DISTRIBUTIONALLY ROBUST OPTIMIZATION WITH INFINITELY CONSTRAINED AMBIGUITY SETS

CHEN, ZHI

NATIONAL UNIVERSITY OF SINGAPORE

2017

DISTRIBUTIONALLY ROBUST OPTIMIZATION WITH INFINITELY CONSTRAINED AMBIGUITY SETS

CHEN, ZHI

B.Eng, Tsinghua University (2012)

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ANALYTICS & OPERATIONS

NATIONAL UNIVERSITY OF SINGAPORE

2017

Supervisors:

Professor Melvyn Sim, Main Supervisor

Associate Professor Huan Xu, Co-Supervisor

Examiners:

Professor Andrew Ee Beng Lim

Dr Zhenyu Hu

Associate Professor Karthik Natarajan, Singapore University of Technology and Design

DECLARATION

I hereby declare that the thesis is my original

work and it has been written by me in its entirety.

I have duly acknowledged all the sources of
information which have been used in the thesis.

This thesis has also not been submitted for any
degree in any university previously.

CHEN, ZHI
17 April 2017

ACKNOWLEDGEMENT

No words overstate my deepest and sincerest gratitude to my supervisor, Melvyn Sim. His burning passion for research and innovative ideas shed light on my road to research. Whenever I approached him for research advice, he could always provide lucid explanations while not losing any academic rigor. Without his invaluable guidance, constant patience and encouragement, and unceasing support, I would not have been able to finish my Ph.D. journey. To me, he is a wise academic mentor, a trustful friend, and a role model. It is my great fortune and pleasure to know him, work with him, and learn from him. Thank you, Melvyn! I would also like to extend my gratitude to my co-supervisor, Huan Xu, for his constant accessibility and support. I benefit immensely from his enlightening advice and ideas and his attitude towards research.

My great appreciation goes to the members of the committee: Zhenyu Hu, Andrew Lim, and Karthik Natarajan. Their insightful comments and suggestions significantly improved the quality of this thesis. It is my greatest privilege to work with my coauthors (in the timeline of my Ph.D. journey): Zhenyu Hu, Qinshen Tang, Daniel Kuhn, Wolfram Wiesemann, Peng Xiong, Long He, William Haskell, and Pengqian Yu. Without their tireless support and remarkable contributions, I could not complete this thesis. I am deeply indebted to Zhuoyu Long and Jin Qi, for kindly sharing their personal experience and providing sincere advice in academia. I am also fortunate to collaborate with Yuan Gao and Jianzhe Zhen, from whose expertise I benefit sharply.

The Department of Analytics & Operations is a wonderful home to me. Besides my supervisor, I have also learned a lot from other remarkable faculty members. I am grateful to Chung-Piaw Teo, Andrew Lim, Jie Sun, Hanqin Zhang, Lucy Chen, Mable Chou, Jussi Keppo, Tong Wang, Yaozhong Wu, Long He, Zhenyu Hu, and Wenjie Tang. Thank you for creating such a warm environment for us graduate students. I would like to thank Qizhang Liu, Ser Aik Quek, and Kok Choon Tan, I learn from their art in teaching. I want to thank Chwee Ming Lee, Dorothy Tan, and Siew Geok Teng, for managing our department.

I would like to thank those who have been studying together with me: Xuchuan Yuan, Junfei Huang, Meilin Zhang, Zhichao Zheng, Li Xiao, Jeremy Chen, Weijia Gu, Baiyu Li, Sheng Zhao, Shasha Han, Guodong Lyu, Shanshan Huang, Yuanguang Zhong, Qingshan Kong, Yu Zhang, and Liuxi Li. Special thanks are attributed to Yini Gao and Zhenzhen Yan, for all the memorable times in our shared office.

NUS is an excellent place for us students. I will never forget the batch of my loving buddies in NUS business School: Jiameng Ma, Jared Nai, Zhenyu Liao, Katrina Lin, Immanuel Lingga, Yukun Liu, Xin Zou, and Man Zhang. Thank you for sharing the journey together. I want to thank Cheow Loo Lim, Cynthia Goh and Hamidah Bte Rabu for making the program a great environment for PhD students. I appreciate the joyful moments with Xiaojun Meng, Yun-Hsiang Wang, Ran Wei, and Xuan Yang, who boarded with me at the same time. To Aziz, Firoz, Ricky Goh, Loke brothers, Nizam, Perumal, Raseef, Rizal, Yue Yang, Zaid, and many other anonymous players, for those enjoyable futsal and soccer games.

I gratefully acknowledge my mother, Panyu Fan, for the nurturing and unconditional love. She is my heroine, who would always conquer challenges in life that are not less harder than research problems. To her, this thesis is dedicated. I would also thank my family, for unceasing support and love.

Last but not least, I owe lots of gratitude to my wife, Xin Wang. Her constant encouragement, company, and support make my life. She is my everything. I am so grateful for growing with her.

Knowledge is infinite, so is love; together, they drive our world.

CONTENTS

ABSTRACT

Uncertainty is ubiquitous. For a wide variety of real-world optimization problems, decision makers usually have to make their decisions before uncertainty is revealed. As a result, solutions obtained from solving deterministic optimization models are often inferior and questionable. Optimization problems across a broad range of industrial fields also span along the time horizon with sequentially revealed uncertainties. Such problems are usually formulated as multi-stage optimization problems and involve adaptive decisions.

In recent years, distributionally robust optimization has grown into one of the common approaches in addressing optimization problems subject to uncertainty. In distributionally robust optimization, uncertainty is modeled as a random variable, whose probability distribution, however, is ambiguous. Instead, the probability distribution of uncertainty is believed to belong to a so-called ambiguity set, which is essentially a family of probability distributions that share common statistical properties such as support, mean, higher order moments and deviation measures, and structural properties such as unimodality and stochastic independence. With the introduction of an ambiguity set, the distributionally robust optimization approach can be viewed as a generic combination of the classical robust optimization approach and the stochastic programming approach, both of which are among the most popular methods in modeling optimization problems under uncertainty. More importantly, applying duality on probability distributions, distributionally robust optimization models can be reduced to finite dimensional conic programs, which are often computationally tractable.

In this thesis, we first motivate and introduce a new class of ambiguity sets, which we call the infinitely constrained ambiguity set. The key feature of an infinitely constrained ambiguity set is the potential of encompassing a possibly infinite number of

expectation constraints. We demonstrate how an infinitely constrained ambiguity set, with infinitely many expectation constraints, is enriched in characterizing uncertainty in its description. We then study the static distributionally robust optimization models with infinitely constrained ambiguity sets, which in general, may not necessarily lead to tractable reformulations. Nevertheless, we propose an algorithm that iteratively solves a sequence of tractable distributionally robust optimization models. Each of these tractable models approximates the original model by considering a finitely constrained relaxation of the infinitely constrained ambiguity set. In each iteration, the algorithm also solves a separation problem that tightens the relaxation of the infinitely constrained ambiguity set and will ultimately improve the approximation. We show favorable results in our computational study that this approach converges reasonably well.

For a broader interest, we also investigate the adaptive distributionally robust optimization model with an infinitely constrained ambiguity set that has a wide range of applications in multi-stage problems. Following a similar scheme, we study the corresponding relaxed models that consider a finitely constrained ambiguity set and iteratively refine such a relaxation. In general, multi-stage optimization problems are much harder than their static counterparts. Specifically, the difficulty is inherent within the arbitrary functional form over an infinite-dimensional space of the optimal adaptive decisions. To overcome the difficulty, we adopt the linear decision rule approximation approach that is popular in literature. In particular, we consider the extended linear decision rule that further incorporates auxiliary random variables that arise from the lifted ambiguity set in correspondence to a finitely constrained ambiguity set. We show the benefits of including auxiliary random variables in the extended linear decision rule. Besides, we demonstrate the payoffs obtained from our extra efforts in refining the relaxation.

Last but not least, we present a unified and tractable framework for distributionally robust optimization with data that could encompass a variety of statistical information including, among other things, constraints on expectation, conditional expectation, and disjoint confidence sets with uncertain probabilities defined by ϕ-divergence. In particular, we also show that the Wasserstein-based ambiguity set has an equivalent formulation

via our newly proposed ambiguity set, which would enable us to tractably approximate a Wasserstein-based distributionally robust optimization problem with recourse. To address a distributionally robust optimization problem with recourse, we introduce the *tractable adaptive recourse scheme* (TARS), which is based on the classical linear decision rule and can be also be applied in situations where the recourse decisions are discrete. Subproblems arising from aforementioned models can be cast as special examples in the format of the new framework: the relaxation of the infinitely constrained ambiguity set is an typical example of the newly proposed ambiguity set, and the extended linear decision rule is encompasses by the TARS.

Through numerical experiments in several classes of practical applications in appointment scheduling, portfolio selection, multi-product newsvendor, and inventory control, we showcase the modeling capability of our framework and present that it can provide competitive solutions.

Thesis Main Supervisor: Melvyn Sim

Title: Professor and Provost's Chair, Department of Analytics & Operations

Thesis Co-Supervisor: Huan Xu

Title: Associate Professor, Department of Industrial Systems Engineering and Management

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Motivation and literature review

Real-world optimization problems are often contaminated by uncertainty, which can severely influence the quality and feasibility of solutions obtained from solving optimizations models ignoring uncertainty. In recent years, *distributionally robust optimization* has grown into a popular approach to address optimization problems under uncertainty. The attractive features of distributionally robust optimization include not only its flexibility in the specification of uncertainty beyond a fixed probability distribution but also its ability to obtain computationally tractable models. Its idea can be traced back to *minimax stochastic programming*, where optimal solutions are evaluated under the worst-case expectation in response to a family of probability distributions of uncertain parameters (see, for instance, Scarf 1958, Dupačová 1987, Gilboa and Schmeidler 1989, Breton and El Hachem 1995, and Shapiro and Kleywegt 2002).

In distributionally robust optimization, the following problem is possibly the most key ingredient.

$$\rho(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] \tag{1.1}$$

where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^M$ is the *here-and-now* decision taken before the uncertainty $\boldsymbol{z} \in \mathbb{R}^N$ realizes, $f(\cdot, \cdot)$: $\mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$ is a given cost function, and $\mathcal{F}$ is called the *ambiguity set* that is assumed to contain the true probability distribution of uncertainty. Given the decision $\boldsymbol{x}$, the problem $\rho(\boldsymbol{x})$ evaluates the corresponding worst-case expected cost over the ambiguity set $\mathcal{F}$.

Distributionally robust optimization constitutes a generalization of *classical robust optimization* and *stochastic programming*, both of which are popular approaches for optimization problems under uncertainty in literature. Indeed, when the ambiguity set only contains the support information, i.e.,

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{F}_0(\mathbb{R}^N) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},$$

the evaluation criterion (1.1) reduces to the worst-case cost

$$\rho(\boldsymbol{x}) = \sup_{\boldsymbol{z} \in \mathcal{W}} f(\boldsymbol{x}, \boldsymbol{z})$$

as in classical robust optimization (see, e.g., Soyster 1973, Ben-Tal and Nemirovski 1998, Ben-Tal and Nemirovski 1999, Ben-Tal and Nemirovski 2000, Bertsimas and Sim 2004, El Ghaoui and Lebret 1997, and El Ghaoui et al. 1998). In contrast, when the ambiguity set is a singleton, i.e., $\mathcal{F} = \{\mathbb{P}_0\}$, then the evaluation criterion (1.1) recovers the expected cost

$$\rho(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$$

considered in stochastic programming (see Birge and Louveaux 2011 and Ruszczynski and Shapiro 2003).

Distributionally robust optimization has attracted considerable interest in various contexts. For example, distributionally robust combinatorial and mixed integer problems (see, e.g., Natarajan et al. 2011 and Li et al. 2014), distributionally robust chance-constrained problems (for instance, Calafiore and El Ghaoui 2006, Zymler et al. 2013, Cheng et al. 2014, and Hanasusanto et al. 2015c), distributionally robust Markov decision processes and dynamic programming (see, for example, Delage and Mannor 2010, Xu and Mannor 2012, Kim and Lim 2015, and Gotoh et al. 2015), and the interplay between distributionally robust optimization and risk measures (for instance, Brown and Sim 2009, Natarajan et al. 2009, Ben-Tal et al. 2010, and Postek et al. 2016).

The introduction of an ambiguity set leads to a greater modeling power and allows us to incorporate information of the uncertainty such as support and descriptive statistics such as moments (see, for instance, El Ghaoui et al. 2003, Chen et al. 2007, Chen and Sim 2009, Delage and Ye 2010, Wiesemann et al. 2014, Hanasusanto et al. 2015c). Recent studies have led to significant progress on computationally tractable reformulations for distributionally robust optimization, which closely relates to the interplay between the ambiguity set and the cost function. In particular, Delage and Ye (2010) propose tractable reformulations for distributionally robust counterparts, where the ambiguity set is specified by the first and second-order moments, which themselves could be uncertain. Wiesemann et al. (2014) generalize the result by proposing a conic based ambiguity set and identifying conditions under which the distributionally robust counterpart has an explicit tractable reformulation. It is also worth mentioning that under specific assumptions (see detailed discussions in Delage and Ye 2010 and Wiesemann et al. 2014), distributionally robust optimization inherits computational tractability from classical robust optimization. Consequently, a bunch of results on the design of uncertainty set in classical robust optimization (for example, Bertsimas et al. 2011a and Bertsimas et al. 2013) can provide an insightful guideline on the design of support in the ambiguity set; besides, techniques that derive robust counterparts of uncertain inequalities apply to the case of distributionally robust counterparts (see, for instance, Ben-Tal and Nemirovski 2001 and Ben-Tal et al. 2015).

Quite notably, a large class of statistical-distance-based ambiguity sets and distributionally robust optimization with this type of ambiguity sets have also been studied in literature. Statistical-distance-based ambiguity sets consider ambiguous probability distributions to be close to a nominal probability distribution in the sense of a chosen statistical distance. Choices of the statistical distance are the ϕ-divergence (Ben-Tal et al. 2013) that includes the Kullback-Leibler divergence (see, e.g., Hu and Hong 2013 and Jiang and Guan 2016) as a specific example, and the Wasserstein distance (see, for instance, Esfahani and Kuhn 2015, Zhao and Guan 2015, Gao and Kleywegt 2016, and Hanasusanto and Kuhn 2016).

Most of the ambiguity sets in literature are characterized by only a finite number of expectation constraints. Mehrotra and Papp (2014) study a sequence of moment-robust optimization problems, each of which is a distributionally robust optimization problem with an ambiguity set that specifies the k-th moment constraint of the uncertainty up to some positive k. Their study is close to the realm of distributionally robust moment problems (for example, Bertsimas and Popescu 2005, Popescu 2005, and Bertsimas et al. 2006). As the order of moment can go arbitrary large, the ambiguity set may include an arbitrarily large number of expectation constraints. However, to the best of our knowledge, there is no specific research on the design of ambiguity sets that can be characterized by an infinite number of expectation constraints.

Motivated by the desire to fill the void above, we propose a new class of infinitely constrained ambiguity sets, whose novelty is the potential to encompass possibly infinitely many expectation constraints. Specifically, we study a generic form of infinitely constrained ambiguity sets and some concrete examples of it. Then we investigate the distributionally robust optimization models with this class of ambiguity sets. For a broader interest, we study both static problems and adaptive problems, and we provide solution procedures, respectively. We present a unified and tractable framework for distributionally robust optimization with data and propose a new solution scheme for models involving recourse decisions that can even be discrete.

1.2 Structure of the thesis

The rest of the thesis is organized as follows.

