymbol{x} \in \mathcal{S} \Leftrightarrow \exists \boldsymbol{y} \in \mathbb{R}^p, \boldsymbol{z} \in \mathbb{Z}^d, \text{s.t. } (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{M};$$

(ii) (An MICP-R Set, definition 1.2 in [24]) A set $\mathcal{S} \in \mathbb{R}^n$ is MICP representable (MICP-R) if there exists a closed convex set $\mathcal{M}$ and two positive integers p and d that induce an MICP formulation of set $\mathcal{S}$;

(iii) (An MICP-R Function) A function $f : \mathcal{S} \to \mathbb{R}$ is MICP-R if both its domain $\mathcal{S}$ and its epigraph are MICP-R; and

(iv) (An MICP Formulation) A mathematical program is MICP-R if both its feasible region and objective function are MICP-R.

The definition of not being MICP-R is simply the opposite of being MICP-R, which is, unfortunately, difficult to verify in practice. Fortunately, the authors in [24] provided a simple and sufficient condition to prove that a set is not MICP-R.

Lemma 1 (lemma 4.1 in [24]) A set $\mathcal{S} \in \mathbb{R}^n$ is not MICP-R if there exists an infinite sequence $\{\widehat{\boldsymbol{x}}^j\}_j$ such that $\widehat{\boldsymbol{x}}^{j_1} \ne \widehat{\boldsymbol{x}}^{j_2} \in \mathcal{S}$ for all $j_1 \ne j_2$ and $1/2(\widehat{\boldsymbol{x}}^{j_1} + \widehat{\boldsymbol{x}}^{j_2}) \notin \mathcal{S}$.

For brevity of notation, we also introduce the McCormick representation [26] of a simple bilinear set having a binary variable.

Definition 3 (McCormick Representation of a Simple Bilinear Set, [26]) The bilinear set $\{(s, \lambda, \gamma) \in \mathbb{R} \times \{\lambda_l, \lambda_u\} \times [\gamma_l, \gamma_u] : s = \lambda\gamma\}$ admits the following mixed-integer linear programming (MILP) McCormick representation:

$$\mathcal{MI}(\lambda_l, \lambda_u, \gamma_l, \gamma_u) = \left\{ (s, \lambda, \gamma) : \begin{matrix} s \in \mathbb{R}, \lambda \in \{\lambda_l, \lambda_u\}, \gamma_l \le \gamma \le \gamma_u, \\ s \ge \lambda_l \gamma + \lambda \gamma_l - \lambda_l \gamma_l, s \ge \lambda_u \gamma + \lambda \gamma_u - \lambda_u \gamma_u, \\ s \le \lambda_u \gamma + \lambda \gamma_l - \lambda_u \gamma_l, s \le \lambda \gamma_u + \lambda_l \gamma - \lambda_l \gamma_u \end{matrix} \right\},$$

where λ_l, λ_u and γ_l, γ_u are the known lower and upper bounds for λ and γ, respectively.

According to Definition 2 and Definition 3, the following result shows that the reverse norm function $f(\boldsymbol{x}) = -\|\boldsymbol{x}\|_p + \chi_{\mathcal{X}}(\boldsymbol{x})$ can be either MICP-R or not MICP-R, which depends on the norm $-\|\cdot\|_p$ (recall that set $\mathcal{X}$ is compact and has a nonempty relative interior based on Assumption 1).

Lemma 2 *The reverse norm function* $f(\boldsymbol{x}) = -\|\boldsymbol{x}\|_p + \chi_{\mathcal{X}}(\boldsymbol{x})$ *is MICP-R if* $p \in \{1, \infty\}$ *and is not MICP-R if* $p \in (1, \infty)$.

Proof We focus on the MICP-R formulation of the epigraph of the function $f(\cdot)$, which reads as

$$\operatorname{epi}(f) = \{(\boldsymbol{x}, t)\colon -\|\boldsymbol{x}\|_p \leq t, \boldsymbol{x} \in \mathcal{X}\}. \tag{5a}$$

Next, we split the proof into three cases based on the choice of p.
Case 1: When $p = 1$, i.e., the norm is L_1, we have

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t)\colon \sum_{i \in [n]} |x_i| \geq -t, \boldsymbol{x} \in \mathcal{X}\right\},$$

which is equivalent to

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t)\colon \max_{\boldsymbol{z} \in \{-1,1\}^n} \sum_{i \in [n]} x_i z_i \geq -t, \boldsymbol{x} \in \mathcal{X}\right\},$$

or

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t)\colon \sum_{i \in [n]} x_i z_i \geq -t, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \{-1,1\}^n\right\}.$$

Since set $\mathcal{X}$ is compact, we can assume that $\mathcal{X} \subseteq [\boldsymbol{l}, \boldsymbol{u}]$, i.e., given $\boldsymbol{x} \in \mathcal{X}$, we have $x_i \in [l_i, u_i]$ for each $i \in [n]$. We can apply the following McCormick inequalities (see more details in Definition 3) to linearize the bilinear term $\{s_i := x_i z_i\}_{i \in [n]}$, i.e., $(s_i, z_i, x_i) \in \mathcal{MI}(-1, 1, l_i, u_i)$ for each $i \in [n]$. Thus, $\operatorname{epi}(f)$ is MICP-R, i.e.,

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t)\colon \begin{array}{l} \exists \boldsymbol{s} \in \mathbb{R}^n, \sum_{i \in [n]} s_i \geq -t, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \{-1,1\}^n, \\ (s_i, z_i, x_i) \in \mathcal{MI}(-1, 1, l_i, u_i), \forall i \in [n] \end{array}\right\}.$$

Case 2: When $p \in (1, \infty)$, i.e., the norm is neither L_1 nor L_∞, since set $\mathcal{X}$ is compact and has a nonempty relative interior, there exists an open ball $B(\bar{\boldsymbol{x}}, r)$ centered at $\bar{\boldsymbol{x}}$ and a positive radius $r > 0$ such that the intersection of $B(\bar{\boldsymbol{x}}, r)$ and the affine space of set $\mathcal{X}$ is contained in set $\mathcal{X}$. Therefore, set

$$\mathcal{S} := \{\boldsymbol{x} \in X : \|\boldsymbol{x}\|_p = \|\bar{\boldsymbol{x}}\|_p\}$$

has a nonempty relative interior. Thus, we can pick a sequence of distinct elements from set $\mathcal{S}$ (e.g., all the possible rational elements) $\{\widehat{\boldsymbol{x}}^j\}_j$. Since $(\widehat{\boldsymbol{x}}^j, \|\bar{\boldsymbol{x}}\|_p) \in \operatorname{epi}(f)$ for each j and function $\|\boldsymbol{x}\|_p$ is strictly convex for any $p \in (1, \infty)$, for any pair (j_1, j_2) with $j_1 \neq j_2$, we must have

$$\frac{1}{2}(\widehat{\boldsymbol{x}}^{j_1}, \|\bar{\boldsymbol{x}}\|_p) + \frac{1}{2}(\widehat{\boldsymbol{x}}^{j_2}, \|\bar{\boldsymbol{x}}\|_p) \notin \operatorname{epi}(f).$$

Since $\{\widehat{\boldsymbol{x}}^j\}_j$ is an infinite sequence, according to Lemma 1, set $\operatorname{epi}(f)$ is not MICP-R.
Case 3: When $p = \infty$, i.e., the norm is L_∞, set $\operatorname{epi}(f)$ reduces to

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t)\colon \max_{i \in [n]} |x_i| \geq -t, \forall i \in [n], \boldsymbol{x} \in \mathcal{X}\right\},$$

which can be reformulated in the form of the following disjunction [6]:

$$\mathrm{epi}(f) = \bigvee_{i \in [n]} \{(\boldsymbol{x}, t) \colon x_i \geq -t, \boldsymbol{x} \in \mathcal{X}\} \bigvee_{i \in [n]} \{(\boldsymbol{x}, t) \colon -x_i \geq -t, \boldsymbol{x} \in \mathcal{X}\}.$$

Since set $\mathcal{X}$ is compact, the MICP-R formulation of set $\mathrm{epi}(f)$ follows the well-known results from disjunctive programming [6]. □

Note that when $p = \infty$, although being MICP-R, the optimization of the function $f(\boldsymbol{x})$ can be done efficiently by solving $2n$ tractable convex programs. The result in Lemma 2 is useful to prove that many DFO problems can be MICP-R. Moreover, it is worth mentioning that Lemma 2 is also applicable in analyzing the MICP-R formulation of many distributional robust optimization problems. For example, [21] discussed the MICP-R formulation in distributionally robust chance constrained programs under the general Wasserstein ambiguity set.

3 sDFO with Piecewise Affine Functions

Motivated from Section 2, we provide sufficient conditions for DFO (1) to be tractable or MICP-R. In this section, we first explore sDFO (2) as a special case of DFO (1). Similar to works in robust optimization literature [8, 14, 41], we focus on the function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ being convex or concave piecewise affine in $\boldsymbol{x}$, respectively.

3.1 sDFO (2) with Concave Piecewise Affine Functions

We first consider that the function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ is the minimum of K piecewise affine functions $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x})$ with affine mappings $\boldsymbol{a}_k(\boldsymbol{x}) = \widehat{\boldsymbol{A}}_k \boldsymbol{x} + \widehat{\boldsymbol{a}}_k \in \mathbb{R}^m$ with $\widehat{\boldsymbol{A}}_k \in \mathbb{R}^{m \times n}, \widehat{\boldsymbol{a}}_k \in \mathbb{R}^m$ and $b_k(\boldsymbol{x}) = \widehat{\boldsymbol{B}}_k^\top \boldsymbol{x} + \widehat{b}_k \in \mathbb{R}$ with $\widehat{\boldsymbol{B}}_k \in \mathbb{R}^n, \widehat{b}_k \in \mathbb{R}$ for each $k \in [K]$, i.e., $Q(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{k \in [K]}[\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x})]$. Suppose that the uncertainty set is defined as the ball $\mathcal{U} = \{\boldsymbol{\xi} : \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta\}$ with the known parameter $\boldsymbol{\xi}^0$ and the radius $\theta \geq 0$. It is worth mentioning that the reformulations and complexity analyses can be simply extended to more general uncertainty sets, such as polyhedral and ellipsoidal (see, e.g., [8]). In this subsection, for brevity, we focus on the ball uncertainty set $\mathcal{U}$. In this setting, sDFO (2) becomes

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{\xi} \in \mathcal{U}} \min_{k \in [K]} \left\{ \boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right\}. \tag{6}$$

Switching the first minimum operator with the third one and invoking the definition of dual norm, problem (6) is further equivalent to

$$v^* = \min_{k \in [K]} \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} \right\},$$

which can be solved by selecting the lowest objective value within these K mathematical programs, that is,

$$v^* = \min_{k \in [K]} \left\{ v_k^* := \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} \right\} \right\}. \tag{7}$$

Note that the inner minimization of sDFO (7) is a concave minimization problem and, in general, can be difficult. However, by exploring the properties of the dual norm, there are some conditions under which sDFO (7) can be tractable.

Theorem 1 *sDFO* (7) *can be tractable if either condition holds:*

(i) If $\|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} := C_k$ *is constant for each* $k \in [K]$ *and* $\boldsymbol{x} \in \mathcal{X}$, *sDFO* (7) *is equivalent to solving* K *tractable convex programs, i.e.,* $v^* = \min_{k \in [K]} v_k^*$, *where for each* $k \in [K]$, *we have*

$$v_k^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta C_k \right\};$$

(ii) If $p = 1$, sDFO (7) is equivalent to solving $2mK$ tractable convex programs, i.e., $v^* = \min_{k \in [K], i \in [m], \ell \in [2]} v^*_{ik\ell}$, where for each $k \in [K]$ and $i \in [m]$, we have

$$v^*_{ik1} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) + \theta a_{ki}(\boldsymbol{x}) \right\}, \quad v^*_{ik2} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta a_{ki}(\boldsymbol{x}) \right\}.$$

Proof We split the proof into two parts by checking these two conditions separately.

(i) When $\|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*}$ is a constant for each $k \in [K]$, i.e., $\|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} = C_k$ for $k \in [K]$, then the objective function of sDFO (7) is linear and optimizing it is equivalent to solving K convex programs;

(ii) When $p = 1$, i.e., when the dual norm is L_∞, then $\theta \|\boldsymbol{a}_k(\boldsymbol{x})\|_\infty = \theta \max_{i \in [m]} \max \{a_{ki}(\boldsymbol{x}), -a_{ki}(\boldsymbol{x})\}$. That is, sDFO (7) can be simplified as

$$\min_{k \in [K]} \min_{i \in [m]} \min \left\{ \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) + \theta a_{ki}(\boldsymbol{x}) \right\}, \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta a_{ki}(\boldsymbol{x}) \right\} \right\},$$

which is equivalent to solving $2mK$ convex programs and selecting the best one with the lowest optimal value. $\qquad\square$

In the following complexity analysis, we focus on the non-trivial cases where $\|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*}$ is not a constant for some $k \in [K]$. Unfortunately, when $p \in (1, \infty]$, solving sDFO (7), in general, is NP-hard with the reduction to the well-known NP-hard problem — maximizing a norm over a polytope.

