Bounds for Multistage Mixed-Integer Distributionally Robust Optimization

Güzin Bayraksan* Francesca Maggioni† Daniel Faccini‡ Ming Yang§

May 20, 2023

Abstract: Multistage mixed-integer distributionally robust optimization (DRO) forms a class of extremely challenging problems since their size grows exponentially with the number of stages. One way to model the uncertainty in multistage DRO is by creating sets of conditional distributions (the so-called conditional ambiguity sets) on a finite scenario tree and requiring that such distributions remain close to nominal conditional distributions according to some measure of similarity/distance (*e.g.*, ϕ-divergences or Wasserstein distance). In this paper, new bounding criteria for this class of difficult decision problems are provided through scenario grouping using the ambiguity sets associated with various commonly used ϕ-divergences and the Wasserstein distance. Our approach does not require any special problem structure such as linearity, convexity, stagewise independence, and so forth. Therefore, while we focus on multistage mixed-integer DRO, our bounds can be applied to a wide range of DRO problems including two-stage and multistage, with or without integer variables, convex or nonconvex, and nested or non-nested formulations. Numerical results on a multistage mixed-integer production problem show the efficiency of the proposed approach through different choices of partition strategies, ambiguity sets, and levels of robustness.

Keywords: Multistage Distributionally Robust Optimization, Bounding, Dynamic Measures of Risk, Phi-Divergences, Wasserstein Distance.

*Bayraksan.1@osu.edu
Department of Integrated Systems Engineering, The Ohio State University, USA
†Francesca.Maggioni@unibg.it
Department of Management, Information and Production Engineering, University of Bergamo, Italy
‡Daniel.Faccini@unibg.it
Department of Management, Information and Production Engineering, University of Bergamo, Italy
§Yang.3149@osu.edu
Department of Integrated Systems Engineering, The Ohio State University, USA

1 Introduction

Multistage stochastic programming has been widely used to solve important problems arising in various fields including finance [19], transportation [8], energy [28, 45] and the environment [47], among others. Despite their wide applicability, this class of problems suffers from two main issues. First, traditional models assume that the underlying stochastic process that governs the uncertain parameters is known. This is rarely true in real life. Second, multistage stochastic programs—particularly those involving mixed-integer variables and nonlinear terms—are notoriously difficult to solve. To alleviate the first issue, Distributionally Robust Optimization (DRO) can be used, where the assumed-known distribution is replaced by an ambiguity set of distributions [43]. Unfortunately, the resulting multistage problem is still extremely challenging to solve due to the exponential growth of the problem in the number stages, and can become even more challenging depending on the type of DRO used. Furthermore, many real-world problems lack special structures (*e.g.*, convexity, stagewise independence, binary state variables) that prevent efficient solution algorithms. Therefore, approximation techniques that provide bounds on the optimal value for multistage DRO problems can be very useful in practice. In these situations, easy-to-compute bounds and approximations are desirable.

Most of the existing literature on DRO focuses on static, two-stage, or chance-constrained settings [35]. There is relatively little work on multistage DRO, most of which investigates different ways of forming ambiguity sets such as moment-based [6, 44, 46], nested Wasserstein [29], modified χ^2 distance [30], general ϕ-divergences [27], L_∞-norm [15], Wasserstein [10] and ∞-Wasserstein distance [5]. Most papers assume linear models with continuous decision variables [31, 43], except for [5, 46], which consider mixed-integer decision variables. The majority of the existing works also focus on solution methods through nested Benders' decomposition or its sampling-based variant, the Stochastic Dual Dynamic Programming (SDDP) [10, 15, 27, 30, 46]. Linear decision rules have also been used to approximately solve these problems [5, 6].

In this paper we investigate bounds via scenario tree decomposition for a class of multistage DRO minimization problems formed using ϕ-divergences and Wasserstein distance. The considered approach divides the sample space into subgroups which, being of smaller size, can be solved more efficiently than the original DRO problem. Then, the optimal values of subgroups can be combined, *e.g.*, in a distributionally robust manner, to form lower bounds (LBs) on the optimal value. We provide conditions on ways to combine

the optimal values of the subgroups to obtain LBs for ambiguity sets formed via many commonly used ϕ-divergences and Wasserstein distance. Because obtaining LBs for minimization problems is typically more challenging than obtaining upper bounds (UBs), the majority of the paper focuses on LBs. UBs, which are briefly discussed, are calculated by fixing some decisions obtained through the LBs, finding a feasible policy, and using the cost of that policy; see, *e.g.,* [21, 25].

Bounding techniques have a rich history in the stochastic programming literature, and these have been successfully applied to traditional multistage stochastic programs with expected-value objectives; see, *e.g.*, [11, 12, 18]. Bounds for multistage stochastic linear programs via scenario tree decomposition, the approach also used in this work, were proposed for the first time in [20]. In [21], the authors extend the bounding approach of [7, 20, 40] for stochastic multistage mixed-integer linear programs, solving a sequence of group subproblems made by a subset of reference scenarios plus a subset of scenarios from the finite support. They show the monotonicity of the chains of LBs in terms of the cardinality of reference scenarios and of the remaining scenarios in each subgroup. A scalable bounding framework for general multistage stochastic programs, extending the work of [40], has been investigated in [41]. Recently, [2] introduced sampling scenario set partition dual bounds for multistage stochastic programs.

Bounds for risk-averse multistage mixed-integer stochastic programs, which are related to the DRO problems studied in this paper (see Section 2.4), via scenario tree decomposition were first proposed by [23] and [25]. In particular, [23] considers multistage convex problems with concave risk functional applied to the total cost over the planning horizon. New refinement chains of LBs are constructed, where each bound can be computed by solving sets of group subproblems less complex than the original one. A monotonically nondecreasing behavior in the cardinality of scenarios of each subproblem is proved. LBs for replacing the scenario process by its expectation are also considered, as well as UBs based on inserting feasible solutions derived from smaller subproblems. In [25], the authors consider a dynamic risk functional in the objective function, formed by a convex combination of mean and Conditional Value-at-Risk (mean-CVaR). LBs by using convolution of mean-CVaRs with different parameters are obtained through various scenario partition strategies, and a solution algorithm for mean-CVaR multistage mixed-integer stochastic problems is provided based on an inner approximation of the dual set of the considered coherent measure of risk; see also [26] for algorithmic use of these bounds. An

alternative approach to bound the original risk-averse multistage stochastic program uses barycentric discretizations [13, 14, 17]. In [24], the authors generalize the bounding ideas of [13, 14, 17] to not necessarily Markovian scenario processes and derive valid LBs and UBs for the convex case based on the concepts of first order and convex order stochastic dominance.

However, contributions based on bounding techniques in the distributionally robust optimization literature are very scarce and to the best of our knowledge the only one is provided in [9], where the authors propose inner and outer approximations for two-stage DRO problems with moment-based ambiguity sets and a combined ambiguity set including Wasserstein distance and moment information. The employed methodology to get LBs relies on splitting a random vector into smaller sub-vectors and is parameterized by the number of split pieces. Nevertheless, these results are limited to two-stage problems, so neither multistage settings nor ϕ-divergences, which are considered in this paper, are investigated.

This paper, to the best of our knowledge for the first time, introduces new LB criteria for multistage DRO through scenario tree decomposition. UBs are also examined. Our work is similar in spirit to [21, 23, 25], but we consider a large class of DRO formed on finite scenario trees, where the ambiguity sets are constructed using a ϕ-divergence (*e.g.*, variation distance, Cressie-Read power divergence, J-divergence, etc.) or Wasserstein distance on a finite support [10]. We provide conditions on how to choose the ambiguity sets' radii of the subgroup problems as well as the radius of the ambiguity set to combine the optimal values of subgroup problems in order to obtain LBs.

Compared to the bounding approach proposed in the stochastic programming literature via scenario tree decomposition [2, 7, 20, 21, 23, 40, 41], where LBs were obtained mainly by relaxing the nonanticipacitivity constraints or refining the filtration, in the DRO setting, additional conditions on how to choose the ambiguity sets' radii are needed. The derivation of bounds in the DRO setting thus requires to adopt a new methodology, mainly based on an inner approximation of the ambiguity sets associated with the chosen measure of similarity between distributions. Results show that in the DRO setting, unlike previous works, the monotonicity of the chains of LBs in terms of the cardinality of the reference scenarios and of the remaining ones in each subgroup does not hold anymore. The proposed approach is more general than the one employed in [25], where the LBs were based on an inner approximation of the mean-CVaR risk envelope. Indeed, our proof through

ambiguity set, although being equivalent to the one using risk envelope, does not require the knowledge of the risk functional's explicit expression which is often difficult to assess. Furthermore, in the multistage setting, we propose a way to dissect the scenario tree and combine the subgroups in a nested fashion which has not been done before.

Our results are applicable to a broad class of problems including two- and multistage, with or without integer variables, nested *vs.* non-nested formulations. Therefore, after recalling background information in Section 2, in Section 3, we first present our results in the two-stage setting and then discuss how to apply these LBs in the multistage setting. UBs are also derived. We then investigate the effectiveness of the proposed bounds on a multistage mixed-integer production planning problem and its two-stage variant in Section 4, discussing the insights gained from these experiments. Section 5 concludes the paper and outlines future research directions. We end by noting that the proposed approach has the important advantage to split a given problem into independent scenario groups. This allows to tackle problems for which simple linear relaxations leave large optimality gaps, problems lacking special structure like convexity, stagewise independence, etc. that prevent efficient solution methods, and large-scale multistage problems that are not solvable by commercial solvers.

2 Basic facts and notation

2.1 Multistage DRO

We consider a finite-horizon sequential decision making problem under uncertainty, where decisions are made at discrete stages $t \in \mathcal{T} := \{0, 1, \ldots, T\}$ and T denotes the planning horizon. The decision process begins with initial decision $\mathbf{x}_0 \in \mathbb{R}^{n_0}_+ \times \mathbb{Z}^{n'_0}_+$ at stage $t = 0$, called the *first-stage* or *here-and-now* decision, followed by sequential decisions $\mathbf{x}_t \in \mathbb{R}^{n_t}_+ \times \mathbb{Z}^{n'_t}_+$ at stages $t \in \mathcal{T} \setminus \{0\}$. The history of the decision process, at a given point in time, is denoted by $\mathbf{x}^t := (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t)$, $t \in \mathcal{T}$. The uncertainty is described by a random process $\boldsymbol{\xi} := \{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T\} \in \mathbb{R}^{d_0} \times \ldots \times \mathbb{R}^{d_T}$ defined on a measurable space $(\Omega, \mathscr{F})$. We assume $\boldsymbol{\xi}_0$ is a constant and $\boldsymbol{\xi}$ is a random parameter evolving as a discrete-time stochastic process with finite support. The filtration associated with $\boldsymbol{\xi}$ is denoted by $\{\emptyset, \Omega\} = \mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}_T = \mathscr{F}$, where $\mathscr{F}_t, t \in \mathcal{T}$—a σ-subalgebra of $\mathscr{F}$—models the information available so far. Each $\boldsymbol{\xi}_t$, $t \in \mathcal{T}$, has finite support $\Omega_t \in \mathbb{R}^{d_t}$ and nominal distribution Q_t. Support of $\boldsymbol{\xi}$ is given by $\Omega = \bigtimes_{t=0}^{T} \Omega_t$, and the history of the random

process up to stage t is denoted by $\boldsymbol{\xi}^t := (\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_t)$, $t \in \mathcal{T}$. The following represents the nested formulation of a multistage DRO (see [43]):

$$\min_{\mathbf{x}_0 \in \mathcal{X}_0(\boldsymbol{\xi}_0)} c_0(\mathbf{x}_0, \boldsymbol{\xi}_0) + \max_{P_1|\boldsymbol{\xi}^0 \in \mathcal{P}_{1|\boldsymbol{\xi}^0}} \mathbb{E}_{P_1|\boldsymbol{\xi}^0}\Bigg[\min_{\mathbf{x}_1 \in \mathcal{X}_1(\mathbf{x}_0, \boldsymbol{\xi}_1)} c_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \max_{P_2|\boldsymbol{\xi}^1 \in \mathcal{P}_{2|\boldsymbol{\xi}^1}} \mathbb{E}_{P_2|\boldsymbol{\xi}^1}\Big[\ldots \\ \cdots + \max_{P_T|\boldsymbol{\xi}^{T-1} \in \mathcal{P}_{T|\boldsymbol{\xi}^{T-1}}} \mathbb{E}_{P_T|\boldsymbol{\xi}^{T-1}}\Big[\min_{\mathbf{x}_T \in \mathcal{X}_T(\mathbf{x}_{T-1}, \boldsymbol{\xi}_T)} c_T(\mathbf{x}_T, \boldsymbol{\xi}_T)\Big] \ldots \Big]\Bigg], \tag{1}$$

where the mixed-integer first-stage feasibility set is given by $\mathcal{X}_0 \subseteq \mathbb{R}_+^{n_0} \times \mathbb{Z}_+^{n_0'}$ and, for $t \in \mathcal{T} \setminus \{0\}$, $\mathcal{X}_t : \mathbb{R}_+^{n_{t-1}} \times \mathbb{Z}_+^{n_{t-1}'} \times \mathbb{R}^{d_t} \to \mathbb{R}_+^{n_t} \times \mathbb{Z}_+^{n_t'}$ are $\mathscr{F}_t$-measurable mixed-integer point-to-set mappings. The possibly nonlinear cost functions are given by $c_0 : \mathbb{R}_+^{n_0} \times \mathbb{Z}_+^{n_0'} \to \mathbb{R}$ in the first stage and by $c_t : \mathbb{R}_+^{n_t} \times \mathbb{Z}_+^{n_t'} \times \mathbb{R}^{d_t} \to \mathbb{R}$ in stages $t \in \mathcal{T} \setminus \{0\}$, which are $\mathscr{F}_t$-measurable. We assume all relevant optimization problems in the paper have finite optimal solutions. Set $\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}}$ denotes the conditional ambiguity set at period $t \in \mathcal{T} \setminus \{0\}$, conditioned on the history $\boldsymbol{\xi}^{t-1}$, and it is defined as

$$\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}} := \Big\{ P_{t|\boldsymbol{\xi}^{t-1}} \in \mathcal{M}(\Omega_{t|\boldsymbol{\xi}^{t-1}}) : \ \Delta(P_{t|\boldsymbol{\xi}^{t-1}}, Q_{t|\boldsymbol{\xi}^{t-1}}) \leq \rho_t \Big\}, \tag{2}$$

where $\rho_t \geq 0$ is a given *radius*, also called the *level of robustness*. Above, $\mathcal{M}(\Omega_{t|\boldsymbol{\xi}^{t-1}})$ represents a class of probability distributions defined on the finite support $\Omega_{t|\boldsymbol{\xi}^{t-1}}$, $Q_{t|\boldsymbol{\xi}^{t-1}}$ denotes the nominal conditional probability measure at stage t, $t \in \mathcal{T} \setminus \{0\}$, conditioned on the history of the process $\boldsymbol{\xi}^{t-1}$, and $\Delta(\cdot, \cdot)$ denotes a measure of similarity or distance between $P_{t|\boldsymbol{\xi}^{t-1}}$ and $Q_{t|\boldsymbol{\xi}^{t-1}}$. We are interested in building ambiguity sets using existing data via ϕ-divergences and Wasserstein distance, which we recall in the next sections. Before we do so, let us define additional notation used in the paper.

Scenario tree and nominal probability notation. Because we assume $\boldsymbol{\xi}$ has finite support, the information structure can be described in the form of a *scenario tree* $\mathfrak{T}$ with $T+1$ levels (stages). Let Ω_t be the set of ordered nodes of the tree $\mathfrak{T}$ at stage $t \in \mathcal{T}$ and let $\Omega := \Omega_1 \times \ldots \times \Omega_T$. By assumption, we have a discrete number $|\Omega_t|$ of nodes at each stage $t \in \mathcal{T}$. Each stage-t ($t > 0$) node n is connected to a unique node at stage $t-1$, called *ancestor* and denoted $a(n)$. Similarly, each stage-t ($t < T$) node n is connected to nodes at stage $t+1$ called *successors* or *children*, where $\mathscr{B}(n)$ denotes the set of children nodes of n. With $q_{a(n),n}$ we denote the conditional nominal probability of the random process at node n given its history up to the ancestor node $a(n)$. A *scenario* ω_i, $i = 1, \ldots, |\Omega_T|$ is a path through nodes from the root node at $t = 0$ to a leaf node

at $t = T$. We indicate with q_{ω_i} the nominal probability of a scenario ω_i passing through nodes $n_0, n_1, \ldots, n_T$ (where n_t, $t = 0, \ldots, T$ represent generic nodes at stage t), defined as $q_{\omega_i} := q_{n_0,n_1} \cdot q_{n_1,n_2} \cdot \ldots \cdot q_{n_{T-1},n_T}$. We also indicate with q_t^n the nominal probability of node n at stage t. So, if node n at stage t is reachable through node n_0 at stage 0, node n_1 at stage 1, ..., node n_{t-1} at stage $t-1$, then $q_t^n := q_{n_0,n_1} \cdot q_{n_1,n_2} \cdot \ldots \cdot q_{n_{t-1},n_t}$. Moreover, $\sum_{n\in\Omega_t} q_t^n = 1,\ t \in \mathcal{T}$ and $\sum_{m\in\mathscr{B}(n)} q_{n,m} = 1,\ n \in \Omega_t,\ \ t = 0, \ldots, T-1$.

Set notation. We use shorthand notation $[m]$ to denote the set $\{1, 2, \ldots, m\}$.

For simplicity, from here until Section 3.5, we consider two-stage DRO and only point to changes for multistage case. That is, we set $T = 1$ in (1), drop $\boldsymbol{\xi}_0$ as it is a constant, and let $\boldsymbol{\xi}_1 \equiv \boldsymbol{\xi}$ be defined on a probability space $(\Omega, \mathscr{F}, Q)$ with sample space $\Omega := \{\omega_1, \omega_2, \ldots, \omega_{|\Omega|}\}$, σ-algebra $\mathscr{F}$ and nominal probability Q. The nominal probability of scenario $\omega_i \in \Omega$ can be specified as $q_{\omega_i} \geq 0$ with $\sum_{i\in[|\Omega|]} q_{\omega_i} = 1$. Similarly, we simply use P with probability of scenario ω_i as $p_{\omega_i} \geq 0$ to define ambiguity set (2). So, (2) becomes $\mathcal{P} = \{P :\ \Delta(P, Q) \leq \rho,\ \sum_{i\in[|\Omega|]} p_{\omega_i} = 1,\ p_{\omega_i} \geq 0, \forall \omega_i \in \Omega\}$.

2.2 ϕ-divergences

For this class of ambiguity sets, Δ in (2) is given by

$$\Delta_\phi(P, Q) := \sum_{i\in[|\Omega|]} q_{\omega_i} \phi\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right),$$

where the convex ϕ-divergence function $\phi(u) \geq 0$ takes value 0 when both $p_{\omega_i} > 0$ and $q_{\omega_i} > 0$ have the same value; *i.e.*, $\phi(1) = 0$. When $q_{\omega_i} = 0$, it holds that $0 \cdot \phi(p_{\omega_i}/0) = p_{\omega_i} \lim_{u\to\infty}(\phi(u)/u)$ and $0 \cdot \phi(0/0) = 0$. Accordingly, ambiguity set $\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}}$ in (2) can be built using some of the well-known ϕ-divergences described in Table 1 and Table 2. These include Variation Distance (VD) and J-divergence, along with two families of ϕ-divergences, namely, the Cressie-Read (CR) power divergence family and the χ-divergence family of order $a > 1$. CR power divergence family includes some of the most widely used ϕ-divergences as a special case—*e.g.*, the modified χ^2 distance and the Kullback-Leibler divergence—when its parameter θ takes specific values or when the limit of θ tends to 0 or 1. These special cases are listed in Table 2. Equivalence of the well-known divergences in Table 2 and the CR power divergence family in Table 1 is achieved when the radius ρ_t in the ambiguity set (2) formed via a divergence in Table 2 is set to an adjusted value $c \cdot \rho_{t,CR}^\theta$, where $\rho_{t,CR}^\theta$ is the radius of the CR divergence in Table 1. Values of coefficient

c corresponding to certain θ are listed in the last column of Table 2. For example, when the radius of the modified χ^2 distance, denoted $\rho_{t,m\chi^2}$, equals $2 \cdot \rho_{t,CR}^{\theta=2}$, where $\rho_{t,CR}^{\theta=2}$ represents the radius formed via CR power divergence with $\theta = 2$, the two ambiguity sets are equivalent.

Divergence	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$
Variation Distance	ϕ_v	$\lvert u-1\rvert$	$\sum \lvert p_{\omega_i} - q_{\omega_i}\rvert$
Cressie-Read Power Divergence	ϕ_{CR}^θ	$\frac{1-\theta+\theta u-u^\theta}{\theta(1-\theta)}, \theta \neq 0,1$	$\frac{1-\sum p_{\omega_i}^\theta q_{\omega_i}^{1-\theta}}{\theta(1-\theta)}, \theta \neq 0,1$
J-Divergence	ϕ_J	$(u-1)\log u$	$\sum (p_{\omega_i} - q_{\omega_i}) \log\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$
χ-Divergence of order $a > 1$	ϕ_χ^a	$\lvert u-1\rvert^a$	$\sum q_{\omega_i} \left\lvert 1 - \frac{p_{\omega_i}}{q_{\omega_i}}\right\rvert^a$

Table 1: Common ϕ-divergences.