- **Chapter 2: Infinitely Constrained Ambiguity Sets.** We motivate and propose the new class of infinitely const rained ambiguity sets that would allow ambiguous distributions to be characterized by potentially an infinite number of expectation constraints. We elucidate the generality of this type of ambiguity sets for characterizing distributional ambiguity sets for characterizing distributional ambiguity set. Some interesting examples of this class of ambiguity sets are also presented: for instance, the covariance dominance ambiguity set, the stochastic dominance ambiguity set, and the entropic dominance ambiguity set.

- **Chapter 3: Static Distributionally Robust Optimization.** We study static distributionally robust optimization models with infinitely constrained ambiguity sets. Though the optimization problem of this type is intractable in general, we propose an algorithm to obtain its solution by solving a sequence of subproblems. Each of these subproblems is a tractable distributionally robust optimization problem, in which the ambiguity set is finitely constrained and is a relaxation of the infinitely constrained ambiguity set. At each iteration, we also solve a separation problem that would enable us to obtain a tighter relaxation of the infinitely constrained ambiguity set. We demonstrate applications of this class of models by investigating the case of covariance dominance ambiguity set and the case of entropic dominance ambiguity set. Our numerical studies not only suggest the algorithm converges reasonably well and could provide competitive solutions, but also support the potential of infinitely constrained ambiguity sets.

- **Chapter 4: Adaptive Distributionally Robust Optimization.** In this chapter, we focus on adaptive distributionally robust optimization models with infinitely constrained ambiguity sets. To obtain tractable reformulation of the optimization problem, we adopt the commonly used linear decision rule approximation approach in literature. Specifically, we first consider the relaxed ambiguity set and its lifted counterpart that involves auxiliary random variables and then adopt the extended linear decision rule approximation first proposed by Bertsimas et al. (2017). We propose an algorithm that iteratively tightens the relaxation of relaxed ambiguity set, thus ultimately improves the quality of linear decision rule approximation. Numerical experiments in a distributionally robust newsvendor problem and a distributionally robust inventory control problem illustrate the quality of the approximate solutions obtained by our method.

- **Chapter 5: Tractable Distributionally Robust Optimization with Data.** In this chapter, we present a unified and tractable framework for distributionally robust optimization that could encompass a variety of statistical information including: constraints on expectation, conditional expectation, and disjoint confidence sets with uncertain probabilities defined by ϕ-divergence. In particular, we also show that the Wasserstein-based ambiguity set has an equivalent formulation via our proposed ambiguity set, which would enable us to tractably approximate a Wasserstein-based distributionally robust optimization problem with recourse. To address a distributionally robust optimization problem with recourse, we introduce the *tractable adaptive recourse scheme* (TARS), which is based on the classical linear decision rule and can also be applied in situations where the recourse decisions are discrete.

- **Chapter 6: Conclusions.** This chapter provides conclusions of the thesis, which summarizes key findings and highlights future research.

1.3 Notation

Throughout this thesis, we use boldface uppercase and lowercase characters to represent matrices and vector respectively. Special vectors of appropriate dimension include $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{e}_n$, which respectively correspond to the vector of zeros, the vector of ones and the n-th standard unit basis in the real space. We denote by $[N]$ the set of positive running indices up to N, that is, $[N] = \{1, 2, \dots, N\}$. We use $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions on $\mathbb{R}^I$. A random variable, $\tilde{\boldsymbol{z}}$ is denoted with a tilde sign and we use $\tilde{\boldsymbol{z}} \sim \mathbb{P}$, $\mathbb{P}_0 \in \mathcal{R}^I$ to define $\tilde{\boldsymbol{z}}$ as an I dimensional random variable with probability distribution $\mathbb{P}$. For a probability distribution $\mathbb{P}$, we use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to denote the corresponding expectation. For a proper cone $\mathcal{K} \subseteq \mathbb{R}^N$, the set constraint $\boldsymbol{x} - \boldsymbol{y} \in \mathcal{K}$ is equivalent to the general inequality $\boldsymbol{x} \succeq_{\mathcal{K}} \boldsymbol{y}$. We denote $\mathcal{K}^*$ as the corresponding dual cone of $\mathcal{K}$ such that $\mathcal{K}^* = \{\boldsymbol{y} \mid \boldsymbol{y}'\boldsymbol{x} \geq 0, \ \boldsymbol{x} \in \mathcal{K}\}$. The second-order cone is denoted by $\mathcal{K}_{soc}$, i.e., $(a, \boldsymbol{b}) \in \mathcal{K}_{soc}$ refers to a second-order conic constraint $||\boldsymbol{b}||_2 \leq a$. The exponential cone $\mathcal{K}_{exp}$ and its dual cone $\mathcal{K}^*_{exp}$ are respectively defined as $\mathcal{K}_{exp} = \text{cl}\left\{(x, y, z) \mid y > 0, \ ye^{x/y} \leq z\right\}$ and $\mathcal{K}^*_{exp} = \text{cl}\left\{(u, v, w) \,\middle|\, u < 0, \ -ue^{v/u} \leq ew\right\}$, where **cl** refers to the closure and $e = \exp(1)$. The symmetric positive semidefinite cone is denoted by $\mathbb{S}^N_+$, hence $\boldsymbol{X} \in \mathbb{S}^N_+$ is equivalent to $\boldsymbol{X} \succeq \mathbf{0}$ or $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{y} \geq 0$ for any $\boldsymbol{y} \in \mathbb{R}^N$. For a given set $\mathcal{C}$, we use cl and CH to denote respectively its closure and convex hull, and we denote $\mathrm{ri}(\mathcal{C})$ and $\Pi_{\boldsymbol{x}}\mathcal{C}$ as the relative interior and the projection onto $\boldsymbol{x}$. Given a set $\mathcal{C}$, we define the cone $\mathcal{K}(\mathcal{C})$ as its perspective, i.e., $\mathcal{K}(\mathcal{C}) = \mathrm{cl}\{(\boldsymbol{x}, t) \mid \boldsymbol{x}/t \in \mathcal{C}, t > 0\}$.

2. INFINITELY CONSTRAINED AMBIGUITY SETS

2.1 Motivation and introduction

Wiesemann et al. (2014) propose a class of conic representable ambiguity sets that results in tractable reformulations of distributionally robust counterparts for convex piecewise linear functions in the form of conic optimization problems. Notwithstanding its generality, it can lead to optimization formats, such as semidefinite optimization problem that is harder to solve in practice and is beyond the scope of tractable conic optimization problems in this section. Thus, for greater computational tractability, we restrict ourselves to the following tractable conic ambiguity set

$$\mathcal{F}_T = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\tilde{\boldsymbol{z}})\right] \leq h_i, \quad \forall i \in [I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{2.1}$$

with parameters $\boldsymbol{G} \in \mathbb{R}^{L\times N}$, $\boldsymbol{\mu} \in \mathbb{R}^L$, $\boldsymbol{h} \in \mathbb{R}^I$, and functions $g_i: \mathbb{R}^N \mapsto \mathbb{R}$, $\forall i \in [I]$. The support set $\mathcal{W}$ is a tractable conic representable set and the functions g_i are also tractable conic representable functions. By tractable cones, we refer to non-negative orthants, second order cone, power cone, exponential cone, and their Cartesian product, or to a lesser degree, positive semidefinite cone. Consequently, the optimization problems involve tractable conic representable constraints are scalable and polynomial time solvable with high efficiency. We refer to Ben-Tal and Nemirovski (2001) for an excellent introduction to conic representable sets and functions.

In the above ambiguity set $\mathcal{F}_T$, the equality expectation constraints specify the mean values of $\boldsymbol{G}\tilde{\boldsymbol{z}}$. Meanwhile, the inequality expectation constraints on those tractable conic representable functions g_i provide characterizations of distributional ambiguity such as bounds on means, variances, expected utility and others. Among aforementioned tractable cones, positive semidefinite cones is less scalable than others. If we exclude the positive semidefinite cone from the domain of scalable tractable cones, the above ambiguity set would be, however, incapable of incorporating covariance information as in the following covariance dominance ambiguity set.

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}. \tag{2.2}$$

Motivated by the desire to address the mentioned issue (and we will present how to address this shortly), we propose the infinitely constrained ambiguity set as follows, by extending the tractable conic ambiguity set to incorporate potentially an infinite number of expectation constraints.

$$\mathcal{F}_I = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\boldsymbol{q}_i, \tilde{\boldsymbol{z}})\right] \leq h_i(\boldsymbol{q}_i), \quad \forall \boldsymbol{q}_i \in \mathcal{Q}_i, \forall i \in [I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{2.3}$$

where parameters $\boldsymbol{G} \in \mathbb{R}^{L\times N}$, $\boldsymbol{\mu} \in \mathbb{R}^L$, and for any $i \in [I]$, sets $\mathcal{Q}_i \subseteq \mathbb{R}^{M_i}$, functions $h_i : \mathcal{Q}_i \mapsto \mathbb{R}$, and functions $g_i : \mathcal{Q}_i \times \mathbb{R}^N \mapsto \mathbb{R}$. The bounded and non-empty support set $\mathcal{W}$ is a tractable conic representable set and for any given $\boldsymbol{q}_i \in \mathcal{Q}_i$, the function $g_i(\boldsymbol{q}_i, \boldsymbol{z})$ is tractable conic representable with respect to $\boldsymbol{z}$.

The following result elucidates the generality of the infinitely constrained ambiguity set.

Theorem 1. Let $\mathcal{X}$ be any infinite set in the real space. Suppose for any given $\boldsymbol{x} \in \mathcal{X}$, the function $r(\boldsymbol{x}, \boldsymbol{z}) : \mathcal{X} \times \mathbb{R}^N \mapsto \mathbb{R}$ is tractable conic representable in the variable $\boldsymbol{z}$, then for any ambiguity set $\mathcal{F} \subseteq \mathcal{P}_0\left(\mathbb{R}^N\right)$, there exists an infinitely constrained ambiguity set $\mathcal{F}_I \subseteq \mathcal{P}_0\left(\mathbb{R}^N\right)$ such that

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right], \ \forall \boldsymbol{x} \in \mathcal{X}.$$

Proof. We consider the following ambiguity set

$$\mathcal{F}_I = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right] \leq h(\boldsymbol{q}), \ \forall \boldsymbol{q} \in \mathcal{Q} \end{array} \right\},$$

where we define the set $\mathcal{Q} := \mathcal{X}$, and the functions $h(\boldsymbol{q}) := \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right]$ and $g(\boldsymbol{q}, \boldsymbol{z}) := f(\boldsymbol{q}, \boldsymbol{z})$. Observe that the function, $g(\boldsymbol{q}, \boldsymbol{z})$ is tractable conic representable in $\boldsymbol{z}$ for any given $\boldsymbol{q} \in \mathcal{Q}$ and the set $\mathcal{Q}$ is infinite, hence the constructive set, $\mathcal{F}_I$ is an infinitely constrained ambiguity set.

For any $\mathbb{P} \in \mathcal{F}_I$, we have

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] = \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq h(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right], \ \forall \boldsymbol{x} \in \mathcal{X},$$

which further implies

$$\sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right], \ \forall \boldsymbol{x} \in \mathcal{X}.$$

Conversely, for any $\mathbb{P} \in \mathcal{F}$, observe that

$$\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] = \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] = h(\boldsymbol{x}), \ \forall \boldsymbol{x} \in \mathcal{X},$$

thus $\mathbb{P} \in \mathcal{F}_I$, and therefore, $\mathcal{F} \subseteq \mathcal{F}_I$. As a result, we have

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \leq \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \ \forall \boldsymbol{x} \in \mathcal{X}.$$

Consequently, we arrive at

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \ \forall \boldsymbol{x} \in \mathcal{X}.$$

This completes our proofs. $\square$

With Theorem 1, we can represent distributionally robust optimization problem (1.1) with any ambiguity set $\mathcal{F}$ as one with an infinitely constrained ambiguity set, as long as for any fixing $\boldsymbol{x} \in \mathcal{X}$, the objective function $f(\boldsymbol{x},\boldsymbol{z})$ is tractable conic representable in $\boldsymbol{z}$. This observation reveals the generality of the infinitely constrained ambiguity set for characterizing distributional ambiguity in distributionally robust optimization models. Note that the tractable conic ambiguity set $\mathcal{F}_T$, which contains only a finite number of expectation constraints, is a particular case of the infinitely constrained ambiguity set $\mathcal{F}_I$. Therefore, the infinitely constrained ambiguity set is capable of providing characterizations of any distributional ambiguity that can be achieved by the tractable conic ambiguity set. Beyond that, with an infinite number of expectation constraints, the infinitely constrained ambiguity set has a greater modeling flexibility in characterizing certain properties of uncertain probability distributions, while a tractable conic ambiguity set, with only a finite number of expectation constraints, would fail to do so. We provide some interesting examples as follows for demonstration.

2.2 Covariance dominance & Copositive dominance

As mentioned earlier, considering only scalable tractable cones and including only a finite number of expectation constraints, the tractable conic ambiguity set (2.1) is incapable of fully capturing covariance information among uncertain components as in $\mathcal{F}_C$. It is well known that a symmetric matrix $\boldsymbol{\Sigma}$ is positive semidefinite, i.e., $\boldsymbol{\Sigma} \succeq \mathbf{0}$, if and only if that for all $\boldsymbol{q} \in \mathbb{R}^N$, $\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q} \geq 0$. Therefore, the covariance dominance ambiguity set (2.2) can also be defined in the form of the infinitely constrained ambiguity set as follows:

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\right)^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \ \forall \boldsymbol{q} \in \mathbb{R}^N : \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}. \tag{2.4}$$

Similarly, since a symmetric matrix $\boldsymbol{\Sigma}$ is copositive, denoted as $\boldsymbol{\Sigma} \succeq_{co}$, if $\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q} \geq 0, \forall \boldsymbol{q} \geq \mathbf{0}$, the following copositive dominance ambiguity set

$$\mathcal{F}_{co} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right] \preceq_{co} \boldsymbol{\Sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}$$

can be recast as

$$\mathcal{F}_{co} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\right)^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \ \forall \boldsymbol{q} \geq \mathbf{0} : \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}.$$

Note that convex quadratic functions are second order conic representable. As we will show later, modeling covariance dominance from such an infinitely constrained perspective, we can attain the scalable second order conic relaxation for the distributionally robust optimization problem of interest.

2.3 Kurtosis dominance

The Kurtosis provides a measure of tailedness of the probability distribution of a random variable $\tilde{z} \in \mathbb{R}$ and is defined by the following normalized fourth moment

$$\text{Kurt}[\tilde{z}] = \frac{\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}-\mu)^4\right]}{\left(\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}-\mu)^2\right]\right)^2}$$

with μ being the mean of $\tilde{z}$. For a class of independently distributed random variables $\{\tilde{z}_n\}_{n\in[N]}$ with identical zero mean and one standard deviation, the Kurtosis of their linear combination, $\sum_{n\in[N]} q_n \tilde{z}_n = \boldsymbol{q}'\tilde{\boldsymbol{z}}$, is given by

$$\text{Kurt}\left[\boldsymbol{q}'\tilde{\boldsymbol{z}}\right] = \sum_{i,j\in[N]} 6q_i^2 q_j^2 - 5\sum_{n\in[N]} q_n^4.$$

Note that the above class of random variables is not uncommon: the underlying random factors of the popular factor-based model is a typical example. The Kurtosis dominance ambiguity set that encompasses the family of probability distributions of this class of random variables can be specified by

$$\mathcal{F}_K = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'\tilde{\boldsymbol{z}}\right)^4\right] \leq \displaystyle\sum_{i,j\in[N]} 6q_i^2q_j^2 - 5\sum_{n\in[N]} q_n^4, \;\forall \boldsymbol{q}\in\mathbb{R}^N : \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}, \tag{2.5}$$

where the support set $\mathcal{W}$ describes the nonidentical ranges of $\tilde{z}_n, n \in [N]$. It is not hard to see that $\mathcal{F}_K \subseteq \mathcal{F}_C$ for some covariance dominance ambiguity set $\mathcal{F}_C$ with $\boldsymbol{\mu} = \boldsymbol{0}$. Hence, the Kurtosis dominance ambiguity set can provide a better characterization by further exploring the Kurtosis.