Proposition 2 *For any $p \in (1, \infty]$, solving sDFO (7), in general, is NP-hard even with $K = 1$.*

Proof Let us consider an NP-hard problem — Norm maximization over a polytope (see theorem 1 in [15]), which asks

Norm maximization over a polytope. *Given the polytope $\{\boldsymbol{x} : \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}\}$, where $\boldsymbol{D} \in \mathbb{R}^{\tau \times n}$ and $\boldsymbol{d} \in \mathbb{R}^\tau$, what is the optimal value of the problem $\max_{\boldsymbol{x}} \{\|\boldsymbol{x}\|_{p^*} : \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}\}$ with $p \in (1, \infty]$?*

Consider a special case of sDFO (7), where $K = 1$, $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = 0$, $\boldsymbol{\xi}^0 = \boldsymbol{0}$, $\theta = 1$, and set $\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}\}$. In this case, sDFO (7) can be written as

$$\max_{\boldsymbol{x}} \{\|\boldsymbol{x}\|_{p^*} : \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}\},$$

which is exactly the desirable norm maximization problem over a polytope for any $p \in (1, \infty]$. Thus, solving sDFO (7), in general, is NP-hard for any $p \in (1, \infty]$. $\qquad\square$

The complexity result suggests that the tractable result in Theorem 1 with $p = 1$ is the best one that we could expect.

Next, for the intractable case, we study the MICP-R formulation of the objective function of sDFO (7). As an extension of Lemma 2, we notice that when $p = \infty$, the objective function of sDFO (7) is MICP-R; otherwise, when $p \in (1, \infty)$, it is not.

Theorem 2 *When $p = \infty$, the objective function of sDFO (7) with domain $\mathcal{X}$ is MICP-R; otherwise, when $p \in (1, \infty)$, the objective function of sDFO (7) with domain $\mathcal{X}$ may not be MICP-R.*

Proof We first rewrite the objective function of sDFO (7) with domain $\mathcal{X}$ as $f_k(\boldsymbol{x}) = \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} + \chi_{\mathcal{X}}(\boldsymbol{x})$ for each $k \in [K]$. We then focus on the MICP-R formulation of the epigraph of the function $f_k(\cdot)$, which reads as

$$\mathrm{epi}(f_k) = \left\{ (\boldsymbol{x}, t) : \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} \le t, \boldsymbol{x} \in \mathcal{X} \right\}. \tag{8}$$

Next, we split the proof into two cases based on the choice of p.

Case 1: When $p = \infty$, i.e., the dual norm is L_1, set $\mathrm{epi}(f_k)$ can be written as

$$\mathrm{epi}(f_k) = \left\{ (\boldsymbol{x}, t) \colon \sum_{i \in [m]} |a_{ki}(\boldsymbol{x})| - \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) - b_k(\boldsymbol{x}) \geq -t, \boldsymbol{x} \in \mathcal{X} \right\}.$$

Then we have

$$\mathrm{epi}(f_k) = \left\{ (\boldsymbol{x}, t) \colon \max_{\boldsymbol{z} \in \{-1, 1\}^m} \sum_{i \in [m]} a_{ki}(\boldsymbol{x}) z_i - \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) - b_k(\boldsymbol{x}) \geq -t, \boldsymbol{x} \in \mathcal{X} \right\},$$

which is equivalent to

$$\mathrm{epi}(f_k) = \left\{ (\boldsymbol{x}, t) \colon \sum_{i \in [m]} a_{ki}(\boldsymbol{x}) z_i - \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) - b_k(\boldsymbol{x}) \geq -t, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \{-1, 1\}^m \right\}.$$

Since set $\mathcal{X}$ is compact, we can apply the McCormick inequalities (see Definition 3) to linearize the bilinear terms $\{a_{ki}(\boldsymbol{x}) z_i\}_{i \in [m]}$. Thus, $\mathrm{epi}(f_k)$ is MICP-R for each $k \in [K]$.

Case 2: When $p \in (1, \infty)$, i.e., when the dual norm is neither L_1 nor L_∞, suppose that $K = 1$, $\boldsymbol{\xi}^0 = \boldsymbol{0}$, $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, and $b_1(\boldsymbol{x}) = 0$, then set (8) reduces to

$$\mathrm{epi}(f_1) = \{ (\boldsymbol{x}, t) \colon -\|\boldsymbol{x}\|_{p^*} \leq t, \boldsymbol{x} \in \mathcal{X} \},$$

which is identical to (5a). According to the result in Lemma 2, when $p \in (1, \infty)$, the objective function of sDFO (7) with domain $\mathcal{X}$, in general, may not be MICP-R. $\qquad \square$

Theorem 2 suggests that the objective function of sDFO (7) with domain $\mathcal{X}$ may not be MICP-R with a general norm, but it is MICP-R when the norm is L_∞. As a direct corollary of Theorem 2, when $p = \infty$, the MICP-R formulation of sDFO (7) can be summarized as follows.

Corollary 1 *When $p = \infty$, suppose that $\mathcal{X} \subseteq [\boldsymbol{l}, \boldsymbol{u}]$ and let $\widehat{l}_{ki} = \sum_{j \in [n]} \min\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ and $\widehat{u}_{ki} = \sum_{j \in [n]} \max\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ for each $i \in [m]$ such that $\boldsymbol{a}_k(\boldsymbol{x}) \in [\widehat{\boldsymbol{l}}_k, \widehat{\boldsymbol{u}}_k]$ for each $k \in [K]$. Then, sDFO (7) is equivalent to solving the following K MICPs, i.e., $v^* = \min_{k \in [K]} v_k^*$, where for each $k \in [K]$, we have*

$$v_k^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{\xi}^{0\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \sum_{i \in [m]} s_{ki} \colon (s_{ki}, z_{ki}, a_{ki}(\boldsymbol{x})) \in \mathcal{MI}\left(-1, 1, \widehat{l}_{ki}, \widehat{u}_{ki}\right), \forall i \in [m] \right\}.$$

3.2 sDFO (2) with Convex Piecewise Affine Functions

In this subsection, we follow the same notation and uncertainty set as the previous subsection (Section 3.1) and consider the maximum of piecewise affine function, that is, the function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ is defined as the maximum of K piecewise affine function $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x})$, i.e., $Q(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{k \in [K]}[\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x})]$. In this setting, sDFO (2) can be recast as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{\xi} \in \mathcal{U}} \max_{k \in [K]} \left\{ \boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right\}. \tag{9}$$

Let us first provide an equivalent reformulation of sDFO (9), which helps establish the tractability and MICP formulation of sDFO (9).

Lemma 3 *sDFO (9) is equivalent to*

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi^0}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right] - \theta \left\| \sum_{k \in [K]} \lambda_k \boldsymbol{a}_k(\boldsymbol{x}) \right\|_{p^*} : \sum_{k \in [K]} \lambda_k = 1 \right\}. \tag{10}$$

Proof Let us first consider the inner minimax of sDFO (9) as

$$\min_{\boldsymbol{\xi}} \left\{ \max_{k \in [K]} \boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) : \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta \right\}.$$

Introducing auxiliary nonnegative variables $\boldsymbol{\lambda}$, the inner minimax of sDFO (9) is equivalent to

$$\min_{\boldsymbol{\xi}} \left\{ \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right] : \sum_{k \in [K]} \lambda_k = 1 \right\} : \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta \right\}.$$

According to Sion's minimax theorem [37], we can interchange the maximum operator with the minimum one as

$$\max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \min_{\boldsymbol{\xi}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right] : \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta \right\} : \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

Invoking the definition of dual norm, we have

$$\max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi^0}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) \right] - \theta \left\| \sum_{k \in [K]} \lambda_k \boldsymbol{a}_k(\boldsymbol{x}) \right\|_{p^*} : \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

This completes the proof.   □

Due to the bilinear terms in the reformulation (10), sDFO (9), in general, can be difficult to solve. However, by exploring the objective function and the properties of the dual norm, we are able to prove conditions under which sDFO (9) can be tractable.

Theorem 3 *sDFO (9) can be tractable if any of the following conditions holds:*

(i) When $\boldsymbol{a}_k(\boldsymbol{x}) := \bar{\boldsymbol{a}}_k$ is constant for all $k \in [K]$ and $\boldsymbol{x} \in \mathcal{X}$, sDFO (9) is equivalent to

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\xi}, \eta} \left\{ \eta : \eta \geq \boldsymbol{\xi}^\top \bar{\boldsymbol{a}}_k + b_k(\boldsymbol{x}), \forall k \in [K], \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta \right\};$$

(ii) When $p = 1$ and $a_k(\boldsymbol{x}) = a_1(\boldsymbol{x})$ for each $k \in [K]$, sDFO (9) is equivalent to solving $2m$ tractable convex programs, and selecting the lowest optimal value, i.e., $v^* = \min_{i \in [m], \ell \in [2]} v_{i\ell}^*$, where for each $i \in [m]$, we have

$$v_{i1}^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta : \eta \geq b_k(\boldsymbol{x}) + \boldsymbol{\xi^0}^\top \boldsymbol{a}_1(\boldsymbol{x}) - \theta a_{1i}(\boldsymbol{x}), \forall k \in [K] \right\},$$

$$v_{i2}^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta : \eta \geq b_k(\boldsymbol{x}) + \boldsymbol{\xi^0}^\top \boldsymbol{a}_1(\boldsymbol{x}) + \theta a_{1i}(\boldsymbol{x}), \forall k \in [K] \right\};$$

(iii) Suppose that $p = 1$, and $\boldsymbol{\xi} := [\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_K]$ such that $\boldsymbol{\xi}^i$ and $\boldsymbol{\xi}^j$ do not overlap for each $i \neq j$, and $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{\xi}^{k\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x})$ for each $k \in [K]$ such that $\|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_\infty = C_k$ is constant for each $k \in [K]$, where $\boldsymbol{\xi}_k \in \mathbb{R}^{m_k}$, $\bar{\boldsymbol{a}}_k(\boldsymbol{x}) = \bar{\boldsymbol{A}}_k \boldsymbol{x} + \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ with $\bar{\boldsymbol{A}}_k \in \mathbb{R}^{m_k \times n}, \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ such that $\bar{\boldsymbol{A}}_i$ and $\bar{\boldsymbol{A}}_j$, $\bar{\boldsymbol{a}}_i$ and $\bar{\boldsymbol{a}}_j$ do not overlap each $i \neq j$ with $\sum_{k \in [K]} m_k = m$ and each m_k is nonnegative. Then, sDFO (9) is equivalent to solving

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \beta, \boldsymbol{\gamma} \geq \boldsymbol{0}} \left\{ \beta : \sum_{k \in [K]} \gamma_k = \theta, \beta \geq \boldsymbol{\xi}_k^{0\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \gamma_k C_k, \forall k \in [K] \right\}.$$

Proof We split the proof into three parts accordingly.

(i) When $\boldsymbol{a}_k(\boldsymbol{x}) = \bar{\boldsymbol{a}}_k$ is constant for all $k \in [K]$, sDFO (9) can be written as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{\xi}} \max_{k \in [K]} \left\{ \boldsymbol{\xi}^\top \bar{\boldsymbol{a}}_k + b_k(\boldsymbol{x}) \colon \|\boldsymbol{\xi} - \boldsymbol{\xi}^0\|_p \leq \theta \right\}.$$

Introducing an auxiliary variable η to linearize the inner maximum, we arrive at Part (i).