θ	*Corresponding Divergence*	$\phi(u)$	$\phi(u), u \geq 0$	$\Delta_\phi(P, Q)$	$\phi_{CR}^\theta(u)$	c
2	*Modified χ^2 Distance*	$\phi_{m\chi^2}$	$(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{q_{\omega_i}}$	$\frac{1}{2}(u^2-2u+1) = \frac{1}{2}(u-1)^2$	2
$\frac{1}{2}$	*Hellinger Distance*	ϕ_H	$(\sqrt{u}-1)^2$	$\sum(\sqrt{p_{\omega_i}} - \sqrt{q_{\omega_i}})^2$	$4(\frac{1}{2} + \frac{1}{2}t - \sqrt{u}) = 2(1-\sqrt{u})^2$	$\frac{1}{2}$
-1	*χ^2 Distance*	ϕ_{χ^2}	$\frac{1}{u}(u-1)^2$	$\sum \frac{(p_{\omega_i}-q_{\omega_i})^2}{p_{\omega_i}}$	$\frac{1}{2}(-2+u+\frac{1}{u}) = \frac{1}{2}(\sqrt{u} - \frac{1}{\sqrt{u}})^2$	2
$\to 1$	*Kullback-Leibler Divergence*	ϕ_{KL}	$u\log u - u + 1$	$\sum p_{\omega_i} \log\left(\frac{p_{\omega_i}}{q_{\omega_i}}\right)$	$u(\log u - 1) + 1$	1
$\to 0$	*Burg Entropy*	ϕ_B	$-\log u + u - 1$	$\sum q_{\omega_i} \log\left(\frac{q_{\omega_i}}{p_{\omega_i}}\right)$	$-\log u + u - 1$	1

Table 2: Some special cases of CR power divergence family. Kullback-Leibler divergence and Burg entropy are obtained by taking the limit of θ to 1 and 0, respectively.

2.3 Wasserstein distance

Let $\boldsymbol{\eta}$ be a random variable on $(\Omega, \mathscr{F}, Q)$ taking values $(\eta_{\omega_1}, \ldots, \eta_{\omega_{|\Omega|}})$. We quantify distributions P close to nominal distribution Q on the same support via Wasserstein distance (see [10]), where Δ in (2) is defined by

$$
\begin{aligned}
\Delta_W(P, Q) := \min_{\{z_{\omega_i,\omega_j}\}_{i,j=1}^{|\Omega|}} \quad & \sum_{j=1}^{|\Omega|} \sum_{i=1}^{|\Omega|} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} & \\
\text{s.t.} \quad & \sum_{i=1}^{|\Omega|} z_{\omega_i,\omega_j} = q_{\omega_j} & j = 1, \ldots, |\Omega| \\
& \sum_{j=1}^{|\Omega|} z_{\omega_i,\omega_j} = p_{\omega_i} & i = 1, \ldots, |\Omega| \\
& z_{\omega_i,\omega_j} \geq 0 & i, j = 1, \ldots, |\Omega|
\end{aligned}
$$

with $d_{\omega_i,\omega_j} := ||\eta_{\omega_i} - \eta_{\omega_j}||_\varsigma$ a distance between the two scenarios ω_i and ω_j using ς-norm (*e.g.*, $\varsigma \in \{1, 2, \infty\}$). Ambiguity set $\mathcal{P}_{t|\boldsymbol{\xi}^{t-1}}$ in (2) can be built accordingly.

2.4 Relation to risk-averse optimization

Because the ambiguity sets considered in this paper are compact convex subsets of (conditional) probability measures and optimal values are assumed to be real-valued, DRO is equivalent to *Risk-Averse Stochastic Optimization* (RASO) with the objective function expressed by a coherent risk measure; see *e.g.,* [3, 36, 43]. Let us recall coherent risk measures. Let $\mathcal{Z} := \mathfrak{L}_\infty(\Omega, \mathscr{F}, Q)$ be the space of bounded and $\mathscr{F}$-measurable random variables with respect to sample space Ω and probability distribution Q, and let $\boldsymbol{\eta} \in \mathcal{Z}$ be a random variable taking values $(\eta_{\omega_1}, \ldots, \eta_{\omega_{|\Omega|}})$. First defined by [1], a function $\mathcal{R}(\boldsymbol{\eta}) : \mathcal{Z} \to \mathbb{R}$ is called a *coherent measure of risk* if it satisfies the following properties:

1. *Convexity*: $\mathcal{R}(\lambda\boldsymbol{\eta}^1 + (1-\lambda)\boldsymbol{\eta}^2) \leq \lambda\mathcal{R}(\boldsymbol{\eta}^1) + (1-\lambda)\mathcal{R}(\boldsymbol{\eta}^2)$ for all $\boldsymbol{\eta}^1, \boldsymbol{\eta}^2 \in \mathcal{Z}$ and $\lambda \in [0,1]$;

2. *Monotonocity*: $\boldsymbol{\eta}^1 \geq \boldsymbol{\eta}^2$ implies $\mathcal{R}(\boldsymbol{\eta}^1) \geq \mathcal{R}(\boldsymbol{\eta}^2)$ for all $\boldsymbol{\eta}^1, \boldsymbol{\eta}^2 \in \mathcal{Z}$;

3. *Translation Equivariance*: $\mathcal{R}(\boldsymbol{\eta} + \lambda) = \mathcal{R}(\boldsymbol{\eta}) + \lambda$ for all $\lambda \in \mathbb{R}$ and $\boldsymbol{\eta} \in \mathcal{Z}$;

4. *Positive Homogeneity*: $\mathcal{R}(\lambda \cdot \boldsymbol{\eta}) = \lambda \cdot \mathcal{R}(\boldsymbol{\eta})$ for all $\lambda > 0$ and $\boldsymbol{\eta} \in \mathcal{Z}$.

Coherent measures of risk can be interpreted as worst-case expectations from a compact convex set of probability measures through their dual representation: $\mathcal{R}(\boldsymbol{\eta}) := \max_{P \in \mathcal{P}} \mathbb{E}_P[\boldsymbol{\eta}]$. Therefore, it follows that a RASO can be re-written as a DRO

$$\min_{\mathbf{x} \in \mathcal{X}} \mathcal{R}\big(c(\mathbf{x}, \boldsymbol{\xi})\big) := \min_{\mathbf{x} \in \mathcal{X}} \max_{P \in \mathcal{P}} \mathbb{E}_P\big[c(\mathbf{x}, \boldsymbol{\xi})\big]. \tag{3}$$

The above conclusion is extended to the multistage setting by recursively using conditional ambiguity sets, which we recall in Section 3.5; see *e.g.,* [37, 38, 43] for nested coherent composite risk measures in the multistage setting.

3 Lower bounds for DRO

The aim of this section is to provide LBs for DRO formed by ϕ-divergences and the Wasserstein distance. For this purpose, instead of dealing with the whole sample space Ω, whose large cardinality may lead to computational concerns, the LB is achieved by dividing the sample space Ω into subgroups that can be considered separately and then combining the optimal values of the subgroups using an ambiguity set with possibly another radius. To perform such a division, we consider the approaches presented in [23] and summarized

below. For ease of presentation (of probabilities, computations, etc.), below we set each subgroup to have the same cardinality. However, subgroups with different cardinalities are possible.

3.1 Dissecting the scenario tree

We construct a collection of m_l subsets, each of cardinality l, of the sample space Ω:

$$\left(\Omega_1^{(l)}, \Omega_2^{(l)}, \ldots, \Omega_{m_l}^{(l)}\right),$$

with the property that their union covers the whole space $\Omega = \cup_{g\in[m_l]} \Omega_g^{(l)}$. For each $\Omega_g^{(l)}$, $g \in [m_l]$, there corresponds a probability measure $Q_g^{(l)}$. Therefore

$$\left(Q_1^{(l)}, Q_2^{(l)}, \ldots, Q_{m_l}^{(l)}\right)$$

represents a dissection of the nominal probability measure $Q = \sum_{g\in[m_l]} \pi_g^{(l)} Q_g^{(l)}$ with $\sum_{g\in[m_l]} \pi_g^{(l)} = 1$ and $\pi_g^{(l)} \geq 0$ for all $g \in [m_l]$. For instance, when $l = 1$, then $\Omega_g^{(1)} = \{\omega_g\}$, $Q_g^{(1)} = \delta_{\omega_g}$, $g \in [|\Omega|]$, where δ_{ω_g} represents the Dirac measure at scenario ω_g; hence, subgroup g has nominal probability equal to one for scenario ω_g and nominal probability equal to zero for all other scenarios. Each measure $Q_g^{(l)}$ in the collection is given by $Q_g^{(l)} = \sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} \cdot \delta_{\omega_i}$, where $(q_{\omega_i})_g^{(l)}$ denotes the nominal probability of scenario ω_i within subgroup g with $\sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} = 1$. Below, we provide details of the measures $Q_g^{(l)}$—and hence details of $(q_{\omega_i})_g^{(l)}$—based on different constructions from [23].

Fixed scenarios. We first consider the case where one or more scenarios appear in all subsets. Without loss of generality, we assume that the first $f < l$ scenarios of Ω ($\Omega_f = \{\omega_1, \ldots, \omega_f\}$) are fixed. Consequently the number of subgroups with cardinality l is $m_l = \frac{|\Omega|-f}{l-f} \in \mathbb{N}$. Notice that since m_l represents the number of subgroups of a given dissection, it must be an integer. Then, the probability measures $Q_g^{(l)}$ can be calculated as follows:

$$Q_g^{(l)} := \sum_{\omega_i\in\Omega_f} q_{\omega_i} \cdot \delta_{\omega_i} + \sum_{\omega_i\in\Omega_g^{(l)}\setminus\Omega_f} \frac{q_{\omega_i}}{\pi_g^{(l)}} \cdot \delta_{\omega_i}$$

with weights $\pi_g^{(l)} := \dfrac{\sum_{\omega_i\in\Omega_g^{(l)}\setminus\Omega_f} q_{\omega_i}}{1 - \sum_{\omega_i\in\Omega_f} q_{\omega_i}}$, for all subgroups $g \in [m_l]$.

Disjoint partitions. Alternatively, we also consider disjoint partitions: $\Omega = \cup_{g\in[m_l]} \Omega_g^{(l)}$ with $\Omega_{g_1}^{(l)} \cap \Omega_{g_2}^{(l)} = \emptyset$ for $g_1 \neq g_2$. Consequently, the number of subgroups with cardinality

l is $m_l = \frac{|\Omega|}{l} \in \mathbb{N}$. In this case, probability measures $Q_g^{(l)}$ are given by

$$Q_g^{(l)} := \sum_{\omega_i \in \Omega_g^{(l)}} \frac{q_{\omega_i}}{\pi_g^{(l)}} \cdot \delta_{\omega_i}$$

with weights $\pi_g^{(l)} := \sum_{\omega_i \in \Omega_g^{(l)}} q_{\omega_i}$, for all subgroups $g \in [m_l]$.

In the multistage setting, for every scenario $\omega_i \in \Omega_g^{(l)}$ passing through nodes n_0, $n_1, \ldots, n_T$ (with n_t, $t \in \mathcal{T}$ a generic node at stage t), conditional nominal probabilities are adjusted as follows:

$$\left(q_{n_{t-1},n_t}\right)_g^{(l)} := \frac{\sum\limits_{\omega \in \Omega_g^{(l)}(n_t)} (q_\omega)_g^{(l)}}{\sum\limits_{\omega \in \Omega_g^{(l)}(n_{t-1})} (q_\omega)_g^{(l)}} \qquad t \in \mathcal{T} \setminus \{0\},$$

where $\Omega_g^{(l)}(n_t)$ denotes the set of scenarios of subtree $\Omega_g^{(l)}$ passing through node n_t.

Example 1. Figure 1a displays a sample space $\Omega = \{\omega_i\}_{i=1}^{15}$ with 15 scenarios. We divide it into 7 subsets $\Omega_g^{(3)}, g \in [7]$ each of them of cardinality $l = 3$ with scenario ω_1 fixed. That is, $\Omega_f = \{\omega_1\}$, $\Omega_1^{(3)} = \{\omega_1, \omega_2, \omega_3\}$, $\Omega_2^{(3)} = \{\omega_1, \omega_4, \omega_5\}$, and so forth. Assuming equal nominal probability for each scenario, *i.e.*, $q_{\omega_i} = \frac{1}{15}$, $i \in [15]$, the nominal probability of fixed scenario ω_1 within each scenario subgroup g is $(q_{\omega_1})_g^{(3)} = \frac{1}{15}$ and the nominal probability of other two scenarios in the same subgroup is $(q_{\omega_i})_g^{(3)} = \frac{7}{15}$. The weight of each subgroup is $\pi_g^{(3)} = \frac{1}{7}$, $g \in [7]$.

Remark. Besides fixing scenarios and disjoint partitions, there are many in-between cases that can be used to construct subgroups. For instance, one could split the sample space Ω such that only a finite number of subgroups share common scenarios, or such that subgroups have different cardinalities. All these kind of dissections are still valid as long as, given nominal probabilities q_{ω_i}, conditional nominal probabilities $(q_{\omega_i})_g$ and groups weights π_g are chosen such that

$$q_{\omega_i} = \sum_{g:\ \omega_i \in \Omega_g} \pi_g (q_{\omega_i})_g.$$

Nonetheless, this way of constructing collection of subgroups requires a specific knowledge of the problem under investigation that we believe is difficult to asses. For this reason, in this paper we focus on the disjoint and scenario fixing constructions.

3.2 Convolution of risk measures to obtain lower bounds

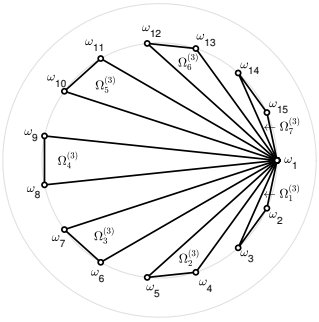
Given a collection of subsets of the scenario tree, our next step is to solve the resulting subgroup problems and combine them judiciously to form LBs on the optimal value of DRO. Toward this end, we use *convolution* (also referred to as *composition*) of risk measures induced by the considered ambiguity sets [25]. We describe this process next.

Let $\mathscr{G}$ be the σ-algebra generated by the collection of subsets $\Omega = \cup_{g\in[m_l]} \Omega_g^{(l)}$, where each subset $\Omega_g^{(l)}$ corresponds to an elementary event of $\mathscr{G}$. First, we solve each subgroup g using DRO given in (3) with intra-group radius $\bar{\rho}_g$. This induces a risk measure denoted $\mathcal{R}_{\bar{\rho}_g}^{(l)}$ on each subgroup $g \in [m_l]$, where $\mathcal{P}_{\bar{\rho}_g}^{(l)}$ is the corresponding ambiguity set with radius $\bar{\rho}_g$. Collectively, we denote these intra-group radii as $\bar{\boldsymbol{\rho}} = \{\bar{\rho}_g\}_{g=1}^{m_l}$. Also, combined together, we represent the one-step conditional risk measure as $\widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}$ and the associated ambiguity set as $\widetilde{\mathcal{P}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}} := \cup_{g\in[m_l]}\mathcal{P}_{\bar{\rho}_g}^{(l)}$. Note that $\widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}$ can be represented in terms of $\mathcal{R}_{\bar{\rho}_g}^{(l)}$, $\forall\ g \in [m_l]$.

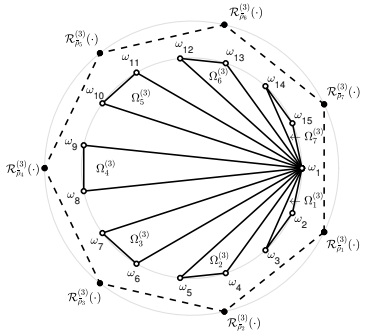
Let $z_g^{*(l)}$ be the optimal value of subgroup $g \in [m_l]$. Consider a $\mathscr{G}$-measurable random variable $\boldsymbol{\zeta}_{LB}$ taking values $\{z_g^{*(l)}\}_{g=1}^{m_l}$ with nominal probabilities $\pi_g^{(l)}$ for each $g \in [m_l]$. For instance, in Example 1, $\boldsymbol{\zeta}_{LB}$ has support $\{z_1^{*(3)}, z_2^{*(3)}, \ldots, z_7^{*(3)}\}$, consisting of optimal values of the 7 subgroups, each with nominal probability $\pi_g^{(l)} = \frac{1}{7}$. We create an ambiguity set $\widetilde{\mathcal{P}}_{\bar{\bar{\rho}}}^{\mathscr{G}}$ around this nominal distribution using inter-groups radius $\bar{\bar{\rho}}$. We then aim to obtain a LB on the optimal value of DRO by calculating $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB}) := \max_{\bar{\bar{P}}\in\widetilde{\mathcal{P}}_{\bar{\bar{\rho}}}^{\mathscr{G}}} \mathbb{E}_{\bar{\bar{P}}}[\boldsymbol{\zeta}_{LB}]$.

This process results in the risk measure $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}(\cdot) := \left(\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}}^{\mathscr{G}} \circ \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}\right)(\cdot)$, which is the convolution of the one-step conditional risk measure $\widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}} : \mathcal{Z} \to \mathfrak{L}_\infty(\Omega, \mathscr{G}, Q)$ and the risk measure on the collection of subsets $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}}^{\mathscr{G}} : \mathfrak{L}_\infty(\Omega, \mathscr{G}, Q) \to \mathbb{R}$. The ambiguity set corresponding to $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}(\cdot)$ is denoted by $\widetilde{\mathcal{P}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}$ so that for any $\boldsymbol{\eta} \in \mathcal{Z}$, $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}(\boldsymbol{\eta}) = \max_{P'\in\widetilde{\mathcal{P}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}} \mathbb{E}_{P'}[\boldsymbol{\eta}]$ (see [25]). Above, we focused on how to use convolution of risk measures to obtain LBs for DRO. This involves an "optimization" step in $\mathbf{x}$ in (3). For convolution of risk measures on random variables, one can consider a "fixed" $\mathbf{x}$ in (3). With fixed $\mathbf{x}$, the term $c(\mathbf{x}, \boldsymbol{\xi})$ in (2.4) becomes only a function of the random variable $\boldsymbol{\xi}$ and hence is a random variable itself. This induced random variable, denoted $\boldsymbol{\eta}_{\mathbf{x}}$, depends on the $\mathbf{x}$ used; *i.e.*, $\boldsymbol{\eta}_{\mathbf{x}} = c(\mathbf{x}, \boldsymbol{\xi})$. In the next two sections, we present LB criteria on general random variables $\boldsymbol{\eta} \in \mathcal{Z}$. These results are directly applicable to static/two-stage DRO because minimization preserves the LBs. Their application to multistage DRO will be discussed in Section 3.5.

The ambiguity sets mentioned above can be formulated as follows. First, given the



(a) A graphical representation of the sample space Ω, with $|\Omega| = 15$ scenarios, divided into $m_3 = 7$ subsets of cardinality $l = 3$, with one fixed scenario $\Omega_f = \{\omega_1\}$.



(b) Computation of the risk measure $\tilde{\mathcal{R}}^{\mathscr{G}}_{\bar{\bar{\rho}}}(\cdot)$ (dashed line) combining the optimal values of subgroups obtained using risk measures $\mathcal{R}^{(l)}_{\bar{\rho}_g}(\cdot)$, $g \in [7]$ induced by DRO with $\mathcal{P}^{(l)}_{\bar{\rho}_g}$.