2.4 Stochastic dominance

Stochastic dominance, particularly in the second order, is a form of stochastic ordering that is prevalent in decision theory and economics.

Definition 1. A random outcome $\tilde{z}$ dominates another random outcome $\tilde{z}^\dagger$ in the second order, denoted as $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$, if $\mathbb{E}_{\mathbb{P}}[u(\tilde{z})] \geq \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$ for every concave nondecreasing function $u(\cdot)$, for which these expected values are finite.

The celebrated *expected utility hypothesis*, introduced by Von Neumann and Morgenstern (1947), states that: for every rational decision maker, there exists a utility function $u(\cdot)$ such that she prefers the random outcome $\tilde{z}$ over some random outcome $\tilde{z}^\dagger$ if and only if $\mathbb{E}_{\mathbb{P}}[u(\tilde{z})] > \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$. However, in practice, it is almost impossible for a decision maker to elicit her utility function. Instead, Dentcheva and Ruszczynski (2003) propose a safe relaxation of the constraint $\mathbb{E}_{\mathbb{P}}[u(\tilde{z})] > \mathbb{E}_{\mathbb{P}^\dagger}\left[u(\tilde{z}^\dagger)\right]$ by considering the second order stochastic dominance constraint $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$. This dominance constraint ensures that without articulating her utility function, the decision maker will prefer the payoff $\tilde{z}$ over the benchmark payoff $\tilde{z}^\dagger$, as long as her utility function $u(\cdot)$ is concave and nondecreasing. The proposal is practically useful in many applications. For example, in portfolio optimization problems, one can construct the reference portfolio from past return rates and seek for a portfolio under the constraint that it dominates the reference portfolio (see, e.g., Dentcheva and Ruszczyński 2006).

The second order stochastic dominance $\tilde{z} \succeq_{(2)} \tilde{z}^\dagger$ has an equivalent representation as follows:

$$\mathbb{E}_{\mathbb{P}}\left[(q-\tilde{z})^+\right] \leq \mathbb{E}_{\mathbb{P}^\dagger}\left[(q-\tilde{z}^\dagger)^+\right],\ \forall q \in \mathbb{R},$$

which can be readily incorporated in our infinitely constrained ambiguity set.

It is also worth mentioning that the modeling capability of our infinitely constrained ambiguity set extends to higher order stochastic dominance relations, which we introduce as follows and refer interested researchers to Dentcheva and Ruszczynski (2003) for complete details. Let $G_z(2,\eta) = \int_{-\infty}^{\eta} F_z(\alpha)d\alpha$ and $G_{z^\dagger}(2,\eta) = \int_{-\infty}^{\eta} F_{z^\dagger}(\alpha)d\alpha$. For

$k \in \mathbb{Z}_+, k \geq 3$, we define recursively the functions

$$G_z(k, \eta) = \int_{-\infty}^{\eta} G_z(k-1, \alpha) d\alpha,$$

and

$$G_{z^\dagger}(k, \eta) = \int_{-\infty}^{\eta} G_{z^\dagger}(k-1, \alpha) d\alpha.$$

Definition 2 (Higher order stochastic dominance). A random outcome $\tilde{z}$ dominates another random outcome $\tilde{z}^\dagger$ in the k-th order ($k = 3, 4, 5, ...$), denoted as $\tilde{z} \succeq_{(k)} \tilde{z}^\dagger$, if for any $\eta \in \mathbb{R}$, $G_z(k, \eta) \leq G_{z^\dagger}(k, \eta)$.

Observe that for $k \geq 2$, we can rewrite the recursive functions as follows:

$$\begin{aligned} G_z(k, \eta) &= \int_{-\infty}^{\eta} G_z(k-1, \alpha) d\alpha \\ &= \int_{-\infty}^{\eta} \frac{(\eta - x)^{k-1}}{(k-1)!} f_z(z) dz \\ &= \frac{\mathbb{E}_{\mathbb{P}}\left[((\eta - \tilde{z})^+)^{k-1}\right]}{(k-1)!}. \end{aligned} \tag{2.6}$$

Thus the k-th order stochastic dominance $\tilde{z} \succeq_{(k)} \tilde{z}^\dagger$ if and only if

$$\mathbb{E}_{\mathbb{P}}\left[(\eta - \tilde{z})^+)^{k-1}\right] \leq \mathbb{E}_{\mathbb{P}^\dagger}\left[((\eta - \tilde{z}^\dagger)^+)^{k-1}\right], \ \forall \eta \in \mathbb{R}.$$

Since the function $((\eta - z)^+)^k$ is tractable conic representable with respect to z for any given $\eta \in \mathbb{R}$ and $k \in \mathbb{Z}_+, k \geq 3$, we can therefore incorporate higher order stochastic dominance in the infinitely constrained ambiguity set.

2.5 Entropic dominance

We now introduce a new entropic dominance ambiguity set defined by

$$\mathcal{F}_E = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))\right] \leq \phi(\boldsymbol{q}),\ \forall \boldsymbol{q} \in \mathbb{R}^N \end{array} \right. \right\}, \tag{2.7}$$

where $\phi : \mathbb{R}^N \mapsto \mathbb{R}$ is some convex twice continuously differentiable function that satisfies $\phi(\mathbf{0}) = 0$ and $\nabla\phi(\mathbf{0}) = \mathbf{0}$, where $\nabla\phi(\cdot)$ is the gradient of $\phi(\cdot)$. As it will become clear subsequently, we generally choose the value of $\boldsymbol{\mu}$ to be the mean of the random variable $\tilde{\boldsymbol{z}}$. Hence, the entropic dominance ambiguity set virtually bounds from above the log moment generating function of the random variable $\tilde{\boldsymbol{z}}$ adjusted by its mean $\boldsymbol{\mu}$. As a result, the entropic dominance ambiguity set can encompass common random variables used in statistics, such as *sub-Gaussian* (see, Wainwright 2015), which is a class of random variables with exponentially decaying tails defined as follows.

Definition 3. A random variable $\tilde{z}$ with mean μ is *sub-Gaussian* with deviation parameter $\sigma > 0$ if

$$\ln \mathbb{E}_{\mathbb{P}}\left[\exp(q(\tilde{z} - \mu))\right] \leq \frac{q^2\sigma^2}{2},\ \forall q \in \mathbb{R}.$$

In the case of normal distribution, the deviation parameter σ corresponds to the standard deviation, and we have equality in the above relation. Thus, if $\{\tilde{z}_n\}_{n\in[N]}$ are independently distributed sub-Gaussian random variables with means $\{\mu_n\}_{n\in[N]}$ and deviation parameters $\{\sigma_n\}_{n\in[N]}$, then we can specify the entropic dominance ambiguity set (2.7) by letting

$$\phi(\boldsymbol{q}) = \frac{1}{2}\sum_{n\in[N]} q_n^2\sigma_n^2,\ \forall \boldsymbol{q} \in \mathbb{R}^N.$$

Notably, given the function $\phi(\cdot)$, the entropic dominance ambiguity set implicitly specify the mean, an upper bound on the covariance, and the support for any probability distribution in this family.

Theorem 2. Consider any probability distribution $\mathbb{P} \in \mathcal{F}_E$, we have

$$\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'] \preceq \nabla^2 \phi(\boldsymbol{0}),$$

where $\nabla^2\phi(\cdot)$ is the Hessian matrix of $\phi(\cdot)$. In addition, the support set is given by

$$\mathcal{W} = \left\{ \boldsymbol{z} \in \mathbb{R}^N \;\middle|\; \boldsymbol{q}'(\boldsymbol{z} - \boldsymbol{\mu}) \leq \lim_{\alpha\to\infty} \frac{1}{\alpha}\phi(\alpha\boldsymbol{q}),\ \forall \boldsymbol{q} \in \mathbb{R}^N \right\}.$$

Moreover, if $\phi(\boldsymbol{q})$ has an additive form given by

$$\phi(\boldsymbol{q}) = \sum_{n\in[N]} \phi_n(q_n),$$

where for any $n \in [N]$, $\phi_n(0) = 0$, we then have the simplified results:

$$\nabla^2\phi(0) = \operatorname{diag}\left(\phi_1''(0), \ldots, \phi_N''(0)\right)$$

and

$$\mathcal{W} = \left\{ \boldsymbol{z} \in \mathbb{R}^N \;\middle|\; z_n \in \left[\mu_n + \lim_{\alpha\to-\infty} \frac{1}{\alpha}\phi_n(\alpha), \mu_n + \lim_{\alpha\to\infty}\frac{1}{\alpha}\phi_n(\alpha)\right],\ \forall n \in [N] \right\}.$$

Proof. Consider any probability distribution, $\mathbb{P} \in \mathcal{F}_E$. Observe that

$$\mathbb{E}_{\mathbb{P}}[\tilde{z}_n - \mu_n] = \lim_{\alpha\to 0} \frac{\mathbb{E}_{\mathbb{P}}[\exp(\alpha(\tilde{z}_n - \mu_n))] - 1}{\alpha}.$$

Let $\alpha \to 0^+(0^-)$, we arrive at

$$\lim_{\alpha\to 0^+(0^-)} \frac{\mathbb{E}_{\mathbb{P}}[\exp(\alpha(\tilde{z}_n - \mu_n))] - 1}{\alpha} \leq (\geq) \lim_{\alpha\to 0^+} \frac{\exp(\phi(\alpha\boldsymbol{e}_n)) - \exp(\phi(\boldsymbol{0}))}{\alpha} = \left.\frac{\partial\phi(\boldsymbol{q})}{\partial q_n}\right|_{\boldsymbol{q}=\boldsymbol{0}}.$$

Thus, for every $n \in [N]$, we have

$$\mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n - \mu_n\right] = \left.\frac{\partial \phi\left(\boldsymbol{q}\right)}{\partial q_n}\right|_{\boldsymbol{q}=\boldsymbol{0}} = 0,$$

which implies $\mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu}$.

For any $\alpha \in \mathbb{R}$ and $\boldsymbol{q} \in \mathbb{R}^N$, we have $\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha \boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)\right] \leq \exp\left(\phi\left(\alpha \boldsymbol{q}\right)\right)$, which implies

$$\lim_{\alpha \to 0} \frac{\mathbb{E}_{\mathbb{P}}\left[\exp\left(\alpha \boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)\right] - \exp\left(\phi\left(\alpha \boldsymbol{q}\right)\right)}{\alpha^2} \leq 0, \ \forall \boldsymbol{q} \in \mathbb{R}^N. \tag{2.8}$$

Taking Taylor expansion up to second order at 0, the first term in the numerator of (2.8) becomes

$$1 + \alpha \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right] + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)^2\right] + o\left(\alpha^2\right) = 1 + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)^2\right] + o\left(\alpha^2\right),$$

while the second term becomes

$$\begin{aligned} \exp\left(\phi(\alpha \boldsymbol{q})\right) &= \exp\left(\phi(\boldsymbol{0}) + \alpha \nabla \phi(\boldsymbol{0})' \boldsymbol{q} + \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q} + o(\alpha^2)\right) \\ &= \exp\left(\frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q} + o(\alpha^2)\right) \\ &= 1 + \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q} + o(\alpha^2). \end{aligned}$$

Therefore, inequality (2.8) becomes

$$\begin{aligned} & \lim_{\alpha \to 0} \frac{1 + \frac{1}{2}\alpha^2 \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)^2\right] - 1 - \frac{1}{2}\alpha^2 \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q} + o\left(\alpha^2\right)}{\alpha^2} \\ = & \lim_{\alpha \to 0} \frac{\alpha^2 \left(\mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)^2\right] - \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q}\right) + o\left(\alpha^2\right)}{2\alpha^2} \\ = & \frac{\mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)^2\right] - \boldsymbol{q}' \nabla^2 \phi(\boldsymbol{0}) \boldsymbol{q}}{2} \leq 0, \ \forall \boldsymbol{q} \in \mathbb{R}^N, \end{aligned}$$

which implies $\mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right] \preceq \nabla^2 \phi(\boldsymbol{0})$.

It is well known that (see, for instance, Kaas et al. 2008)

$$\lim_{\alpha\to\infty}\frac{1}{\alpha}\ln\mathbb{E}_{\mathbb{P}}\left[\exp(\alpha\tilde{z})\right]=\sup\left(\tilde{z}\right).$$

The support then follows from

$$\sup\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})\right)=\lim_{\alpha\to\infty}\frac{1}{\alpha}\ln\mathbb{E}_{\mathbb{P}}\left[\exp(\alpha\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))\right]\leq\lim_{\alpha\to\infty}\frac{1}{\alpha}\phi(\alpha\boldsymbol{q}).$$

When $\phi(\boldsymbol{q})$ has the additive form, we have $\dfrac{\partial\phi\left(\boldsymbol{q}\right)}{\partial q_n}=\phi_n'(q_n),\forall n\in[N]$. Therefore,

$$\frac{\partial^2\phi\left(\boldsymbol{q}\right)}{\partial q_l\partial q_n}=\begin{cases}\phi_n''(q_n) & \forall l,n\in[N],l=n\\ 0 & \text{otherwise.}\end{cases}$$

and $\nabla^2\phi(0)=\operatorname{diag}\left(\phi_1''(0),\ldots,\phi_N''(0)\right)$ naturally follows. Besides, the class of constraints in the support set is satisfied if and only if

$$\max_{\boldsymbol{q}\in\mathbb{R}^N}\left\{\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu})-\lim_{\alpha\to\infty}\frac{1}{\alpha}\sum_{n\in[N]}\phi_n(\alpha q_n)\right\}\leq 0,$$

which is equivalent to

$$\sum_{n\in[N]}\left(\max_{q_n\in\mathbb{R}}\left\{q_n(z_n-\mu_n)-\lim_{\alpha\to\infty}\frac{\phi_n(\alpha q_n)}{\alpha}\right\}\right)\leq 0, \tag{2.9}$$

because the maximization operates additively in $\boldsymbol{q}$. Observe that

$$q_n(z_n-\mu_n)-\lim_{\alpha\to\infty}\frac{\phi_n(\alpha q_n)}{\alpha}=\begin{cases}q_n\left(z_n-\mu_n-\lim\limits_{\alpha\to\infty}\dfrac{\phi_n(\alpha)}{\alpha}\right) & q_n\geq 0\\ q_n\left(z_n-\mu_n-\lim\limits_{\alpha\to-\infty}\dfrac{\phi_n(\alpha)}{\alpha}\right) & q_n\leq 0,\end{cases}$$

thus the inequality (2.9) holds if and only if

$$\mu_n + \lim_{\alpha \to -\infty} \frac{\phi_n(\alpha)}{\alpha} \le z_n \le \mu_n + \lim_{\alpha \to \infty} \frac{\phi_n(\alpha)}{\alpha}, \ \forall n \in [N].$$

This completes our proof. $\square$

Theorem 2 implies that every entropic dominance ambiguity set is a subset of some covariance dominance ambiguity set, which, as we will elaborate in the subsequent sections, can be helpful in developing an improved characterization of stochastic independence among uncertain components. We remark that Theorem 2 only works with the infinite number of expectation constraints in $\mathcal{F}_E$, and the choice of the function $\phi(\cdot)$ may lead to invalid support, say, $\mathcal{W} = \mathbb{R}^N$. Therefore, we suggest explicitly specify the mean, the covariance, and the support in the entropic dominance ambiguity set for practical implementations.