(ii) When $\boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{a}_1(\boldsymbol{x})$ for each $k \in [K]$, we rewrite sDFO (10) as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) + \boldsymbol{\xi}^{0\top} \boldsymbol{a}_1(\boldsymbol{x}) - \theta \|\boldsymbol{a}_1(\boldsymbol{x})\|_{p^*} \colon \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

Taking the dual of the inner maximization problem and using strong duality from linear programming, we have

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta \colon \eta \geq b_k(\boldsymbol{x}) + \boldsymbol{\xi}^{0\top} \boldsymbol{a}_1(\boldsymbol{x}) - \theta \|\boldsymbol{a}_1(\boldsymbol{x})\|_{p^*}, \forall k \in [K] \right\}.$$

When $p = 1$, i.e., when the dual norm is L_∞, then $\theta \|\boldsymbol{a}_1(\boldsymbol{x})\|_\infty = \theta \max_{i \in [m]} \max\{a_{1i}(\boldsymbol{x}), -a_{1i}(\boldsymbol{x})\}$. Therefore, sDFO (10) is equivalent to solving $2m$ tractable problems, and selecting the best one with the lowest optimal value, i.e., $v^* = \min_{i \in [m], \ell \in [2]} v_{i\ell}^*$, where for each $i \in [m]$, we have

$$v_{i1}^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta \colon \eta \geq b_k(\boldsymbol{x}) + \boldsymbol{\xi}^{0\top} \boldsymbol{a}_1(\boldsymbol{x}) - \theta a_{1i}(\boldsymbol{x}), \forall k \in [K] \right\},$$

$$v_{i2}^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta \colon \eta \geq b_k(\boldsymbol{x}) + \boldsymbol{\xi}^{0\top} \boldsymbol{a}_1(\boldsymbol{x}) + \theta a_{1i}(\boldsymbol{x}), \forall k \in [K] \right\}.$$

(iii) Since $\boldsymbol{\xi}^i$ and $\boldsymbol{\xi}^j$ do not overlap for each $i \neq j$ and $\bar{\boldsymbol{a}}_j(\boldsymbol{x})$ and $\bar{\boldsymbol{a}}_j(\boldsymbol{x})$ do not overlap for each $i \neq j$ as well, when $p = 1$ and $\|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_\infty = C_k$ is constant for each $k \in [K]$, the dual norm term in (10) can be simplified as

$$\left\| \sum_{k \in [K]} \lambda_k \boldsymbol{a}_k(\boldsymbol{x}) \right\|_\infty = \max_{k \in [K]} \lambda_k \|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_\infty = \max_{k \in [K]} \lambda_k C_k.$$

Then, sDFO (10) can be written as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}_k^{0\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) \right] + \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) - \theta \max_{k \in [K]} \lambda_k C_k \colon \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

Introducing one variable η to linearize the term $\max_{k \in [K]} \lambda_k C_k$, then we have

$$\max_{\boldsymbol{\lambda} \geq \boldsymbol{0}, \eta} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}_k^{0\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) \right] + \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) - \theta \eta \colon \sum_{k \in [K]} \lambda_k = 1, \lambda_k C_k - \eta \leq 0, \forall k \in [K] \right\}.$$

Taking the dual of the inner maximization problem with dual variables $\beta, \boldsymbol{\gamma}$ and using strong duality from linear programming, we have

$$\min_{\beta, \boldsymbol{\gamma} \geq \boldsymbol{0}} \left\{ \beta \colon \sum_{k \in [K]} \gamma_k = \theta, \beta \geq \boldsymbol{\xi}_k^{0\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \gamma_k C_k, \forall k \in [K] \right\}.$$

This completes the proof. □

The results in Theorem 3 may be the best ones that we could expect. In general, solving sDFO (9) is NP-hard for any convex L_p norm.

Proposition 3 *For any $p \in [1, \infty]$, solving sDFO (9), in general, is NP-hard.*

Proof Note that when $K = 1$, the sDFO (9) is equivalent to formulation (7). Thus, the complexity results in Proposition 2 hold, i.e., solving the sDFO (9) is, in general, NP-hard for $p \in (1, \infty]$.

It remains to show that solving the sDFO (9) is also NP-hard when $p = 1$. Let us consider the NP-complete problem — feasibility problem of a general binary program, which asks

> ***Feasibility of a binary program.*** *Given an integer matrix $\boldsymbol{D} \in \mathbb{Z}^{\tau \times K}$, and integer vector $\boldsymbol{d} \in \mathbb{Z}^\tau$, is there a vector $\boldsymbol{x} \in \{-1, 1\}^K$ such that $\boldsymbol{Dx} \le \boldsymbol{d}$?*

Let us consider the following special case of the sDFO (9). We first suppose $\theta = 1$ and $\boldsymbol{\xi}^0 = \boldsymbol{0}$, then the uncertainty set becomes

$$\mathcal{U} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^K : \|\boldsymbol{\xi}\|_1 \le 1 \right\}.$$

Next, let us consider the following function

$$Q(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{k \in [K]} \max \left\{ \xi_k^k x_k - 1, -\xi_k^k x_k + 1 \right\},$$

and the set $\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{Dx} \le \boldsymbol{d}, -1 \le x_k \le 1, \forall k \in [K]\}$. Under this special setting, the sDFO (9) reduces to

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{\xi}} \left\{ \max_{k \in [K]} \left\{ |\xi_k^k x_k - 1| \right\} : \boldsymbol{Dx} \le \boldsymbol{d}, \boldsymbol{x} \in [-1, 1]^K, \boldsymbol{\xi}^k \in [-1, 1]^K, \forall k \in [K] \right\}. \tag{11}$$

We observe that the optimal value $v^* = 0$ in (11) if and only if $\xi_k^k x_k = 1$ for all $k \in [K]$, i.e., if and only if there exists a binary feasible solution $\boldsymbol{x} \in \{-1, 1\}^K$ such that $\boldsymbol{Dx} \le \boldsymbol{d}$. Thus, solving problem (11) is NP-hard, so is the sDFO (9). □

Proposition 3 motivates us to investigate the MICP-R formulation of the objective function of sDFO (9) with domain $\mathcal{X}$. Unfortunately, in most cases, the objective function of sDFO (9) with domain $\mathcal{X}$ may not be MICP-R.

Proposition 4 *For any $p \in (1, \infty)$, the objective function of sDFO (9) with domain $\mathcal{X}$ may not be MICP-R.*

Proof Note that when $K = 1$, sDFO (9) is equivalent to sDFO (7). Therefore, according to the result in Theorem 2, when $p \in (1, \infty)$, the objective function of sDFO (9) with domain $\mathcal{X}$ may not be MICP-R. □

When $p \notin (1, \infty)$, in Theorem 3, we show that when $p = 1$, there exist special cases such that sDFO (9) can be tractable. Next, we show special cases under which the objective function of sDFO (9) with domain $\mathcal{X}$ can be MICP-R when $p = \infty$.

Theorem 4 *When $p = \infty$, the objective function of sDFO (9) with domain $\mathcal{X}$ is MICP-R if one of the following conditions holds:*

(i) *When $\boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{a}_1(\boldsymbol{x})$ for each $k \in [K]$; or*
(ii) *When $\boldsymbol{\xi} := [\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_K]$ such that $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ do not overlap for each $i \ne j$, and $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{\xi}_k^\top \bar{\boldsymbol{a}}_k(\boldsymbol{x})$ for each $k \in [K]$, where $\boldsymbol{\xi}_k \in \mathbb{R}^{m_k}$, $\bar{\boldsymbol{a}}_k(\boldsymbol{x}) = \bar{\boldsymbol{A}}_k \boldsymbol{x} + \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ with $\bar{\boldsymbol{A}}_k \in \mathbb{R}^{m_k \times n}, \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ such that $\bar{\boldsymbol{A}}_i$ and $\bar{\boldsymbol{A}}_j$, $\bar{\boldsymbol{a}}_i$ and $\bar{\boldsymbol{a}}_j$ do not overlap each $i \ne j$ with $\sum_{k \in [K]} m_k = m$.*

Proof We split the proof into two parts accordingly.

(i) When $\boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{a}_1(\boldsymbol{x})$ for each $k \in [K]$, we can rewrite sDFO (10) as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) + \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_1(\boldsymbol{x}) - \theta \|\boldsymbol{a}_1(\boldsymbol{x})\|_1 : \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

Taking the dual of the inner maximization problem and using strong duality from linear programming, we have

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta \colon \eta \geq \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_1(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \|\boldsymbol{a}_1(\boldsymbol{x})\|_1, \forall k \in [K] \right\}. \tag{12}$$

According to Part (i) in Theorem 2, sDFO (12) is MICP-R.

(ii) When $p = \infty$ and $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{\xi}_k^\top \bar{\boldsymbol{a}}_k(\boldsymbol{x})$ for each $k \in [K]$, we can rewrite sDFO (10) as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}_k^{0^\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) \right] + \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) - \theta \sum_{k \in [K]} \lambda_k \|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_1 : \sum_{k \in [K]} \lambda_k = 1 \right\},$$

which can be simplified as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \left\{ \sum_{k \in [K]} \lambda_k \left[\boldsymbol{\xi}_k^{0^\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) \right] + \sum_{k \in [K]} \lambda_k b_k(\boldsymbol{x}) - \theta \sum_{k \in [K]} \lambda_k \|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_1 : \sum_{k \in [K]} \lambda_k = 1 \right\}.$$

Taking the dual of the inner maximization problem and using strong duality from linear programming, we have

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \left\{ \eta \colon \eta \geq \left[\boldsymbol{\xi}_k^{0^\top} \bar{\boldsymbol{a}}_k(\boldsymbol{x}) \right] + b_k(\boldsymbol{x}) - \theta \|\bar{\boldsymbol{a}}_k(\boldsymbol{x})\|_1, \forall k \in [K] \right\}. \tag{13}$$

According to Part (i) in Theorem 2, sDFO (13) is MICP-R. $\qquad\square$

The following Corollary 2 shows the MICP-R formulations of the two cases discussed in Theorem 4.

Corollary 2 *When $p = \infty$, suppose that $\mathcal{X} \subseteq [\boldsymbol{l}, \boldsymbol{u}]$.*

(i) If $\boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{a}_1(\boldsymbol{x})$ for each $k \in [K]$, sDFO (9) can be reformulated as the following MICP

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \quad \eta,$$
$$\text{s.t.} \quad \eta \geq \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_1(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \sum_{i \in [m]} s_{1i}, \forall k \in [K],$$
$$(s_{1i}, z_{1i}, a_{1i}(\boldsymbol{x})) \in \mathcal{MI}\left(-1, 1, \widehat{l}_{1i}, \widehat{u}_{1i} \right), \forall i \in [m],$$

where we let $\widehat{l}_{ki} = \sum_{j \in [n]} \min\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ and $\widehat{u}_{ki} = \sum_{j \in [n]} \max\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ for each $i \in [m]$ such that $\boldsymbol{a}_k(\boldsymbol{x}) \in [\widehat{\boldsymbol{l}}_k, \widehat{\boldsymbol{u}}_k]$ for each $k \in [K]$; and

(ii) Suppose $\boldsymbol{\xi} := [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_K]$ such that $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ do not overlap for each $i \neq j$, and $\boldsymbol{\xi}^\top \boldsymbol{a}_k(\boldsymbol{x}) = \boldsymbol{\xi}_k^\top \bar{\boldsymbol{a}}_k(\boldsymbol{x})$ for each $k \in [K]$, where $\boldsymbol{\xi}_k \in \mathbb{R}^{m_k}$, $\bar{\boldsymbol{a}}_k(\boldsymbol{x}) = \bar{\boldsymbol{A}}_k \boldsymbol{x} + \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ with $\bar{\boldsymbol{A}}_k \in \mathbb{R}^{m_k \times n}, \bar{\boldsymbol{a}}_k \in \mathbb{R}^{m_k}$ such that $\bar{\boldsymbol{A}}_i$ and $\bar{\boldsymbol{A}}_j$, $\bar{\boldsymbol{a}}_i$ and $\bar{\boldsymbol{a}}_j$ do not overlap each $i \neq j$ with $\sum_{k \in [K]} m_k = m$ and each m_k is nonnegative. Then, sDFO (9) can be reformulated as the following MICP

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \eta} \quad \eta,$$
$$\text{s.t.} \quad \eta \geq \boldsymbol{\xi}^{0^\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \sum_{i \in [m]} s_{ki}, \forall k \in [K],$$
$$(s_{ki}, z_{ki}, \bar{a}_{ki}(\boldsymbol{x})) \in \mathcal{MI}\left(-1, 1, \bar{l}_{ki}, \bar{u}_{ki} \right), \forall k \in [K], i \in [m],$$

where we let $\bar{l}_{ki} = \sum_{j \in [n]} \min\{\bar{A}_{kij} l_j, \bar{A}_{kij} u_j\} + \bar{a}_{ki}$ and $\bar{u}_{ki} = \sum_{j \in [n]} \max\{\bar{A}_{kij} l_j, \bar{A}_{kij} u_j\} + \bar{a}_{ki}$ for each $i \in [m]$ such that $\bar{\boldsymbol{a}}_k(\boldsymbol{x}) \in [\bar{\boldsymbol{l}}_k, \bar{\boldsymbol{u}}_k]$ for each $k \in [K]$.

We remark that Corollary 2 extends the implications outlined in Theorem 4 by providing the MICP-R formulations of special cases of sDFO (9) when $p = \infty$. Given that sDFO (2) represents the special case of DFO (1), the results derived in this section can directly contribute to the discussions of DFO (1) in the next section.

4 DFO with the Ambiguity Set

Although DFO (1) is generally known to be NP-hard, in this section, we present sufficient conditions under which DFO (1) with different types of ambiguity sets can be tractable or MICP-R. We specifically investigate two representative ambiguity sets– the type-∞ Wasserstein ambiguity set and the ambiguity set with finite support.