Figure 1: Two-step procedure to compute the convolution risk measure $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}(\cdot) = (\widetilde{\mathcal{R}}^{\mathscr{G}}_{\bar{\bar{\rho}}} \circ \widetilde{\mathcal{R}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}})(\cdot)$.

subset $\Omega^{(l)}_g$ for subgroup g, the ambiguity set associated with DRO in (3) using intra-group radius $\bar{\rho}_g \geq 0$ inducing risk measure $\mathcal{R}^{(l)}_{\bar{\rho}_g}$ on this subgroup is

$$\mathcal{P}^{(l)}_{\bar{\rho}_g} := \left\{ \bar{P}^{(l)}_g : \Delta(\bar{P}^{(l)}_g, Q^{(l)}_g) \leq \bar{\rho}_g, \sum_{\omega_i \in \Omega^{(l)}_g} (\bar{p}_{\omega_i})^{(l)}_g = 1, \ (\bar{p}_{\omega_i})^{(l)}_g \geq 0, \forall \omega_i \in \Omega^{(l)}_g \right\},$$

where $(\bar{p}_{\omega_i})^{(l)}_g$ represents the probability $\bar{P}^{(l)}_g$ assumes for scenario $\omega_i \in \Omega^{(l)}_g$. Hence, the ambiguity set corresponding to the one-step conditional risk measure $\widetilde{\mathcal{R}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}}$ is

$$\widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}} := \left\{ \bar{P} : \Delta(\bar{P}^{(l)}_g, Q^{(l)}_g) \leq \bar{\rho}_g, \forall g \in [m_l], \sum_{\omega_i \in \Omega^{(l)}_g} (\bar{p}_{\omega_i})^{(l)}_g = 1, \forall g \in [m_l], (\bar{p}_{\omega_i})^{(l)}_g \geq 0, \forall \omega_i \in \Omega^{(l)}_g, \forall g \in [m_l] \right\}.$$

Above, $\bar{P} := \{\bar{P}^{(l)}_g\}_{g=1}^{m_l}$. Next, the ambiguity set using inter-groups radius $\bar{\bar{\rho}} \geq 0$ inducing risk measure $\widetilde{\mathcal{R}}^{\mathscr{G}}_{\bar{\bar{\rho}}}$ on the collection of subsets is

$$\widetilde{\mathcal{P}}^{\mathscr{G}}_{\bar{\bar{\rho}}} := \left\{ \bar{\bar{P}} : \Delta(\bar{\bar{P}}, \bar{\bar{Q}}) \leq \bar{\bar{\rho}}, \sum_{g \in [m_l]} \bar{\bar{p}}^{(l)}_g = 1, \ \bar{\bar{p}}^{(l)}_g \geq 0, \forall g \in [m_l] \right\},$$

where $\bar{\bar{Q}}$ is the nominal distribution composed of the weights $\pi^{(l)}_g$ detailed in Section 3.1 and $\bar{\bar{p}}^{(l)}_g$ represents the probability $\bar{\bar{P}}$ assumes for the subgroup g with cardinality l. Finally, the ambiguity set corresponding to the convolution $\widetilde{\mathcal{P}}^{\mathscr{G}}_{\bar{\bar{\rho}}} \circ \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}}$ with associated risk measure $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}}$ becomes

$$\begin{aligned} \widetilde{\mathcal{P}}_{\bar{\bar{\rho}},\bar{\boldsymbol{\rho}}} := \Big\{ P' : \ & p'_{\omega_i,g} = \bar{\bar{p}}^{(l)}_g \cdot (\bar{p}_{\omega_i})^{(l)}_g, \ \forall \omega_i \in \Omega_f, g \in [m_l], \ p'_{\omega_i} = \sum_{g \in [m_l]} p'_{\omega_i,g}, \ \forall \omega_i \in \Omega_f, \\ & \text{and } p'_{\omega_i} = \bar{\bar{p}}^{(l)}_g \cdot (\bar{p}_{\omega_i})^{(l)}_g, \ \forall \omega_i \in \Omega^{(l)}_g \setminus \Omega_f, g \in [m_l], \ \bar{\bar{P}} \in \widetilde{\mathcal{P}}^{\mathscr{G}}_{\bar{\bar{\rho}}}, \ \bar{P} \in \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}} \Big\}. \end{aligned} \tag{4}$$

Recall Ω_f denotes the set of fixed scenarios (Section 3.1), and if $\Omega_f = \emptyset$, disjoint partitions are used. For any fixed scenario $\omega_i \in \Omega_f$, its probability p'_{ω_i} is found by summing up its group probabilities $p'_{\omega_i,g}$ for all subgroups $g \in [m_l]$. For instance, in Example 1, the probability of scenario ω_1 after convolution is found by $p'_{\omega_1} = \sum_{g=1}^{7} \bar{\bar{p}}_g^{(3)} \cdot (\bar{p}_{\omega_1})_g^{(3)}$, whereas for all other scenarios we have $p'_{\omega_2} = \bar{\bar{p}}_1^{(3)} \cdot (\bar{p}_{\omega_2})_1^{(3)}, \ldots,\ p'_{\omega_5} = \bar{\bar{p}}_2^{(3)} \cdot (\bar{p}_{\omega_5})_2^{(3)},\ \ldots,\ p'_{\omega_{15}} = \bar{\bar{p}}_7^{(3)} \cdot (\bar{p}_{\omega_{15}})_7^{(3)}$ (see Figure 1b). Notice that in (4) the condition $\sum_{\omega_i \in \Omega} p'_{\omega_i} = 1$ always holds because the respective ambiguity sets $\widetilde{\mathcal{P}}^{\mathscr{G}}_{\bar{\bar{\rho}}}$ and $\widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\bar{\boldsymbol{\rho}}}$ require $\sum_{g\in[m_l]} \bar{\bar{p}}_g^{(l)} = 1$ and $\sum_{\omega_i \in \Omega_g^{(l)}} (\bar{p}_{\omega_i})_g^{(l)} = 1$ for all subgroups $g \in [m_l]$:

$$\begin{aligned}\sum_{\omega_i\in\Omega} p'_{\omega_i} &= \sum_{\omega_i\in\Omega_f} p'_{\omega_i} + \sum_{\omega_i\in(\Omega_f)^C} p'_{\omega_i} = \sum_{\omega_i\in\Omega_f}\sum_{g\in[m_l]} \bar{\bar{p}}_g^{(l)} \cdot (\bar{p}_{\omega_i})_g^{(l)} + \sum_{g\in[m_l]}\sum_{\omega_i\in\Omega_g^{(l)}\setminus\Omega_f} \bar{\bar{p}}_g^{(l)} \cdot (\bar{p}_{\omega_i})_g^{(l)} \\ &= \sum_{g\in[m_l]}\sum_{\omega_i\in\Omega_g^{(l)}} \bar{\bar{p}}_g^{(l)} \cdot (\bar{p}_{\omega_i})_g^{(l)} = \sum_{g\in[m_l]} \left(\bar{\bar{p}}_g^{(l)} \cdot \sum_{\omega_i\in\Omega_g^{(l)}} (\bar{p}_{\omega_i})_g^{(l)}\right) = \sum_{g\in[m_l]} \left(\bar{\bar{p}}_g^{(l)} \cdot 1\right) = 1,\end{aligned}$$

where the complement set with respect to Ω is denoted by $(\cdot)^C$.

In the rest of this section, we denote the nominal probabilities of the *fixed* scenarios after dissection as $q_{\omega_i,g} = \pi_g^{(l)} \cdot (q_{\omega_i})_g^{(l)}, \forall \omega_i \in \Omega_f,\ g \in [m_l]$, where $q_{\omega_i} = \sum_{g\in[m_l]} q_{\omega_i,g}$ for any $\omega_i \in \Omega_f$. We use $\bar{\rho}_{max}$ to denote the maximal value of intra-group radii $\bar{\rho}_g$ among subgroups $g \in [m_l]$ (*i.e.*, $\bar{\rho}_{max} = \max_{g\in[m_l]} \bar{\rho}_g$). Subscript ϕ_{CR}^{θ} is used to represent all relevant ambiguity sets and risk measures induced by CR power divergence family with parameter θ. For instance, $\mathcal{R}_{\phi_{CR}^{\theta}(\rho)}$ denotes the risk measure induced by the CR power divergence with ambiguity set $\mathcal{P}_{\phi_{CR}^{\theta}(\rho)}$ using radius ρ of the original DRO, and $\widetilde{\mathcal{R}}_{\phi_{CR}^{\theta}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})} = \left(\widetilde{\mathcal{R}}^{\mathscr{G}}_{\phi_{CR}^{\theta}(\bar{\bar{\rho}})} \circ \widetilde{\mathcal{R}}^{\mathscr{F}|\mathscr{G}}_{\phi_{CR}^{\theta}(\bar{\boldsymbol{\rho}})}\right)$ denotes the risk measure after convolution using intra-group radii $\bar{\bar{\rho}}$ and inter-groups radius $\bar{\boldsymbol{\rho}}$, and so forth. Similarly, we use subscripts ϕ_v, ϕ_J, ϕ_χ^a, and W to denote VD, J-divergence, χ-divergence of order $a > 1$ and Wasserstein distance, respectively. These notations are used in the subsequent results and their proofs.

3.3 Lower-bound criteria for ϕ-divergences

We now present LB criteria for DRO formed via commonly used ϕ-divergences listed in Table 1 through scenario grouping. We begin with CR power divergence family and present the proof in detail. The LB criteria for the special cases of the CR power divergence family in Table 2 can be acquired from the below result.

Proposition 1. **(LB criteria for CR power divergences).** Consider the convolution formed by CR power divergence with parameter $\theta \neq 0, 1$. For $0 < \theta < 1$, suppose the radii

$\rho,\ \bar{\rho}_g,\ \bar{\bar{\rho}} \in \left[0,\ \frac{1}{\theta}+\frac{1}{(1-\theta)}\right],\ g \in [m_l]$. For $\theta < 0$ or $\theta > 1$, suppose the radii $\rho,\ \bar{\rho}_g,\ \bar{\bar{\rho}} \geq 0$ and the sample space Ω is dissected by disjoint partitions (*i.e.*, $\Omega_f = \emptyset$). If

$$\begin{cases} \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho & \text{when} \quad \theta \in (0,1), \\ \bar{\bar{\rho}} + \bar{\rho}_{max} - \theta(1-\theta)\cdot\bar{\bar{\rho}}\cdot\bar{\rho}_{max} \leq \rho & \text{when} \quad \theta < 0 \text{ or } \theta > 1 \text{ and } \Omega_f = \emptyset, \\ \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho & \text{when} \quad \theta \to 0 \text{ or } \theta \to 1, \end{cases}$$

then $\widetilde{\mathcal{R}}_{\phi^\theta_{CR}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{\phi^\theta_{CR}(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

Proof. Let $P' \in \widetilde{\mathcal{P}}_{\phi^\theta_{CR}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}$. Then there exists $\bar{\bar{P}} \in \widetilde{\mathcal{P}}^{\mathscr{G}}_{\phi^\theta_{CR}(\bar{\bar{\rho}})}$ and $\bar{P} \in \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\phi^\theta_{CR}(\bar{\boldsymbol{\rho}})}$ such that $\sum_{g\in[m_l]} \bar{\bar{p}}^{(l)}_g = 1$, $\sum_{\omega_i\in\Omega^{(l)}_g} (\bar{p}_{\omega_i})^{(l)}_g = 1$ and, by the definition of $\Delta_{\phi^\theta_{CR}}$ of Table 1, satisfying

$$\frac{1-\sum\limits_{g\in[m_l]}\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)} \leq \bar{\bar{\rho}}, \text{ and } \frac{1-\sum\limits_{\omega_i\in\Omega^{(l)}_g}\left((\bar{p}_{\omega_i})^{(l)}_g\right)^\theta\left((q_{\omega_i})^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)} \leq \bar{\rho}_g,\ \forall g \in [m_l]. \tag{5}$$

We now show the steps to find the criteria for $\Delta_{\phi^\theta_{CR}}(P', Q) \leq \rho$:

1. When $\theta \in (0,1)$, writing out $\Delta_{\phi^\theta_{CR}}(P', Q)$, we obtain

$$\begin{aligned}
&\frac{1-\sum\limits_{\omega_i\in\Omega}\left(p'_{\omega_i}\right)^\theta(q_{\omega_i})^{1-\theta}}{\theta(1-\theta)} = \frac{1-\sum\limits_{\omega_i\in\Omega_f}\left(p'_{\omega_i}\right)^\theta(q_{\omega_i})^{1-\theta} - \sum\limits_{\omega_i\in(\Omega_f)^C}\left(p'_{\omega_i}\right)^\theta(q_{\omega_i})^{1-\theta}}{\theta(1-\theta)} \\
=&\frac{1-\sum\limits_{\omega_i\in\Omega_f}\left(\sum\limits_{g\in[m_l]}p'_{\omega_i,g}\right)^\theta\left(\sum\limits_{g\in[m_l]}q_{\omega_i,g}\right)^{1-\theta} - \sum\limits_{\omega_i\in(\Omega_f)^C}\left(p'_{\omega_i}\right)^\theta(q_{\omega_i})^{1-\theta}}{\theta(1-\theta)} \\
\leq&\frac{1-\sum\limits_{\omega_i\in\Omega_f}\sum\limits_{g\in[m_l]}\left(p'_{\omega_i,g}\right)^\theta(q_{\omega_i,g})^{1-\theta} - \sum\limits_{\omega_i\in(\Omega_f)^C}\left(p'_{\omega_i}\right)^\theta(q_{\omega_i})^{1-\theta}}{\theta(1-\theta)} \qquad (6) \\
=&\frac{1-\sum\limits_{g\in[m_l]}\sum\limits_{\omega_i\in\Omega^{(l)}_g}\left(\bar{\bar{p}}^{(l)}_g(\bar{p}_{\omega_i})^{(l)}_g\right)^\theta\left(\pi^{(l)}_g(q_{\omega_i})^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)} + \left(\frac{-\sum\limits_{g\in[m_l]}\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta} + \sum\limits_{g\in[m_l]}\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)}\right) \qquad (7) \\
=&\frac{1-\sum\limits_{g\in[m_l]}\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)} + \sum_{g\in[m_l]}\left[\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}\frac{1-\sum\limits_{\omega_i\in\Omega^{(l)}_g}\left((\bar{p}_{\omega_i})^{(l)}_g\right)^\theta\left((q_{\omega_i})^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)}\right] \\
\leq&\bar{\bar{\rho}} + \sum_{g\in[m_l]}\left[\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}\bar{\rho}_g\right] \leq \bar{\bar{\rho}} + \sum_{g\in[m_l]}\left[\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}\right]\cdot\bar{\rho}_{max} \qquad (8) \\
=&\bar{\bar{\rho}} + \bar{\rho}_{max} - \theta(1-\theta)\frac{1-\sum\limits_{g\in[m_l]}\left(\bar{\bar{p}}^{(l)}_g\right)^\theta\left(\pi^{(l)}_g\right)^{1-\theta}}{\theta(1-\theta)}\cdot\bar{\rho}_{max}, \qquad (9)
\end{aligned}$$

where inequality (6) follows from Hölder's inequality applied on the fixed scenarios. The

first inequality in (8) follows from (5) and the second from definition of $\bar{\rho}_{max}$. Let us denote the right-hand side of (9) as A. Since $-\theta(1-\theta)$ is negative, $A \leq \bar{\bar{\rho}} + \bar{\rho}_{max}$. Therefore, if $\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$, we have $\widetilde{\mathcal{P}}_{\phi^{\theta}_{CR}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})} \subseteq \mathcal{P}_{\phi^{\theta}_{CR}(\rho)}$. This implies that for any $\boldsymbol{\eta} \in \mathcal{Z}$, we obtain $\widetilde{\mathcal{R}}_{\phi^{\theta}_{CR}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) = \max_{P' \in \widetilde{\mathcal{P}}_{\phi^{\theta}_{CR}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}} \mathbb{E}_{P'}[\boldsymbol{\eta}] \leq \max_{P \in \mathcal{P}_{\phi^{\theta}_{CR}(\rho)}} \mathbb{E}_{P}[\boldsymbol{\eta}] = \mathcal{R}_{\phi^{\theta}_{CR}(\rho)}(\boldsymbol{\eta})$.

2. When $\theta < 0$ or $\theta > 1$, we can no longer apply Hölder's inequality on the fixed scenarios in (6). However, when $\Omega_f = \emptyset$ (*i.e.*, disjoint partitions are used), we can directly start from (7) and follow the steps to (9). Since $-\theta(1-\theta)$ is positive, $A \leq \bar{\bar{\rho}} + \bar{\rho}_{max} - \theta(1-\theta) \cdot \bar{\bar{\rho}} \cdot \bar{\rho}_{max}$ by (5). Therefore, if $\bar{\bar{\rho}} + \bar{\rho}_{max} - \theta(1-\theta) \cdot \bar{\bar{\rho}} \cdot \bar{\rho}_{max} \leq \rho$, the result follows.

3. When $\theta \to 1$, the CR power divergence is equivalent to Kullback-Leibler divergence. Detailed proof is provided in Appendix A.1.

4. When $\theta \to 0$, the proof is similar to the $\theta \to 1$ case and hence omitted. □

Proof of Proposition 1 and Appendix A.1 reveal that when $\theta \in (0,1)$ or in the limit cases of Kullback-Leibler divergence ($\theta \to 1$) and Burg entropy ($\theta \to 0$), the sample space Ω can be dissected in any way, either using disjoint partitions or fixed scenarios. However, when $\theta < 0$ or $\theta > 1$, the above result is valid for disjoint partitions.

LB criteria for other ϕ-divergences in Table 1 can be obtained using a similar proof technique. Below, we provide the results and relegate the proofs to Appendix A.

Proposition 2. **(LB criterion for variation distance).** Consider the convolution formed by variation distance, where radii $\rho, \bar{\rho}_g, \bar{\bar{\rho}} \in [0,2],\ g \in [m_l]$. If

$$\bar{\bar{\rho}} \cdot \bar{\rho}_{max} + \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho,$$

then $\widetilde{\mathcal{R}}_{\phi_v(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{\phi_v(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

Proposition 3. **(LB criterion for J-divergence).** Consider the convolution formed by J-divergence, where radii $\rho,\ \bar{\rho}_g,\ \bar{\bar{\rho}} \geq 0,\ g \in [m_l]$. If

$$\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho,$$

then $\widetilde{\mathcal{R}}_{\phi_J(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{\phi_J(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

Proposition 4. **(LB criterion for χ-divergence of order $a > 1$).** Consider the convolution formed by χ-divergence of order $a > 1$, where radii $\rho,\ \bar{\rho}_g,\ \bar{\bar{\rho}} \geq 0,\ g \in [m_l]$ and the sample space Ω is dissected by disjoint partitions (*i.e.*, $\Omega_f = \emptyset$). If

$$\left[(\bar{\bar{\rho}})^{\frac{1}{a}} + (\bar{\rho}_{max})^{\frac{1}{a}} + (\bar{\bar{\rho}} \cdot \bar{\rho}_{max})^{\frac{1}{a}} \right]^a \leq \rho,$$

then $\widetilde{\mathcal{R}}_{\phi^a_\chi(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{\phi^a_\chi(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

Note that Proposition 4 is a general result that applies to all values of $a > 1$. For certain values of a, a tighter inequality might be available. For instance, when $a = 2$, by directly using $\Delta_{\phi^a_\chi}(P', Q) = \sum_{\omega_i \in \Omega} \frac{(p'_{\omega_i})^2}{q_{\omega_i}} - 1$ instead of Minkowski's inequality (see Appendix A.4), we obtain the LB criterion $\bar{\bar{\rho}} \cdot \bar{\rho}_{max} + \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$.

3.4 Lower-bound criterion for Wasserstein distance

We now provide a LB criterion for Wasserstein distance through scenario decomposition. The main idea is the same: to find criteria that guarantee the convoluted ambiguity set being a subset of the ambiguity set of the original problem. Recall that Wasserstein distance needs a distance d_{ω_i,ω_j} between any two scenarios ω_i and ω_j. To apply it to scenario groups, we need a distance between subgroups as well. We provide such a distance between subgroups and a criterion for radii $\bar{\bar{\rho}}$, $\bar{\boldsymbol{\rho}}$ to ensure LBs below.