Before we proceed, we showcase yet another possible application of the entropic dominance ambiguity set in providing a new bound for the following expected surplus of an affine function of a set of independently distributed random variables $\{\tilde{z}_n\}_{n \in [N]}$,

$$\rho_0(x_0, \boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right]. \tag{2.10}$$

This term frequently appears in operation management such as the newsvendor problem and the inventory control problem, and risk management. Note that computing the value of $\rho_0(x_0, \boldsymbol{x})$ exactly involves high dimensional integration and it is recently known to be $\#P$-hard even for the case of $\{\tilde{z}_n\}_{n \in [N]}$ all being uniformly distributed (Hanasusanto et al. 2016a). Nevertheless, its upper bound has been useful for providing tractable approximations to stochastic programming and chance constrained optimization problems (see, e.g., Nemirovski and Shapiro 2006, Chen et al. 2008, Goh and Sim 2010, Chen et al. 2010). A well-known approach that provides an upper bound of Problem (2.10) is based on the observation that $\omega^+ \le \eta \exp(\omega/\eta - 1), \forall \eta > 0$. Consequently, under stochastic independence, we can then obtain an upper bound of Problem (2.10) by optimizing the problem

$$\rho_B(x_0, \boldsymbol{x}) = \inf_{\eta>0} \mathbb{E}_{\mathbb{P}_0}\left[\frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] = \inf_{\eta>0} \frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\boldsymbol{\mu}}{\eta} + \sum_{n\in[N]} \phi_n\left(\frac{x_n}{\eta}\right)\right),$$

where for any $n \in [N]$, $\phi_n(q) = \ln \mathbb{E}_{\mathbb{P}_n}\left[\exp\left(q(\tilde{z}_n - \mu_n)\right)\right]$ with μ_n and $\mathbb{P}_n$ respectively being the mean and the marginal distribution of $\tilde{z}_n$. Based on the entropic dominance ambiguity set, we present a new approach for bounding $\rho_0(x_0, \boldsymbol{x})$ in the following result.

Proposition 1. Consider the entropic dominance ambiguity set

$$\mathcal{F}_E = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})\right)\right] \le \sum_{n\in[N]} \phi_n(f_n), \ \forall \boldsymbol{q} \in \mathbb{R}^N \end{array} \right\},$$

and denote $\rho_E(x_0, \boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}_E} \mathbb{E}_{\mathbb{P}}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right]$. Then for all $\boldsymbol{x} \in \mathbb{R}^N$, we have

$$\rho_0(x_0, \boldsymbol{x}) \le \rho_E(x_0, \boldsymbol{x}) \le \rho_B(x_0, \boldsymbol{x}).$$

Proof. Since $\tilde{z}_n$, $n \in [N]$ are independent, we have $\mathbb{P}_0 \in \mathcal{F}_E$, which implies

$$\rho_0(x_0, \boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_0}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] \le \sup_{\mathbb{P}\in\mathcal{F}_E} \mathbb{E}_{\mathbb{P}}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] = \rho_E(x_0, \boldsymbol{x}).$$

Following from $\omega^+ \le \eta \exp\left(\omega/\eta - 1\right), \forall \eta > 0$, we have

$$\mathbb{E}_{\mathbb{P}}\left[(x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] \le \mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0 + \boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right], \ \forall \mathbb{P} \in \mathcal{F}_E, \forall \eta > 0.$$

By definition of the entropic dominance ambiguity set, we have

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] \le \exp\left(\frac{\boldsymbol{x}'\boldsymbol{\mu}}{\eta} + \sum_{n\in[N]} \phi_n\left(\frac{x_n}{\eta}\right)\right), \ \forall \mathbb{P} \in \mathcal{F}_E,$$

which implies that for any $\mathbb{P} \in \mathcal{F}_E$,

$$\mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right] \leq \frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\boldsymbol{\mu}}{\eta}+\sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right),\ \forall \eta > 0.$$

Combining these inequalities together, we arrive

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{F}_E}\mathbb{E}_{\mathbb{P}}\left[(x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}})^+\right] &\leq \sup_{\mathbb{P}\in\mathcal{F}_E}\mathbb{E}_{\mathbb{P}}\left[\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\tilde{\boldsymbol{z}}}{\eta}\right)\right]\\ &\leq \inf_{\eta>0}\frac{\eta}{e}\exp\left(\frac{x_0+\boldsymbol{x}'\boldsymbol{\mu}}{\eta}+\sum_{n\in[N]}\phi_n\left(\frac{x_n}{\eta}\right)\right),\end{aligned}$$

i.e., $\rho_E(x_0, \boldsymbol{x}) \leq \rho_B(x_0, \boldsymbol{x})$. □

Note that the value of $\rho_E(x_0, \boldsymbol{x})$ is generally not the same as $\rho_B(x_0, \boldsymbol{x})$, as illustrated as follows.

Proposition 2. Let $N = 1$ and $\mathbb{P}_0$ be the standard normal distribution, then

$$\rho_0(0,1) = \frac{1}{\sqrt{2\pi}},\ \rho_E(0,1) = \frac{1}{2},\ \rho_B(0,1) = \frac{1}{\sqrt{e}}.$$

Proof. We can determine $\rho_0(0,1)$ and $\rho_B(0,1)$ as follows:

$$\rho_0(0,1) = \int_{-\infty}^{\infty} z^+\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \int_0^{\infty} z\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \frac{1}{2\sqrt{2\pi}}\int_0^{\infty} e^{-y/2}dy = \frac{1}{\sqrt{2\pi}},$$

and

$$\rho_B(0,1) = \inf_{\eta>0}\mathbb{E}_{\mathbb{P}_0}\left[\frac{\eta}{e}\exp\left(\frac{\tilde{z}}{\eta}\right)\right] = \inf_{\eta>0}\frac{\eta}{e}\exp\left(\frac{1}{2\eta^2}\right) = \frac{1}{\sqrt{e}}.$$

With $\mathbb{P}_0$ being the standard normal distribution, the entropic dominance ambiguity set becomes

$$\mathcal{F}_E = \left\{\mathbb{P}\in\mathcal{F}(\mathbb{R}) \ \middle|\ \begin{array}{l}\tilde{z}\sim\mathbb{P}\\ \ln\mathbb{E}_{\mathbb{P}}\left[\exp(q\tilde{z})\right] \leq \frac{q^2}{2},\ \forall q\in\mathbb{R}\end{array}\right\},$$

which by Theorem 2, implies $\mathbb{E}_{\mathbb{P}}[\tilde{z}] = 0$ and $\mathbb{E}_{\mathbb{P}}[\tilde{z}^2] \leq 1$. Thus, for all $\mathbb{P} \in \mathcal{F}_E$, we have

$$\mathbb{E}_{\mathbb{P}}\left[\tilde{z}^+\right] = \frac{\mathbb{E}_{\mathbb{P}}[|\tilde{z}| + \tilde{z}]}{2} = \frac{\mathbb{E}_{\mathbb{P}}[|\tilde{z}|]}{2} \leq \frac{\sqrt{\mathbb{E}_{\mathbb{P}}[\tilde{z}^2]}}{2} \leq \frac{1}{2},$$

which implies $\rho_E(0,1) \leq 1/2$. Observe that this inequality is binding with a two-point distribution that takes values in $\{1,-1\}$ with equal probability. To conclude $\rho_E(0,1) = 1/2$. it is sufficient to show that the mentioned two-point distribution lies in $\mathcal{F}_E$. To this end, we show

$$\ln \frac{e^q + e^{-q}}{2} \leq \frac{q^2}{2}, \ \forall q \in \mathbb{R}.$$

In fact, considering Taylor series, we have

$$\frac{e^q + e^{-q}}{2} = \sum_{n=0}^{\infty} \frac{q^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{q^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{q^2}{2}\right)^n = \exp\left(\frac{q^2}{2}\right), \ \forall q \in \mathbb{R},$$

where the inequality follows from the fact that $(2n)! = 2n(2n-1)\cdots(n+1)n! \geq 2^n n!, \forall n \in \mathbb{Z}_+$. $\square$

We easily extend this approach to the case of distributional ambiguity if the underlying random variables are independently distributed. Particularly, suppose $\tilde{z}_n$ has the ambiguous marginal distribution $\mathbb{P}_n \in \mathcal{F}_n$, we define the function ϕ_n as the tightest upper bound of the log moment generating function of the random variable adjusted by its mean, given by $\phi_n(q) = \sup_{\mathbb{P} \in \mathcal{F}_n} \ln \mathbb{E}_{\mathbb{P}}[\exp(q(\tilde{z}_n - \mu_n))]$. For instance, in the case of $\tilde{z}_n$ being ambiguous sub-Gaussian random variable with mean μ_n and deviation parameter σ_n, we would define $\phi_n(q) = (q^2\sigma_n^2)/2$.

3. STATIC DISTRIBUTIONALLY ROBUST OPTIMIZATION

3.1 Introduction

In this chapter, we examine how we can obtain the optimal solution or approximate ones to the distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \min_{\boldsymbol{x}\in\mathcal{X}} \rho_I(\boldsymbol{x}) \tag{3.1}$$

where the feasible set of decision, $\mathcal{X}$ is tractable conic representable, and $\rho_I(\boldsymbol{x})$ is the worst-case expected objective over the convex and compact infinitely constrained ambiguity set, $\mathcal{F}_I$. Observe that Problem (3.1) may not necessarily be tractable since s-tandard reformulation approaches based on duality results (see, Isii 1962, Shapiro 2001, and Bertsimas and Popescu 2005) could lead to potentially infinitely many dual variables associated with the expectation constraints. Instead, we consider the following relaxed ambiguity set

$$\mathcal{F}_R = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right]=\boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\boldsymbol{q}_i,\tilde{\boldsymbol{z}})\right]\leq h_i(\boldsymbol{q}_i), \quad \forall \boldsymbol{q}_i\in\bar{\mathcal{Q}}_i, \forall i\in[I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}}\in\mathcal{W}\right]=1 \end{array} \right. \right\}, \tag{3.2}$$

where for every $i\in[I]$, $\bar{\mathcal{Q}}_i=\left\{\boldsymbol{q}_{ij}\mid j\in[J_i]\right\}$ for some $\boldsymbol{q}_{ij}\in\mathcal{Q}_i$, $j\in[J_i]$. Note that for any $i\in[I]$, the set $\bar{\mathcal{Q}}_i$ is a finite approximation of (potentially infinite) $\mathcal{Q}_i$. Consequently, this asserts $\mathcal{F}_R$ is essentially a relaxation of $\mathcal{F}_I$, i.e., $\mathcal{F}_I\subseteq\mathcal{F}_R$. To obtain explicit formulation in a tractable conic optimization format, we define the epigraphs of g_i together with the support set $\mathcal{W}$ as

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^N \times \mathbb{R}^J \;\middle|\; \boldsymbol{z} \in \mathcal{W},\ g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) \leq u_{ij},\ \forall i \in [I], \forall j \in [J_i] \right\},$$

where $J = \sum_{i \in [I]} J_i$, and we utilize the concept of conic representation and assume:

Assumption 1. The conic representation of the following system

$$\begin{aligned} & \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{G}\boldsymbol{z} = \boldsymbol{\mu} \\ & g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) \leq h_i(\boldsymbol{q}_{ij}), \quad \forall i \in [I], j \in [J_i] \end{aligned} \tag{3.3}$$

satisfies the Slater's condition (see, Theorem 1.4.2 in Ben-Tal and Nemirovski 2001).

In association with the relaxed ambiguity set $\mathcal{F}_R$, we now consider the relaxed distributionally robust optimization problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] = \min_{\boldsymbol{x} \in \mathcal{X}} \rho_R(\boldsymbol{x}), \tag{3.4}$$

where $\rho_R(\boldsymbol{x})$ is the relaxed worst-case expected objective. Observe that $\rho_I(\boldsymbol{x}) \leq \rho_R(\boldsymbol{x})$ and our aim is to sequentially tighten the relaxed ambiguity set to achieve better solutions. We will focus on the following convex and piecewise affine objective function:

$$f(\boldsymbol{x}, \boldsymbol{z}) = \max_{k \in [K]} f_k(\boldsymbol{x}, \boldsymbol{z}) = \max_{k \in [K]} \left\{ \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}) \right\},$$

where for each $k \in [K]$, $\boldsymbol{a}_k : \mathcal{X} \mapsto \mathbb{R}^N$ and $b_k : \mathcal{X} \mapsto \mathbb{R}$ are some affine functions. The requirement on objective function can be relaxed in several ways and can be extended to a richer class of functions. In such cases, the resultant formulation of Problem (3.4) would vary accordingly. We refer interested readers to Wiesemann et al. (2014) for detailed discussions.

3.2 Solution procedure

By considering the relaxed ambiguity set and solving the relaxed distributionally robust optimization problem, we have an approximation that provides an upper bound to the original problem. In this section, we demonstrate how to improve this upper bound in detail.

To proceed, we first recast $\rho_R(\boldsymbol{x})$ as a minimization problem, which afterwards, can be performed jointly with the outer minimization over $\boldsymbol{x}$.

Theorem 3. Given the relaxed ambiguity set $\mathcal{F}_R$, the relaxed worst-case expected objective, $\rho_R(\boldsymbol{x})$ is the same as the optimal value of the following tractable conic optimization problem:

$$\begin{array}{lll} \inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \displaystyle\sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}) & \\ \text{s.t.} & \alpha - b_k(\boldsymbol{x}) - t_k \geq 0, & \forall k \in [K] \\ & \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0}, & \forall k \in [K] \\ & \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0}, & \forall k \in [K] \\ & (\boldsymbol{r}_k, \boldsymbol{s}_k, t_k) \in \mathcal{K}^*, & \forall k \in [K] \\ & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^L, \boldsymbol{\gamma} \in \mathbb{R}^J_+, & \end{array} \tag{3.5}$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K} = \mathrm{cl}\left\{(\boldsymbol{z}, \boldsymbol{u}, t) \in \mathbb{R}^N \times \mathbb{R}^J \times \mathbb{R} \;\middle|\; \left(\frac{\boldsymbol{z}}{t}, \frac{\boldsymbol{u}}{t}\right) \in \bar{\mathcal{W}}, t > 0\right\}$. Moreover, Problem (3.5) is solvable, i.e., its optimal value is attainable.

Proof. Introducing dual variables α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ that respectively correspond to the probability and expectation constraints in $\mathcal{F}_R$, we obtain the dual of $\rho_R(\boldsymbol{x})$ as follows:

$$\begin{array}{rll} \rho_0(\boldsymbol{x}) = & \inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \displaystyle\sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}) \\ & \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{G}\boldsymbol{z} + \displaystyle\sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) \geq f(\boldsymbol{x}, \boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^L, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \end{array} \tag{3.6}$$

which provides an upper bound of $\rho_R(\boldsymbol{x})$. Indeed, consider any $\mathbb{P} \in \mathcal{F}_R$ and any feasible solution $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in the dual, the robust counterpart in the dual implies that

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \alpha + \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\beta}'\boldsymbol{G}\tilde{\boldsymbol{z}}\right] + \sum_{i\in[I]}\sum_{j\in[J_i]} \mathbb{E}_{\mathbb{P}}\left[\gamma_{ij} g_i(\boldsymbol{q}_{ij}, \tilde{\boldsymbol{z}})\right] \leq \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}).$$

Thus, weak duality follows, i.e., $\rho_R(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x})$. We next proceed to establish $\rho_R(\boldsymbol{x}) = \rho_0(\boldsymbol{x})$.