4.1 DFO (1) with Type-∞ Wasserstein Ambiguity Set

Our first goal is to expand upon the results in Section 3, particularly those related to tractability, complexity, and MICP-R formulations. We aim to apply and adapt these insights specifically to DFO (1). In particular, we focus on type-∞ Wasserstein ambiguity set, which is defined as $\mathcal{P}_\infty^W = \{\mathbb{P}: \mathbb{P}\{\tilde{\boldsymbol{\xi}} \in \mathcal{U}\} = 1, W_\infty(\mathbb{P}, \mathbb{P}_{\tilde{\zeta}}) \leq \theta\}$, where $\mathbb{P}_{\tilde{\zeta}}$ is a discrete empirical reference distribution of random parameters $\tilde{\boldsymbol{\zeta}}$ generated by N i.i.d. samples such that $\mathbb{P}_{\tilde{\zeta}}\{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i\} = 1/N$, i.e., $\mathbb{P}_{\tilde{\zeta}} = 1/N \sum_{i \in [N]} \delta_{\boldsymbol{\zeta}^i}$ and $\delta_{\boldsymbol{\zeta}^i}$ is the Dirac function that places unit mass on the realization $\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i$ for each $i \in [N]$, and $\theta \geq 0$ is the Wasserstein radius. The type-∞ Wasserstein distance between two probability distributions $\mathbb{P}_1, \mathbb{P}_2$ is defined as

$$W_\infty(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \text{ess.sup}_\mathbb{Q} \left\| \boldsymbol{\xi}^1 - \boldsymbol{\xi}^2 \right\|_p : \begin{array}{c} \mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}^1 \text{ and } \tilde{\boldsymbol{\xi}}^2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \end{array} \right\}.$$

In this setting, DFO (1) admits the following representation (see, e.g., [9, 41]):

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \inf_{\mathbb{P} \in \mathcal{P}_\infty^W} \mathbb{E}_\mathbb{P} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] \right\} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \frac{1}{N} \sum_{i \in [N]} \left[\inf_{\boldsymbol{\xi}} \left\{ Q(\boldsymbol{x}, \boldsymbol{\xi}): \left\| \boldsymbol{\xi} - \boldsymbol{\zeta}^i \right\|_p \leq \theta \right\} \right] \right\}. \tag{14}$$

Similar to the discussions in Section 3.1 and Section 3.2, we then consider function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ to be convex and concave piecewise affine, respectively.

Special Case I. Concave Piecewise Function. We first consider the concave piecewise affine function. Following the same notation as Section 3.1, i.e., function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ is the minimum of piecewise affine functions, DFO (14) can be written as

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{N} \sum_{i \in [N]} \left[\min_{k \in [K]} \boldsymbol{\zeta}^{i\top} \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \left\| \boldsymbol{a}_k(\boldsymbol{x}) \right\|_{p^*} \right]. \tag{15}$$

We notice that if there is only one sample available in the empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ (i.e., $N = 1$), then DFO (15) reduces to sDFO (7). Thus, according to Theorem 1, if $p \in (1, \infty]$, solving DFO (15) is, in general, NP-hard. It turns out that even with $p = 1$, solving DFO (15) is also NP-hard.

Proposition 5 *For any $p \in [1, \infty]$, solving DFO (15) is, in general, NP-hard.*

Proof For any $p \in (1, \infty]$, DFO (15) reduces to the favorable optimization (9) if there is only one sample available for the empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ (i.e., $N = 1$). Thus, according to Theorem 1, solving DFO (15) is, in general, NP-hard. It remains to show that solving DFO (15) is also NP-hard when $p = 1$. Recall the NP-complete problem - the feasibility problem of a general binary program, which asks

Feasibility of a binary program. *Given an integer matrix $\boldsymbol{D} \in \mathbb{Z}^{\tau \times n}$, and integer vector $\boldsymbol{d} \in \mathbb{Z}^\tau$, is there a vector $\boldsymbol{x} \in \{-1, 1\}^n$ such that $\boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}$?*

Let us consider the following special case of DFO (15). Let set $\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}, -1 \le x_i \le 1, \forall i \in [n]\}$ and suppose $\theta = 0$, $b_k(\boldsymbol{x}) = 1$ for each $k \in [K]$, $\boldsymbol{\zeta}^i = \boldsymbol{e}_i$ for each $i \in [N]$, $N = n$, $K = 2n$, $m = n$, and

$$\boldsymbol{a}_k(\boldsymbol{x}) = \begin{cases} x_k \boldsymbol{e}_k, & k \le n, \\ -x_k \boldsymbol{e}_k, & n+1 \le k \le 2n. \end{cases}$$

Then, the inner minimum $\min_{k \in [K]} {\boldsymbol{\zeta}^i}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x})$ reduces to

$$\min_{k \in [K]} {\boldsymbol{\zeta}^i}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) = \min\{1, 1 - x_i, 1 + x_i\} = \min\{1, 1 - |x_i|\} = 1 - |x_i|.$$

Thus, DFO (15) can be written as

$$v^* = \min_{\substack{\boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}, \\ \boldsymbol{x} \in [-1,1]^n}} \frac{1}{n} \sum_{i \in [n]} [1 - |x_i|]. \tag{16}$$

We observe that the optimal value of DFO (16) $v^* = 0$ if and only if $|x_i| = 1$ for all $i \in [n]$, i.e., if and only if there exists a binary feasible solution $\boldsymbol{x} \in \{-1, 1\}^n$ such that $\boldsymbol{D}\boldsymbol{x} \le \boldsymbol{d}$. Thus, solving problem (16) is NP-hard, and so is DFO (15). $\qquad\square$

Albeit being NP-hard, when $p \in \{1, \infty\}$, next theorem provides MICP-R formulation for DFO (15).

Theorem 5 *Suppose $\mathcal{X} \subseteq [\boldsymbol{l}, \boldsymbol{u}]$, let $\widehat{l}_k^b = \sum_{j \in [n]} \min\{\widehat{B}_{kj} l_j, \widehat{B}_{kj} u_j\} + \widehat{b}_k$ and $\widehat{u}_k^b = \sum_{j \in [n]} \max\{\widehat{B}_{kj} l_j, \widehat{B}_{kj} u_j\} + \widehat{b}_k$ such that $b_k(\boldsymbol{x}) \in [\widehat{l}_k^b, \widehat{u}_k^b]$ for each $k \in [K]$, and let $\widehat{l}_{ki}^a = \sum_{j \in [n]} \min\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ and $\widehat{u}_{ki}^a = \sum_{j \in [n]} \max\{\widehat{A}_{kij} l_j, \widehat{A}_{kij} u_j\} + \widehat{a}_{ki}$ for each $i \in [m]$ such that $\boldsymbol{a}_k(\boldsymbol{x}) \in [\widehat{l}_k^a, \widehat{u}_k^a]$ for each $k \in [K]$. When $p \in \{1, \infty\}$, DFO (15) is MICP-R.*

Proof We first introduce binary variables $\boldsymbol{\lambda}$ to reformulate the inner minimum, that is,

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\lambda}} \quad \frac{1}{N} \sum_{i \in [N]} \sum_{k \in [K]} \lambda_{ki} \left[{\boldsymbol{\zeta}^i}^\top \boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) - \theta \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*} \right],$$

$$\text{s.t.} \quad \sum_{k \in [K]} \lambda_{ki} = 1, \forall i \in [N],$$

$$\lambda_{ki} \in \{0, 1\}, \forall k \in [K], i \in [N].$$

Since set $\mathcal{X}$ is compact, we can apply McCormick inequalities [26] to linearize the terms $\{\lambda_{ki} \boldsymbol{a}_k(\boldsymbol{x})\}_{i \in [N], k \in [K]}$ and $\{\lambda_{ki} b_k(\boldsymbol{x})\}_{i \in [N], k \in [K]}$. It remains to provide the MICP-R formulation for the term $\{\lambda_{ki} \|\boldsymbol{a}_k(\boldsymbol{x})\|_{p^*}\}_{i \in [N], k \in [K]}$. We split the discussions into two parts.

(i) When $p = \infty$, i.e., the dual norm is L_1, the term $\{\lambda_{ki} \|\boldsymbol{a}_k(\boldsymbol{x})\|_1\}$ can be linearized by applying McCormick inequalities twice for each $i \in [N]$ and $k \in [K]$. In this case, DFO (15) is equivalent to the following MICP:

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\lambda}, \boldsymbol{s}^a, \boldsymbol{s}^b, \boldsymbol{\eta}} \quad \frac{1}{N} \sum_{i \in [N]} \sum_{k \in [K]} \sum_{j \in [m]} \zeta_{ij} s_{kij}^a + \frac{1}{N} \sum_{i \in [N]} \sum_{k \in [K]} s_{ki}^b - \frac{\theta}{N} \sum_{i \in [N]} \sum_{k \in [K]} \eta_{ki},$$

$$\text{s.t.} \quad \sum_{k \in [K]} \lambda_{ki} = 1, \forall i \in [N],$$

$$\left(s_{kij}^a, \lambda_{ki}, a_{kj}(\boldsymbol{x}) \right) \in \mathcal{MI}(0, 1, \widehat{l}_{kj}^a, \widehat{u}_{kj}^a), \forall i \in [N], j \in [m], k \in [K],$$

$$\left(s_{ki}^b, \lambda_{ki}, b_k(\boldsymbol{x}) \right) \in \mathcal{MI}(0, 1, \widehat{l}_k^b, \widehat{u}_k^b), \forall i \in [N], k \in [K],$$

$$\eta_{ki} \ge s_{kij}^a, \forall i \in [N], j \in [m], k \in [K],$$

$$\eta_{ki} \ge -s_{kij}^a, \forall i \in [N], j \in [m], k \in [K].$$

(ii) When $p = 1$, i.e., the dual norm is L_∞, we can apply disjunctive programming [6] to the terms $\{\lambda_{ki}\|\boldsymbol{a}_k(\boldsymbol{x})\|_\infty\}$ and then apply McCormick inequalities to linearize $\{\lambda_{ki}a_{kj}(\boldsymbol{x})\}_{i\in[N],k\in[K],j\in[m]}$. In this case, DFO (15) is equivalent to the following MICP:

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\lambda},\boldsymbol{s}^a,\boldsymbol{s}^b,\boldsymbol{\eta}} \quad \frac{1}{N}\sum_{i\in[N]}\sum_{k\in[K]}\sum_{j\in[m]}\zeta_{ij}s^a_{kij} + \frac{1}{N}\sum_{i\in[N]}\sum_{k\in[K]}s^b_{ki} - \frac{\theta}{N}\sum_{i\in[N]}\sum_{k\in[K]}\eta_{ki},$$

$$\text{s.t.} \quad \sum_{k\in[K]}\lambda_{ki} = 1, \forall i\in[N],$$

$$\left(s^a_{kij}, \lambda_{ki}, a_{kj}(\boldsymbol{x})\right) \in \mathcal{MI}(0,1,\widehat{l}^a_{kj},\widehat{u}^a_{kj}), \forall i\in[N], j\in[m], k\in[K],$$

$$\left(s^b_{ki}, \lambda_{ki}, b_k(\boldsymbol{x})\right) \in \mathcal{MI}(0,1,\widehat{l}^b_k,\widehat{u}^b_k), \forall i\in[N], k\in[K],$$

$$\eta_{ki} \geq s^a_{kij}, \forall i\in[N], j\in[m], k\in[K],$$

$$\eta_{ki} \geq -s^a_{kij}, \forall i\in[N], j\in[m], k\in[K].$$

Therefore, according to the result in Lemma 2, DFO (15) is MICP-R with $p\in\{1,\infty\}$. $\qquad\square$

Moreover, notice that sDFO (7) is a special case of DFO (15) with $N = 1$. According to Theorem 2, DFO (15) may not be MICP-R when $p\in(1,\infty)$.

Corollary 3 *When $p\in(1,\infty)$, DFO (15) may not be MICP-R.*

We remark that Corollary 3 shows the conditions under which DFO (15) may not be MICP-R.
Special Case II. Convex Piecewise Function. Following the same notation introduced in Section 3.2, i.e., function $Q(\boldsymbol{x},\boldsymbol{\xi})$ is the maximum of piecewise affine function, DFO (14) can be written as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}}\frac{1}{N}\sum_{i\in[N]}\left[\inf_{\boldsymbol{\xi}}\left\{\max_{k\in[K]}\boldsymbol{\xi}^\top\boldsymbol{a}_k(\boldsymbol{x}) + b_k(\boldsymbol{x}) : \|\boldsymbol{\xi}-\boldsymbol{\zeta}^i\|_p \leq \theta\right\}\right]. \tag{17}$$

In this special case, all the results in Section 3.2 can be naturally extended to DFO (17), as presented below. The proofs and formulations are omitted for brevity.