Proposition 5. **(LB criterion for Wasserstein distance).** Consider the convolution of DRO formed by Wasserstein distance, where radii $\rho,\ \bar{\rho}_g,\ \bar{\bar{\rho}} \geq 0,\ g \in [m_l]$. Let the distance between scenario groups be defined as

$$d_{g_1,g_2} := \begin{cases} \max_{\omega_i \in \Omega^{(l)}_{g_1},\, \omega_j \in \Omega^{(l)}_{g_2}} \{d_{\omega_i,\omega_j}\} & \text{when} \quad g_1 \neq g_2 \\ 0 & \text{when} \quad g_1 = g_2, \end{cases} \tag{10}$$

where $g_1,\ g_2 \in [m_l]$. If

$$\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho,$$

then $\widetilde{\mathcal{R}}_{W(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{W(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

Proof. First assume Ω is dissected using *disjoint* partitions (*i.e.*, $\Omega_f = \emptyset$). Given $P' \in \widetilde{\mathcal{P}}_{W(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}$ formed after scenario grouping and convolution, the Wasserstein distance between P' and Q for the original problem can be written as

$$\Delta_W(P', Q) := \min_{z \geq 0} \left\{ \sum_{\omega_i \in \Omega} \sum_{\omega_j \in \Omega} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} : \sum_{\omega_i \in \Omega} z_{\omega_i,\omega_j} = q_{\omega_j}, \forall \omega_j \in \Omega, \sum_{\omega_j \in \Omega} z_{\omega_i,\omega_j} = p'_{\omega_i}, \forall \omega_i \in \Omega \right\}, \tag{11}$$

and the Wasserstein distance for ambiguity set $\widetilde{\mathcal{P}}^{\mathcal{G}}_{W(\bar{\bar{\rho}})}$ can be written as

$$\Delta_W(\bar{\bar{P}}, \bar{\bar{Q}}) := \min_{y \geq 0} \left\{ \sum_{g_1 \in [m_l]} \sum_{g_2 \in [m_l]} d_{g_1,g_2}\, y_{g_1,g_2} : \sum_{g_1 \in [m_l]} y_{g_1,g_2} = \pi^{(l)}_{g_2}, \forall g_2 \in [m_l], \sum_{g_2 \in [m_l]} y_{g_1,g_2} = \bar{\bar{p}}^{(l)}_{g_1}, \forall g_1 \in [m_l] \right\}. \tag{12}$$

Similarly, for each subgroup $g \in [m_l]$, we have

$$\Delta_W(\bar{P}_g^{(l)}, Q_g^{(l)}) := \min_{x\geq 0} \Bigg\{ \sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j}(x_{\omega_i,\omega_j})_g : \sum_{\omega_i\in\Omega_g^{(l)}} (x_{\omega_i,\omega_j})_g = (q_{\omega_j})_g^{(l)},\ \omega_j \in \Omega_g^{(l)}, \\ \sum_{\omega_j\in\Omega_g^{(l)}} (x_{\omega_i,\omega_j})_g = (\bar{p}_{\omega_i})_g^{(l)},\ \omega_i \in \Omega_g^{(l)} \Bigg\}. \tag{13}$$

Let us now define one way to obtain z_{ω_i,ω_j} in (11) by using the decision variables y_{g_1,g_2} and $(x_{\omega_i,\omega_j})_g$ in (12) and (13), respectively:

$$z_{\omega_i,\omega_j} = y_{g,g}\ (x_{\omega_i,\omega_j})_g, \qquad \forall \omega_i,\omega_j \in \Omega_g^{(l)},\ \forall g \in [m_l] \tag{14}$$

$$\sum_{\omega_i\in\Omega_{g_1}^{(l)}} z_{\omega_i,\omega_j} = y_{g_1,g_2} \sum_{\omega_i\in\Omega_{g_2}^{(l)}} (x_{\omega_i,\omega_j})_{g_2}, \qquad \forall \omega_j \in \Omega_{g_2}^{(l)},\ \forall g_2 \in [m_l] \tag{15}$$

$$\sum_{\omega_j\in\Omega_{g_2}^{(l)}} z_{\omega_i,\omega_j} = y_{g_1,g_2} \sum_{\omega_j\in\Omega_{g_1}^{(l)}} (x_{\omega_i,\omega_j})_{g_1}, \qquad \forall \omega_i \in \Omega_{g_1}^{(l)},\ \forall g_1 \in [m_l]. \tag{16}$$

Above, we also assume $z \geq 0$. (This is automatically satisfied for within-subgroup z_{ω_i,ω_j} by (14) given $y \geq 0$ and $x \geq 0$.) With the transformation above, we can show that the constraints in (11) of the Wasserstein distance $\Delta_W(P', Q)$ are all satisfied, even though z_{ω_i,ω_j} formed through (14)–(16) may not be optimal to $\Delta_W(P', Q)$. For instance, the first set of constraints in (11) for all $\omega_j \in \Omega$ (or equivalently all $\omega_j \in \Omega_g^{(l)},\ g \in [m_l]$) are satisfied by (15) and the first sets of constraints in (12) and (13):

$$\begin{aligned}
\sum_{\omega_i\in\Omega} z_{\omega_i,\omega_j} &= \sum_{g_1\in[m_l]} \sum_{\omega_i\in\Omega_{g_1}^{(l)}} z_{\omega_i,\omega_j} \\
&= \sum_{g_1\in[m_l]} \left(y_{g_1,g} \sum_{\omega_i\in\Omega_g^{(l)}} (x_{\omega_i,\omega_j})_g \right) \\
&= \left(\sum_{g_1\in[m_l]} y_{g_1,g} \right) \left(\sum_{\omega_i\in\Omega_g^{(l)}} (x_{\omega_i,\omega_j})_g \right) \\
&= \pi_g^{(l)}\ (q_{\omega_j})_g^{(l)} = q_{\omega_j},
\end{aligned}$$

where g denotes the subgroup scenario ω_j belongs to. The second set of constraints in (11) can be shown similarly by using (16) and the second sets of constraints in (12) and (13). Hence, all feasible solutions to constraints in (12) and (13) are also feasible to the constraints in (11).

Let $\mathbb{Z}$, $\mathbb{Y}$, and $\mathbb{X}$ denote the feasible regions given by the constraints in (11), (12), and (13), respectively, each supplemented with their nonnegativity constraints $z \geq 0$, $y \geq 0$, and $x \geq 0$. We now show steps to find criteria for $\Delta_W(P', Q) \leq \rho$:

$$\Delta_W(P',Q) = \min_{z\in\mathbb{Z}} \sum_{\omega_i\in\Omega}\sum_{\omega_j\in\Omega} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j}$$

$$= \min_{z\in\mathbb{Z}} \sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} \left(\sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} + \sum_{\omega_j\in\left(\Omega_g^{(l)}\right)^C} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} \right)$$

$$\leq \min_{\substack{z\in\mathbb{Z}\\ y\in\mathbb{Y}\\ x\in\mathbb{X}}} \sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} \left(\sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j} y_{g,g}(x_{\omega_i,\omega_j})_g + \sum_{\omega_j\in\left(\Omega_g^{(l)}\right)^C} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} \right) \tag{a}$$

$$\leq \min_{\substack{z\in\mathbb{Z}\\ y\in\mathbb{Y}\\ x\in\mathbb{X}}} \sum_{g\in[m_l]} \left(\sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j} y_{g,g}(x_{\omega_i,\omega_j})_g + \sum_{g_2\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_{g2}^{(l)}} d_{g,g_2} z_{\omega_i,\omega_j} \right) \tag{b}$$

$$\leq \min_{\substack{y\in\mathbb{Y}\\ x\in\mathbb{X}}} \sum_{g\in[m_l]} \left(y_{g,g} \sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j} (x_{\omega_i,\omega_j})_g \right) + \sum_{g\in[m_l]} \sum_{g_2\in[m_l]} d_{g,g_2} y_{g,g_2} \tag{c}$$

$$\leq \sum_{g\in[m_l]} \left(\pi_g^{(l)} \min_{x\in\mathbb{X}} \sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_g^{(l)}} d_{\omega_i,\omega_j} (x_{\omega_i,\omega_j})_g \right) + \min_{y\in\mathbb{Y}} \sum_{g\in[m_l]} \sum_{g_2\in[m_l]} d_{g,g_2} y_{g,g_2} \tag{d}$$

$$= \sum_{g\in[m_l]} \pi_g^{(l)} \Delta_W(\bar{P}_g^{(l)}, Q_g^{(l)}) + \Delta_W(\bar{\bar{P}}, \bar{\bar{Q}}) \leq \sum_{g\in[m_l]} \pi_g^{(l)} \cdot \bar{\rho}_g + \bar{\bar{\rho}} \leq \bar{\bar{\rho}} + \bar{\rho}_{max}, \tag{e}$$

where (a) follows from (14). Note that this is an inequality because decisions z obtained through this transformation may not be optimal. Inequality (b) follows (10). By summing over $\omega_j \in \Omega_{g2}^{(l)}$ on both sides of (15), we can show $y_{g,g_2} = \sum_{\omega_i\in\Omega_g^{(l)}} \sum_{\omega_j\in\Omega_{g2}^{(l)}} z_{\omega_i,\omega_j}$, $\forall g, g_2 \in [m_l]$. Then (c) follows. Inequality (d) follows from $y_{g,g} \leq \pi_g^{(l)}$ (see (12)) and the fact that the resulting problem is separable. Finally, the equality in (e) follows from the definition of $\Delta_W(\bar{P}_g^{(l)}, Q_g^{(l)})$ and $\Delta_W(\bar{\bar{P}}, \bar{\bar{Q}})$, and the inequality in (e) follows by construction. Therefore, similar to the statement at the end of the proof of part 1 of Proposition 1, when $\theta \in (0,1)$, if $\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$, then $\widetilde{\mathcal{R}}_{W(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}(\boldsymbol{\eta}) \leq \mathcal{R}_{W(\rho)}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathcal{Z}$.

We now consider the case when Ω is dissected using *fixed* scenarios ($\Omega_f \neq \emptyset$). Recall in (4), for any fixed scenario $\omega_i \in \Omega_f$, we have $p'_{\omega_i,g} = \bar{\bar{p}}_g^{(l)} \ (\bar{p}_{\omega_i})_g^{(l)}$, $g \in [m_l]$ and $p'_{\omega_i} = \sum_{g\in[m_l]} p'_{\omega_i,g}$. Hence splitting the two constraints in (11), we have

$$\sum_{\omega_i\in\Omega} z_{\omega_i,\omega_j} = q_{\omega_j} = \sum_{g\in[m_l]} q_{\omega_j,g} = \sum_{g\in[m_l]} \pi_g^{(l)} \ (q_{\omega_j})_g^{(l)}, \qquad \omega_j \in \Omega_f, \tag{17}$$

$$\sum_{\omega_i\in\Omega} z_{\omega_i,\omega_j} = q_{\omega_j} = \pi_g^{(l)} \ (q_{\omega_j})_g^{(l)}, \qquad \omega_j \in (\Omega_f)^C, \tag{18}$$

$$\sum_{\omega_j\in\Omega} z_{\omega_i,\omega_j} = p'_{\omega_i} = \sum_{g\in[m_l]} p'_{\omega_i,g} = \sum_{g\in[m_l]} \bar{\bar{p}}_g^{(l)} \ (\bar{p}_{\omega_i})_g^{(l)}, \qquad \omega_i \in \Omega_f, \tag{19}$$

$$\sum_{\omega_j\in\Omega} z_{\omega_i,\omega_j} = p'_{\omega_i} = \bar{\bar{p}}_g^{(l)} \ (\bar{p}_{\omega_i})_g^{(l)}, \qquad \omega_i \in (\Omega_f)^C. \tag{20}$$

Define a finite expanded space $\widetilde{\Omega} := \{\omega_{1_{(1)}}, \omega_{1_{(2)}}, \ldots, \omega_{1_{(m_l)}}, \omega_{2_{(1)}}, \omega_{2_{(2)}}, \ldots, \omega_{2_{(m_l)}}, \ldots, \omega_{f_{(1)}}, \omega_{f_{(2)}}, \ldots, \omega_{f_{(m_l)}}, \omega_{f+1}, \omega_{f+2}, \ldots, \omega_{|\Omega|}\}$, where the fixed scenarios $\omega_i \in \Omega_f = \{\omega_1, \ldots, \omega_f\}$ in different subgroups are considered to have different "atoms" and the rest of the scenarios $\omega_i \in (\Omega_f)^C$ are left as before. We again use the same subgroups $\Omega_g^{(l)}$ but on the expanded space $\widetilde{\Omega}$. Then, we have the following Wasserstein distance on $\widetilde{\Omega}$:

$$\widetilde{\Delta}_W(P', Q) := \min_{\tilde{z} \geq 0} \Big\{ \sum_{\omega_i \in \widetilde{\Omega}} \sum_{\omega_j \in \widetilde{\Omega}} d_{\omega_i,\omega_j} \tilde{z}_{\omega_i,\omega_j} : \quad \sum_{\omega_i \in \widetilde{\Omega}} \tilde{z}_{\omega_i,\omega_j} = \pi_g^{(l)} \, (q_{\omega_j})_g^{(l)}, \forall \omega_j \in \widetilde{\Omega}, \quad \sum_{\omega_j \in \widetilde{\Omega}} \tilde{z}_{\omega_i,\omega_j} = \bar{\bar{p}}_g^{(l)} \, (\bar{p}_{\omega_i})_g^{(l)}, \forall \omega_i \in \widetilde{\Omega} \Big\}. \tag{21}$$

For any $\tilde{z} \geq 0$ feasible to Wasserstein distance $\widetilde{\Delta}_W(P', Q)$—including the optimal $\tilde{z}$—on the expanded space $\widetilde{\Omega}$, we can generate a feasible solution $z \geq 0$ to (17)–(20). First, for any non-fixed scenarios $\omega_i, \omega_j \in (\Omega_f)^C$, we set $z_{\omega_i,\omega_j} = \tilde{z}_{\omega_i,\omega_j}$ and observe the constraints in (21) are the same as (18) and (20). Next, for any fixed scenario $\omega_i \in \Omega_f$ and non-fixed scenario $\omega_j \in (\Omega_f)^C$, we set (i) $z_{\omega_i,\omega_i} = \sum_{g_1 \in [m_l]} \sum_{g_2 \in [m_l]} \tilde{z}_{\omega_{i_{(g_1)}},\omega_{i_{(g_2)}}}$, (ii) $z_{\omega_i,\omega_j} = \sum_{g \in [m_l]} \tilde{z}_{\omega_{i_{(g)}},\omega_j}$, and (iii) $z_{\omega_j,\omega_i} = \sum_{g \in [m_l]} \tilde{z}_{\omega_j,\omega_{i_{(g)}}}$. Then (17) and (19) are also satisfied. Furthermore, with this z, the objective functions of two Wasserstein distances coincide: $\sum_{\omega_i \in \Omega} \sum_{\omega_j \in \Omega} d_{\omega_i,\omega_j} z_{\omega_i,\omega_j} = \sum_{\omega_i \in \widetilde{\Omega}} \sum_{\omega_j \in \widetilde{\Omega}} d_{\omega_i,\omega_j} \tilde{z}_{\omega_i,\omega_j}$ because any distance involving $\omega_{i_{(g)}} \in \widetilde{\Omega}$ is equivalent to distance involving $\omega_i \in \Omega_f$, *e.g.*, $d_{\omega_{i_{(g_1)}},\omega_{i_{(g_2)}}} = 0$ for all $\omega_i \in \Omega_f$, $g_1, g_2 \in [m_l]$. As a result, $\Delta_W(P', Q) \leq \widetilde{\Delta}_W(P', Q)$. Observe $\widetilde{\Delta}_W(P', Q)$ is obtained as "disjoint" partitions on $\widetilde{\Omega}$. Then, following similar steps to the disjoint partition, we show $\widetilde{\Delta}_W(P', Q) \leq \bar{\bar{\rho}} + \bar{\rho}_{max}$. Therefore, if $\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$ the result follows. □

Remark. Although the above propositions use $\bar{\rho}_{max}$, these results can also be obtained using the individual intra-group radius $\bar{\rho}_g$ values for each scenario group $g \in [m_l]$. For instance, the condition for Wasserstein would be $\bar{\bar{\rho}} + \sum_{g \in [m_l]} \pi_g^{(l)} \cdot \bar{\rho}_g \leq \rho$ and the result for J-divergence would be $\bar{\bar{\rho}} + \sum_{g \in [m_l]} \left(\bar{\bar{p}}_g^{(l)} \cdot \bar{\rho}_g \right) \leq \rho$. In our numerical results, we use the same value for each group, *i.e.*, $\bar{\rho}_g = \bar{\rho}_{max}$ for all $g \in [m_l]$.

3.5 Lower bounds for multistage problems

We now extend the LBs obtained in Sections 3.3 and 3.4 to multistage DRO. Due to the correspondence between DRO and RASO, here we focus on a RASO formulation and present our results and proofs using the properties of conditional coherent risk measures.

We first recall the results in [37]. For a multistage decision horizon with stages $t \in \mathcal{T}$ let $\mathcal{Z}_t := \mathfrak{L}_\infty(\Omega_t, \mathscr{F}_t, Q_t)$. The mapping $\mathcal{R}_{\rho_{t+1}}^{\mathscr{F}_{t+1}|\mathscr{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is called one-step conditional

risk measure if it satisfies the properties presented in Section 2.4 for corresponding spaces $\mathcal{Z}_t$ and $\mathcal{Z}_{t+1}$ for all $t \in \{0, \ldots, T-1\}$. The risk involved in a sequence of random variables $\boldsymbol{\eta}_t \in \mathcal{Z}_t$, $t \in \mathcal{T}$ adapted to the filtration $\mathscr{F}_t$, $t \in \mathcal{T}$ can be evaluated by a time-consistent dynamic risk measure $\mathfrak{R}_{\boldsymbol{\rho}}$ induced by a measure of similarity between distributions Δ using radii $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_T)$, defined as follows:

$$\mathfrak{R}_{\boldsymbol{\rho}}(\boldsymbol{\eta}_0, \ldots, \boldsymbol{\eta}_T) := \boldsymbol{\eta}_0 + \mathcal{R}_{\rho_1}^{\mathscr{F}_1|\mathscr{F}_0}\Big(\boldsymbol{\eta}_1 + \mathcal{R}_{\rho_2}^{\mathscr{F}_2|\mathscr{F}_1}(\boldsymbol{\eta}_2 + \ldots + \mathcal{R}_{\rho_T}^{\mathscr{F}_T|\mathscr{F}_{T-1}}(\boldsymbol{\eta}_T))\Big). \tag{22}$$

In general, it is not necessary to use the same Δ at each stage of the problem, but we do so in this paper for simplicity. Also, by changing the radii ρ_t we can choose how close we remain to the nominal distributions at different stages. Setting $\boldsymbol{\eta}_t := c_t(\mathbf{x}_t, \boldsymbol{\xi}_t)$ at stages $t \in \mathcal{T}$ and using (22), the multistage RASO problem can be formulated as

$$\min_{\mathbf{x}_0 \in \mathcal{X}_0(\boldsymbol{\xi}_0)} \; c_0(\mathbf{x}_0, \boldsymbol{\xi}_0) + \mathcal{R}_{\rho_1}\Big(\mathcal{Q}_1(\mathbf{x}_0, \boldsymbol{\xi}^0)\Big) \tag{23}$$

where

$$\mathcal{Q}_1(\mathbf{x}_0, \boldsymbol{\xi}^0) := \min_{\mathbf{x}_t \in \mathcal{X}_t(\mathbf{x}_{t-1}, \boldsymbol{\xi}_t),\, t \in \mathcal{T}\setminus\{0\}} \; \mathfrak{R}_{\rho_2, \ldots, \rho_T}\left(c_1(\mathbf{x}_1, \boldsymbol{\xi}_1), \ldots, c_T(\mathbf{x}_T, \boldsymbol{\xi}_T)\right). \tag{24}$$

Let $\mathbf{x}_0^*$ and z^* be an optimal first-stage solution and the optimal value of (23)–(24), respectively.

We now introduce two approaches that provide different types of LBs to the original multistage DRO problem. Our first approach, which we refer to as *first-level LB*, allows changes to only the first-stage radius ρ_1; hence the name 'first-level'. This approach applies to risk measures induced by any ϕ-divergence, and the scenario tree can be dissected in any form. Our second approach, referred to as *multi-level LB*, instead allows changes to several layers of ρ_t, based on how the scenario tree is dissected, and then applies the convolution step to combine the subgroups' optimal values in a nested fashion following the structure of the scenario tree. Because of the nested application of the convolution step, the scenario tree needs to be dissected in a specific way. The multi-level LB applies to both the Wasserstein distance and the ϕ-divergences. We now examine these approaches and their differences in more detail.

3.5.1 First-level bounding scheme

Our first approach is formed as follows, similar to the two-stage case. Consider the collection of subsets $\Omega = \cup_{g \in [m_l]} \Omega_g^{(l)}$ and its induced σ-algebra $\mathscr{G}$. On each subgroup $g \in [m_l]$, we solve problem (23)–(24) with sample space $\Omega_g^{(l)}$, where ρ_1 is replaced by $\bar{\rho}_g$, while ρ_t for $t > 1$ are unchanged. Let $z_g^{*(l)}$ be its optimal value. Also let $\boldsymbol{\zeta}_{LB} := \{z_g^{*(l)}\}_{g=1}^{m_l}$

be a $\mathscr{G}$-measurable random variable with probabilities $\pi_g^{(l)}$, $\forall g \in [m_l]$. A LB on z^* can be obtained by applying the LB risk measure $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}$ introduced in the previous section to $\boldsymbol{\zeta}_{LB}$, hence computing $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$. In this approach, only the radius at the first stage is replaced, and there is no need to change the radii from stage 2 to stage T. Below we show that this approach provides a valid LB for multistage DRO formed by ϕ-divergences.

Proposition 6.**(First-level LB for multistage DRO).** Given problem (23)–(24), assume the risk measure at each stage is induced by a ϕ-divergence. Consider the risk measure $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}} : \mathfrak{L}_\infty(\Omega, \mathscr{G}, Q) \to \mathbb{R}$ to combine the subgroups and the one-step conditional risk measure on the subgroups $\widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}} : \mathfrak{L}_\infty(\Omega, \mathscr{F}, Q) \to \mathfrak{L}_\infty(\Omega, \mathscr{G}, Q)$ (*i.e.*, $\mathcal{R}_{\bar{\rho}_g}^{(l)} : \mathfrak{L}_\infty(\Omega, \sigma(\Omega_g^{(l)}), Q) \to \mathbb{R}, \forall g \in [m_l]$ with $\sigma(\Omega_g^{(l)})$ the σ-algebra on $\Omega_g^{(l)}$), where these risk measures are induced by the same type of ϕ-divergence. If $\bar{\rho}$ and $\bar{\rho}_{max}$ satisfy the corresponding criteria from one of the Propositions 1-4 with $\rho = \rho_1$, then $z^* \geq \widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$.