The robust counterpart can be written more compactly as a set of robust counterparts as follows:

$$\alpha + \boldsymbol{\beta}'\boldsymbol{G}\boldsymbol{z} + \sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} g_i(\boldsymbol{q}_{ij}, \boldsymbol{z}) - \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} - b_k(\boldsymbol{x}) \geq 0, \ \forall \boldsymbol{z} \in \mathcal{W}, \forall k \in [K].$$

Introducing the auxiliary random variable $\tilde{\boldsymbol{u}}$, each robust counterpart is not violated if and only if

$$\begin{aligned} \inf \quad & \left(\boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x})\right)' \boldsymbol{z} + \sum_{i\in[I]}\sum_{j\in[J_i]} \gamma_{ij} u_{ij} \\ \text{s.t.} \quad & (\boldsymbol{z}, \boldsymbol{u}, 1) \in \mathcal{K}, \end{aligned}$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. The dual of the above problem is given by

$$\begin{aligned} \sup \quad & -t_k \\ \text{s.t.} \quad & \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0} \\ & \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0} \\ & (\boldsymbol{r}_k, \boldsymbol{s}_k, t_k) \in \mathcal{K}^*. \end{aligned}$$

Reinjecting the dual formulation into Problem (3.6), we arrive at

$$
\begin{aligned}
\rho_0(\boldsymbol{x}) \leq \rho_1(\boldsymbol{x}) = \quad \inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{i \in [I]} \sum_{j \in [J_i]} \gamma_{ij} h_i(\boldsymbol{q}_{ij}) \\
\text{s.t.} \quad & \alpha - b_k(\boldsymbol{x}) - t_k \geq 0, && \forall k \in [K] \\
& \boldsymbol{G}'\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k = \boldsymbol{0}, && \forall k \in [K] \\
& \boldsymbol{\gamma} - \boldsymbol{s}_k = \boldsymbol{0}, && \forall k \in [K] \\
& (\boldsymbol{r}_k, \boldsymbol{s}_k, t_k) \in \mathcal{K}^*, && \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^L, \boldsymbol{\gamma} \in \mathbb{R}^J_+.
\end{aligned}
\tag{3.7}
$$

By conic duality, the dual of Problem (3.7) is given by

$$
\begin{aligned}
\rho_2(\boldsymbol{x}) = \quad \sup \quad & \sum_{k \in [K]} (\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k) \\
\text{s.t.} \quad & \sum_{k \in [K]} \eta_k = 1 \\
& \sum_{k \in [K]} \boldsymbol{G}\boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k \in [K]} [\boldsymbol{\zeta}_k]_{ij} \leq h_i(\boldsymbol{q}_{ij}), && \forall i \in [I], \forall j \in [J_i] \\
& (\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k) \in \mathcal{K}, && \forall k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, && \forall k \in [K],
\end{aligned}
\tag{3.8}
$$

where $\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k, \forall k \in [K]$ are the dual variables associated with the specified constraints respectively, and the conic constraint follows from the fact that $\mathcal{K}^{**} = \mathcal{K}$.

Under Assumption 1, Slater's condition holds for Problem (3.8), hence strong duality holds, i.e., $\rho_1(\boldsymbol{x}) = \rho_2(\boldsymbol{x})$. In fact, there exists a sequence of strictly feasible solution in Problem (3.8) $\left\{(\bar{\boldsymbol{\xi}}^l_k, \bar{\boldsymbol{\zeta}}^l_k, \bar{\eta}^l_k)_{k \in [K]}\right\}_{l \geq 0}$ such that

$$
\lim_{l \to \infty} \sum_{k \in [K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\bar{\boldsymbol{\xi}}_k + b_k(\boldsymbol{x})\bar{\eta}_k\right) = \rho_2(\boldsymbol{x}).
$$

Thus, we can construct a sequence of discrete probability distributions $\left\{\mathbb{P}_l \in \mathcal{P}_0(\mathbb{R}^N)\right\}_{l \geq 0}$

on random variable $\tilde{\boldsymbol{z}} \in \mathbb{R}^N$ in the following way:

$$\mathbb{P}_l\left[\tilde{\boldsymbol{z}} = \frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l}\right] = \bar{\eta}_k^l,\ \forall k \in [K].$$

Since for all $l \geq 0$, $\mathbb{E}_{\mathbb{P}_l}[\boldsymbol{G}\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}$, $\mathbb{E}_{\mathbb{P}_l}\left[g_i(\boldsymbol{q}_{ij}, \tilde{\boldsymbol{z}})\right] \leq h_i(\boldsymbol{q}_{ij}), \forall i \in [I], j \in [J_i]$, and $\boldsymbol{\xi}_k^l/\eta_k^l \in \mathcal{W}, \forall k \in [K]$, we have $\mathbb{P}_l \in \mathcal{F}_R$. Moreover, we have

$$\begin{aligned}
\rho_2(\boldsymbol{x}) = & \lim_{l\to\infty} \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\bar{\boldsymbol{\xi}}_k^l + b_k(\boldsymbol{x})\bar{\eta}_k^l\right) \\
= & \lim_{l\to\infty} \sum_{k\in[K]} \bar{\eta}_k^l \left(\boldsymbol{a}_k'(\boldsymbol{x})\frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l} + b_k(\boldsymbol{x})\right) \\
\leq & \lim_{l\to\infty} \sum_{k\in[K]} \bar{\eta}_k^l \left(\max_{n\in[K]}\left\{\boldsymbol{a}_n(\boldsymbol{x})'\frac{\bar{\boldsymbol{\xi}}_k^l}{\bar{\eta}_k^l} + b_n(\boldsymbol{x})\right\}\right) \\
= & \lim_{l\to\infty} \mathbb{E}_{\mathbb{P}_l}\left[\max_{n\in[K]}\left\{\boldsymbol{a}_n(\boldsymbol{x})'\tilde{\boldsymbol{z}} + b_n(\boldsymbol{x})\right\}\right] \\
\leq & \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \\
= & \rho_R(\boldsymbol{x}).
\end{aligned}$$

Recall that $\rho_R(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x}) \leq \rho_1(\boldsymbol{x})$, we now arrive at $\rho_R(\boldsymbol{x}) \leq \rho_0(\boldsymbol{x}) \leq \rho_1(\boldsymbol{x}) = \rho_2(\boldsymbol{x}) \leq \rho_R(\boldsymbol{x})$. By Theorem 1.4.2 in Ben-Tal and Nemirovski (2001), Problem (3.5), the dual of Problem (3.8), is solvable. □

Theorem 3 provides two equivalent reformations of the worst-case expected objective $\rho_R(\boldsymbol{x})$. The minimization reformation, $\rho_1(\boldsymbol{x})$ enables us to solve the relaxed distributionally robust optimization (3.4) as a single conic optimization problem. The maximization reformation, $\rho_2(\boldsymbol{x})$ enables us to equivalently represent Problem (3.4), which is of the minimax type (Sion et al. 1958), as yet another minimax problem, $\min_{\boldsymbol{x}\in\mathcal{X}} \rho_2(\boldsymbol{x})$. By exploring this equivalent representation, we can determine a worst-case distribution in the relaxed ambiguity set in correspondence to the optimal $\boldsymbol{x}^*$ for Problem (3.4), which plays a crucial role in our proposed algorithm for tightening the relaxation.

Theorem 4. Given the relaxed ambiguity set $\mathcal{F}_R$, let $(\boldsymbol{\xi}_k^*, \eta_k^*)_{k\in[K]}$ be the optimal solution to the following tractable conic optimization problem,

$$
\begin{aligned}
\sup \quad & \rho \\
\text{s.t.} \quad & \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k\right) \geq \rho, \quad \forall \boldsymbol{x} \in \mathcal{X} \\
& \sum_{k\in[K]} \eta_k = 1 \\
& \sum_{k\in[K]} \boldsymbol{G}\boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k\in[K]} [\boldsymbol{\zeta}_k]_{ij} \leq h_i(\boldsymbol{q}_{ij}), && \forall i \in [I], \forall j \in [J_i] \\
& (\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k) \in \mathcal{K}, && \forall k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, && \forall k \in [K].
\end{aligned}
\tag{3.9}
$$

The worst-case distribution in $\mathcal{F}_R$ for the worst-case expected objective $\rho_R(\boldsymbol{x})$ is given by

$$
\mathbb{P}_v\left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*}\right] = \eta_k^*, \ \forall k \in [K] : \eta_k^* > 0.
$$

Proof. For ease of notation, let $(\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = (\boldsymbol{\xi}_k, \boldsymbol{\zeta}_k, \eta_k)_{k\in[K]}$,

$$
\bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}) = \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k\right),
$$

and $\mathcal{A}$ be the convex set of the feasible solution $(\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta})$ in Problem (3.8). Following from Theorem 3, the relaxed distributionally robust optimization problem (3.4) is equivalent to the following minimax problem

$$
\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{A}} \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}), \tag{3.10}
$$

which satisfies

$$
\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{A}} \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}) = \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{A}} \min_{\boldsymbol{x}\in\mathcal{X}} \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}).
$$

By the minimax theorem, $\boldsymbol{x}^* \in \mathcal{X}$ is optimal to Problem (3.10) if and only if there exists $(\boldsymbol{\xi}^*, \boldsymbol{\eta}^*) \in \mathcal{A}$ such that $(\boldsymbol{x}^*, \boldsymbol{\xi}^*, \boldsymbol{\eta}^*)$ is a saddle point, i.e.,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{A}} \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}) = \bar{f}(\boldsymbol{x}^*, \boldsymbol{\xi}^*, \boldsymbol{\eta}^*) = \sup_{(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{A}} \min_{\boldsymbol{x}\in\mathcal{X}} \bar{f}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}).$$

Introducing an auxiliary variable ρ, the right-hand side in the above becomes Problem (3.9), and we have $\rho^* = \rho_R(\boldsymbol{x}^*)$.

We now consider the constructive discrete distribution. Note that if $\eta^*_{k'} = 0$ for some $k' \in [K]$, then $\boldsymbol{\xi}^*_{k'} = \boldsymbol{0}$. Otherwise, since $(\boldsymbol{\xi}^*_{k'}, \boldsymbol{\zeta}^*_{k'}, \eta^*_{k'}) \in \mathcal{K}$, we have $\boldsymbol{\xi}^*_{k'}/\eta^*_{k'} \in \mathcal{W}$, which contradicts the fact that the support set $\mathcal{W}$ is compact. We can verify

$$\mathbb{P}_v\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1,\ \mathbb{E}_{\mathbb{P}_v}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}_v}\left[g_i(\boldsymbol{q}_{ij}, \tilde{\boldsymbol{z}})\right] \le h_i(\boldsymbol{q}_{ij}), \forall i \in [I], \forall j \in [J_i],$$

which implies $\mathbb{P}_v \in \mathcal{F}_R$. Observe that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_v}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] &= \mathbb{E}_{\mathbb{P}_v}\left[\max_{n\in[K]}\left\{\boldsymbol{a}_n(\boldsymbol{x})'\tilde{\boldsymbol{z}} + b_n(\boldsymbol{x})\right\}\right] \\
&= \sum_{k\in[K]} \eta^*_k \max_{n\in[K]}\left\{\boldsymbol{a}_n(\boldsymbol{x})'\frac{\boldsymbol{\xi}^*_k}{\eta^*_k} + b_n(\boldsymbol{x})\right\} \\
&\ge \sum_{k\in[K]} \left(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}^*_k + b_k(\boldsymbol{x})'\eta^*_k\right) \\
&\ge \rho_R(\boldsymbol{x}^*),
\end{aligned}$$

where the last inequality follows from $\rho^* = \rho_R(\boldsymbol{x})$. Therefore, $\mathbb{P}_v$ is the worst-case distribution. $\square$

Theorem 4 asserts that $\mathbb{P}_v$ is the worst-case distribution in $\mathcal{F}_R$. However, $\mathbb{P}_v$ may not belong to the infinitely constrained ambiguity set $\mathcal{F}_I$, because it may violate expectation constraints for some $\boldsymbol{q}^*_i \in \mathcal{Q}_i$, $i \in [I]$. In such a case, the worst-case distribution $\mathbb{P}_v$ is essentially a violating distribution in $\mathcal{F}_I$. In other words, $\mathbb{P}_v$ lies in $\mathcal{F}_I$ if and only if it satisfies all expectation constraints. To check whether each of the i-th set of the infinite expectation constraints is feasible, we solve the following separation problem

$$Z_i(\mathbb{P}_v) = \min_{\boldsymbol{q}\in\mathcal{Q}_i} \left\{ \psi_i\left(h_i(\boldsymbol{q})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v}\left[g_i(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right]\right) \right\} \tag{3.11}$$

for some increasing function $\psi_i : \mathbb{R} \mapsto \mathbb{R}$. In principle, ψ_i could simply be a linear function, while in some cases, having a nonlinear mapping function may improve the precision when evaluating the objective of the separation problem. For instance, for the case of sub-Gaussian, $h_i(\boldsymbol{q})$ would be an exponential function, in which case, based on the 64 bit floating point precision, we would not be able to evaluate the value of $\exp(x)$ for $x \geq 710$. In this case, we can choose ψ_i to be a logarithmic function to improve the precision when solving the separation problem. We formally propose our algorithm that utilizes the separation problems (3.11) for tightening the relaxation as follows.

Algorithm 1. Static Distributionally Robust Optimization

Input: Initial finite subsets $\bar{\mathcal{Q}}_i \subseteq \mathcal{Q}_i, \forall i \in [I]$.

1. Solve Problem (3.4) and obtain an optimal solution $\boldsymbol{x}$.
2. Solve Problem (3.9) and obtain the worst-case distribution $\mathbb{P}_v$.
3. For all $i \in [I]$, solve Problem (3.11): if $Z_i(\mathbb{P}_v) < 0$, obtain the optimal solution $\boldsymbol{q}_i$ and update $\bar{\mathcal{Q}}_i := \bar{\mathcal{Q}}_i \cup \{\boldsymbol{q}_i\}$.
4. If $Z_i(\mathbb{P}_v) \geq 0$ for all $i \in [I]$, then STOP. Otherwise Go to Step 1.

Output: Solution $\boldsymbol{x}$.

Theorem 5. The sequence of solutions in Algorithm 1 converges to the optimal solution of the distributionally robust optimization problem (3.1).