Corollary 4 *For DFO (17), the complexity and tractability results in Section 3.2 directly follow.*

4.2 DFO with the Ambiguity Set of Finite Support

Many robust statistics recovered by DFO (1) can be considered as DFO (1) with the finite-support ambiguity set, i.e., when the support $\mathcal{U} := \{\boldsymbol{\xi}^i\}_{i\in[N]}$ is finite (see more discussions in [22]). In this setting, we cannot obtain any nontrivial tractable results for DFO (1). Furthermore, since evaluating the best case in DFO (1) with a given decision is, in general, NP-hard (see Proposition 1), we instead focus on the MICP-R formulations of DFO (1) with finite support in this subsection. Notably, when the ambiguity set $\mathcal{P}$ is a polytope (i.e., $\mathcal{P} = \left\{\boldsymbol{p}\in\mathbb{R}^N_+ : \boldsymbol{Dp}\leq\boldsymbol{d},\mathbf{e}^\top\boldsymbol{p} = 1\right\}$ is a polytope with $\boldsymbol{D}\in\mathbb{R}^{\ell\times N}$ and $\boldsymbol{d}\in\mathbb{R}^\ell$), we show that DFO (1) is MICP-R by observing that the number of the extreme points of the polyhedral ambiguity set $\mathcal{P}$ is finite. More specifically, DFO (1) with polyhedral ambiguity set $\mathcal{P}$ can be written as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}}\min_{\boldsymbol{p}\geq\boldsymbol{0}}\left\{\sum_{i\in[N]}p_iQ(\boldsymbol{x},\boldsymbol{\xi}^i) : \boldsymbol{Dp}\leq\boldsymbol{d},\mathbf{e}^\top\boldsymbol{p} = 1\right\}. \tag{18}$$

Theorem 6 *Suppose that both set $\mathcal{X}$ and function $Q(\boldsymbol{x},\boldsymbol{\xi})$ are MICP-R. Then, the corresponding DFO (18) under a polyhedral ambiguity set $\mathcal{P} = \left\{\boldsymbol{p}\in\mathbb{R}^N_+ : \boldsymbol{Dp}\leq\boldsymbol{d},\mathbf{e}^\top\boldsymbol{p} = 1\right\}$, is MICP-R.*

Proof Since the polyhedral ambiguity set $\mathcal{P}$ is a polytope, we can enumerate all its extreme points, i.e., $\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^s \in \mathbb{R}_+^N$ are the total s vertices of $\mathcal{P}$. Then, DFO (18) is equivalent to

$$v^* = \min_{j \in [s]} \min_{\boldsymbol{x} \in \mathcal{X}} \sum_{i \in [N]} \gamma_i^j Q(\boldsymbol{x}, \boldsymbol{\xi}^i),$$

which is MICP-R, since both set $\mathcal{X}$ and function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ are MICP-R. $\qquad\square$

We remark that the proof of Theorem 6 relies on the enumeration of extreme points of the polyhedral ambiguity set $\mathcal{P}$, which can be computationally inefficient. Instead of enumerating all the extreme points of the polyhedral ambiguity set $\mathcal{P}$, there exists an alternative MICP-R formulation of the corresponding DFO (18), which is based on the KKT condition.

Theorem 7 *Suppose that $\boldsymbol{D} \in \mathbb{Z}^{\ell \times N}$ in the polyhedral ambiguity set $\mathcal{P}$ and $N \geq \ell + 1$, there exists an $\bar{L}$ such that the largest row encoding length of the matrix $\begin{bmatrix} \boldsymbol{d}^\top & 1 \\ \boldsymbol{D}^\top & \boldsymbol{e} \end{bmatrix}$ is $\bar{L}$. Then, under the same presumptions in Theorem 6, the corresponding DFO (18) can be written as:*

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{v}, t, \boldsymbol{\alpha}, \beta, \boldsymbol{z}, \bar{\boldsymbol{z}}, \boldsymbol{p}} \left\{ t : \begin{array}{c} \boldsymbol{\alpha}^\top \boldsymbol{d} + \beta \leq t, \boldsymbol{D}^\top \boldsymbol{\alpha} + \beta \boldsymbol{e} \leq \boldsymbol{v}, \\ 0 \leq \alpha_j \leq M_{j,1} z_j, \forall j \in [\ell], 0 \leq d_j - \boldsymbol{D}_{j \cdot} \boldsymbol{p} \leq M_{j,2}(1 - z_j), \forall j \in [\ell], \\ 0 \leq p_j \leq \bar{z}_j, \forall i \in [N], \boldsymbol{e}^\top \boldsymbol{p} = 1, \\ 0 \leq v_i - \boldsymbol{D}_{\cdot i}^\top \boldsymbol{\alpha} - \beta \leq \bar{M}_i(1 - \bar{z}_i), \forall i \in [N], \\ v_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \forall i \in [N], \boldsymbol{z} \in \{0,1\}^\ell, \bar{\boldsymbol{z}} \in \{0,1\}^N, \boldsymbol{x} \in \mathcal{X} \end{array} \right\}, \quad (19)$$

where $M_{j,1}$ and $M_{j,2}$ are valid upper bounds for α_j and $d_j - \boldsymbol{D}_{j \cdot} \boldsymbol{p}$, respectively, for each $j \in [\ell]$, and $\bar{M}_i$ is the valid upper bound for $v_i - \boldsymbol{D}_{\cdot i}^\top \boldsymbol{\alpha} - \beta$ for each $i \in [N]$, i.e., letting $U_{i,1} = \max_{\boldsymbol{x} \in \mathcal{X}} Q(\boldsymbol{x}, \boldsymbol{\xi}^i)$ for each $i \in [N]$, and the corresponding big-M coefficients in DFO (19) can be found as:

$$M_{j,1} \geq U_{i,1}(\ell + 1)2^{\bar{L}}, \forall j \in [\ell], M_{j,2} \geq d_j + \sum_{i \in [N]} |D_{ji}|, \forall j \in [\ell],$$

$$\bar{M}_i \geq U_{i,1} \left(\sum_{j \in [\ell]} |D_{ji}| + 1 \right) (\ell + 1)2^{\bar{L}} + \min_{i \in [N]} U_{i,1}(\ell + 1)2^{\bar{L}}, \forall i \in [N].$$

Proof We split the proof into four steps.
Step I. Introducing the slack variables $\boldsymbol{v}$ and t for DFO (18), we have

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{v}, t} \left\{ t : \min_{\boldsymbol{p} \geq \boldsymbol{0}} \left\{ \sum_{i \in [N]} p_i v_i : \boldsymbol{D} \boldsymbol{p} \leq \boldsymbol{d}, \boldsymbol{e}^\top \boldsymbol{p} = 1 \right\} \leq t, v_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \forall i \in [N] \right\}.$$

For the innermost minimization of the problem over $\boldsymbol{p}$, we can use the complementary slackness to obtain the equivalent reformulation with big-M coefficients, that is,

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{v}, t, \boldsymbol{\alpha}, \beta, \boldsymbol{z}, \boldsymbol{p}} \left\{ t : \begin{array}{c} \boldsymbol{\alpha}^\top \boldsymbol{d} + \beta \leq t, \boldsymbol{D}^\top \boldsymbol{\alpha} + \beta \boldsymbol{e} \leq \boldsymbol{v}, \\ 0 \leq \alpha_j \leq M_{j,1} z_j, \forall j \in [\ell], 0 \leq d_j - \boldsymbol{D}_{j \cdot} \boldsymbol{p} \leq M_{j,2}(1 - z_j), \forall j \in [\ell], \\ 0 \leq p_j \leq \bar{z}_j, \forall i \in [N], \boldsymbol{e}^\top \boldsymbol{p} = 1, \\ 0 \leq v_i - \boldsymbol{D}_{\cdot i}^\top \boldsymbol{\alpha} - \beta \leq \bar{M}_i(1 - \bar{z}_i), \forall i \in [N], \\ v_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \forall i \in [N], \boldsymbol{z} \in \{0,1\}^\ell, \bar{\boldsymbol{z}} \in \{0,1\}^N \end{array} \right\}.$$

Step II. We then demonstrate that set

$$\bar{\Theta} = \left\{ (\boldsymbol{\alpha}, \beta) : \begin{bmatrix} \boldsymbol{d}^\top & 1 \\ \boldsymbol{D}^\top & \boldsymbol{e} \\ -\boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \beta \end{bmatrix} \leq \begin{bmatrix} t \\ \boldsymbol{v} \\ \boldsymbol{0} \end{bmatrix} \right\}$$

is nonempty and contains no line. It is evident that the point $(\boldsymbol{\alpha} = \mathbf{0}, \beta = \min_{i \in [N]} v_i \in \bar{\Theta}$, indicating the existence of at least one point in set $\bar{\Theta}$. Hence, set $\bar{\Theta}$ is nonempty. Suppose there exists a vector $\boldsymbol{d}$ such that for every point $(\bar{\boldsymbol{\alpha}}, \bar{\beta})$ within set $\bar{\Theta}$, $(\bar{\boldsymbol{\alpha}}, \bar{\beta}) + \lambda \boldsymbol{d}$ is also in set $\bar{\Theta}$ for all $\lambda \in \mathbb{R}$. That is,

$$
\begin{bmatrix} \boldsymbol{d}^\top & 1 \\ \boldsymbol{D}^\top & \boldsymbol{e} \\ -\boldsymbol{I} & \mathbf{0} \end{bmatrix} \left(\begin{bmatrix} \bar{\boldsymbol{\alpha}} \\ \bar{\beta} \end{bmatrix} + \lambda \boldsymbol{d} \right) \leq \begin{bmatrix} t \\ \boldsymbol{v} \\ \mathbf{0} \end{bmatrix}, \forall \lambda \in \mathbb{R}.
$$

Therefore, we have $\boldsymbol{d} = \mathbf{0}$. Hence, set $\bar{\Theta}$ contains no line, which implies that set $\bar{\Theta}$ contains an extreme point.
Step III. We now proceed to demonstrate that each extreme point in $\bar{\Theta}$ is bounded. Since there exists an $\bar{L}$ such that the row encoding length of every subset of $(\ell + 1)$ rows from the matrix $\begin{bmatrix} \boldsymbol{d}^\top & 1 \\ \boldsymbol{D}^\top & \boldsymbol{e} \end{bmatrix}$ is $\bar{L}$ (see the details in [18]), we have

$$
\mathrm{ext} \left\{ (\boldsymbol{\alpha}, \beta) \colon \begin{bmatrix} \boldsymbol{d}^\top & 1 \\ \boldsymbol{D}^\top & \boldsymbol{e} \\ -\boldsymbol{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \beta \end{bmatrix} \leq \begin{bmatrix} t \\ \boldsymbol{v} \\ \mathbf{0} \end{bmatrix} \right\} \subseteq \left\{ (\boldsymbol{\alpha}, \beta) \colon \alpha_j \leq U_{i,1}(\ell+1)2^{\bar{L}}, \forall j \in [\ell], |\beta| \leq \min_{i \in [N]} U_{i,1}(\ell+1)2^{\bar{L}} \right\}.
$$

Step IV. Next, we determine the values of big-M coefficients. By substituting the upper bounds of $\boldsymbol{\alpha}$ and β, we can determine the explicit values for the big-M coefficients in DFO (19). Specifically, the coefficients can be found as follows:

$$
\begin{aligned}
& M_{j,1} \geq U_{i,1}(\ell+1)2^{\bar{L}} \geq \alpha_j, \forall j \in [\ell], \\
& M_{j,2} \geq d_j + \sum_{i \in [N]} |D_{ji}| \geq d_j - \boldsymbol{D}_{j \cdot} \boldsymbol{p}, \forall j \in [\ell], \\
& \bar{M}_i \geq U_{i,1} \left(\sum_{j \in [\ell]} |D_{ji}| + 1 \right) (\ell+1)2^{\bar{L}} + \min_{i \in [N]} U_{i,1}(\ell+1)2^{\bar{L}} \geq v_i - \boldsymbol{D}_{\cdot i}^\top \boldsymbol{\alpha} - \beta, \forall i \in [N].
\end{aligned}
$$

This completes the proof. $\qquad \square$

We remark that based on the results in Theorem 7, following the same presumptions in Theorem 6 with nonnegative matrix $\boldsymbol{D}$ and positive vector $\boldsymbol{d}$, we can obtain similar results. Below is an example.

Corollary 5 *Under the same presumptions as that in Theorem 6 and the assumptions that $\boldsymbol{D} \geq \mathbf{0}$ and $\boldsymbol{d} > \mathbf{0}$ in the polyhedral ambiguity set $\mathcal{P}$, the corresponding big-M coefficients in DFO* (19) *can be found as:*

$$
\begin{aligned}
& M_{j,1} \geq \frac{1}{d_j} \left[\max_{i \in [N]} U_{i,1} - \min_{i \in [N]} L_{i,1} \right], \forall j \in [\ell], \\
& M_{j,2} \geq d_j, \forall j \in [\ell], \bar{M}_i \geq U_{i,1} - \min_{i \in [N]} L_{i,1}, \forall i \in [N],
\end{aligned}
$$

where $M_{j,1}$ and $M_{j,2}$ are valid upper bounds for α_j and $d_j - \boldsymbol{D}_{j \cdot} \boldsymbol{p}$, respectively, for each $j \in [\ell]$, and $\bar{M}_i$ is the valid upper bound for $v_i - \boldsymbol{D}_{\cdot i}^\top \boldsymbol{\alpha} - \beta$ for each $i \in [N]$, i.e., letting $L_{i,1} = \min_{\boldsymbol{x} \in \mathcal{X}} Q(\boldsymbol{x}, \boldsymbol{\xi}^i)$ for each $i \in [N]$, $U_{i,1} = \max_{\boldsymbol{x} \in \mathcal{X}} Q(\boldsymbol{x}, \boldsymbol{\xi}^i)$ for each $i \in [N]$.