Proof. If $\mathbf{x}_0^*$ is an optimal first-stage solution of (23)–(24), then it is a feasible first-stage solution for each subgroup problem $g \in [m_l]$. Thus, we have $c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \mathcal{R}_{\bar{\rho}_g}^{(l)}(\mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0)) \geq z_g^{*(l)}$, $\forall g \in [m_l]$, or equivalently $c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}(\mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0)) \geq \boldsymbol{\zeta}_{LB}$. Both sides of this inequality are $\mathscr{G}$-measurable, and since $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}$ is a coherent risk measure that satisfies the monotonicity property (see Section 2.4), we obtain $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}\left(c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}(\mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0))\right) \geq \widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$. We can now apply translation equivariance property (see Section 2.4) to the left-hand side of above inequality to get $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}\left(\widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}}^{\mathscr{F}|\mathscr{G}}(c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0))\right) \geq \widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$. Since the criteria from Propositions 1-4 are satisfied, we obtain $\mathcal{R}_{\rho_1}(c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0)) \geq \widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$. Using once more the translation equivariance property, we reach the result $z^* = c_0(\mathbf{x}_0^*, \boldsymbol{\xi}_0) + \mathcal{R}_{\rho_1}(\mathcal{Q}_1(\mathbf{x}_0^*, \boldsymbol{\xi}^0)) \geq \widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}(\boldsymbol{\zeta}_{LB})$. See also [25]. □

In general, first-level LB cannot be applied when the risk measure $\widetilde{\mathcal{R}}_{\bar{\rho}}^{\mathscr{G}}$ is induced by the Wasserstein distance. This is because the distance between subgroups provided in (10) for the two-stage case does not hold anymore in the multistage setting. Recall that we need a distance between subgroups to perform the convolution step. Thus, in the following section, we introduce a multi-level bounding scheme that reduces the computation of the distance between subgroups in the multistage setting to a *recursive* application of (10), which was devised for the two-stage setting.

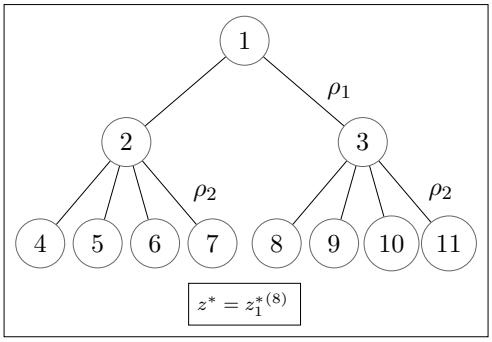
3.5.2 Multi-level bounding scheme

Our second approach applies to risk measures induced by any ϕ-divergence or Wasserstein distance. Let us first introduce the following definition of dissecting a scenario tree.

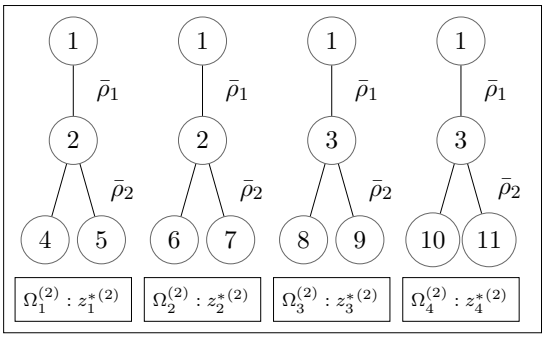
Definition 1. (Dissected up to stage τ). Let $\Omega_{t,g}^{(l)}$ denote set of nodes of subgroup $\Omega_g^{(l)}$ at stage $t \in \mathcal{T}$. A scenario tree $\mathfrak{T}$ is dissected up to stage $\tau \in \mathcal{T} \setminus \{0\}$ if

1. for every subgroup $\Omega_g^{(l)}$, $g \in [m_l]$ all nodes at stage τ of that subgroup have the same ancestor, *i.e.*, $a(n') = a(n'')$, $n', n'' \in \Omega_{\tau,g}^{(l)}$, for all subgroups $g \in [m_l]$;
2. all the children of stage-τ ($\tau < T$) nodes belong to the same subgroup, *i.e.*, $\mathscr{B}(n) \subseteq \Omega_{\tau+1,g}^{(l)}$, for all stage-$\tau$ nodes $n \in \Omega_{\tau,g}^{(l)}$, $\tau \neq T$ for all subgroups $g \in [m_l]$;
3. subgroups are disjoint, i.e., $\Omega_{\tau,g_1}^{(l)} \cap \Omega_{\tau,g_2}^{(l)} = \emptyset$, for any two $g_1, g_2 \in [m_l]$, $g_1 \neq g_2$.

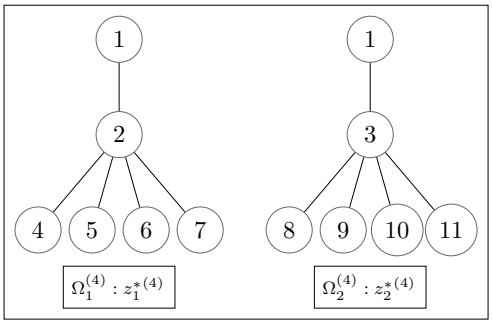
According to Definition 1, we split the ambiguity sets up to stage τ while the ambiguity sets at subsequent stages are not modified. Furthermore, all the subgroups involved in the splitting of a given ambiguity set at stage τ share the same single path up to stage $\tau - 1$. For example, in Figure 2b, the scenario tree depicted in Figure 2a is dissected up to stage $\tau = 2$. Consequently, subgroups $\Omega_1^{(2)}$ and $\Omega_2^{(2)}$ share the same nodes 1 and 2 up to stage $\tau - 1 = 1$. Similarly, subgroups $\Omega_3^{(2)}$ and $\Omega_4^{(2)}$ share the same nodes 1 and 3 up to stage $\tau - 1 = 1$. Furthermore, according to Definition 1, a tree dissected up to stage $\tau \in \mathcal{T} \setminus \{0, 1\}$ forms a refinement of a tree dissected up to stage $\tau - 1$. Compare, for instance, the dissections in Figure 2b and Figure 2c. Similarly, dissection up to stage $\tau = 1$ forms a refinement of the tree $\mathfrak{T}$.



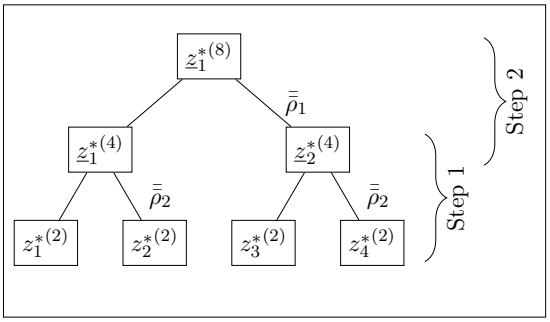
(a) Scenario tree $\mathfrak{T}$ with $\mathcal{T} := \{0, 1, 2\}$.



(b) Tree $\mathfrak{T}$ dissected up to stage $\tau = 2$.



(c) Tree $\mathfrak{T}$ dissected up to stage $\tau = 1$.



(d) Multi-level LB for dissection in part (b).

Figure 2: Visual representation of the multi-level bounding scheme.

The multi-level LB scheme works as follows. Any stage-t radii for $t \leq \tau$ can be changed to the intra-group radii $\bar{\rho}_1, \ldots, \bar{\rho}_\tau$ for subgroup problems, but all intra-group radii for stages $t > \tau$ are kept the same as the original problem $\rho_{\tau+1}, \ldots, \rho_T$. Then, the optimal values of the subgroups are combined in a nested fashion, following the structure of the tree, using the convolution method described in Section 3.2 starting from stage τ up to stage 1 with inter-groups radii $\bar{\bar{\rho}}_\tau, \ldots, \bar{\bar{\rho}}_1$. Before we present the theoretical result, let us first illustrate it with an example.

Example 2. Consider the scenario tree $\mathfrak{T}$ in Figure 2a, which is dissected up to stage $\tau = 2$ in Figure 2b. Viewing subgroups $\Omega_1^{(2)}$ and $\Omega_2^{(2)}$ in Figure 2b as a dissection of the first subset $\Omega_1^{(4)}$ in Figure 2c, we apply the convolution method in Section 3.2 using $\bar{\rho}_2$ and $\bar{\bar{\rho}}_2$. If these radii satisfy the criteria presented in Sections 3.3 and 3.4, we obtain a LB on the optimal value of the first subset $\Omega_1^{(4)}$ in Figure2c: $\underline{z}_1^{*(4)} \leq z_1^{*(4)}$. In the case of Wasserstein distance, the distance between the subgroups $\Omega_1^{(2)}$ and $\Omega_2^{(2)}$ is obtained by $\max\{d_{46}{=}d_{64},\ d_{47}{=}d_{74},\ d_{56}{=}d_{65},\ d_{57}{=}d_{75}\}$, where d_{ij} denotes the distance between $\boldsymbol{\xi}_\tau$ in nodes i and j. The same procedure is performed on the other two subgroups $\Omega_3^{(2)}$ and $\Omega_4^{(2)}$, obtaining a LB for the second subset $\Omega_2^{(4)}$ in Figure 2c: $\underline{z}_2^{*(4)} \leq z_2^{*(4)}$. Now, instead of $z_1^{*(4)}, z_2^{*(4)}$, using their LBs $\underline{z}_1^{*(4)}$, $\underline{z}_2^{*(4)}$, the convolution is applied once again to the subsets $\Omega_1^{(4)}$ and $\Omega_2^{(4)}$ in Figure 2c, which form a dissection of the tree $\mathfrak{T}$ in Figure 2a. Here, the distance between subgroups $\Omega_1^{(4)}$ and $\Omega_2^{(4)}$ is given by $d_{23} = d_{32}$. Figure 2d shows a graphical representation of this process.

Proposition 7. **(Multi-level LB for multistage DRO).** Let scenario tree $\mathfrak{T}$ be dissected up to stage τ in subgroups $\Omega_g^{(l)}$, $g \in [m_l]$, according to Definition 1. Let $z_g^{*(l)}$, $g \in [m_l]$ be the optimal values of problem (23)–(24) with sample space $\Omega_g^{(l)}$, where ρ_t is replaced by $\bar{\rho}_{t,g}$, $t = 1, \ldots, \tau$ and $\bar{\rho}_{t,max} := \max_{g\in[m_l]} \bar{\rho}_{t,g}$. Also let $\boldsymbol{\zeta}_{LB} := \{z_g^{*(l)}\}_{g=1}^{m_l}$ be a $\mathscr{G}$-measurable random variable with probabilities $\pi_g^{(l)}$, $g \in [m_l]$. If $\bar{\rho}_{t,g} = \rho_t$, $g \in [m_l]$, for all $t = \tau+1, \ldots, T$ and $\bar{\bar{\rho}}_t$, $\bar{\rho}_{t,max}$ satisfy the corresponding criteria from one of the Propositions 1-5 with respect to ρ_t for all $t = 1, \ldots, \tau$, then

$$\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_1}^{\mathscr{F}_1|\mathscr{F}_0}\left(\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_2}^{\mathscr{F}_2|\mathscr{F}_1}\left(\cdots\left(\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{G}|\mathscr{F}_{\tau-1}}\left(\boldsymbol{\zeta}_{LB}\right)\right)\right)\right) \leq z^*,$$

with $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{G}|\mathscr{F}_{\tau-1}}\left(\boldsymbol{\zeta}_{LB}\right) := \left\{\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}\left(\boldsymbol{\zeta}_{LB}^s\right)\right\}_{s=1}^{|\Omega_{\tau-1}|}$ and $\boldsymbol{\zeta}_{LB}^s := \{z_g^{*(l)}\}_{\Omega_g^{(l)} \subseteq \Omega_s^{(k)}}$, with $k > l$ and $\Omega_s^{(k)}$, $s \in [|\Omega_{\tau-1}|]$ dissection of the scenario tree $\mathfrak{T}$ up to stage $\tau - 1$.

Proof. Recall $\Omega_g^{(l)}$, $g \in [m_l]$ is a dissection of the scenario tree $\mathfrak{T}$ up to stage τ and $\Omega_{\tau-1}$

denotes the set of nodes in the original tree $\mathfrak{T}$ at stage $\tau-1$. Let $\Omega_s^{(k)}$, $s \in [|\Omega_{\tau-1}|]$ be a dissection of the scenario tree $\mathfrak{T}$ up to stage $\tau-1$. Fix an s. Let $\mathbf{x}_{i,s}^*$, $i=0,\ldots,\tau-1$, the optimal solutions of subgroup $\Omega_s^{(k)}$, be given. Then, they form a feasible solution for each subgroup problem $\Omega_g^{(l)} \subseteq \Omega_s^{(k)}$ at stages $1,\ldots,\tau-1$. Thus, for all $\Omega_g^{(l)} \subseteq \Omega_s^{(k)}$,

$$c_0(\mathbf{x}_{0,s}^*, \boldsymbol{\xi}_0) + \mathcal{R}_{\bar{\rho}_{1,g}}^{(l)}\Big(c_1(\mathbf{x}_{1,s}^*, \boldsymbol{\xi}_1) + \mathcal{R}_{\bar{\rho}_{2,g}}^{(l)}\Big(\ldots + \mathcal{R}_{\bar{\rho}_{\tau,g}}^{(l)}(\mathcal{Q}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}^{\tau-1}))\Big)\Big) \geq z_g^{*(l)}$$

with $\mathcal{Q}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}^{\tau-1}) := \min\{\mathfrak{R}_{\rho_{\tau+1},\ldots,\rho_T}(c_\tau(\mathbf{x}_\tau, \boldsymbol{\xi}_\tau), \ldots, c_T(\mathbf{x}_T, \boldsymbol{\xi}_T)) : \mathbf{x}_\tau \in \mathcal{X}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}_t),\ \mathbf{x}_t \in \mathcal{X}_t(\mathbf{x}_{t-1}, \boldsymbol{\xi}_t), \text{for } t = \tau+1,\ldots,T\}$. Equivalently defining $\boldsymbol{\zeta}_{LB}^s := \{z_g^{*(l)}\}_{\Omega_g^{(l)} \subseteq \Omega_s^{(k)}}$, $\bar{\boldsymbol{\rho}}_t^s := \{\bar{\rho}_{t,g}\}_{\Omega_g^{(l)} \subseteq \Omega_s^{(k)}}$ and $\mathscr{H}^s$ the σ-algebra induced by the collection of subsets $\Omega_s^{(k)} = \bigcup_{\Omega_g^{(l)} \subseteq \Omega_s^{(k)}} \Omega_g^{(l)}$, we have $c_0(\mathbf{x}_{0,s}^*, \boldsymbol{\xi}_0) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_1^s}^{(k)}\Big(c_1(\mathbf{x}_{1,s}^*, \boldsymbol{\xi}_1) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_2^s}^{(k)}\Big(\ldots + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_\tau^s}^{(k)}(\mathcal{Q}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}^{\tau-1}))\Big)\Big) \geq \boldsymbol{\zeta}_{LB}^s$, where we removed notation like $\mathscr{F}_{\tau-1}|\mathscr{H}^s$ on $\widetilde{\mathcal{R}}^{(k)}$ for simplicity. Both sides of the inequality are $\mathscr{H}^s$-measurable. Then, since $\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{H}^s}$ is a coherent risk measure that satisfies monotonicity, we obtain

$$\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{H}^s}\Big(c_0(\mathbf{x}_{0,s}^*, \boldsymbol{\xi}_0) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_1^s}^{(k)}\Big(c_1(\mathbf{x}_{1,s}^*, \boldsymbol{\xi}_1) + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_2^s}^{(k)}\Big(\ldots + \widetilde{\mathcal{R}}_{\bar{\boldsymbol{\rho}}_\tau^s}^{(k)}(\mathcal{Q}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}^{\tau-1}))\Big)\Big)\Big) \geq \widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{H}^s}(\boldsymbol{\zeta}_{LB}^s).$$

Since by hypothesis $\bar{\bar{\rho}}_\tau$, $\bar{\rho}_{\tau,max}$ satisfy the criteria from one of the the Propositions 1-5 with respect to ρ_τ, by translation equivariance, and since by Definition 1 there is a single path up to stage $\tau-1$, we get

$$z_s^{*(k)} = \sum_{i=0}^{\tau-1} c_i(\mathbf{x}_{i,s}^*, \boldsymbol{\xi}_i) + \mathcal{R}_{\rho_\tau}^{(k)}\Big(\mathcal{Q}_\tau(\mathbf{x}_{\tau-1,s}^*, \boldsymbol{\xi}^{\tau-1})\Big) \geq \widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{H}}(\boldsymbol{\zeta}_{LB}^s) := \underline{z}_s^{*(k)}.$$

Repeating for all $s \in [|\Omega_{\tau-1}|]$, we obtain $\Big[z_1^{*(k)}, \ldots, z_{|\Omega_{\tau-1}|}^{*(k)}\Big]^\top \geq \Big[\underline{z}_1^{*(k)}, \ldots, \underline{z}_{|\Omega_{\tau-1}|}^{*(k)}\Big]^\top = \widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{G}|\mathscr{F}_{\tau-1}}(\boldsymbol{\zeta}_{LB})$, where $[\cdot]^\top$ denotes transpose. Let $\Omega_d^{(j)}$, $d \in [|\Omega_{\tau-2}|]$ be a dissection of the scenario tree $\mathfrak{T}$ up to stage $\tau-2$. Let $\mathbf{x}_{i,d}^*$, $i=0,\ldots,\tau-2$, the optimal solutions of subgroup $\Omega_d^{(j)}$, be given. Then, they form a feasible solution for each subproblem $\Omega_s^{(k)} \subseteq \Omega_d^{(j)}$. Following the steps above and defining $\underline{\boldsymbol{\zeta}}_{LB}^d := \{\underline{z}_s^{*(k)}\}_{\Omega_s^{(k)} \subseteq \Omega_d^{(j)}}$ and $\mathscr{D}^d$ the σ-algebra induced by the collection of subsets $\Omega_d^{(j)} = \bigcup_{\Omega_s^{(k)} \subseteq \Omega_d^{(j)}} \Omega_s^{(k)}$ we reach

$$z_d^{*(j)} = \sum_{i=0}^{\tau-2} c_i(\mathbf{x}_{i,d}^*, \boldsymbol{\xi}_i) + \mathcal{R}_{\rho_{\tau-1}}^{(j)}\Big(\mathcal{Q}_{\tau-1}(\mathbf{x}_{\tau-2,d}^*, \boldsymbol{\xi}^{\tau-2})\Big) \geq \widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_{\tau-1}}^{\mathscr{D}^d}(\underline{\boldsymbol{\zeta}}_{LB}^d) := \underline{z}_d^{*(j)}.$$

Repeating for all $d \in [|\Omega_{\tau-2}|]$, we obtain $\Big[z_1^{*(j)}, \ldots, z_{|\Omega_{\tau-2}|}^{*(j)}\Big]^\top \geq \Big[\underline{z}_1^{*(j)}, \ldots, \underline{z}_{|\Omega_{\tau-2}|}^{*(j)}\Big]^\top = \widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_{\tau-1}}^{\mathscr{F}_{\tau-1}|\mathscr{F}_{\tau-2}}\Big(\widetilde{\mathcal{R}}_{\bar{\bar{\rho}}_\tau}^{\mathscr{G}|\mathscr{F}_{\tau-1}}(\boldsymbol{\zeta}_{LB})\Big)$. Repeating the same procedure going backward for other $\tau-2$ times, the result follows. □

Remark. When the multi-level LB scheme is applied to the Wasserstein case, the distance at time $t = \tau, \ldots, 1$ between groups g_1, g_2 is computed as follows:

$$d_{g_1,g_2} = \begin{cases} \max\limits_{i \in \Omega^{(l)}_{t,g_1},\, k \in \Omega^{(l)}_{t,g_2}} \{d_{ik}\} & \text{when} \quad g_1 \neq g_2 \\ 0 & \text{when} \quad g_1 = g_2, \end{cases}$$

where g_1 and g_2 are chosen such that $\forall n_1 \in \Omega^{(l)}_{t,g_1}, n_2 \in \Omega^{(l)}_{t,g_2} : a(n_1) = a(n_2)$ and d_{ik} is the distance between nodes i and k.

Remark. The first-level LB has only two parameters that can be tuned: $\bar{\bar{\rho}}$ and $\bar{\rho}_{max}$. For the multi-level LB, however, as τ grows (*i.e.*, as l decreases), the parameters that can be tuned are $\bar{\bar{\rho}}_t$ and $\bar{\rho}_{t,max}$ for $t = 1, \ldots, \tau$. We expect that this second approach can allow to obtain better results for small subgroup cardinalities l due to having more degrees of freedom to strengthen the LBs. In terms of computations, given the same dissection of a scenario tree, the multi-level LB is only marginally more expensive. This is because the computational bottleneck lies in the solution of subproblems, and the convolution of subgroups' optimal values is computationally very cheap.

3.6 Upper bounds for multistage problems

Finding an UB of an optimization problem is of critical importance when an optimal solution is not available. In general, UBs are obtained by constraining some decision variables to be equal to pre-determined fixed values. In this paper, UBs are obtained by using optimal solutions of scenario subproblems. Using the procedure described before, we solve each single scenario group $\Omega^{(1)}_g$, $g \in [|\Omega_T|]$ obtaining $(\hat{\mathbf{x}}_{0,g}, \ldots, \hat{\mathbf{x}}_{T,g})$ as its optimal solution. Let UB^t_g, $t \in \mathcal{T}$ be the optimal value of the original problem (1) where the variables up to stage t are set to $\hat{\mathbf{x}}_{i,g}$ for $i = 0, \ldots, t$. From an algorithmic perspective, this approach requires us to solve problems of smaller dimension than the original one. The best available UB is obtained by taking the minimum value of UB^t_g over all $g \in [|\Omega_T|]$, *i.e.*, $UB^t := \min_{g \in [|\Omega_T|]} UB^t_g$. See [21] for the formal definition and the proof.