Proof. Let $\mathcal{F}_R^t$ be the ambiguity set at the t-th iteration of Algorithm 1 and respectively denote $\mathbb{P}_v^t$ and $\boldsymbol{x}^t$ as the optimal solutions for Problem (3.9) and Problem (3.4). Observe that $\{\mathcal{F}_R^t\}_{t\geq 1}$ is in a non-increasing sequence of sets that all contain $\mathcal{F}_I$, hence there exists $\mathcal{F}_R^*$ such that $\lim_{t\to\infty} \mathcal{F}_R^t = \mathcal{F}_R^*$ and $\mathcal{F}_I \subseteq \mathcal{F}_R^*$. Steps in Algorithm 1 indicate that $(\boldsymbol{x}^t, \mathbb{P}_v^t)$ is a saddle point of the minimax typed problem $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_R^t} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]$, hence we have

$$\sup_{\mathbb{P}\in\mathcal{F}_R^t} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^t,\tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^t}\left[f(\boldsymbol{x}^t,\tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^t}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \tag{3.12}$$

Since $\mathcal{X}$ and $\mathcal{F}$ are compact, the sequence of $\{\boldsymbol{x}^t, \mathbb{P}_v^t\}_{t\geq 0}$ has a subsequence converging to a limit point $(\boldsymbol{x}^*, \mathbb{P}_v^*)$. Consequently, inequality (3.12) becomes

$$\sup_{\mathbb{P}\in\mathcal{F}_R^*} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],$$

which implies that $(\boldsymbol{x}^*, \mathbb{P}_v^*)$ is a saddle point for $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_R^*} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]$. Note that $\mathcal{F}_I \subseteq \mathcal{F}_R^*$ implies $\sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right] \leq \sup_{\mathbb{P}\in\mathcal{F}_R^*} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right]$, we then have

$$\sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right] \leq \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x}^*,\tilde{\boldsymbol{z}})\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_v^*}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right].$$

Moreover, we can see that $\mathbb{P}_v^* \in \mathcal{F}_I$. Otherwise, there must be some i such that the corresponding separation problem yields $Z_i(\mathbb{P}_v^*) < 0$ for some optimal solution $\boldsymbol{q}^*$. Let $\boldsymbol{q}^t$ be the optimal solution for the separation problem at t-th iteration, i.e.,

$$\boldsymbol{q}^t \in \arg\min_{\boldsymbol{q}\in\mathcal{Q}_i} \left\{ \psi_i\left(h_i(\boldsymbol{q})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q},\tilde{\boldsymbol{z}})\right]\right) \right\}.$$

The procedure of Algorithm 1 ensures that

$$\psi_i\left(h_i(\boldsymbol{q}^t)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\boldsymbol{q}^t,\tilde{\boldsymbol{z}})\right]\right) \geq 0,\ \forall t \geq 1. \tag{3.13}$$

By definition of $\boldsymbol{q}^t$, the following inequality holds for every t:

$$\psi_i\left(h_i(\boldsymbol{q}^t)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q}^t,\tilde{\boldsymbol{z}})\right]\right) \leq \psi_i\left(h_i(\boldsymbol{q}^*)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^t}\left[g_i(\boldsymbol{q}^*,\tilde{\boldsymbol{z}})\right]\right).$$

Let $t \to \infty$, we arrive at $\psi_i\left(h_i(\bar{\boldsymbol{q}})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\bar{\boldsymbol{q}}, \tilde{\boldsymbol{z}})\right]\right) \leq \psi_i\left(h_i(\boldsymbol{q}^*)\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\boldsymbol{q}^*, \tilde{\boldsymbol{z}})\right]\right) = Z_i(\mathbb{P}_v^*)$ for some accumulation point, $\bar{\boldsymbol{q}}$ of the sequence $\boldsymbol{q}^t$. However, by inequality (3.13), we also have $\psi_i\left(h_i(\bar{\boldsymbol{q}})\right) - \psi_i\left(\mathbb{E}_{\mathbb{P}_v^*}\left[g_i(\bar{\boldsymbol{q}}, \tilde{\boldsymbol{z}})\right]\right) \geq 0$, which contradicts $Z_i(\mathbb{P}_v^*) < 0$. Therefore, $(\boldsymbol{x}^*, \mathbb{P}_v^*)$ is a saddle point for $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$, which asserts $\boldsymbol{x}^*$ is an optimal solution of Problem (3.1). □

Since the separation problem (3.11) is generally not convex, it could be computationally challenging to obtain its optimal solution. Nevertheless, for the case of covariance dominance ambiguity set, we can solve the separation problem to optimality in polynomial time because the corresponding separation problem is a minimal eigenvalue problem. For the case of entropic dominance ambiguity set, we can reasonably efficiently obtain solutions that attain negative objective values using the trust region method (Conn et al. 2000). We note that if the solver adopts the primal-dual approach to solve Problem (3.4) and provides the corresponding optimal dual solutions, we can then construct the worst-case distribution using these associated dual variables without solving Problem (3.9).

3.3 Case of covariance dominance ambiguity set

In this section, we consider the following distributionally robust optimization model,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \tag{3.14}$$

with $\mathcal{F}_C$ being the covariance dominance ambiguity set as in (2.2). Under the covariance dominance ambiguity set of the form (2.2), we can express Problem (3.14) as a positive semidefinite conic program as shown in the following result.

Proposition 3. Problem (3.14) is equivalent to

$$
\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} \rangle \\
\text{s.t.} \quad & \alpha - \boldsymbol{b}_k(\boldsymbol{x}) + 2\boldsymbol{\chi}_k'\boldsymbol{\mu} - \delta_k - t_k \geq 0, && \forall k \in [K] \\
& \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - 2\boldsymbol{\chi}_k = \boldsymbol{0}, && \forall k \in [K] \\
& \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_k, && \forall k \in [K] \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}), && \forall k \in [K] \\
& \begin{pmatrix} \delta_k & \boldsymbol{\chi}_k' \\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k \end{pmatrix} \succeq \boldsymbol{0}, && \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{3.15}
$$

where $\mathcal{K}^*(\mathcal{W})$ is the dual cone of $\mathcal{K}(\mathcal{W}) = \operatorname{cl}\left\{(\boldsymbol{z}, t) \in \mathbb{R}^N \times \mathbb{R} \mid \boldsymbol{z}/t \in \mathcal{W}, t > 0\right\}$ and $\langle \cdot, \cdot \rangle$ denotes the trace inner product of two matrices.

Proof. The inner supremum in Problem (3.14) is given by the following classical robust optimization problem:

$$
\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} \rangle \\
\text{s.t.} \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \geq \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}), \quad \forall (\boldsymbol{z}, 1) \in \mathcal{K}_{\mathcal{W}}, \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})', \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\Gamma} \succeq \boldsymbol{0},
\end{aligned}
\tag{3.16}
$$

where we introduce an auxiliary variable $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ and represent the constraint $\boldsymbol{z} \in \mathcal{W}$ by $(\boldsymbol{z}, 1) \in \mathcal{W}$. Using Schur complement, we know that each k-th robust counterpart is not violated if and only if

$$
\begin{aligned}
\inf \quad & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \\
\text{s.t.} \quad & \begin{pmatrix} 1 & (\boldsymbol{z} - \boldsymbol{\mu})' \\ (\boldsymbol{z} - \boldsymbol{\mu}) & \boldsymbol{U} \end{pmatrix} \succeq \boldsymbol{0} \\
& (\boldsymbol{z}, 1) \in \mathcal{K}_{\mathcal{W}}
\end{aligned}
$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. The dual of the above problem is

$$
\begin{aligned}
\sup \quad & 2\boldsymbol{\chi}_k'\boldsymbol{\mu} - \delta_k - t_k \\
\text{s.t.} \quad & \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - 2\boldsymbol{\chi}_k = \mathbf{0} \\
& \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_k \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*_{\mathcal{W}} \\
& \begin{pmatrix} \delta_k & \boldsymbol{\chi}_k' \\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k \end{pmatrix} \succeq \mathbf{0}.
\end{aligned}
$$

Substituting the above dual formulation and performing the outer and inner minimizations jointly, we then have formulation (3.15). □

As mentioned early, a covariance dominance ambiguity set can be represented in the form of an infinitely constrained ambiguity set as in (2.4). With this representation, we now consider the corresponding relaxed ambiguity set as follows:

$$
\mathcal{G}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\right)^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},
$$

where $\bar{\mathcal{Q}} = \left\{\boldsymbol{q}_j \mid j \in [J]\right\}$ for some $\boldsymbol{q}_j \in \mathbb{R}^N$, $\|\boldsymbol{q}_j\|_2 \leq 1, j \in [J]$. The relaxed distributionally robust optimization is then given by

$$
\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{G}_C} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right], \tag{3.17}
$$

which in contrast with Problem (3.14), is equivalent to a second order conic program. In addition, the optimization problem for determining the worst-case probability distribution required for solving the separation problem is also a second order conic program.

Proposition 4. Problem (3.17) is equivalent to

$$
\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j' \boldsymbol{\Sigma} \boldsymbol{q}_j \\
\text{s.t.} \quad & \alpha - b_k(\boldsymbol{x}) + \sum_{j\in[J]} \big(m_{jk} - l_{jk} + 2n_{jk}\boldsymbol{q}_j'\boldsymbol{\mu}\big) - t_k \geq 0, && \forall k \in [K] \\
& \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - \sum_{j\in[J]} 2n_{jk}\boldsymbol{q}_j = \boldsymbol{0}, && \forall k \in [K] \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}), && \forall k \in [K] \\
& \gamma_j = l_{jk} + m_{jk}, && \forall j \in [J], \forall k \in [K] \\
& (l_{jk}, m_{jk}, n_{jk}) \in \mathcal{K}_{soc}, && \forall j \in [J], \forall k \in [K] \\
& \boldsymbol{l}_k, \boldsymbol{m}_k, \boldsymbol{n}_k \in \mathbb{R}^J, && \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \ \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{3.18}
$$

Given the relaxed ambiguity set $\mathcal{F}_R$, let $(\boldsymbol{\xi}_k^*, \eta_k^*)_{k\in[K]}$ be the optimal solution to the following tractable conic optimization problem,

$$
\begin{aligned}
\sup \quad & \rho \\
\text{s.t.} \quad & \sum_{k\in[K]} \big(\boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k\big) \geq \rho, && \forall \boldsymbol{x} \in \mathcal{X} \\
& \sum_{k\in[K]} \eta_k = 1 \\
& \sum_{k\in[K]} \boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k\in[K]} \zeta_{kj} \leq \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j, && \forall j \in [J] \\
& (\boldsymbol{\xi}_k, \eta_k) \in \mathcal{K}(\mathcal{W}), && \forall k \in [K] \\
& \big(\zeta_{kj} + \eta_k, \zeta_{kj} - \eta_k, 2\boldsymbol{q}_j'(\boldsymbol{\xi}_k - \eta_k\boldsymbol{\mu})\big) \in \mathcal{K}_{soc}, && \forall j \in [J], \forall k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, && \forall k \in [K].
\end{aligned}
$$

The worst-case distribution is given by

$$\mathbb{P}_v\left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*}\right] = \eta_k^*,\ \forall k \in [K] : \eta_k^* > 0.$$

Proof. The inner supremum in Problem (3.17) can be reformulated as

$$\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j \\
\text{s.t.} \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j\in[J]} \gamma_j (\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu}))^2 \geq \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}), \quad \forall (\boldsymbol{z},1) \in \mathcal{K}(\mathcal{W}), \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}_+^J,
\end{aligned}$$

where each k-th constraint is not violated if and only if the optimal value of the following problem

$$\begin{aligned}
\inf \quad & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\,\boldsymbol{z} + \boldsymbol{\gamma}'\boldsymbol{u} \\
\text{s.t.} \quad & u_j \geq \left(\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu})\right)^2,\ \forall j \in [J] \\
& (\boldsymbol{z}, 1) \in \mathcal{W}
\end{aligned}$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. In the above formulation, we introduce an auxiliary vector $\boldsymbol{u}$ such that u_j are associated with the epigraphs of quadratic functions $\left(\boldsymbol{q}_j'(\boldsymbol{z}-\boldsymbol{\mu})\right)^2$. The first set of constraints has an equivalent second order conic representation as

$$\left(u_j + 1, u_j - 1, 2\boldsymbol{y}_j'(\boldsymbol{z}-\boldsymbol{\mu})\right) \in \mathcal{K}_{soc},$$

which enables us to derive the dual problem as below:

$$
\begin{aligned}
\sup \quad & \sum_{j\in[J]} (m_{jk} - l_{jk} + 2n_{jk}\boldsymbol{q}_j'\boldsymbol{\mu}) - t_k \\
\text{s.t.} \quad & \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - \sum_{j\in[J]} 2n_{jk}\boldsymbol{q}_j = \boldsymbol{0} \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}) \\
& \gamma_j = l_{jk} + m_{jk}, && \forall j \in [J] \\
& (l_{jk}, m_{jk}, n_{jk}) \in \mathcal{K}_{soc}, && \forall j \in [J] \\
& \boldsymbol{l}_k, \boldsymbol{m}_k, \boldsymbol{n}_k \in \mathbb{R}^J.
\end{aligned}
$$

Note that set of conic constraints follows from the fact that the second order cone is self-dual, i.e., $\mathcal{K}_{soc} = \mathcal{K}_{soc}^*$. Reinjecting this dual into the above reformation of the inner supremum and combining it with the outer minimization over $\boldsymbol{x} \in \mathcal{X}$, we then have reformulation (3.17). The proof for the second item is a direct implementation of Theorem 4 and is thus omitted. $\square$

Given the worst-case distribution, $\mathbb{P}_v$, the corresponding separation problem would be

$$\min_{\boldsymbol{q}:\ \|\boldsymbol{q}\|_2\leq 1} \boldsymbol{q}' \left(\boldsymbol{\Sigma} - \mathbb{E}_{\mathbb{P}_v}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right]\right)\boldsymbol{q},$$

which is the classical minimal eigenvalue problem that can be solved efficiently using numerical techniques. Whenever the minimal eigenvalue is negative, we can add the corresponding eigenvector into the set $\bar{\mathcal{Q}}$, which ultimately would tighten the relaxation of the ambiguity set. Quite interestingly, there exists an ambiguity set, $\mathcal{G}_C$ with $J = N$ such that Problems (3.14) and (3.17) have the same objective value, as elucidated by the following result.

Theorem 6. Consider any function $f : \mathbb{R}^N \mapsto \mathbb{R}$ for which the problem

$$\sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$$

is finitely optimal. Then there exists an ambiguity set, $\mathcal{G}_C$ such that $J = N$ and

$$\sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P}\in\mathcal{G}_C} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right].$$

Proof. Consider the problem $\sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$ and apply duality on probability distributions, we have

$$\begin{aligned} Z_A = \quad & \inf \quad \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} \rangle \\ & \text{s.t.} \quad \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \left\langle \boldsymbol{\Gamma}, (\boldsymbol{z}-\boldsymbol{\mu})(\boldsymbol{z}-\boldsymbol{\mu})' \right\rangle \geq f(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \qquad \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\Gamma} \succeq \boldsymbol{0}, \end{aligned} \tag{3.19}$$

whose optimal solution we denote by $(\alpha^*, \boldsymbol{\beta}^*, \boldsymbol{\Gamma}^*)$. Using the eigendecomposition of a positive semidefinite matrix, we have $\boldsymbol{\Gamma}^* = \sum_{n\in[N]} \lambda_n^* \boldsymbol{q}_n^* (\boldsymbol{q}_n^*)'$, where $(\lambda_n^*)_{n\in[N]}$ are eigenvalues of $\boldsymbol{\Gamma}^*$ and $(\boldsymbol{q}^*)_{n\in[N]}$ are corresponding eigenvectors. Note that we have $\lambda_n^* \geq 0, \forall n \in [N]$ since $\boldsymbol{\Gamma}^* \succeq \boldsymbol{0}$.