Proof Under the conditions that $\boldsymbol{D} \geq \mathbf{0}$, $\boldsymbol{d} > \mathbf{0}$, and set $\mathcal{X}$ is compact, we know t in DFO (19) is bounded by $t \in [\min_{i \in [N]} L_{i,1}, \max_{i \in [N]} U_{i,1}]$, and β in DFO (19) is bounded by $\beta \in [\min_{i \in [N]} L_{i,1}, \max_{i \in [N]} U_{i,1}]$. Then, we obtain explicit values for the valid upper bounds $\{M_{j,1}\}_{j \in [\ell]}$, $\{M_{j,2}\}_{j \in [\ell]}$, $\{\bar{M}_i\}_{i \in [N]}$. This completes the proof. $\qquad \square$

A Special Case: MICP-R Formulation for Interval Polyhedral Ambiguity Set. Building on Theorem 6 and Theorem 7, we provide a compact MICP-R formulation for a particular type of ambiguity set, namely the interval polyhedral ambiguity set, which is defined as $\mathcal{P}_I = \{\boldsymbol{p} = \boldsymbol{p}^0 + \boldsymbol{\psi} \in \mathbb{R}_+^N : \boldsymbol{l} \leq \boldsymbol{\psi} \leq \boldsymbol{u}, \boldsymbol{e}^\top \boldsymbol{\psi} = 0\}$. Here, we let $\boldsymbol{p}^0$ denote the nominal probability vector with $\boldsymbol{p}^0 \geq \boldsymbol{0}$ and $\sum_{i \in [N]} p_i^0 = 1$, the lower bound vector $\boldsymbol{l} \geq -\boldsymbol{p}^0$ and the bounds $l_i = \bar{l}_i/q, u_i = \bar{u}_i/q$ with q being a positive integer and $\bar{l}_i, \bar{u}_i$ being integers for each $i \in [N]$.

Corollary 6 *Suppose both set $\mathcal{X}$ and function $Q(\boldsymbol{x}, \boldsymbol{\xi})$ are MICP-R. Then under the interval polyhedral ambiguity set $\mathcal{P}_I$, the optimal value of the corresponding DFO (1) is $v^* = \min_{j \in [N], \tau \in [\bar{l}_j, \bar{u}_j]} v_{j\tau}^*$ and for each $j \in [N]$ and $\tau \in \{\bar{l}_j, \bar{l}_j + 1, \cdots, \bar{u}_j\}$, the value $v_{j\tau}^*$ can be computed by solving the following MICP-R formulation:*

$$v_{j\tau}^* = \min_{\substack{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\eta}, \boldsymbol{\nu}, \\ \boldsymbol{z}^j \in \{0,1\}^N}} \sum_{i \in [N] \setminus \{j\}} \left[p_i^0 + \bar{l}_i/q\right] \nu_i + \sum_{i \in [N] \setminus \{j\}} (\bar{u}_i/q - \bar{l}_i/q)\eta_i^j + \left(p_j^0 + \tau/q\right) \nu_j,$$

$$\text{s.t.} \quad \nu_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \left(\eta_i^j, z_i^j, \nu_i\right) \in \mathcal{MI}\left(0, 1, L_i, U_i\right), \forall i \in [N], j \in [N],$$

$$- \sum_{i \in [N] \setminus \{j\}} \left(\bar{l}_i + \left(\bar{u}_i - \bar{l}_i\right)z_i^j\right) = \tau,$$

where for each $i \in [N]$, L_i and U_i are the valid lower and upper bounds of the function $Q(\boldsymbol{x}, \boldsymbol{\xi}^i)$, respectively.

Proof DFO (1) is equivalent to

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{\psi}} \left\{ \sum_{i \in [N]} \left[(p_i^0 + \psi_i)Q(\boldsymbol{x}, \boldsymbol{\xi}^i) : \sum_{i \in [N]} \psi_i = 0, \bar{l}_i/q \leq \psi_i \leq \bar{u}_i/q, \forall i \in [N]\right] \right\}. \tag{20}$$

According to the extreme point characterization of the ambiguity set $\mathcal{P}_I$, for any extreme point $\widehat{\boldsymbol{\psi}}$, it has at least $N-1$ components taking values from $\boldsymbol{l}$ or $\boldsymbol{u}$ and one component corresponding to equality constraint. Let us assume that component $j \in [N]$ corresponds to the equality constraint. Accordingly, we can define the binary variable $z_i^j \in \{0,1\}$ for each $i \in [N]$ and $\widehat{\psi}_i = l_i + (u_i - l_i)z_i^j$ for each $i \in [N] \setminus \{j\}$. Since we have $\sum_{i \in [N]} \widehat{\psi}_i = 0$, thus $\widehat{\psi}_j = -\sum_{i \in [N] \setminus \{j\}}(l_i + (u_i - l_i)z_i^j)$. Plugging the extreme point representation into set $\mathcal{P}_I$, we have

$$\mathcal{P}_I = \text{conv}\left[\bigcup_{j \in [N]} \left\{ \boldsymbol{p} = \boldsymbol{p}^0 + \widehat{\boldsymbol{\psi}} \in \mathbb{R}_+^N : \begin{array}{l} \widehat{\psi}_i = l_i + (u_i - l_i)z_i^j, \forall i \in [N] \setminus \{j\}, \\ \widehat{\psi}_j = -\sum_{i \in [N], i \setminus \{j\}}(l_i + (u_i - l_i)z_i^j) \in [l_j, u_j] \end{array} \right\} \right].$$

Plugging the representation of set $\mathcal{P}_I$, DFO (20) is equivalent to $v^* = \min_{j \in [N]} v_j^*$ and for each $j \in [N]$,

$$v_j^* = \min_{\substack{\boldsymbol{x} \in \mathcal{X}, \\ \boldsymbol{z}^j \in \{0,1\}^N}} \sum_{i \in [N] \setminus \{j\}} \left[p_i^0 + l_i + (u_i - l_i)z_i^j\right] Q(\boldsymbol{x}, \boldsymbol{\xi}^i) + \left(p_j^0 - \sum_{i \in [N], i \setminus \{j\}} \left(l_i + (u_i - l_i)z_i^j\right)\right) Q(\boldsymbol{x}, \boldsymbol{\xi}^j),$$

$$\text{s.t.} \quad l_j \leq - \sum_{i \in [N] \setminus \{j\}} \left(l_i + (u_i - l_i)z_i^j\right) \leq u_j.$$

Since $l_i = \bar{l}_i/q, u_i = \bar{u}_i/q$ for each $i \in [N]$, then for each $j \in [N]$, the expression $-\sum_{i \in [N] \setminus \{j\}}(l_i + (u_i - l_i)z_i^j)$ can take values from $\{\tau/q\}_{\tau \in [\bar{l}_j, \bar{u}_j]}$ and τ is an integer. This fact allows us to simplify $v_j^* = \min_{\tau \in [\bar{l}_j, \bar{u}_j]} v_{j\tau}^*$, where

$$v_{j\tau}^* = \min_{\substack{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\nu}, \\ \boldsymbol{z}^j \in \{0,1\}^N}} \sum_{i \in [N] \setminus \{j\}} \left[p_i^0 + \frac{\bar{l}_i}{q} + \frac{1}{q}(\bar{u}_i - \bar{l}_i)z_i^j\right] \nu_i + \left(p_j^0 + \frac{\tau}{q}\right) \nu_j,$$

$$\text{s.t.} \quad -\sum_{i\in[N]\setminus\{j\}}\left(\bar{l}_i + (\bar{u}_i - \bar{l}_i)z_i^j\right) = \tau, \nu_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \forall i \in [N],$$

for each $j \in [N]$ and $\tau \in [\bar{l}_j, \bar{u}_j]$. Since set $\mathcal{X}$ is compact, we can apply the McCormick inequalities [26] to linearize the bilinear terms $\{z_i^j \nu_i\}_{i\in[N], j\in[N]}$, this completes the proof. $\qquad\square$

We remark that as a direct application of Corollary 6, when $l_i = l, u_i = u$ for each $i \in [N]$, the MICP-R formulation of Corollary 6 can be further simplified.

Corollary 7 *Suppose that the premises of Corollary 6 hold and $p_i^0 = 1/N, l_i = -1/N, u_i = u$ for each $i \in [N]$. Then the optimal value of the corresponding DFO (1) is $v^* = \min_{j\in[N]} v_j^*$ and for each $j \in [N]$, the value v_j^* can be computed via the following MICP-R formulation:*

$$v_j^* = \min_{\substack{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{\eta}, \boldsymbol{\nu}, \\ \boldsymbol{z}^j \in \{0,1\}^N}} \sum_{i\in[N]\setminus\{j\}} (u + 1/N)\nu_i + (1 - \lfloor\kappa\rfloor/\kappa)\nu_j,$$

$$\text{s.t.} \quad \nu_i \geq Q(\boldsymbol{x}, \boldsymbol{\xi}^i), \left(\eta_i^j, z_i^j, \nu_i\right) \in \mathcal{MI}\left(0, 1, L_i, U_i\right), \forall i \in [N], j \in [N],$$

$$\sum_{i\in[N]\setminus\{j\}} z_i^j = \lfloor\kappa\rfloor,$$

where $\kappa = N/(uN + 1)$ and for each $i \in [N]$, L_i and U_i are the lower and upper bounds of the function $Q(\boldsymbol{x}, \boldsymbol{\xi}^i)$, respectively.

The result in Corollary 7 will be demonstrated in the numerical study section. It is important to note, however, that the MICP-R result in Theorem 6 does not hold when the ambiguity set with finite support is not polyhedral.

Proposition 6 *Suppose that the ambiguity set is $\mathcal{P} = \{\boldsymbol{p} : \|\boldsymbol{p} - \boldsymbol{p}^0\|_2 \leq \theta, \sum_{i\in[N]} p_i = 1, \boldsymbol{p} \geq \boldsymbol{0}\}$, where $\boldsymbol{p}^0 = \boldsymbol{e}/N$ denotes the nominal probability. When $0 < \theta \leq \sqrt{1/(N(N-1))}$, DFO (1) may not be MICP-R.*

Proof Let us consider a simple function $Q(\boldsymbol{x}, \boldsymbol{\xi}^i) = x_i$ for each $i \in [N]$ and $\boldsymbol{c} = \boldsymbol{0}$. Then, DFO (1) is equivalent to

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \min_{\boldsymbol{p}\geq\boldsymbol{0}} \left[\sum_{i\in[N]} p_i x_i : \sum_{i\in[N]} p_i = 1, \left\|\boldsymbol{p} - \frac{1}{N}\boldsymbol{e}\right\|_2 \leq \theta\right]. \tag{21a}$$

Let us focus on simplifying the inner minimization of DFO (21a) and define $\boldsymbol{y} = \boldsymbol{x} - (\boldsymbol{x}^\top \boldsymbol{e})\boldsymbol{e}/N$. Then, by the definition, we must have $\sum_{i\in[N]} y_i = 0$ and

$$\sum_{i\in[N]} p_i y_i = \sum_{i\in[N]} p_i x_i - \frac{1}{N}\sum_{i\in[N]} (x_i - 1).$$

DFO (21a) is equivalent to

$$v^* = \min_{\substack{\boldsymbol{x}\in\mathcal{X}, \\ \boldsymbol{y}=\boldsymbol{x}-(\boldsymbol{x}^\top\boldsymbol{e})\boldsymbol{e}/N}} \frac{1}{N}\sum_{i\in[N]} (x_i - 1) + \min_{\boldsymbol{p}\geq\boldsymbol{0}} \left\{\sum_{i\in[N]} p_i y_i : \sum_{i\in[N]} p_i = 1, \left\|\boldsymbol{p} - \frac{1}{N}\boldsymbol{e}\right\|_2 \leq \theta\right\}. \tag{21b}$$

Letting $\widehat{\boldsymbol{p}} = \boldsymbol{p} - \boldsymbol{e}/N$, DFO (21b) is simplified as

$$v^* = \min_{\boldsymbol{y}} \min_{\widehat{\boldsymbol{p}}\geq-\boldsymbol{e}/N} \left\{\sum_{i\in[N]} \widehat{p}_i y_i : \sum_{i\in[N]} \widehat{p}_i = 0, \|\widehat{\boldsymbol{p}}\|_2 \leq \theta, \sum_{i\in[N]} y_i = 0\right\}. \tag{21c}$$