4 Case study: a mixed-integer production problem

4.1 Problem description

To show the effectiveness of the proposed approach, we consider a mixed-integer variant of the inventory management problem introduced in [24], which now includes binary variables

to indicate machinery start-ups. The problem can be summarized as follows. Consider a single product inventory system, comprised of a warehouse and a factory equipped with production machinery. At each time step $t = 0, \ldots, T-1$, production can be performed by starting up machinery. Random demands coming from customers have to be satisfied from the existing inventory. If the random demand exceeds the stock, it will be satisfied by rapid orders from a different source that come at a higher price. The goal is to minimize the total costs of the factory for the entire planning period.

The problem formulation is similar to the one presented in [24], except for (i) the binary variables at each node of the scenario tree indicating machine start-ups, (ii) the corresponding start-up costs, and (iii) the typical changes in the constraints that ensure no production takes place if machinery is not started in that period. In the following, we first assume a two-stage ($T = 1$) scenario tree with 100 scenarios. Later, we deal with a six-stage ($T = 5$) scenario tree with 5 branches from the root, 4 from each of the second stage nodes, and 3 from each of the third, fourth, fifth stages nodes, resulting in $|\Omega_T| = 5 \times 4 \times 3 \times 3 \times 3 = 540$ scenarios and 806 nodes. The scenario tree and the nominal distribution are generated as a dependent process across stages; see [23] for details. The value of the demand at the root node ($n = 1$) is $\xi_1 = 65$. At each period $t = 0, \ldots, T-1$, the maximal production capacity of the factory is 567 units and the machinery start-up cost is $k_t = 15$ for the two-stage problem or $k_t = 75$ for the six-stage one. The initial inventory is 10, the final value of the inventory is 2 per unit, and the values of production price c_t, selling price s_t, inventory holding cost h_t and procurement cost b_t at time period t are presented in Table 3 (data for $t > 1$ are disregarded for the two-stage problem).

t	0	1	2	3	4	5
c_t	3.5	3.6	2.3	2.8	3	-
s_t	-	10.7	10.5	10.9	10.6	10

t	0	1	2	3	4	5
h_t	2	1.9	2.1	2.2	2.1	-
b_t	-	4	3.1	4.9	7	7.5

Table 3: Production price c_t, selling price s_t, holding cost h_t from time t to time $t+1$, and procurement cost b_t for extra stock from another retailer at time t.

4.2 Computation of lower bounds

This section presents computations of LBs on a DRO version of the production problem described above, considering different ambiguity sets using the ϕ-divergences VD and the modified χ^2 distance, and the Wasserstein distance. In addition to dissections of the scenario tree formed either by *disjoint* partitions or *fixed* scenarios, we also consider three possible partition strategies to form subgroups (see [25]): *sequential*, *similar*, and

different. The strategy *sequential* follows the structure of the scenario tree, so that the first scenarios are assigned to the first group and so on until last scenarios are assigned to the last group. Under the strategy *similar* we assign scenarios having the largest values of the uncertain parameter to the first group; then, scenarios having the second largest values of the uncertain parameter are assigned to the second group, and so on. Under this strategy the dispersion within each group is low while the dispersion among groups is high. Under the strategy *different*, we assign the scenario with the largest uncertain parameter to the first group, the scenario with the second largest uncertain parameter to the second group, and so on. Under this partition strategy, the dispersion within each group is high.

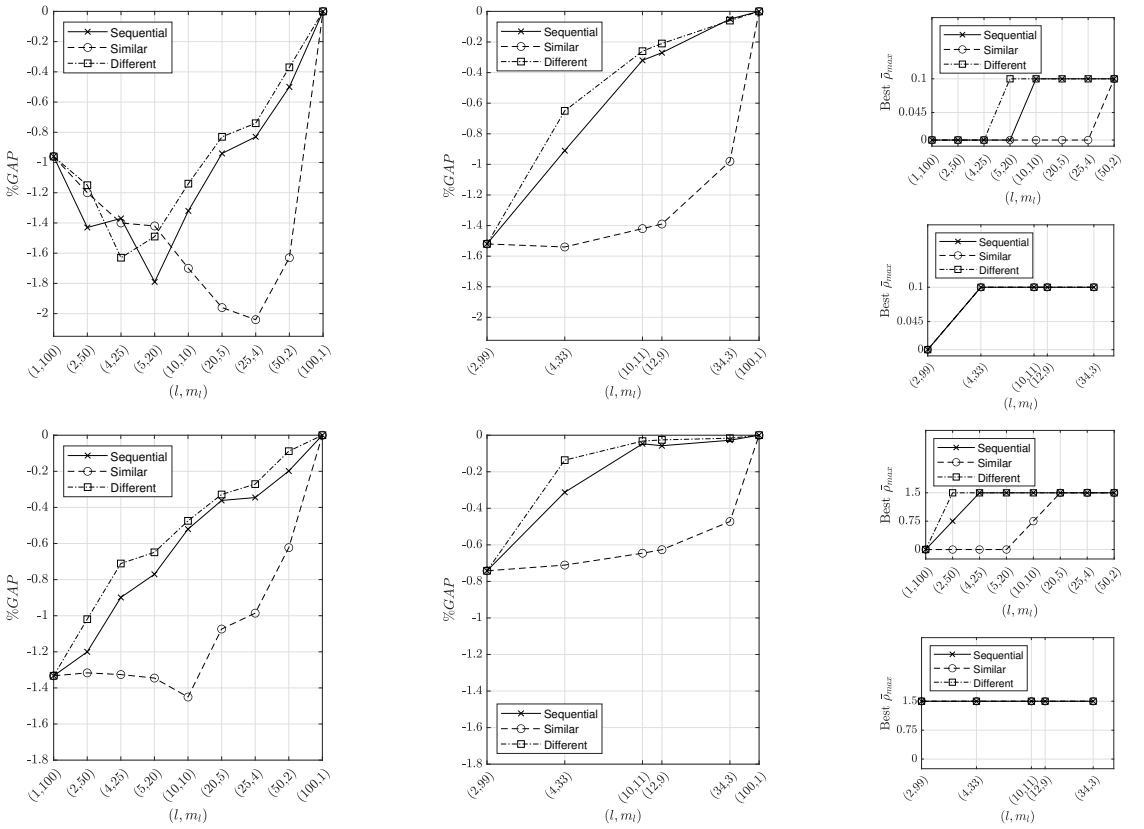
We first study the two-stage variant of the production problem to examine in detail the three partition strategies *sequential*, *similar*, and *different*. Then, we consider the multistage variant to investigate the proposed bounds in the multistage setting. For both variants, we obtain results using Proposition 6 (first-level LB) by setting the intra-group radii $\bar{\rho}_g = \bar{\rho}_{max}$, $g \in [m_l]$, selecting the inter-groups radius $\bar{\bar{\rho}}$ to be either the smallest value (0), a middle value, or the largest value (*i.e.*, radius ρ_1 of the original problem), and choosing the combinations $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ satisfying the corresponding criteria from Proposition 1, Proposition 2, and Proposition 5 at equality. For the multistage variant, we adopt a similar setup for using Proposition 7 (multi-level LB), with further details provided in Section 4.2.2. Note that for modified χ^2 distance, we use its typical form presented in Table 2, not its equivalent Cressie-Read form. In experiments where the worst scenario is *fixed* in all subgroups, the cardinality l of each subproblem is chosen to have the number of subgroups $m_l = \frac{|\Omega_T|-1}{l-1}$ to be an integer. The *worst scenario* is defined as the one with maximal objective function of single-scenario problems, *i.e.*, $\omega \in \text{argmax}\{z^{*(1)}_{\omega_i} : \omega_i \in \Omega\}$.

When presenting the results, to measure the quality of the obtained LBs, an optimality gap information is computed as $\%GAP := \frac{LB - z^*}{|z^*|} \cdot 100$. The problems derived from our case study were solved under AMPL environment using the CPLEX solver 12.8.0.0. Computations have been performed on a 64-bit machine with 8 GB of RAM and a 1.8 GHz Intel i7 processor.

4.2.1 Two-stage case

LBs obtained under the subgroup configurations *sequential*, *similar*, and *different* described above and considering *disjoint* partitions or dissections with *fixed* scenarios are

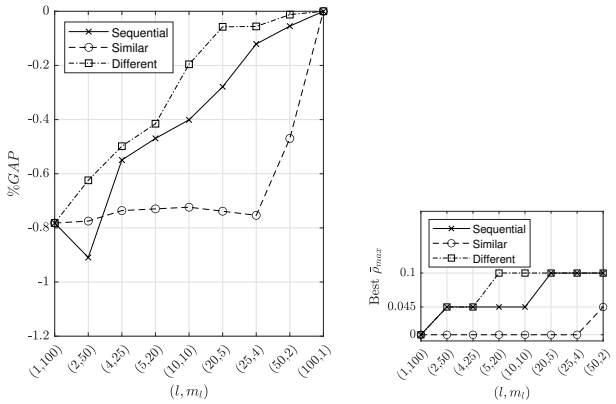
depicted in Figure 3 and Figure 4. The ambiguity sets are built via VD or modified χ^2 with radius $\rho = 0.10$ and Wasserstein distance with radius $\rho = 1.50$. These radii are selected so that the optimal objective function values of all DROs are similar ($-446.50, -445.21, -445.43$ for VD, mod. χ^2 and Wasserstein distance, respectively) to enable a fair comparison. Figure 3a and Figure 3b plot the best performing LBs for disjoint and fixed dissections, respectively, and Figure 3c depicts the best intra-group radii $\bar{\rho}_{max}$ that lead to these best performing LBs as the cardinality l and the number m_l of subgroups vary for the two-stage production problem with VD (top row) and Wasserstein distance (bottom row). Figure 4 plots the same for modified χ^2 distance, except that, by Proposition 1, we only consider disjoint partitions.



(a) Best LB – Disjoint (b) Best LB – Fixed (c) Best $\bar{\rho}_{max}$

Figure 3: Best LBs for the two-stage production problem obtained under different strategies and for disjoint (Figure 3a) or worst-scenario fixed (Figure 3b) partitions. Figure 3c compares the best $\bar{\rho}_{max}$ for disjoint (*above*) or worst-scenario fixed partitions (*below*). Top row: VD. Bottom row: Wasserstein distance.

Examining the best performing LBs, we make the following observations. First, among the three partition strategies, the strategy *similar* generally records the poorest performances, and the strategies *different* and *sequential* behave similarly and often lead to sig-



(a) Best LB – Disjoint (b) Best $\bar{\rho}_{max}$

Figure 4: Best LBs for the two-stage production problem obtained under different strategies (Figure 4a), and the $\bar{\rho}_{max}$ values that lead to these best performing LBs (Figure 4b) for disjoint partitions with mod. χ^2 distance.

nificantly better LBs than the strategy *similar*. For our problem, the strategy *sequential* also increases the dispersion within each subgroup, so it performs like *different*; however, this may not be true for all problems. Second, as the size of subgroups l increases, better LBs are obtained. Note that as l increases, the dispersion within each subgroup increases regardless of the partition strategy chosen. Third, adding the worst scenario typically improves the LBs for VD and Wasserstein distance. Overall, opting for larger cardinalities of subgroups with high levels of intra-group dispersion and when possible strengthening the subgroups with the worst scenario raises the quality of the obtained LBs.

With respect to the $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ combinations that led to the best performing LBs, we observe that when subgroups are constructed such that the dispersion within each of them is high yet the dispersion among them is low (as in strategy *different* and larger values of cardinality l), then more weight should be assigned to the intra-group radius $\bar{\rho}_{max}$ at the expense of the inter-groups radius $\bar{\bar{\rho}}$. Contrariwise, when collections are built such that the dispersion within subgroups is low but the dispersion among them is high (as in strategy *similar* and low values of cardinality l), then more weight should be assigned to the inter-groups radius $\bar{\bar{\rho}}$ at the expense of the intra-group radius $\bar{\rho}_{max}$. Since the considered problems are risk-averse, fixing the worst scenario typically strengthens the subgroups. Therefore, to utilize the strengthened subgroups for dissections with worst scenarios *fixed*, it is favorable to assign more weight to the intra-group radius $\bar{\rho}_{max}$.

These results imply that increasing the diversity of the scenarios within the subgroups (either by choosing *different* or *sequential* strategies or by increasing the cardinality l) or

strengthening the subgroups by incorporating the worst scenarios improves the LBs. Note that increasing the size of the subgroups l comes at a computational cost because larger problems need to be solved. We will investigate computation time, which poses a heavier burden on multistage problems, in the next section.

4.2.2 Multistage case

For the multistage case we restrict our attention to the strategy *sequential* that, for the two-stage case, proved to generate sharp LBs with no undue effort. We choose instead not to implement neither the strategy *similar* as leading to the poorest results, nor the strategy *different* as requiring additional effort on scenario listing with no remarkable improvements. To form the ambiguity sets, we use the same value of the radii ρ_t over all stages $t \in \mathcal{T} \setminus \{0\}$, and set $\rho_t = 0.50$ for VD or modified χ^2 distance and $\rho_t = 4$ for Wasserstein distance. Again, these radii are selected so that the optimal objective values of all multistage DROs are similar. We compute bounds for ten different realizations of the scenario tree, and present averages over these multiple randomly generated instances. In the subsequent tables, for each cardinality l, the best $\%GAPs$ are highlighted in bold.

Variation distance case. Table 4 lists the LBs obtained by applying Proposition 6 (first-level LB) using VD. We choose subgroups $\Omega_g^{(l)}$ to be disjoint with $l = 1, 3, 9, 27, 54, 108$. The instance $l = 540$ refers to the original problem, which we report as a benchmark. The overall problem, *i.e.*, the full tree with 540 scenarios, is solved on average within 15.235 seconds.

From numerical results of Table 4 obtained considering disjoint subgroups, we observe that the tightest bounds are achieved for greater values of $\bar{\bar{\rho}}$ and l, at the cost of increasing CPU running times. Indeed, overall, the best calculated LB on average is given at $\bar{\bar{\rho}} = 0.50$, $\bar{\rho}_{max} = 0.00$ when $l = 108$ and $m_l = 5$, while the worst LB on average is given at $\bar{\bar{\rho}} = 0.00$, $\bar{\rho}_{max} = 0.50$ when $l = 1$ and $m_l = 540$, which is the partition into atoms. Results show monotonic increases in average CPU time per subproblem with both the dimension of each subproblem (cardinality l) and the values of inter-group radius $\bar{\bar{\rho}}$. These results also show that very high-quality LBs can be obtained, saving considerable time with respect to the original DRO problem. For example, when $l = 54$ an average $\%GAP = -0.83\%$ can be achieved with about 4 times faster overall computation time.

Table 4 shows that the best $(\bar{\bar{\rho}}, \bar{\rho}_{max})$ combinations always assign $\bar{\rho}_{max} = 0$ regardless of the size of the subgroups l. At first glance, this seems to contradict the two-stage

results from Section 4.2.1. However, recall that in the multistage setting, the scenario tree has 5 scenarios in the first stage, and each first-stage node has 4 children, followed by each subsequent stage-t node for $t = 3, 4, 5$ having 3 children. Therefore, for every case of "l" reported in Table 4, all dissections have only one scenario at the first stage, where the first-level LB is applied. So, all cases in Table 4 essentially correspond to cardinality "$l = 1$" in the two-stage setting. Furthermore, with only 1 scenario present, there is not really an ambiguity set of distributions, and so it is best not to waste any of the robustness budget ρ_1 on the intra-group radius $\bar{\rho}_{max}$. Thus, in this case, the best LBs are obtained by setting $(\bar{\bar{\rho}}, \bar{\rho}_{max}) = (\rho_1, 0)$.

l	m_l	$\bar{\bar{\rho}}$	$\bar{\rho}_{max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP*
540	1	0.00	0.50	15.235	15.235	-
108	5	0.00	0.50	4.748	0.950	−1.80%
		0.25	0.20	5.514	1.103	−0.96%
		0.50	**0.00**	**6.114**	**1.223**	**−0.28%**
54	10	0.00	0.50	3.178	0.318	−4.82%
		0.25	0.20	3.482	0.348	−2.64%
		0.50	**0.00**	**3.623**	**0.362**	**−0.83%**
27	20	0.00	0.50	1.975	0.099	−7.00%
		0.25	0.20	2.290	0.114	−3.89%
		0.50	**0.00**	**2.575**	**0.129**	**−1.70%**

l	m_l	$\bar{\bar{\rho}}$	$\bar{\rho}_{max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP*
9	60	0.00	0.50	3.355	0.056	−12.42%
		0.25	0.20	3.562	0.059	−7.67%
		0.50	**0.00**	**3.800**	**0.063**	**−3.96%**
3	180	0.00	0.50	8.631	0.048	−17.56%
		0.25	0.20	10.351	0.058	−11.45%
		0.50	**0.00**	**12.849**	**0.071**	**−6.78%**
1	540	0.00	0.50	22.224	0.041	−21.32%
		0.25	0.20	24.501	0.045	−14.25%
		0.50	**0.00**	**26.785**	**0.050**	**−8.71%**

Table 4: Average $\%GAPs$ of LBs with disjoint subsets obtained by Proposition 6 (first-level LB) to the multistage inventory problem with VD.

Table 5 provides detailed results obtained by keeping the worst scenario *fixed* in all subsets $\Omega_g^{(l)}$. VD focuses on a convex combination of CVaR and the worst case [16, 34]. So when the worst-case scenario is fixed at each subgroup, as we saw in Section 4.2.1, we may get better LBs. Results show that fixing the worst scenario in the multistage setting improves the quality of the LB for those partitions with small cardinalities l and greater values of intra-group radius $\bar{\rho}_{max}$ (see for instance $l = 1, 3, 9$ in Table 4 and $l = 2, 8$ in Table 5). As already observed for the two-stage case, when the worst scenario is fixed, although the tightest bounds are still obtained for greater values of l, tighter bounds are obtained by setting progressively smaller value of inter-groups radius $\bar{\bar{\rho}}$ and larger intra-group radius $\bar{\rho}_{max}$ (see, for instance, $l = 50$ and $l = 78$ in Table 5). Note that the subgroups in the *fixed* dissections reported in Table 5, unlike the ones reported in Table 4, can have multiple scenarios at stage $t = 1$.

Modified χ^2 case. In Table 6, we construct collections of LBs applying Proposition 6 (first-level LB) using modified χ^2 distance and disjoint subsets $\Omega_g^{(l)}$. For all 10 instances, the overall problem, *i.e.*, the full tree with 540 scenarios, was unsolvable within a time limit of 86400 CPU seconds (or 24 hours) and values of percentage deviations ($\%GAP$)

l	m_l	$\bar{\bar{\rho}}$	$\bar{\rho}_{max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP*
78	7	**0.00**	**0.50**	**5.149**	**0.736**	**−1.35%**
		0.25	0.20	5.381	0.769	−1.58%
		0.50	0.00	5.630	0.804	−1.93%
50	11	0.00	0.50	4.573	0.416	−2.52%
		0.25	**0.20**	**4.746**	**0.431**	**−2.46%**
		0.50	0.00	4.945	0.450	−2.47%
12	49	0.00	0.50	4.524	0.092	−6.44%
		0.25	0.20	4.718	0.096	−5.49%
		0.50	**0.00**	**4.988**	**0.102**	**−4.12%**
8	77	0.00	0.50	6.818	0.089	−7.75%
		0.25	0.20	7.186	0.093	−6.53%
		0.50	**0.00**	**7.737**	**0.100**	**−4.71%**
2	539	0.00	0.50	32.560	0.060	−13.10%
		0.25	0.20	35.682	0.066	−11.10%
		0.50	**0.00**	**37.304**	**0.069**	**−8.01%**

Table 5: Average $\%GAPs$ of LBs obtained by keeping the worst scenario fixed in all subsets and applying Proposition 6 (first-level LB) to the multistage inventory problem with VD.

with respect to the optimal value could not be explicitly computed. Therefore, to measure the quality of the obtained LB, a new optimality gap is computed as follows: $\%GAP^* := \frac{LB-LB^*}{|LB^*|} \cdot 100$, with LB^* representing the best observed LB for that instance. Being a problem too large to be solved exactly within the prespecified time limit, the bounding methodology proposed in this paper is particularly helpful. It is worth noting that when $l = 108$ the solver also could not solve the subproblems within the time limit, and therefore $l = 108$ results are not reported for ease of presentation.