Let $\mathcal{G}_C$ be the particular relaxed ambiguity set with $\bar{\mathcal{Q}} = \{\boldsymbol{q}_1^*, \ldots, \boldsymbol{q}_N^*\}$ and consider the problem $\sup_{\mathbb{P}\in\mathcal{G}_C} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$. Observe that this problem is equivalent to the following dual problem.

$$\begin{aligned} Z_B = \quad & \inf \quad \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{n\in[N]} \gamma_n (\boldsymbol{q}_n^*)' \boldsymbol{\Sigma} \boldsymbol{q}_n^* \\ & \text{s.t.} \quad \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{n\in[N]} \gamma_n ((\boldsymbol{q}_n^*)'(\boldsymbol{z}-\boldsymbol{\mu}))^2 \geq f(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \qquad \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}^N_+. \end{aligned} \tag{3.20}$$

Based on the optimal solution of Problem (3.19), we can construct a feasible solution $(\alpha^B, \boldsymbol{\beta}^B, \boldsymbol{\gamma}^B)$ to Problem (3.20) by letting: $\alpha^B = \alpha^*, \boldsymbol{\beta}^B = \boldsymbol{\beta}^*$, and $\boldsymbol{\gamma}^B = \boldsymbol{\lambda}^*$. Observe that

$$\langle \boldsymbol{\Gamma}^*, \boldsymbol{\Sigma} \rangle = \sum_{n\in[N]} \langle \lambda_n^* \boldsymbol{q}_n^* (\boldsymbol{q}_n^*)', \boldsymbol{\Sigma} \rangle = \sum_{n\in[N]} \gamma_n (\boldsymbol{q}_n^*)' \boldsymbol{\Sigma} \boldsymbol{q}_n^*,$$

hence we have $Z_B \leq Z_A$. On the other hand, we already know that $Z_A \leq Z_B$ for any relaxed ambiguity set. Therefore, our claim holds. □

For a covariance dominance ambiguity set $\mathcal{F}_C$, we call such an ambiguity set $\mathcal{G}_C$ in Theorem 6 as the corresponding finite reduction. As in the proof of Theorem 6, we show the existence of a finite reduction by construction. However, it seems that not until the distributionally robust optimization problem with a covariance dominance ambiguity set has been solved, can we obtain the finite reduction. Nevertheless, we believe that the concept of finite reduction has potential in practice and we will give a numerical example in Section 3.4.1.

3.3.1 An application in appointment scheduling

Distributionally robust appointment scheduling problems under various forms of ambiguity sets have been proposed and studied in literature (see, for instance, Kong et al. 2013, Mak et al. 2014, Qi 2016, Bertsimas et al. 2017). In our computational study, we adopt the covariance dominance ambiguity set and investigate the efficiency of our proposed optimization procedure via Algorithm 1. We consider a service system with one server and K participants (with indices $k \in [K]$) that arrive according to a pre-determined service sequence. Service time for the k-th participant is uncertain and denoted by $\tilde{z}_k$. Thus, we only need to determine the time allowance x_k (or equivalently the appointment time) for each participant. We require the desired server's completion time to be upper bounded by T so that the feasible scheduling plan $\boldsymbol{x} = (x_1, \cdots, x_K)$ is restricted to

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \mathbb{R}_+^K \;\middle|\; \sum_{k=1}^{K} x_k \leq T \right\}.$$

To obtain a tractable optimization model, we adopt the approach of Qi (2016) and consider the following formulation that splits the total waiting time in the system into short delays among every participant.

$$\min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \sum_{k=1}^{K} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[f_k(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] + \rho \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[f_{K+1}(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] \right\},$$

where

$$f_k(\boldsymbol{x}, \boldsymbol{z}) = \max\left\{0, (z_{k-1} - x_{k-1}), \cdots, \sum_{r=2}^{k}(z_{r-1} - x_{r-1})\right\}.$$

Note that $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f_k(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$ denotes the worst-case expected waiting objective experienced by the participant for $k \in [K]$, or the worst-case expected overtime objective of the server for $k = K + 1$. The parameter $\rho > 0$ represents the relative tradeoff between a participant's delay and the server's over time. We consider numerical settings similar to Mak et al. (2014), where uncertain service times are independent. The number of jobs is eight and the unit overtime objective ρ is 2. For each instance, we generate μ_k from the uniform distribution over $[30, 60]$ and $\sigma_k = \epsilon \cdot \mu_k$, where ϵ is randomly generated from uniform distribution over $[0, 0.3]$. We set the desired server's completion time as

$$T = \sum_{k=1}^{K} \mu_k + 0.5\sqrt{\sum_{k=1}^{K} \sigma_k^2}.$$

Inspired by Kong et al. (2013), we consider the following covariance dominance ambiguity set

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^K\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}_+^K) = 1 \end{array} \right\}.$$

We also consider the marginal moment ambiguity set inspired by Mak et al. (2014) as follows:

$$\mathcal{F}_M = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^K\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{z}_k - \mu_k)^2\right] \leq \boldsymbol{e}_k'\boldsymbol{\Sigma}\boldsymbol{e}_k, \quad \forall k \in [K] \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}_+^K) = 1 \end{array} \right\},$$

which is a superset of $\mathcal{F}_C$. We initialize Algorithm 1 with the above marginal moment ambiguity set and iteratively improve the approximate second order conic solution with 15 iterations. We compare the second order conic approach against the positive semidefinite conic approach and report the iteratively improved relaxation of the former approach.

In Table 1, we report the objective values attained for five randomly generated instances using the positive semidefinite conic approach (denoted by Z_C) and the second order conic approach at different iterations (denoted by Z_M and Z_C^i, where i represents the number of iterations). The difference between Z_M and Z_C reveals the significance of cross-moment information (or more specifically, uncorrelated service times) in this distributionally robust appointment scheduling problem. The third to seventh rows show the iterative improvement of second order conic solution. In Figure 1, we can see that the performance ratio, defined as the relative gap between the second order conic solution and positive semidefinite conic solution, diminishes to zero quickly.

Instance Index	1	2	3	4	5
Z_M	75.88	100.51	106.49	126.18	143.21
Z_C^2	68.93	94.04	97.64	116.83	133.17
Z_C^3	66.55	92.34	94.43	112.13	130.07
Z_C^5	64.81	89.68	91.81	105.83	126.51
Z_C^{10}	64.15	88.19	90.93	104.11	124.91
Z_C^{15}	64.01	87.78	90.75	103.69	124.53
Z_C	63.87	87.61	90.70	103.48	124.29

Tab. 3.1: Objectives values of second order conic approach and positive semidefinite conic approach.

3.3.2 An experiment on the initialization of Algorithm 1

In the application of appointment scheduling, we naturally choose the marginal moment ambiguity set as the initial relaxed covariance dominance ambiguity set. In this subsection, to better elaborate the impact of the initialization of Algorithm 1, we conduct an experiment using randomly generated instances. Specifically, we study the following

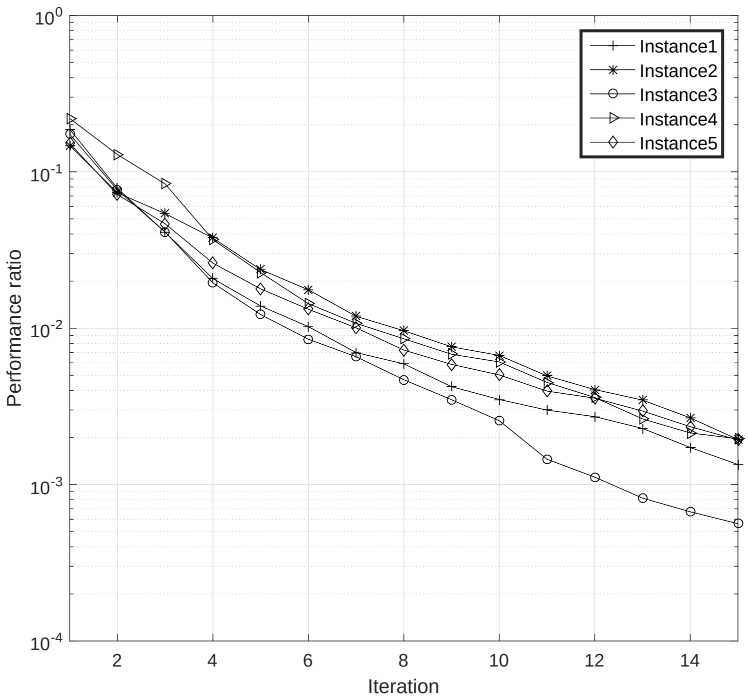

Fig. 3.1: Relative gap between second order conic approach and positive semidefinite conic approach.

optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]$$

and we use Algorithm 1 to obtain a sequence of approximate solutions.

The setting of this experiment is described as follows. The set of feasible decision is a simplex, that is,

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \mathbb{R}^M_+ \;\middle|\; \sum_{m\in[M]} x_m = 1, \right\}.$$

The cost function is given by

$$f(\boldsymbol{x},\boldsymbol{z}) = \max_{k\in[K]} \left\{ (\boldsymbol{S}_k'\boldsymbol{x} + \boldsymbol{t}_k)'\boldsymbol{z} + \boldsymbol{s}_k'\boldsymbol{x} + t_k \right\},$$

and the covariance dominance ambiguity set is given by

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{P}\left[\underline{\boldsymbol{z}} \leq \tilde{\boldsymbol{z}} \leq \bar{\boldsymbol{z}}\right] = 1 \end{array} \right\}.$$

We consider uncertain components to be independent. For different sizes of the problem (determined by K, M, N), we randomly generate 50 instances as follows. We generate μ_n from the uniform distribution over $[0, 30]$ and $\sigma_n = \epsilon\mu_n$, where ϵ is randomly drawn from uniform distribution over $[0, 0.3]$. We generate components in $(\boldsymbol{S}_k, \boldsymbol{s}_k, \boldsymbol{t}_k, t_k)_{k\in[K]}$ from the standard normal distribution. The upper bounds and lower bounds of components are identical and are given by 30 and -30, respectively. For the initialization of Algorithm 1, we consider different relaxed covariance dominance ambiguity sets as below.

- (**Marginal moment ambiguity set**): $\mathcal{F}_M$ such that $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n | n \in [N]\}$.
- (**Partial cross moment ambiguity set**): relaxed covariance dominance ambiguity set $\mathcal{G}_C$ such that $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n | n \in [N]\} \cup \{\mathbf{1}\}$.
- (**Random relaxed ambiguity set**): relaxed covariance dominance ambiguity set $\mathcal{G}_C$ such that $|\bar{\mathcal{Q}}| = N$ and each element in $\bar{\mathcal{Q}}$ is randomly generated from a standard normal distribution.

We terminate Algorithm 1 once the approximate second order conic solution is within 0.1% accuracy of the optimal positive semidefinite conic solution, or after 50 or 100 iterations according to the size of the problem. We summarize the results of different initializations among randomly generated instances in Table 3.2. The initialization with the marginal moment ambiguity set and that with the partial cross moment ambiguity set perform almost identical. This is intuitive because the difference between them is only one expectation. Initialization with these two ambiguity sets slightly outperform that with the random relaxed ambiguity set. This observation suggests that the initialization of Algorithm 1 does not strongly influence its final performance. In fact, we have a similar observation in the case of entropic dominance ambiguity set. Nevertheless, at least in the case of covariance dominance ambiguity set, we believe that the marginal moment ambiguity set or the partial cross moment ambiguity set are good candidates for the initial relaxed ambiguity set because they are meaningful in the practical sense.

(K, M, N)	Marginal moment ambiguity set	Partial cross moment ambiguity set	Random relaxed ambiguity set
$(3, 4, 5)$	0.09 $(0.05, 0.17, 1.42)$	0.09 $(0.04, 0.17, 1.42)$	0.09 $(0.05, 0.24, 3.21)$
$(12, 10, 8)$	0.10 $(0.06, 0.14, 0.08)$	0.10 $(0.05, 0.13, 0.76)$	0.09 $(0.04, 0.14, 0.67)$
$(6, 8, 10)$	1.49 $(0.08, 2.86, 15.27)$	1.25 $(0.08, 2.56, 14.84)$	1.23 $(0.10, 2.62, 16.33)$
$(10, 12, 15)$	2.25 $(0.25, 3.41, 16.29)$	2.37 $(0.25, 3.42, 16.57)$	2.55 $(0.23, 3.71, 25.21)$
$(15, 10, 20)$	2.28 $(0.45, 2.84, 23.56)$	2.26 $(0.47, 2.83, 23.59)$	2.53 $(0.45, 3.47, 32.07)$
$(20, 15, 25)$	1.25 $(0.15, 1.50, 7.80)$	1.24 $(0.15, 1.49, 7.85)$	1.28 $(0.17, 1.50, 8.96)$

Tab. 3.2: Median (minimal, average, maximal) relative gap (%) to the exact solution for different initializations.

For the case of $K = 20, M = 15, N = 25$, we terminate Algorithm 1 after at most 100 iterations. For other cases, we terminate Algorithm 1 after at most 50 iterations.

Remark 1. The aim of our numerical experiments in Section 3.3.1 and Section 3.3.2 is to showcase the convergence of our proposed algorithm, while by no means, to suggest to solve a sequence of second order programs to approximate a positive semidefinite program, which itself can be efficiently solved using Mosek, SDPT3, and SeDuMi when the problem size is not very large. Besides, there has been some improvement in solving larger scale positive semidefinite programs, for example, using SDPNAL+ (`http://www.math.nus.edu.sg/~mattohkc/SDPNALplus.html`) developed by researches at National University of Singapore.

3.4 Case of entropic dominance ambiguity set

In this section, we present the modeling power and potential applications of the entropic dominance ambiguity set $\mathcal{F}_E$ in (2.7). Particularly, we consider the relaxed distributionally robust optimization problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{G}_E} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \tag{3.21}$$

with the relaxed entropic dominance ambiguity set as follows:

$$\mathcal{G}_E = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^N\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})'\right] \preceq \boldsymbol{\Sigma} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})\right)\right] \leq \phi(\boldsymbol{q}), \quad \forall \boldsymbol{q}\in\bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}}\in\mathcal{W}\right] = 1 \end{array} \right. \right\},$$

where $\bar{\mathcal{Q}} = \left\{\boldsymbol{q}_j \middle|\, j\in[J]\right\}$ for some $\boldsymbol{q}_j \in \mathbb{R}^N, j\in[J]$. Here in the relaxed entropic dominance ambiguity set, as mentioned earlier, we explicitly specify the support, the mean, and the variance of $\tilde{\boldsymbol{z}}$ because they are not implied by only a finite number of expectation constraints in $\mathcal{G}_E$. We will show that Problem (3.21) can be formulated as a conic optimization problem involving the exponential cone. Before we proceed, we remark that an exponential conic constraint can be approximated fairly accurately via a small number of second order conic constraints (see Appendix B in Chen and Sim 2009). This kind of successive approximation methods is supported in Matlab toolboxes, such as ROME (Goh and Sim 2011) and CVX (Grant and Boyd 2008 and Grant et al. 2008). Alternatively, an exponential conic programs can be efficiently and exactly solved by interior-point methods (see, e.g., Chares 2009 and Skajaa and Ye 2015).