According to the Hölder's inequality, the inner minimization of DFO (21c) can be lower bounded by

$$\min_{\widehat{\boldsymbol{p}} \geq -\boldsymbol{e}/N} \left\{ \sum_{i \in [N]} \widehat{p}_i y_i \colon \sum_{i \in [N]} \widehat{p}_i = 0, \|\widehat{\boldsymbol{p}}\|_2 \leq \theta \right\} \geq -\theta \|\boldsymbol{y}\|_2.$$

In fact, the above equality can be achieved by the solution $\widehat{p}_i^* = -\theta y_i / \|\boldsymbol{y}\|_2$ for all $i \in [N]$. Since $\sum_{i \in [N]} \widehat{p}_i^* = 0$ and $\|\widehat{\boldsymbol{p}}^*\|_2 = \theta$, it suffices to show that

$$-|\widehat{p}_i^*| = -\theta |y_i| / \|\boldsymbol{y}\|_2 \geq -\frac{1}{N}, \tag{21d}$$

for all $\sum_{i \in [N]} \widehat{p}_i^* = 0$. That is, we need to show that

$$\max_{\sum_{i \in [N]} y_i = 0} \frac{|y_i|}{\|\boldsymbol{y}\|_2} \leq \frac{1}{N\theta}. \tag{21e}$$

Without loss of generality, suppose that $y_\ell \neq 0$. Letting $y_i' = y_i / y_\ell$ for each $i \in [N] \setminus \{\ell\}$, the condition $\sum_{i \in [N]} y_i = 0$ is equivalent to $\sum_{i \in [N]} y_i' = -1$. Thus,

$$\max_{\sum_{i \in [N]} y_i = 0} \frac{|y_i|}{\|\boldsymbol{y}\|_2} = \min_{\boldsymbol{y}'} \left\{ \sqrt{1 + \sum_{i \in [N] \setminus \ell} y_i'^2} \colon \sum_{i \in [N]} y_i' = -1 \right\} = \sqrt{1 + \left(\frac{1}{\sqrt{N-1}}\right)^2} = \sqrt{\frac{N}{N-1}}.$$

Hence, the inequality in (21e) must be satisfied since

$$\max_{\sum_{i \in [N]} y_i' = -1} \frac{1}{\sqrt{1 + \sum_{i \in [N] \setminus \ell} y_i'^2}} = \sqrt{\frac{N-1}{N}} \leq \frac{1}{N\theta}$$

and $0 < \theta \leq \sqrt{1/(N(N-1))}$. Therefore, plugging in $\boldsymbol{y} = \boldsymbol{x} - (\boldsymbol{x}^\top \boldsymbol{e})\boldsymbol{e}/N$, DFO (21c) is equivalent to

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{N} \sum_{i \in [N]} (x_i - 1) - \frac{\theta}{N} \left\| N\boldsymbol{x} - (\boldsymbol{x}^\top \boldsymbol{e})\boldsymbol{e} \right\|_2.$$

Using the fact that set $\mathcal{X}$ has a nonempty relative interior and following the similar proof as that of Lemma 2, we conclude that this DFO cannot be MICP-R. □

The findings in Proposition 6 reveal that even when DFO (1) is based on a finite-support ambiguity set, it may not always be MICP-R. This highlights the necessity of carefully selecting an appropriate ambiguity set for DFO problems, especially in the context of data-driven decision-making. Additionally, it is important to note that the insights gained in this subsection are not limited to static uncertainty; they can be extended to encompass decision-dependent uncertainty. This includes applications in two-stage stochastic programs with decision-dependent uncertainty, as explored in recent works such as [25, 40, 42, 43].

5 Numerical Study

To demonstrate the value of the MICP-R formulations, we consider the interval polyhedral ambiguity set $\mathcal{P}_I$ and apply the result in Corollary 7. All instances in this section are coded in Python 3.9 with calls to solver Gurobi (version 9.5.2 with default settings) on a personal PC with an Apple M1 Pro processor and 16GB of memory. We set the time limit of each instance to be 3600s.

In particular, we consider a two-stage resource allocation (TRA) problem (also studied by [13]), which consists of a set of facilities, denoted by $s \in [n]$, that can be used to meet the demand from the customer sites, denoted by $j \in [n_1]$. In the TRA problem, the first-stage decision is to distribute a single type of

commodity across these facilities. Once the allocation for each facility is determined, we then meet the demand in the second stage at the lowest possible cost. When the supply is insufficient, a large unit penalty (i.e., outsourcing) cost ρ will be incurred for unsatisfied demand. Conversely, surplus supply at any facility has to bear a unit holding cost, h. The TRA problem can be formulated as

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right], \tag{22a}$$

where for a realization $\boldsymbol{\xi}$,

$$Q(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{u} \geq \boldsymbol{0}, \boldsymbol{v} \geq \boldsymbol{0}} \left\{ \sum_{s \in [n]} \sum_{j \in [n_1]} q_{sj} y_{sj} + \rho \sum_{j \in [n_1]} u_j + h \sum_{s \in [n]} v_s : \begin{array}{l} \sum_{j \in [n_1]} y_{sj} + v_s = x_s, \forall s \in [n], \\ \sum_{s \in [n]} y_{sj} + u_j \geq \xi_j, \forall j \in [n_1] \end{array} \right\}. \tag{22b}$$

In the TRA (22), for each $s \in [n]$, the variable x_s denotes the supply allocated to facility s. For $s \in [n]$ and $j \in [n_1]$, variable y_{sj} represents the amount of the demand from the customer site j satisfied by facility s, with the associated transportation cost denoted by q_{sj}. The parameters $\tilde{\boldsymbol{\xi}}$ are random, where $\tilde{\xi}_j$ denotes the random demand of customer site $j \in [n_1]$.

In the numerical experiments, we solve TRA (22) under finite support by generating random instances with varying sample sizes N. All the random variables (i.e., the customer demands $\tilde{\boldsymbol{\xi}}$) are truncated to be nonnegative. For each instance, we assume that the transportation cost vector $\boldsymbol{q}$ components are i.i.d. truncated Gaussian with mean 1 and variance 0.2. The components of the customer demand $\tilde{\boldsymbol{\xi}}$ are i.i.d. truncated Gaussian random variables with means $\bar{d}/n_1$ and variances, $0.005 \times \bar{d}$ with $\bar{d} = 1000$. We also assume some outliers exist in the customer demand information, denoted by $\tilde{\boldsymbol{\xi}}^o$. We assume the components of random vector $\tilde{\boldsymbol{\xi}}^o$ are i.i.d. truncated Gaussian with mean $\bar{d}/n_1$ and variance $0.01 \times \bar{d}$. The observed demand vector follows the following distribution $0.95\tilde{\boldsymbol{\xi}} + 0.05\tilde{\boldsymbol{\xi}}^o$. We set the number of potential facilities $n = 8$, the number of customers $n_1 = 20$, the unit penalty cost $\rho = 10$, and the unit cost for holding inventory $h = 1$.

Experiment 1. Value of MICP-R from Corollary 7. In the numerical implementation, we use DFO (1) to reduce the effect of outliers in the original SAA problem (22a). Particularly, we consider the formulation in Corollary 7, that is,

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \inf_{\mathbb{P} \in \mathcal{P}_I} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right], \tag{23}$$

We set $\mu = 1/(N - N\varepsilon) - 1/N$ with $\varepsilon \in (0, 1)$ and $N\varepsilon$ being a rational number but not an integer in the interval polyhedral ambiguity set $\mathcal{P}_I$. According to Corollary 7, the proposed DFO (1) is still MICP-R. Alternatively, the two-stage program (22a) with the interval polyhedral ambiguity set $\mathcal{P}_I$ admits a naive bilinear formulation, which can be solved directly by Gurobi. Since we cannot solve the bilinear model to optimality within the time limit, we use GAP to denote its optimality gap as $\text{GAP}(\%) = (|\text{UB} - \text{LB}|)/|\text{LB}| \times 100$, where "UB" and "LB" denotes the best upper bound and the best lower bound found by Gurobi. We repeat the solution process 5 times and display the average performance result in Table 2. We find that the MICP-R formulation can improve the running time significantly, even for small-scale instances, which shows the effectiveness of exploring the MICP-R formulation.

Table 2 Comparisons Between DFO (1) in Corollary 7 and Its Bilinear Counterpart.

N	$\varepsilon = 0.16$			$\varepsilon = 0.18$		
	Bilinear		DFO (1) in Corollary 7	Bilinear		DFO (1) in Corollary 7
	Time (s)	GAP(%)	Time (s)	Time (s)	GAP(%)	Time (s)
45	3600	3.04	32.73	3600	4.39	46.94
55	3600	5.35	45.92	3600	6.24	57.93
65	3600	7.49	58.38	3600	8.93	72.96

Experiment 2. Value of MICP-R using Hurwicz Model. Based on Corollary 7, we can also provide MICP-R formulations under the Hurwicz criterion for the TRA problem (22). That is, we consider the following problem:

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \left\{ \bar{\lambda} \inf_{\mathbb{P} \in \mathcal{P}_I} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] + (1 - \bar{\lambda}) \sup_{\mathbb{P} \in \mathcal{P}_I} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right] \right\}, \tag{24}$$

where $\bar{\lambda} \in [0, 1]$ is a known level of optimism. Notice that when $\bar{\lambda} = 0$ in Hurwicz (24), we obtain the DRO formulation for TRA problem (22) as below

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \sup_{\mathbb{P} \in \mathcal{P}_I} \mathbb{E}_{\mathbb{P}} \left[Q(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) \right]. \tag{25}$$

In this experiment, we compare the solutions from DFO (23), DRO (25), and Hurwicz (24) via out-of-sample performances. Specifically, after solving the corresponding DFO (23), DRO (25), and Hurwicz (24), we generate additional 100 random testing instances to evaluate the solution performances, i.e., to assess the performance of the first-stage decision of each model. Training and test instances are generated in the same manner, i.e., the components of the customer demand $\tilde{\boldsymbol{\xi}}$ are i.i.d. truncated Gaussian with mean $\bar{d}/n_1$ and variance $0.005 \times \bar{d}$, $\bar{d} = 1000$. We record all the $50\%, 60\%, 70\%, 80\%, 90\%$ quantiles of the second-stage values, respectively. We then report each quantile's 95% asymptotic confidence interval (C.I.) among these 100 testing instances. We set $\bar{\lambda} = \{0.2, 0.4, 0.6, 0.8\}$ in Hurwicz (24) and consider the training sample size $N = 45$ with $\epsilon = 0.16$. The results are shown in Table 3 and Figure 1. In this case, DFO (23) is consistently better than other methods when comparing $50\%, 60\%, 70\%$, and 80% quantiles of the second-stage values. Hurwicz (24) with $\bar{\lambda} = 0.8$ performs better than other methods when comparing 90% quantile of the second-stage values. By carefully choosing an optimism level $\bar{\lambda}$, it is seen that the Hurwicz (24) can reduce the conservatism (in this case, we can choose $\bar{\lambda} = 0.8$).

Table 3 Quantile Comparisons among DFO, DRO, and Hurwicz Models in Experiment 2.

Model	Quantile				
	50%	60%	70%	80%	90%
DRO (25)	[1826.63, 1826.95]	[1828.39, 1828.73]	[1830.33, 1830.66]	[1832.75, 1833.14]	[1835.82,1836.23]
Hurwicz (24) with $\lambda = 0.2$	[1771.41, 1771.74]	[1773.23, 1773.58]	[1775.22, 1775.56]	[1777.67, 1778.07]	[1780.75,1781.16]
Hurwicz (24) with $\lambda = 0.4$	[1771.21, 1771.55]	[1773.07, 1773.42]	[1775.07, 1775.40]	[1777.53, 1777.93]	[1780.63,1781.03]
Hurwicz (24) with $\lambda = 0.6$	[1766.25, 1766.60]	[1768.11, 1768.47]	[1770.13, 1770.47]	[1772.61, 1773.00]	[1775.70,1776.11]
Hurwicz (24) with $\lambda = 0.8$	[1763.06, 1763.41]	[1764.94, 1765.31]	[1767.02, 1767.37]	[1769.49, 1769.90]	[1772.86,1773.43]
DFO (23)	[1757.02, 1757.37]	[1758.95, 1759.34]	[1761.22, 1761.67]	[1768.88, 1769.01]	[1774.94,1776.10]

Experiment 3. Model Comparisons When the Testing Distribution is Different From the Training One. We follow the same procedure described in Experiment 2, i.e., we record all the $50\%, 60\%, 70\%, 80\%, 90\%$ quantiles in the second-stage scenarios for each method (e.g., DFO, DRO, and Hurwicz models) in each testing instance, respectively, and report the average of each quantile among these 100 randomly generated testing instances. The testing and training setting are the same as that of Experiment 2, except that we assume that the components of the customer demand $\tilde{\boldsymbol{\xi}}$ are i.i.d. truncated Gaussian with mean $1000/n_1$ and variances 0.01×1000. We repeat the procedure in Experiment 2 to solve Hurwicz (24) with $\bar{\lambda} = \{0, 0.2, 0.4, 0.6, 0.8\}$ to better capture the influence of the optimism level. The results are shown in Table 4 and Figure 2. As anticipated, Hurwicz (24) can alleviate conservatism and improve out-of-sample performance when the testing distribution has perturbations by selecting the level of optimism $\bar{\lambda}$ (in this case, we can choose $\bar{\lambda} = 0.2$ or 0.4).