Once again, all dissections reported in Table 6 have only one scenario at stage $t = 1$. Therefore, the best strategy to get tighter LBs is to set the cardinality l and the inter-groups radius $\bar{\bar{\rho}}$ as large as possible. Indeed, overall, the best calculated LB is obtained at $\bar{\bar{\rho}} = 0.50$, $\bar{\rho}_{max} = 0.00$ when $l = 54$ and $m_l = 10$, although it requires considerable effort in terms of CPU time (24482.303 sec.s overall on average). A drastic reduction in computation time can be obtained by using smaller subgroups of cardinality $l = 27$ without sacrificing the quality of the LB too much. By setting $\bar{\bar{\rho}} = 0.50$, $\bar{\rho}_{max} = 0.00$, a LB within 2.11% of the best LB is obtained in approximately 24.8 times faster overall computation time. On the other hand, the worst LB is obtained at $\bar{\bar{\rho}} = 0.00$, $\bar{\rho}_{max} = 0.50$ when $l = 1$ and $m_l = 540$ in just 208.287 average CPU seconds.

l	m_l	$\bar{\bar{\rho}}$	$\bar{\rho}_{max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP**
540	1	0.00	0.50	-	-	-
54	10	0.00	0.50	24015.705	2401.571	−4.54%
		0.25	0.20	24221.180	2422.118	−1.29%
		0.50	**0.00**	**24482.303**	**2448.230**	**-**
27	20	0.00	0.50	930.780	46.539	−7.16%
		0.25	0.20	964.556	48.228	−3.47%
		0.50	**0.00**	**986.520**	**49.326**	**−2.11%**
9	60	0.00	0.50	115.022	1.917	−13.23%
		0.25	0.20	128.926	2.149	−7.85%
		0.50	**0.00**	**140.064**	**2.334**	**−5.84%**
3	180	0.00	0.50	130.976	0.728	−19.07%
		0.25	0.20	158.267	0.879	−12.63%
		0.50	**0.00**	**174.913**	**0.972**	**−10.20%**
1	540	0.00	0.50	208.287	0.386	−23.36%
		0.25	0.20	231.623	0.429	−16.32%
		0.50	**0.00**	**243.876**	**0.452**	**−13.66%**

Table 6: Average $\%GAPs$ of LBs with disjoint subsets obtained by Proposition 6 (first-level LB) to the multistage inventory problem with mod. χ^2 (time limit = 86400 CPU sec.s).

We now apply the bounding scheme proposed in Proposition 7 (multi-level LB) to the multistage inventory problem with modified χ^2 distance. The results are reported in Table

l	m_l	τ	$\{\bar{\bar{\rho}}_t\}_{t=1}^{\tau-1}$	$\{\bar{\rho}_{t,max}\}_{t=1}^{\tau-1}$	$\bar{\bar{\rho}}_\tau$	$\bar{\rho}_{\tau,max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP**
540	1	-	-	-	-	-	-	-	-
54	10	2	0.50	0.00	0.00	0.50	27940.491	2794.049	−2.85%
					0.25	**0.20**	**28316.308**	**2831.631**	**−0.56%**
					0.50	0.00	28927.738	2892.774	−0.92%
27	20	2	0.50	0.00	0.00	0.50	1024.180	51.209	−5.39%
					0.25	0.20	1050.987	52.549	−2.01%
					0.50	**0.00**	**1081.875**	**54.094**	**−0.77%**
9	60	3	0.50	0.00	0.00	0.50	143.759	2.396	−6.61%
					0.25	0.20	148.468	2.474	−2.99%
					0.50	**0.00**	**152.087**	**2.535**	**−1.64%**
3	180	4	0.50	0.00	0.00	0.50	160.423	0.891	−7.20%
					0.25	0.20	162.355	0.902	−4.04%
					0.50	**0.00**	**166.230**	**0.923**	**−2.88%**
1	540	5	0.50	0.00	0.00	0.50	246.403	0.456	−7.00%
					0.25	0.20	252.071	0.467	−4.46%
					0.50	**0.00**	**256.903**	**0.476**	**−3.53%**

Table 7: Average $\%GAPs$ of LBs with disjoint subsets obtained by Proposition 7 (multi-level LB) to the multistage inventory problem with mod. χ^2 (time limit = 86400 CPU sec.s).

7. Due to the way the scenario tree is dissected for the multi-level LB, for $t = 1, \ldots, \tau - 1$, all stage-t ambiguity sets contain only one scenario in all subgroups at any $\tau > 1$ and any cardinality l. Therefore, the best strategy to get tighter LBs is to set inter-groups radii $\bar{\bar{\rho}}_t = \rho_t$ for all $t = 1, \ldots, \tau - 1$ (which are therefore the only results we report, for ease of exposition). We allow, instead, changes in inter-groups radius $\bar{\bar{\rho}}_\tau$ (and hence intra-group radius $\bar{\rho}_{\tau,max}$) at stage τ, taking values $0.00, 0.25$ and 0.50. When $l = 54$, there are 2 scenarios present in all subgroups at stage $\tau = 2$ and all other cases reported in Table 7 contain a single scenario at stage τ. Overall, the best LB is obtained by setting $\bar{\bar{\rho}}_2 = 0.25$ and $\bar{\rho}_{2,max} = 0.20$ when $l = 54$, $m_l = 10$ and $\tau = 2$.

Comparing these results with bounds of Table 6, we conclude that the multi-level bounding technique becomes particularly useful by allowing to get tighter LBs as partitions progressively contain smaller-dimensional subproblems. For instance, at lower values of l, better LBs are obtained with similar overall computation times. This is because with the first-level LB, regardless of the size l of the subgroups, the only two parameters that can be tuned are $\bar{\bar{\rho}}$ and $\bar{\rho}_{max}$. On the other hand, with the multi-level LB the number of robustness parameters that can be adjusted gradually increases with the stage τ—namely, $\bar{\bar{\rho}}_t$ and $\bar{\rho}_{t,max}$ for $t = 1, \ldots, \tau$. Therefore, as τ grows (*i.e.*, as l decreases, being the scenario tree dissected further) the degree of freedom of the multi-level bounding scheme increases too and there are more opportunities to strengthen the bounds in a nested fashion. This supports the evidence that sharper LBs are obtained under the multi-level scheme for dissections with smaller cardinalities.

Wasserstein distance case. In Table 8, we construct collections of LBs by applying Proposition 7 (multi-level LB) for the Wasserstein distance. Given τ, the last stage where

the tree is dissected, according to Proposition 7 we set inter-groups radii $\bar{\bar{\rho}}_t = \rho_t$, $t = 1, \ldots, \tau - 1$ and choose the combinations $(\bar{\bar{\rho}}_\tau, \bar{\rho}_{\tau,max})$ with $\bar{\bar{\rho}}_\tau \in \{0, 2, 4\}$ and the intra-group radius $\bar{\rho}_{\tau,max} = \rho_\tau - \bar{\bar{\rho}}_\tau$. For $t = \tau+1, \ldots, T$ we set $\bar{\rho}_{t,g} = \rho_t$, $g \in [m_l]$. When $\tau = 1$, we work directly with $(\bar{\bar{\rho}}_1, \bar{\rho}_{1,max})$. Note that the first-level and multi-level LB schemes coincide when the decision tree is dissected up to stage $\tau = 1$ according to Definition 1 (see Table 8 with $\tau = 1$). The overall problem, *i.e.*, the full tree with 540 scenarios, is solved on average within 93.492 seconds.

From the results in Table 8, overall, the best calculated LB on average is obtained by setting $\bar{\bar{\rho}}_1 = 4$, $\bar{\rho}_{1,max} = 0$ when $l = 108$ and $m_l = 5$. On the contrary, the worst LB on average is given at $\bar{\bar{\rho}}_5 = 0$, $\bar{\rho}_{5,max} = 4$ when $l = 1$ and $m_l = 540$. Results still show monotonic increases in average CPU time per subproblem with both the dimension l of each subproblem and the values of $\bar{\bar{\rho}}_\tau$ for this problem.

l	m_l	τ	$\{\bar{\bar{\rho}}_t\}_{t=1}^{\tau-1}$	$\{\bar{\rho}_{t,max}\}_{t=1}^{\tau-1}$	$\bar{\bar{\rho}}_\tau$	$\bar{\rho}_{\tau,max}$	*Avg. CPU time overall*	*Avg. CPU time/subpr.*	*Avg. %GAP*
540	1	-	-	-	-	-	93.492	93.492	-
108	5	1	-	-	0	4	10.222	2.044	−3.20%
					2	2	10.489	2.098	−1.21%
					4	**0**	**10.801**	**2.160**	**−0.18%**
54	10	2	4	0	**0**	**4**	**6.290**	**0.629**	**−1.35%**
					2	2	6.625	0.662	−1.72%
					4	0	7.019	0.702	−2.27%
27	20	2	4	0	0	4	5.351	0.268	−4.14%
					2	2	5.655	0.283	−2.53%
					4	**0**	**5.997**	**0.300**	**−1.03%**
9	60	3	4	0	0	4	9.067	0.151	−5.63%
					2	2	9.684	0.161	−4.05%
					4	**0**	**10.234**	**0.171**	**−2.49%**
3	180	4	4	0	0	4	11.788	0.065	−7.08%
					2	2	12.237	0.068	−5.78%
					4	**0**	**12.679**	**0.070**	**−4.51%**
1	540	5	4	0	0	4	28.427	0.053	−7.69%
					2	2	30.241	0.056	−6.77%
					4	**0**	**31.191**	**0.058**	**−5.85%**

Table 8: Average $\%GAPs$ of LBs with disjoint subsets obtained by Proposition 7 (multi-level LB) to the multistage inventory problem with Wasserstein distance.

4.2.3 Lower bounds benchmarking

We now benchmark our LBs with other approaches available. We consider the following relation applied to a time-consistent dynamic risk measure given in (22)

$$\begin{aligned} &\boldsymbol{\eta}_0 + \mathcal{R}_{\rho_1}^{\mathscr{F}_1|\mathscr{F}_0}\Big(\boldsymbol{\eta}_1 + \mathcal{R}_{\rho_2}^{\mathscr{F}_2|\mathscr{F}_1}\Big(\ldots + \mathcal{R}_{\rho_T}^{\mathscr{F}_T|\mathscr{F}_{T-1}}(\boldsymbol{\eta}_T)\Big)\Big) \\ &\geq \boldsymbol{\eta}_0 + \mathcal{R}_{\rho_1}^{\mathscr{F}_1|\mathscr{F}_0}\Big(\boldsymbol{\eta}_1 + \mathcal{R}_{\rho_2}^{\mathscr{F}_2|\mathscr{F}_1}\Big(\ldots + \mathcal{R}_{\rho_{i-1}}^{\mathscr{F}_{i-1}|\mathscr{F}_{i-2}}\big(\boldsymbol{\eta}_{i-1} + \mathbb{E}_{Q_i|\boldsymbol{\xi}^{i-1}}(\ldots + \mathbb{E}_{Q_T|\boldsymbol{\xi}^{T-1}}(\boldsymbol{\eta}_T))\big)\Big)\Big) \end{aligned} \tag{25}$$

with $i \in \mathcal{T} \setminus \{0\}$. Relation (25) is obtained by noting that expectation with respect to the nominal distribution is always less than or equal to the maximization over an ambiguity

set of distributions that contains that nominal distribution, and a recursive application of the monotonicity property of coherent risk measures (see [42]).

We report in Table 9 the $\%GAPs$ obtained by applying relation (25) to the multistage VD case, where for every instance the $\%GAP$ is computed with respect to the optimal value of the full problem. Comparing these results with the ones proposed in our paper, which appear in Table 4 and Table 5, we observe that our bounding techniques enable strengthening even the sharpest LB obtained in Table 9 (*i.e.*, *Avg.* $\%GAP = -0.28\%$ *vs. Avg.* $\%GAP = -2.54\%$) with less computational effort (6.114 Avg. CPU seconds *vs.* 13.517). The same conclusions can be drawn for the Wasserstein distance case: *Avg.* $\%GAP = -0.18\%$ obtained within 10.801 Avg. CPU seconds with the proposed methods *vs. Avg.* $\%GAP = -1.93\%$ obtained within 85.488 seconds on average using (25). For the modified χ^2 case, inequality (25) allows us to compute LBs only when $i = 1$; for all the other instances, we cannot solve any of the problems to optimality within a time frame of 86400 CPU seconds. As before, *Avg.* $\%GAP^*$ of the incumbent objective function are computed with respect to the best LB obtained overall. Once again, our methods prove to be very effective in computing sharper bounds in reasonable time.

	Variation Distance		*Modified* χ^2		*Wasserstein Distance*	
	Avg. CPU time	*Avg. %GAP*	*Avg. CPU time*	*Avg. %GAP**	*Avg. CPU time*	*Avg. %GAP*
$i = 5$	13.517	-2.54%	86400	-	85.488	-1.93%
$i = 4$	10.418	-5.58%	86400	-	12.465	-4.59%
$i = 3$	8.989	-9.34%	86400	-	8.955	-7.88%
$i = 2$	8.294	-13.16%	86400	-	8.538	-10.68%
$i = 1$	8.132	-14.89%	8.132	-16.82%	8.132	-13.49%

Table 9: Average $\%GAPs$ of LBs obtained by inequality (25) for various values of $i \in \mathcal{T} \setminus \{0\}$ to the multistage inventory problem with VD, mod. χ^2, and Wasserstein distance.

4.3 Computation of upper bounds

This section presents computations of UBs on the multistage production problem under the methodology described in Section 3.6. Results are reported in Table 10. Percentage gaps have been calculated as $\%GAP := \frac{UB - z^*}{|z^*|} \cdot 100$, and for the modified χ^2 distance as $\%GAP^* := \frac{UB - LB^*}{|LB^*|} \cdot 100$. These values are particularly insightful for the modified χ^2 distance case since the problem was too large to be solved exactly.

	Variation Distance		*Modified* χ^2		*Wasserstein Distance*	
	Avg. CPU time	*Avg. %GAP*	*Avg. CPU time*	*Avg. %GAP**	*Avg. CPU time*	*Avg. %GAP*
$t = 0$	4613.639	0.01%	86400	-	44132.710	0.01%
$t = 1$	2908.818	0.20%	86400	-	34063.400	0.23%
$t = 2$	1245.456	1.20%	86400	-	9666.558	0.96%
$t = 3$	424.866	2.59%	68849.500	2.60%	532.211	2.73%
$t = 4$	10.541	5.53%	130.324	5.55%	10.530	5.64%

Table 10: UBs for multistage inventory problem with VD, mod. χ^2 and Wasserstein distance.

For VD and Wasserstein distance, all the UBs are obtained within the time limit and show improvements by fixing solutions at earlier stages only, although at larger computational costs. For the modified χ^2 case, the only UBs we were able to compute are UB^3 (within 68849.500 CPU seconds on average) and UB^4 (within 130.324 CPU seconds on average). Although UB^3 performs better than UB^4, it requires a considerably larger computational effort. All the other UBs (UB^t, $t = 0, 1, 2$) exceeded the set time limit (86400 CPU seconds) because the number of fixed variables was not enough to reduce the dimension of the scenario tree to a computationally tractable size. The difference between the best obtained UB and LBs (2.60% of the LB on average) gives information about the range where we should expect to find the total cost of the full DRO problem. Results confirm the effectiveness of the proposed LBs for this case.

4.4 Discussion

Some insights gained from the numerical experiments are as follows. Generally, increasing the dispersion within subgroups raises the quality of the LBs, and the greater the number of scenarios per subproblem, the sharper the obtained LBs.

The stages where the bounding procedure is applied play an important role. For the first-level LB, which pertains to both two-stage and multistage DRO, the LB procedure is applied at stage $t = 1$, and for the multi-level LB, at stages $t = 1, \ldots, \tau$. When there is only a single scenario present in the subgroups at the stage where the bounding procedure is applied, then the best LBs are obtained by setting the intra-group radius $\bar{\rho}_{max} = 0$ and using the largest possible value of inter-groups radius $\bar{\bar{\rho}}$ at that stage (see, *e.g.*, Figure 3c and Figure 4b when $l = 1$, Table 4, Table 6, and Tables 7-8 when $l = 1, 3, 9, 27, 108$). This is because, with a single scenario, increasing $\bar{\rho}_{\max}$ has no effect on the subgroup problem but takes away from the value of $\bar{\bar{\rho}}$. For the multi-level LB, by Definition 1, this means that for all stages before the last stage of dissection, *i.e.,* for all $t = 1, \ldots, \tau - 1$, it is best to set $(\bar{\bar{\rho}}, \bar{\rho}_{t,\max}) = (\rho_t, 0)$. Otherwise, when there are multiple scenarios at stage $t = 1$ for the first-level LB or stage $t = \tau$ for the multi-level LB, as the dispersion within subgroups increases and the dispersion among subgroups decreases (*e.g.*, when using *different* or *sequential* strategies or when there are smaller number of subgroups to be convoluted), then progressively more importance should be assigned to the intra-group radius $\bar{\rho}_{max}$ at the expense of the inter-groups radius $\bar{\bar{\rho}}$ (see, *e.g.*, Figure 3c and Figure 4b with these strategies and when l is relatively large, Table 5 when $l = 50, 78$, and Table

7-8 when $l = 54$). Contrariwise, when there is less dispersion within subgroups and more dispersion among subgroups (*e.g.*, when using *similar* strategy or when there are relatively large number of subgroups to be convoluted), then more importance should be assigned to inter-groups radius $\bar{\bar{\rho}}$ at the expense of the intra-group radius $\bar{\rho}_{max}$ (see, *e.g.*, Figure 3c and Figure 4b with *similar* strategy and/or when l is relatively small, and Table 5 when $l = 2, 8, 12$).

For dissections with smaller cardinality l, the multi-level bounding scheme described in Proposition 7 appears to be more effective than the first-level bounding scheme described in Proposition 6, which is instead more useful when the number of scenarios in each subgroup is larger. At smaller cardinalities l (hence larger values of τ), the multi-level scheme provides more opportunities to improve the LB, instead of only once as in the first-level scheme.

We observe some gains in fixing the worst scenario in the two-stage case using VD and Wasserstein distance and in the multistage case using VD at small cardinalities l and higher values of intra-group radius $\bar{\rho}_{max}$, with a slight increase in computation time due to having slightly larger subproblems to solve. This strategy can be useful, *e.g.*, when using measures of similarity that can pop scenarios [3]. Measures of similarity that can pop scenarios (like VD, CR divergence with $\theta < 1$, $\theta \to 0$, and Wassertein distance) can make the worst-case scenarios to have positive probabilities even if they have a nominal probability of zero. Thus, in large-scale versions of such problems, it may be possible to obtain better LBs by fixing the worst-case scenarios. Finally, even though we empirically observe monotonicity of LBs in the subgroups' cardinality l for fixed values of $(\bar{\bar{\rho}}, \bar{\rho}_{max})$, we found cases (not shown here for brevity) where monotonicity in l is not satisfied. This is in contrast to the risk-neutral stochastic optimization setting, where the monotonicity of the LBs in l is guaranteed [21].

5 Conclusions

In this paper new LB criteria for multistage mixed-integer DRO problems—formed by creating ambiguity sets associated with various commonly used ϕ-divergences and the Wasserstein distance on a finite scenario tree—are derived. Conditions on the way the scenario tree is dissected and the convolutions are formed to ensure a LB on the optimal value are established. The scenario tree can be dissected either by disjoint partitions or by

fixing certain scenarios in each subgroup, except for CR power divergences with $\theta < 0$ and $\theta > 1$ and χ-divergence of order $a > 1$, for which the results are established under disjoint partitions. Two ways to implement these results in the multistage setting are devised: first-level and multi-level. A comparison with classical UBs and an alternative LB scheme shows the effectiveness of the proposed LB approach. Our results do not require any structural properties, and thus they are applicable to a broad class of problems. Numerical results on a multistage production problem show that high-quality LBs can be obtained with a small computation time using the proposed bounding methodology.

Future work includes combining the proposed bounds with sampling-based bounds (*e.g.*, [32]) or using them within solution algorithms (*e.g.*, [26]). Devising new hybrid sampling-based algorithms that could utilize the proposed bounds to be used within SDDP type algorithms (*e.g.*, [10, 30]) merit further research. It would also be interesting to investigate the concept of ineffective and effective scenarios, defined in [34, 33], in this context. Ineffective scenarios can be removed from the problem without altering the optimal value. Therefore, if such scenarios can be identified, these might significantly speed up the proposed bounds. Effective scenarios might be used as fixed scenarios on each subgroup to improve the bounds. Another area of future research includes identifying an optimal method for partitioning a given scenario tree (see [39]) and create optimized chains of LBs. Finally, it should be highlighted that the proposed approach has the important advantage of dividing a given problem into subproblems, the solution of which are independent from one another and might be easily parallelized. Such parallel implementations might significantly decrease running times and therefore merit further computational research.

References

[1] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[2] I. Bakir, N. Boland, B. Dandurand, and A. Erera. Sampling scenario set partition dual bounds for multistage stochastic programs. *INFORMS J. Comput.*, 32(1):145–163, 2020.

[3] G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. In *The Operations Research Revolution*, pages 1–19. INFORMS, 2015.

[4] A. Ben-Tal, D. D. Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Manag. Sci.*, 59:341–357, 2013.

[5] D. Bertsimas, S. Shtern, and B. Sturt. A data-driven approach to multistage stochastic linear optimization. *Manag. Sci.*, 2022.

[6] D. Bertsimas, M. Sim, and M. Zhang. Adaptive distributionally robust optimization. *Manag. Sci.*, 65(2):604–618, 2019.

[7] J. R. Birge. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Math. Program.*, 24(1):314–325, 1982.

[8] R. Cavagnini, L. Bertazzi, and F. Maggioni. A rolling horizon approach for a multistage stochastic fixed-charge transportation problem with transshipment. *Eur. J. Oper. Res.*, 301(3):912–922, 2022.

[9] M. Cheramin, J. Cheng, R. Jiang, and K. Pan. Computationally efficient approximations for distributionally robust optimization under moment and wasserstein ambiguity. *INFORMS J. Comput.*, 2022.

[10] D. Duque and D. P. Morton. Distributionally robust stochastic dual dynamic programming. *SIAM J. Optim.*, 30(4):2841–2865, 2020.

[11] N. Edirisinghe. Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation. *Ann. Oper. Res.*, 85:103–127, 1999.

[12] N. Edirisinghe and W. Ziemba. Tight bounds for stochastic convex programs. *Oper. Res.*, 40(4):660–677, 1992.

[13] K. Frauendorfer, D. Kuhn, and M. Schürle. *Barycentric Bounds in Stochastic Programming: Theory and Application*, pages 67–96. Springer New York, New York, NY, 2011.

[14] K. Frauendorfer and M. Schürle. *Multistage Stochastic Programming: Barycentric Approximation*, pages 2527–2531. Springer US, Boston, MA, 2009.

[15] J. Huang, K. Zhou, and Y. Guan. A study of distributionally robust multistage stochastic optimization. *arXiv preprint arXiv:1708.07930*, 2017.

[16] R. Jiang and Y. Guan. Risk-averse two-stage stochastic program with distributional ambiguity. *Oper. Res.*, 66(5):1390–1405, 2018.

[17] D. Kuhn. *Generalized Bounds for Convex Multistage Stochastic Programs*, volume 548 of *Lecture Notes in Economics and Mathematical Systems*. Spinger Verlag Berlin Heidelberg, 2005.

[18] D. Kuhn. Aggregation and discretization in multistage stochastic programming. *Math. Program.*, 113(1):61–94, 2008.

[19] D. Lauria, G. Consigli, and F. Maggioni. Optimal chance-constrained pension fund management through dynamic stochastic control. *OR Spectrum*, pages 1–41, 2022.

[20] F. Maggioni, E. Allevi, and M. Bertocchi. Bounds in multistage linear stochastic programming. *J. Optim. Theory Appl.*, 163(1):200–229, 2014.

[21] F. Maggioni, E. Allevi, and M. Bertocchi. Monotonic bounds in multistage mixed-integer stochastic programming. *Comput. Manag. Sci.*, 13(3):423–457, 2016.

[22] F. Maggioni and R. Cavagnini. Optimization driven monotonic bounds for stochastic programs. *Work in progress*, 2022.

[23] F. Maggioni and G. C. Pflug. Bounds and approximations for multistage stochastic programs. *SIAM J. Optim.*, 26(1):831–855, 2016.

[24] F. Maggioni and G. C. Pflug. Guaranteed bounds for general nondiscrete multistage risk-averse stochastic optimization programs. *SIAM J. Optim.*, 29(1):454–483, 2019.

[25] A. İ. Mahmutoğulları, Ö. Çavuş, and M. S. Aktürk. Bounds on risk-averse mixed-integer multi-stage stochastic programming problems with mean-CVaR. *Eur. J. Oper. Res.*, 266(2):595–608, 2018.

[26] A. İ. Mahmutoğulları, Ö. Çavuş, and M. S. Aktürk. An exact solution approach for risk-averse mixed-integer multi-stage stochastic programming problems. *Ann. Oper. Res.*, pages 1–22, 2019.

[27] J. Park and G. Bayraksan. A multistage distributionally robust optimization approach to water allocation under climate uncertainty. *Eur. J. Oper. Res.*, 306(2):849–871, 2022.

[28] M. V. Pereira and L. M. Pinto. Multi-stage stochastic optimization applied to energy planning. *Math. Program.*, 52(1-3):359–375, 1991.

[29] G. C. Pflug and A. Pichler. *The Problem of Ambiguity in Stochastic Optimization*, pages 229–255. Springer International Publishing, Cham, 2014.

[30] A. B. Philpott, V. L. de Matos, and L. Kapelevich. Distributionally robust SDDP. *Comput. Manag. Sci.*, 15(3-4):431–454, 2018.

[31] A. Pichler and A. Shapiro. Mathematical foundations of distributionally robust multistage optimization. *SIAM J. Optim.*, 31(4):3044–3067, 2021.

[32] P. Pierre-Louis, D. Morton, and G. Bayraksan. A combined deterministic and sampling-based sequential bounding method for stochastic programming. In *Proc. 2011 Winter Simulation Conf.*, pages 4172–4183, Piscataway, New Jersey, 2011.

[33] H. Rahimian, G. Bayraksan, and T. H. De-Mello. Effective scenarios in multistage distributionally robust optimization with a focus on total variation distance. *SIAM J. Optim.*, 32(3):1698–1727, 2022.

[34] H. Rahimian, G. Bayraksan, and T. Homem-de Mello. Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Math. Program.*, 173(1-2):393–430, 2019.

[35] H. Rahimian and S. Mehrotra. Frameworks and results in distributionally robust optimization. *Open J. of Math. Optim.*, 3:1–85, 2022. Article no.4.

[36] R. T. Rockafellar. Coherent approaches to risk in optimization under uncertainty. In *OR Tools and Applications: Glimpses of Future Technologies*, pages 38–61. INFORMS, 2007.

[37] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Math. Oper. Res.*, 31(3):544–561, 2006.

[38] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Math. Oper. Res.*, 31(3):433–452, 2006.

[39] K. Ryan, S. Ahmed, S. S. Dey, D. Rajan, A. Musselman, and J.-P. Watson. Optimization-driven scenario grouping. *INFORMS J. Comput.*, 32(3):805–821, 2020.

[40] B. Sandıkçı, N. Kong, and A. J. Schaefer. A hierarchy of bounds for stochastic mixed-integer programs. *Math. Program.*, 138(1):253–272, 2013.

[41] B. Sandikçi and O. Y. Özaltin. A scalable bounding method for multistage stochastic programs. *SIAM J. Optim.*, 27(3):1772–1800, 2017.

[42] A. Shapiro. Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research*, 219(3):719–726, 2012.

[43] A. Shapiro. Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *Eur. J. Oper. Res.*, 288:1–13, 2021.

[44] A. Shapiro and L. Xin. Technical note—time inconsistency of optimal policies of distributionally robust inventory models. *Oper. Res.*, 68(5):1576–1584, 2020.

[45] S. W. Wallace and S.-E. Fleten. Stochastic programming models in energy. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 637–677. Elsevier, 2003.

[46] X. Yu and S. Shen. Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets. *Math. Program.*, 2020.

[47] W. Zhang, H. Rahimian, and G. Bayraksan. Decomposition algorithms for risk-averse multistage stochastic programs with application to water allocation under uncertainty. *INFORMS J. Comput.*, 28(3):385–404, 2016.

A Proofs

A.1 Proof of LB criterion for Kullback-Leibler divergence (Part of Proposition 1)

We prove the LB criteria for Kullback-Leibler (KL) divergence in Proposition 1 by directly working with the ambiguity set of KL divergence.

Proof. Let $P' \in \widetilde{\mathcal{P}}_{\phi_{KL}(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}$. Then there exists $\bar{\bar{P}} \in \widetilde{\mathcal{P}}^{\mathscr{G}}_{\phi_{KL}(\bar{\bar{\rho}})}$ and $\bar{P} \in \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\phi_{KL}(\bar{\boldsymbol{\rho}})}$ such that $\sum_{g\in[m_l]}\left[\bar{\bar{p}}_g^{(l)}\log\left(\frac{\bar{\bar{p}}_g^{(l)}}{\pi_g^{(l)}}\right)\right]\leq\bar{\bar{\rho}}$, $\sum_{g\in[m_l]}\bar{\bar{p}}_g^{(l)}=1$ and $\sum_{\omega_i\in\Omega_g^{(l)}}\left[(\bar{p}_{\omega_i})_g^{(l)}\log\left(\frac{(\bar{p}_{\omega_i})_g^{(l)}}{(q_{\omega_i})_g^{(l)}}\right)\right]\leq\bar{\rho}_g$, $\sum_{\omega_i\in\Omega_g^{(l)}}(\bar{p}_{\omega_i})_g^{(l)}=1$ for all subgroups $g\in[m_l]$ with

$$\Delta_{\phi_{KL}}(P',Q)=\sum_{\omega_i\in\Omega}\left[p'_{\omega_i}\log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right]=\sum_{\omega_i\in\Omega_f}\left[p'_{\omega_i}\log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right]+\sum_{\omega_i\in(\Omega_f)^C}\left[p'_{\omega_i}\log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right]$$

$$
\begin{aligned}
&= \sum_{\omega_i \in \Omega_f} \left[p'_{\omega_i} \log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right] + \sum_{\omega_i \in (\Omega_f)^C} \left[p'_{\omega_i} \log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right] \\
&= \sum_{\omega_i \in \Omega_f} \left[\sum_{g\in[m_l]} p'_{\omega_i,g} \log\left(\frac{\sum_{g\in[m_l]} p'_{\omega_i,g}}{\sum_{g\in[m_l]} q_{\omega_i,g}}\right)\right] + \sum_{\omega_i \in (\Omega_f)^C} \left[p'_{\omega_i} \log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right] \\
&\leq \sum_{\omega_i \in \Omega_f} \sum_{g\in[m_l]} \left[p'_{\omega_i,g} \log\left(\frac{p'_{\omega_i,g}}{q_{\omega_i,g}}\right)\right] + \sum_{\omega_i \in (\Omega_f)^C} \left[p'_{\omega_i} \log\left(\frac{p'_{\omega_i}}{q_{\omega_i}}\right)\right] \\
&= \sum_{g\in[m_l]} \left[\bar{\bar{p}}_g^{(l)} \log\left(\frac{\bar{\bar{p}}_g^{(l)}}{\pi_g^{(l)}}\right) \sum_{\omega_i \in \Omega_g^{(l)}} (\bar{p}_{\omega_i})_g^{(l)} \right] + \sum_{g\in[m_l]} \left[\bar{\bar{p}}_g^{(l)} \sum_{\omega_i \in \Omega_g^{(l)}} \left((\bar{p}_{\omega_i})_g^{(l)} \log\left(\frac{(\bar{p}_{\omega_i})_g^{(l)}}{(q_{\omega_i})_g^{(l)}}\right)\right)\right] \\
&\leq \sum_{g\in[m_l]} \left[\bar{\bar{p}}_g^{(l)} \log\left(\frac{\bar{\bar{p}}_g^{(l)}}{\pi_g^{(l)}}\right) 1\right] + \sum_{g\in[m_l]} \left[\bar{\bar{p}}_g^{(l)} \bar{\rho}_g\right] \leq \bar{\bar{\rho}} + \sum_{g\in[m_l]} \left[\bar{\bar{p}}_g^{(l)}\right] \cdot \bar{\rho}_{max} = \bar{\bar{\rho}} + \bar{\rho}_{max}
\end{aligned}
$$

where the first inequality (on the third line) follows from the log sum inequality applied to fixed scenarios $\omega_i \in \Omega_f$ and the last set of inequalities follow from the facts that $\Delta_{\phi_{KL}}(\bar{P}_g^{(l)}, Q_g^{(l)}) \leq \bar{\rho}_g$ for all subgroups $g \in [m_l]$, the definition of $\bar{\rho}_{max}$, $\Delta_{\phi_{KL}}(\bar{\bar{P}}, \bar{\bar{Q}}) \leq \bar{\bar{\rho}}$, and $\sum_{\omega_i \in \Omega_g^{(l)}} (\bar{p}_{\omega_i})_g^{(l)} = \sum_{g\in[m_l]} \bar{\bar{p}}_g^{(l)} = 1$. Therefore, if $\bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$, the result follows. □

A.2 Proof of Proposition 2 (LB criterion for variation distance)

Proof. Let $P' \in \widetilde{\mathcal{P}}_{\phi_v(\bar{\bar{\rho}}, \bar{\boldsymbol{\rho}})}$. Then there exists $\bar{\bar{P}} \in \widetilde{\mathcal{P}}^{\mathscr{G}}_{\phi_v(\bar{\bar{\rho}})}$ and $\bar{P} \in \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\phi_v(\bar{\boldsymbol{\rho}})}$ such that $\sum_{g\in[m_l]} |\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}| \leq \bar{\bar{\rho}}$, $\sum_{g\in[m_l]} \bar{\bar{p}}_g^{(l)} = 1$ and $\sum_{\omega_i \in \Omega_g^{(l)}} |(\bar{p}_{\omega_i})_g^{(l)} - (q_{\omega_i})_g^{(l)}| \leq \bar{\rho}_g$, $\sum_{\omega_i \in \Omega_g^{(l)}} (\bar{p}_{\omega_i})_g^{(l)} = 1$ with

$$
\begin{aligned}
&\Delta_{\phi_v}(P', Q) = \sum_{\omega_i \in \Omega} |p'_{\omega_i} - q_{\omega_i}| = \sum_{\omega_i \in \Omega_f} \Big| \sum_{g\in[m_l]} p'_{\omega_i,g} - \sum_{g\in[m_l]} q_{\omega_i,g} \Big| + \sum_{\omega_i \in (\Omega_f)^C} |p'_{\omega_i} - q_{\omega_i}| \\
&\leq \sum_{\omega_i \in \Omega_f} \sum_{g\in[m_l]} |p'_{\omega_i,g} - q_{\omega_i,g}| + \sum_{\omega_i \in (\Omega_f)^C} |p'_{\omega_i} - q_{\omega_i}| = \sum_{g\in[m_l]} \sum_{\omega_i \in \Omega_g^{(l)}} \left| \bar{\bar{p}}_g^{(l)} (\bar{p}_{\omega_i})_g^{(l)} - \pi_g^{(l)} (q_{\omega_i})_g^{(l)} \right| \\
&\leq \sum_{g\in[m_l]} \sum_{\omega_i \in \Omega_g^{(l)}} \left(\left|\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}\right| \left|(\bar{p}_{\omega_i})_g^{(l)} - (q_{\omega_i})_g^{(l)}\right| + \left|\left(\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}\right) (q_{\omega_i})_g^{(l)}\right| + \left|\pi_g^{(l)} \left((\bar{p}_{\omega_i})_g^{(l)} - (q_{\omega_i})_g^{(l)}\right)\right| \right) \\
&= \sum_{g\in[m_l]} \left(\left|\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}\right| \sum_{\omega_i \in \Omega_g^{(l)}} \left|(\bar{p}_{\omega_i})_g^{(l)} - (q_{\omega_i})_g^{(l)}\right| \right) + \sum_{g\in[m_l]} \left(\left|\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}\right| \sum_{\omega_i \in \Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} \right) + \sum_{g\in[m_l]} \left(\pi_g^{(l)} \sum_{\omega_i \in \Omega_g^{(l)}} \left|(\bar{p}_{\omega_i})_g^{(l)} - (q_{\omega_i})_g^{(l)}\right| \right) \\
&\leq \sum_{g\in[m_l]} \left(\left|\bar{\bar{p}}_g^{(l)} - \pi_g^{(l)}\right| \bar{\rho}_g \right) + \bar{\bar{\rho}} + \sum_{g\in[m_l]} \left(\pi_g^{(l)} \bar{\rho}_g \right) \leq \bar{\bar{\rho}} \cdot \bar{\rho}_{max} + \bar{\bar{\rho}} + \bar{\rho}_{max},
\end{aligned}
$$

where the equality on the second line above follows from the way $p'_{\omega_i,g}$, $q_{\omega_i,g}$ for fixed and p'_{ω_i}, q_{ω_i} for non-fixed scenarios are defined in set (4) and in Sections 3.1 and 3.2. The inequality on the third line follows from, for any numbers a, b, c, d, that we have $|ac - bd| = |(a-b)(c-d) + (a-b)d + b(c-d)| \leq |(a-b)||(c-d)| + |(a-b)d| + |b(c-d)|$. The final set of inequalities follow from the facts that $\Delta_{\phi_v}(\bar{P}_g^{(l)}, Q_g^{(l)}) \leq \bar{\rho}_g$ for all subgroups $g \in [m_l]$, the definition of $\bar{\rho}_{max}$, $\Delta_{\phi_v}(\bar{\bar{P}}, \bar{\bar{Q}}) \leq \bar{\bar{\rho}}$, and $\sum_{\omega_i \in \Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} = \sum_{g\in[m_l]} \pi_g^{(l)} = 1$. Therefore, if $\bar{\bar{\rho}} \cdot \bar{\rho}_{max} + \bar{\bar{\rho}} + \bar{\rho}_{max} \leq \rho$ the result follows. □

A.3 Sketch of Proof of Proposition 3 (LB criterion for J-divergence)

J-divergence is the sum of KL divergence and Burg entropy [4], and Burg entropy is similar to the KL divergence with q_{ω_i} and p_{ω_i} exchanged (see Table 1). Therefore, the proof of Burg entropy follows by first splitting the J-divergence as sum of KL divergence and Burg entropy and then following along similar lines as the proof for KL divergence in Appendix A.1. Hence, it is skipped for brevity.

A.4 Proof of Proposition 4 (LB criterion for χ-divergence of $a > 1$)

Proof. Set $x_g^{(l)} = 1 - \frac{\bar{\bar{p}}_g^{(l)}}{\pi_g^{(l)}}$ and $(y_{\omega_i})_g^{(l)} = 1 - \frac{(\bar{p}_{\omega_i})_g^{(l)}}{(q_{\omega_i})_g^{(l)}}$. Let $P' \in \widetilde{\mathcal{P}}_{\phi_\chi^a(\bar{\bar{\rho}},\bar{\boldsymbol{\rho}})}$. Then there exists $\bar{\bar{P}} \in \widetilde{\mathcal{P}}^{\mathscr{G}}_{\phi_\chi^a(\bar{\bar{\rho}})}$ and $\bar{P} \in \widetilde{\mathcal{P}}^{\mathscr{F}|\mathscr{G}}_{\phi_\chi^a(\bar{\boldsymbol{\rho}})}$ such that $\sum_{g\in[m_l]} \pi_g^{(l)} |x_g^{(l)}|^a \le \bar{\bar{\rho}}$ and $\sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} |(y_{\omega_i})_g^{(l)}|^a \le \bar{\rho}_g$ for all subgroups $g \in [m_l]$. Since the scenario tree Ω is dissected using disjoint partitions (*i.e.*, $\Omega_f = \emptyset$), we have $q_{\omega_i} = \pi_g^{(l)} (q_{\omega_i})_g^{(l)}$, for all $\omega_i \in \Omega_g^{(l)}$ and $g \in [m_l]$. Then,

$$
\begin{aligned}
&\Delta_{\phi_\chi^a}(P', Q) = \sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} q_{\omega_i} \left| 1 - \frac{\bar{\bar{p}}_g^{(l)} (\bar{p}_{\omega_i})_g^{(l)}}{\pi_g^{(l)} (q_{\omega_i})_g^{(l)}} \right|^a = \sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} q_{\omega_i} \left| x_g^{(l)} + (y_{\omega_i})_g^{(l)} - x_g^{(l)} (y_{\omega_i})_g^{(l)} \right|^a \\
&\le \left[\left(\sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} q_{\omega_i} \left| x_g^{(l)} \right|^a \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} q_{\omega_i} \left| (y_{\omega_i})_g^{(l)} \right|^a \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \sum_{\omega_i\in\Omega_g^{(l)}} q_{\omega_i} \left| x_g^{(l)} (y_{\omega_i})_g^{(l)} \right|^a \right)^{\frac{1}{a}} \right]^a \\
&= \left[\left(\sum_{g\in[m_l]} \pi_g^{(l)} \left| x_g^{(l)} \right|^a \sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \pi_g^{(l)} \sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} \left| (y_{\omega_i})_g^{(l)} \right|^a \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \pi_g^{(l)} \left| x_g^{(l)} \right|^a \sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} \left| (y_{\omega_i})_g^{(l)} \right|^a \right)^{\frac{1}{a}} \right]^a \\
&\le \left[\left(\sum_{g\in[m_l]} \pi_g^{(l)} \left| x_g^{(l)} \right|^a \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \pi_g^{(l)} \bar{\rho}_g \right)^{\frac{1}{a}} + \left(\sum_{g\in[m_l]} \pi_g^{(l)} \left| x_g^{(l)} \right|^a \bar{\rho}_g \right)^{\frac{1}{a}} \right]^a \le \left[(\bar{\bar{\rho}})^{\frac{1}{a}} + (\bar{\rho}_{max})^{\frac{1}{a}} + (\bar{\bar{\rho}} \cdot \bar{\rho}_{max})^{\frac{1}{a}} \right]^a,
\end{aligned}
$$

where the first inequality (on the third line) follows from Minkowski inequality and the last set of inequalities follow from the facts that $\Delta_{\phi_\chi^a}(\bar{\bar{P}}, \bar{\bar{Q}}) \le \bar{\bar{\rho}}$ and $\Delta_{\phi_\chi^a}(\bar{P}_g^{(l)}, Q_g^{(l)}) \le \bar{\rho}_g$ for all subgroups $g \in [m_l]$, the definition of $\bar{\rho}_{max}$, and $\sum_{\omega_i\in\Omega_g^{(l)}} (q_{\omega_i})_g^{(l)} = \sum_{g\in[m_l]} \pi_g^{(l)} = 1$. Therefore, if $\left[(\bar{\bar{\rho}})^{\frac{1}{a}} + (\bar{\rho}_{max})^{\frac{1}{a}} + (\bar{\bar{\rho}} \cdot \bar{\rho}_{max})^{\frac{1}{a}} \right]^a \le \rho$ the result follows. □