6 Conclusion

This paper provided sufficient and necessary conditions where DFO can be tractable or intractable. Even though DFO is NP-hard to solve in general, we demonstrated that many DFO problems can be mixed-integer convex programming representable, which can be solved by off-the-shelf solvers. We numerically
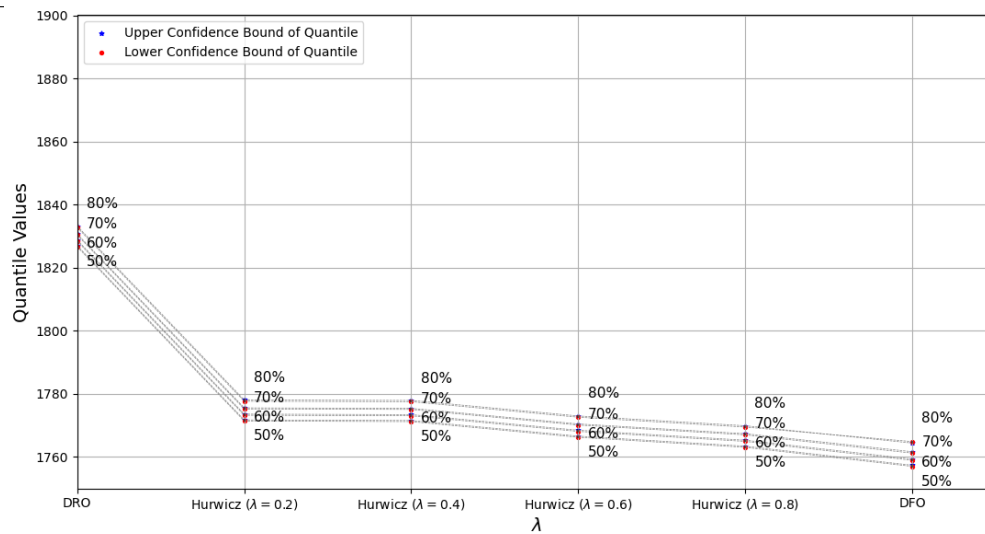
Fig. 1 Illustration of Quantile Comparisons among DFO, DRO, and Hurwicz Models in Experiment 2.

Table 4 Quantile Comparisons among DFO, DRO, and Hurwicz Models in Experiment 3.

| Model	Quantile				
	50%	60%	70%	80%	90%
DRO (25)	[1827.91, 1828.75]	[1831.56, 1832.45]	[1835.51, 1836.45]	[1840.37, 1841.31]	[1846.95,1847.93]
Hurwicz (24) with $\lambda = 0.2$	[1772.65, 1773.51]	[1776.35, 1777.27]	[1780.36, 1781.34]	[1785.69, 1786.85]	[1815.39,1833.73]
Hurwicz (24) with $\lambda = 0.4$	[1772.59, 1773.47]	[1776.38, 1777.29]	[1780.49, 1781.47]	[1785.76, 1786.92]	[1815.58,1833.99]
Hurwicz (24) with $\lambda = 0.6$	[1767.90, 1768.81]	[1771.77, 1772.71]	[1776.25, 1777.55]	[1785.61, 1792.29]	[1900.45,1927.38]
Hurwicz (24) with $\lambda = 0.8$	[1764.88, 1765.81]	[1768.92, 1769.95]	[1773.96, 1778.12]	[1812.27, 1828.57]	[1972.28,2000.03]
DFO (23)	[1759.27, 1760.42]	[1765.92, 1770.52]	[1804.41, 1824.36]	[1931.37, 1956.21]	[2111.82,2139.61]
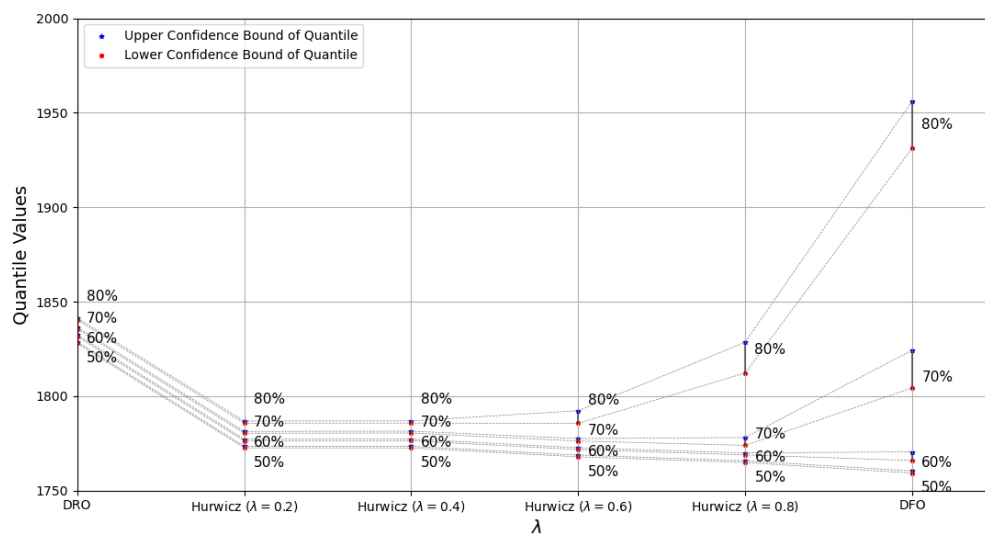
Fig. 2 Illustration of Quantile Comparisons in Experiment 3

demonstrated the effectiveness of using MICP-R formulations. One future direction is to extend the results to the two-stage stochastic programs under decision-dependent uncertainty. It is also interesting to investigate the theoretical advantages of the Hurwicz criterion.

Acknowledgments

References

1. Tobias Achterberg, Robert E Bixby, Zonghao Gu, Edward Rothberg, and Dieter Weninger. Presolve reductions in mixed integer programming. *INFORMS Journal on Computing*, 32(2):473–506, 2020.
2. Tobias Achterberg and Roland Wunderling. Mixed integer programming: Analyzing 12 years of progress. In *Facets of combinatorial optimization: Festschrift for martin grötschel*, pages 449–481. Springer, 2013.
3. Rishabh Agarwal, Dale Schuurmans, and Mohammad Norouzi. An optimistic perspective on offline reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2020.
4. Kenneth J Arrow and Leonid Hurwicz. An optimality criterion for decision-making under ignorance. *Uncertainty and expectations in economics*, 1, 1972.
5. Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.
6. Egon Balas. Disjunctive programming. *Annals of Discrete Mathematics*, 5:3–51, 1979.
7. Amir Beck and Aharon Ben-Tal. Duality in robust optimization: primal worst equals dual best. *Operations Research Letters*, 37(1):1–6, 2009.
8. Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.
9. Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. A data-driven approach to multistage stochastic linear optimization. *Management Science*, 69(1):51–74, 2023.
10. Zhi Chen, Melvyn Sim, and Peng Xiong. Robust stochastic optimization made easy with rsome. *Management Science*, 66(8):3329–3339, 2020.
11. Zhi Chen and Weijun Xie. Regret in the newsvendor model with demand and yield randomness. *Production and Operations Management*, 30(11):4176–4197, 2021.
12. Raphael Chinchilla and João P Hespanha. Stochastic programming using expected value bounds. *IEEE Transactions on Automatic Control*, 2022.
13. Daniel Duque, Sanjay Mehrotra, and David P Morton. Distributionally robust two-stage stochastic programming. *SIAM Journal on Optimization*, 32(3):1499–1522, 2022.
14. Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.
15. Dongdong Ge, Xiaoye Jiang, and Yinyu Ye. A note on the complexity of l p minimization. *Mathematical programming*, 129(2):285–299, 2011.
16. Vikas Goel and Ignacio E Grossmann. A class of stochastic programs with decision dependent uncertainty. *Mathematical programming*, 108(2-3):355–394, 2006.
17. Jun-ya Gotoh, Michael Jong Kim, and Andrew EB Lim. A Data-Driven Approach to Beating SAA Out of Sample. *Operations Research*, 2023.
18. Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.
19. Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.
20. Leonid Hurwicz. The generalized bayes minimax principle: a criterion for decision making under uncertainty. *Cowles Comm. Discuss. Paper Stat*, 335:1950, 1951.
21. Nan Jiang and Weijun Xie. ALSO-X#: Better Convex Approximations for Distributionally Robust Chance Constrained Programs. *arXiv preprint arXiv:2302.01737*, 2023.
22. Nan Jiang and Weijun Xie. Distributionally Favorable Optimization: A Framework for Data-driven Decision-making with Endogenous Outliers. *SIAM Journal on Optimization*, 34(1):419–458, 2024.
23. Çagıl Koçyigit, Garud Iyengar, Daniel Kuhn, and Wolfram Wiesemann. Distributionally robust mechanism design. *Management Science*, 66(1):159–189, 2020.
24. Miles Lubin, Juan Pablo Vielma, and Ilias Zadik. Mixed-integer convex representability. *Mathematics of Operations Research*, 47(1):720–749, 2022.

25. Fengqiao Luo and Sanjay Mehrotra. Distributionally robust optimization with decision dependent ambiguity sets. *Optimization Letters*, 14:2565–2594, 2020.
26. Garth P McCormick. Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.
27. Peyman Mohajerin Esfahani, Soroosh Shafieezadeh-Abadeh, Grani A Hanasusanto, and Daniel Kuhn. Data-driven inverse optimization with imperfect information. *Mathematical Programming*, 167:191–234, 2018.
28. Viet Anh Nguyen, Soroosh Shafieezadeh Abadeh, Man-Chung Yue, Daniel Kuhn, and Wolfram Wiesemann. Calculating optimistic likelihoods using (geodesically) convex optimization. In *Advances in Neural Information Processing Systems*, pages 13942–13953, 2019.
29. Viet Anh Nguyen, Soroosh Shafieezadeh Abadeh, Man-Chung Yue, Daniel Kuhn, and Wolfram Wiesemann. Optimistic distributionally robust optimization for nonparametric likelihood approximation. In *Advances in Neural Information Processing Systems*, pages 15872–15882, 2019.
30. Viet Anh Nguyen, Nian Si, and Jose Blanchet. Robust bayesian classification using an optimistic score ratio. In *International Conference on Machine Learning*, pages 7327–7337. PMLR, 2020.
31. Omid Nohadani and Kartikey Sharma. Optimization under decision-dependent uncertainty. *SIAM Journal on Optimization*, 28(2):1773–1795, 2018.
32. Matthew Norton, Akiko Takeda, and Alexander Mafusalov. Optimistic robust optimization with applications to machine learning. arXiv preprint arXiv:1711.07511, 2017.
33. Jin Qi, Melvyn Sim, Defeng Sun, and Xiaoming Yuan. Preferences for travel time under risk and ambiguity: Implications in path selection and network equilibrium. *Transportation Research Part B: Methodological*, 94:264–284, 2016.
34. Johannes O Royset, Louis L Chen, and Eric Eckstrand. Rockafellian relaxation in optimization under uncertainty: Asymptotically exact formulations. *arXiv preprint arXiv:2204.04762*, 2022.
35. Soroosh Shafieezadeh Abadeh, Peyman M Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. *Advances in Neural Information Processing Systems*, 28, 2015.
36. Alexander Shapiro and Anton Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.
37. Maurice Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.
38. Jun Song and Chaoyue Zhao. Optimistic distributionally robust policy optimization. arXiv preprint arXiv:2006.07815, 2020.
39. Luying Sun, Weijun Xie, and Tim Witten. Distributionally robust fair transit resource allocation during a pandemic. *Transportation science*, 57(4):954–978, 2023.
40. Phebe Vayanos, Angelos Georghiou, and Han Yu. Robust optimization with decision-dependent information discovery. *arXiv preprint arXiv:2004.08490*, 2020.
41. Weijun Xie. Tractable reformulations of two-stage distributionally robust linear programs over the type-∞ Wasserstein ball. *Operations Research Letters*, 48(4):513–523, 2020.
42. Bo Zeng and Wei Wang. Two-stage robust optimization with decision dependent uncertainty. *arXiv preprint arXiv:2203.16484*, 2022.
43. Qi Zhang and Wei Feng. A unified framework for adjustable robust optimization with endogenous uncertainty. *AIChe journal*, 66(12):e17047, 2020.