Proposition 5. Problem (3.21) is equivalent to

$$
\begin{array}{lll}
\inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[j]} \gamma_j + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} \rangle & \\
\text{s.t.} & \alpha - b_k(\boldsymbol{x}) + \sum_{j\in[J]} l_{kj}\left(\boldsymbol{q}_j'\boldsymbol{\mu} + \phi\left(\boldsymbol{q}_j\right)\right) - \sum_{j\in[J]} m_{kj} + 2\boldsymbol{\chi}'\boldsymbol{\mu} - \delta_k - t_k \geq 0, & \forall k \in [K] \\
& \boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}) - \boldsymbol{r}_k - 2\boldsymbol{\chi} - \sum_{j\in[J]} l_{kj}\boldsymbol{q}_j = \boldsymbol{0}, & \forall k \in [K] \\
& \boldsymbol{\gamma} - \boldsymbol{n}_k = \boldsymbol{0}, & \forall k \in [K] \\
& \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_k, & \forall k \in [K] \\
& (l_{kj}, m_{kj}, n_{kj}) \in \mathcal{K}_{exp}^*, & \forall j \in [J], \forall k \in [K] \\
& (\boldsymbol{r}_k, t_k) \in \mathcal{K}^*(\mathcal{W}), & \forall k \in [K] \\
& \begin{pmatrix} \delta_k & \boldsymbol{\chi}_k' \\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k \end{pmatrix} \succeq \boldsymbol{0}, & \forall k \in [K] \\
& \boldsymbol{l}_k, \boldsymbol{m}_k, \boldsymbol{n}_k \in \mathbb{R}^J, & \forall k \in [K] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}_+^J,\ \boldsymbol{x} \in \mathcal{X} &
\end{array}
$$

with $\mathcal{K}_{exp}^*$ being the dual cone of the exponential cone, where the exponential cone is defined by $\mathcal{K}_{exp} = \mathrm{cl}\left\{(d_1, d_2, d_3) \,\middle|\, d_2 > 0,\ d_2 e^{d_1/d_2} \leq d_3\right\}$, and $\mathcal{K}_{exp}^*$ is explicitly expressed as $\mathcal{K}_{exp}^* = \mathrm{cl}\left\{(d_1, d_2, d_3) \,\middle|\, d_1 < 0,\ -d_1 e^{d_2/d_1} \leq e d_3\right\}$.

Proof. Observe that the set of expectation constraints can be equivalently represented as

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) - \phi(\boldsymbol{q})\right)\right] \leq 1.$$

The inner supremum in Problem (3.21) can be reformulated as the following optimization problem:

$$
\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j \in [J]} \gamma_j + \langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} \rangle \\
\text{s.t.} \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j \in [J]} \gamma_j \exp\left(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}) - \phi(\boldsymbol{q})\right) + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \geq f(\boldsymbol{x}, \boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W}, \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})' \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}^J_+.
\end{aligned}
\tag{3.22}
$$

The first constraint is equivalent to a set of robust counterparts such that

$$
\alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{j \in [J]} \gamma_j \exp\left(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi\left(\boldsymbol{q}_j\right)\right) + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \geq \boldsymbol{a}_k(\boldsymbol{x})'\boldsymbol{z} + b_k(\boldsymbol{x}), \forall \boldsymbol{z} \in \mathcal{W}, \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})'
$$

for all $k \in [K]$, each of which is satisfied if and only if the optimal value of the problem

$$
\inf_{\boldsymbol{z} \in \mathcal{W}, \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})'} \left\{ (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \sum_{j \in [J]} \gamma_j \exp\left(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi\left(\boldsymbol{q}_j\right)\right) + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \right\} \tag{3.23}
$$

is not less than $b_k(\boldsymbol{x}) - \alpha$. Introducing an auxiliary vector $\boldsymbol{v} \in \mathbb{R}^J$ and re-expressing $\boldsymbol{z} \in \mathcal{W}$ by $(\boldsymbol{z}, 1) \in \mathcal{K}(\mathcal{W})$, we then have an equivalent representation to Problem (3.23), given by

$$
\begin{aligned}
\inf \quad & (\boldsymbol{\beta} - \boldsymbol{a}_k(\boldsymbol{x}))'\boldsymbol{z} + \boldsymbol{\gamma}'\boldsymbol{v} + \langle \boldsymbol{\Gamma}, \boldsymbol{U} \rangle \\
\text{s.t.} \quad & v_j \geq \exp\left(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi\left(\boldsymbol{q}_j\right)\right), \quad \forall j \in [J] \\
& \boldsymbol{U} \succeq (\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})' \\
& (\boldsymbol{z}, 1) \in \mathcal{K}(\mathcal{W}).
\end{aligned}
\tag{3.24}
$$

The above exponential constraints can be represented set constraints

$$
\left(\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}) - \phi\left(\boldsymbol{q}_j\right), 1, v_j\right) \in \mathcal{K}_{exp},
$$

with which the dual of Problem (3.24) then can be obtained by

$$
\begin{aligned}
\sup \quad & \sum_{j\in[J]} l_{kj}\left(\boldsymbol{q}_j'\boldsymbol{\mu}+\phi\left(\boldsymbol{q}_j\right)\right)-\sum_{j\in[J]} m_{kj}+2\boldsymbol{\chi}'\boldsymbol{\mu}-\delta_k-t_k\\
\text{s.t.} \quad & \boldsymbol{\beta}-\boldsymbol{a}_k(\boldsymbol{x})-\boldsymbol{r}_k-2\boldsymbol{\chi}_k-\sum_{j\in[J]} l_{kj}\boldsymbol{q}_j=\boldsymbol{0}\\
& \boldsymbol{\gamma}-\boldsymbol{n}_k=\boldsymbol{0}\\
& \boldsymbol{\Gamma}=\boldsymbol{\Gamma}_k\\
& (l_{kj},m_{kj},n_{kj})\in\mathcal{K}^*_{exp},\ \forall j\in[J]\\
& (\boldsymbol{r}_k,t_k)\in\mathcal{K}^*(\mathcal{W})\\
& \begin{pmatrix}\delta_k & \boldsymbol{\chi}_k'\\ \boldsymbol{\chi}_k & \boldsymbol{\Gamma}_k\end{pmatrix}\succeq\boldsymbol{0}\\
& \boldsymbol{l}_k,\boldsymbol{m}_k,\boldsymbol{n}_k\in\mathbb{R}^J.
\end{aligned}
$$

Substituting this dual reformulation in (3.22) and performing the outer and inner minimizations jointly, we then have the conic reformulation. □

In addition, we can obtain the worst-case distribution by solving an exponential conic optimization problem.

Proposition 6. Given the relaxed ambiguity set $\mathcal{F}_R$, let $(\boldsymbol{\xi}_k^*,\eta_k^*)_{k\in[K]}$ be the optimal solution to the following tractable conic optimization problem,

$$
\begin{aligned}
\sup \quad & \rho \\
\text{s.t.} \quad & \sum_{k \in [K]} (\boldsymbol{a}_k(\boldsymbol{x})' \boldsymbol{\xi}_k + b_k(\boldsymbol{x})\eta_k) \geq \rho, && \forall \boldsymbol{x} \in \mathcal{X} \\
& \sum_{k \in [K]} \eta_k = 1 \\
& \sum_{k \in [K]} \boldsymbol{\xi}_k = \boldsymbol{\mu} \\
& \sum_{k \in [K]} \boldsymbol{\zeta}_k \leq \mathbf{1} \\
& \sum_{k \in [K]} \boldsymbol{\Lambda}_k = \boldsymbol{\Sigma} \\
& \left(\boldsymbol{q}_j'(\boldsymbol{\xi}_k - \eta_k \boldsymbol{\mu}) - \eta_k \phi(\boldsymbol{q}_j), \eta_k, \zeta_{kj}\right) \in \mathcal{K}_{exp}, && \forall j \in [J], k \in [K] \\
& (\boldsymbol{\xi}_k, \eta_k) \in \mathcal{K}(\mathcal{W}), && \forall k \in [K] \\
& \begin{pmatrix} \eta_k & (\boldsymbol{\xi}_k - \boldsymbol{\mu})' \\ (\boldsymbol{\xi}_k - \boldsymbol{\mu}) & \boldsymbol{\Lambda}_k \end{pmatrix} \succeq \mathbf{0}, && \forall k \in [K] \\
& \boldsymbol{\xi}_k \in \mathbb{R}^N, \boldsymbol{\zeta}_k \in \mathbb{R}^J, \eta_k \in \mathbb{R}_+, \boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}, && \forall k \in [K].
\end{aligned}
$$

The worst-case distribution is given by

$$
\mathbb{P}_v \left[\tilde{\boldsymbol{z}} = \frac{\boldsymbol{\xi}_k^*}{\eta_k^*} \right] = \eta_k^*, \ \forall k \in [K] : \eta_k^* > 0.
$$

Proof. The proof follows from Theorem 4 and is thus omitted. □

In the case of entropic dominance ambiguity set, the corresponding separation problem is unconstrained and takes the following form:

$$
\min_{\boldsymbol{q}} \Phi(\boldsymbol{q}) = \min_{\boldsymbol{q}} \left\{ \phi(\boldsymbol{q}) - \ln \mathbb{E}_{\mathbb{P}_v} \left[\exp(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})) \right] \right\} = \min_{\boldsymbol{q}} \left\{ \phi(\boldsymbol{q}) - \ln \sum_{k \in [K]} p_k \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right) \right\},
$$

where the worst-case distribution $\mathbb{P}_v$ takes value $\boldsymbol{z}^k$ with probability p_k. Unfortunately, since the objective function is non-quadratic, the above separation problem is non-convex and is generally difficult to obtain the optimal solution. Nevertheless, as we will present in the coming numerical example, the trust region method (TRM) is practically useful for finding local minimum. TRM is one of the most important numerical optimization methods in solving nonlinear programming problems. TRM works in a fashion that solves a sequence of subproblems, in each of which, it approximates the original objective function by a quadratic function (usually obtained by taking Taylor series up to second order) and searches an improving direction in a region around the current solution. Note that solving an unconstrained nonlinear optimization problem by TRM is well supported in Matlab (Mathworks 2017). To use the trust region method, we compute the gradient and Hessian of the objective function as follows:

$$\frac{\partial \Phi(\boldsymbol{q})}{\partial q_n} = \frac{\partial \phi(\boldsymbol{q})}{\partial q_n} - \frac{\sum_{k\in[K]} p_k (z_n^k - \mu_n) \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right)}{C}, \ \forall n \in [N]$$

and

$$\frac{\partial^2 \Phi(\boldsymbol{q})}{\partial q_l \partial q_n} = \frac{\partial^2 \phi(\boldsymbol{q})}{\partial q_l \partial q_n} + \frac{A_l A_n}{C^2} - \frac{B_{ln}}{C}, \ \forall l, n \in [N],$$

where A_n, B_{ln}, and C are respectively given by

$$A_n = \sum_{k\in[K]} p_k (z_n^k - \mu_n) \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right), \ \forall n \in [N],$$

$$B_{ln} = \sum_{k\in[N]} p_k (z_l^k - \mu_l)(z_n^k - \mu_n) \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right), \ \forall l, n \in [N],$$

and

$$C = \sum_{k\in[K]} p_k \exp\left(\boldsymbol{q}'(\boldsymbol{z}^k - \boldsymbol{\mu})\right).$$

3.4.1 An application in portfolio selection

In our numerical example, we illustrate how our new bound for the expected surplus $\mathbb{E}_{\mathbb{P}}\left[(\cdot)^+\right]$ can be leveraged in practice. In particular, we study the distributionally robust portfolio optimization problem under the worst-case CVaR measure of Rockafellar and

Uryasev (2002). We consider N assets, each with independently distributed random return premium $\mu_n + \sigma_n \tilde{z}_n$, influenced by the uncertainty $\tilde{z}_n$ with mean 0 and support $[-1, 1]$. The parameters used in our study are $N = 50$,

$$\mu_n = \frac{n}{250}, \ \sigma_n = \frac{N\sqrt{2n}}{1000}.$$

Thus, the asset with higher return premium is at same time more risky. The feasible set of investment is a simplex given by

$$\boldsymbol{x} \in \mathcal{X} = \left\{ \boldsymbol{x} \in \mathbb{R}^N_+ \ \middle| \ \sum_{n \in [N]} x_n = 1 \right\}.$$

The total return premium obtained from an investment $\boldsymbol{x}$ under realization $\boldsymbol{z}$ is $L(\boldsymbol{x}, \boldsymbol{z}) = \sum_{n \in [N]} x_n(\mu_n + \sigma_n z_n)$. We consider the following optimization model that seeks an investment plan with minimized worst-case CVaR.

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\text{-CVaR}_{1-\epsilon} \left(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right), \tag{3.25}$$

where the CVaR at level ϵ with respect to a probability distribution $\mathbb{P}$ is

$$\mathbb{P}\text{-CVaR}_{1-\epsilon} \left(L(\boldsymbol{x}, \boldsymbol{x}) \right) = \min_{\theta \in \mathbb{R}} \left\{ \theta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

By the stochastic min-max theorem by Shapiro and Kleywegt (2002), we can re-express the worst-case CVaR in the objective function of Problem (3.25) as

$$\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\text{-CVaR}_{1-\epsilon} \left(L(\boldsymbol{x}, \boldsymbol{x}) \right) = \min_{\theta \in \mathbb{R}} \left\{ \theta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

Consequently, the worst-case CVaR portfolio selection problem (3.25) becomes

$$\min_{\boldsymbol{x} \in \mathcal{X}, \theta \in \mathbb{R}} \left\{ \theta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[(-L(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \theta)^+ \right] \right\}.$$

We consider the following covariance dominance ambiguity set that encompasses the family of distributions of these independently distributed uncertainties $\tilde{z}_n$

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right] \leq \boldsymbol{q}'\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \mathbb{R}^N : \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},$$

where the support set is modeled as $\mathcal{W} = \left\{\boldsymbol{z} \in \mathbb{R}^N \mid \|\boldsymbol{z}\|_\infty \leq 1\right\}$. Observe that $\tilde{z}_n$ is sub-Gaussian with zero mean and unit deviation parameter for all $n \in [N]$. We then also consider the following entropic ambiguity set that captures the sub-Gaussianity of these uncertainties

$$\mathcal{F}_G = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'\tilde{\boldsymbol{z}}\right)\right] \leq \dfrac{1}{2}\boldsymbol{q}'\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \mathbb{R}^N \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},$$

where we explicitly specify the support set $\mathcal{W}$. The reason is that after applying Theorem 2, the chosen function $\phi(\cdot)$ only implies the support of $\tilde{\boldsymbol{z}}$ is $\mathbb{R}^N$. As shown in the following result, the entropic dominance ambiguity set improves upon the covariance dominance ambiguity set in capturing independently distributed random variables with known mean and support.

Proposition 7. Let $\tilde{z}_1, \ldots, \tilde{z}_N$ be independently distributed random variables, with zero means and support $\mathcal{W} = \left\{\boldsymbol{z} \in \mathbb{R}^N \mid \|\boldsymbol{z}\|_\infty \leq 1\right\}$. The minimal covariance dominance ambiguity set that encompasses this family of distributions is $\mathcal{F}_C$. On the other hand, the entropic dominance ambiguity set, $\mathcal{F}_G$ provides a better characterization of this family of distributions, that is, $\mathcal{F}_G \subseteq \mathcal{F}_C$.

Proof. Observe that for any $\boldsymbol{q} \in \mathbb{R}^N$,

$$\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right] = \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[N]}\sum_{j\in[N]} q_i q_j \tilde{z}_i \tilde{z}_j\right] = \sum_{n\in[N]} q_n^2 \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right],$$

where the second equality follows from mutual independence and zero means of random variables. To show $\mathcal{F}_C$ is the required minimal covariance dominance ambiguity set, we only need to show

$$\boldsymbol{q}'\boldsymbol{q} = \sum_{n\in[N]} q_n^2 \sup_{\mathbb{P}\in\mathcal{F}_n} \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right], \ \forall \boldsymbol{q} \in \mathbb{R}^N,$$

where for every $n \in [N]$,

$$\mathcal{F}_n = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) \;\middle|\; \begin{array}{l} \tilde{z} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{z}_n] = 0 \\ \mathbb{P}[-1 \leq \tilde{z}_n \leq 1] = 1 \end{array} \right\}.$$

By weak duality, we have

$$\begin{array}{rll} \sup_{\mathbb{P}\in\mathcal{F}_n} \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right] \leq & \inf & \alpha \\ & \text{s.t.} & \alpha + \beta z_n \geq z_n^2, \quad \forall -1 \leq z_n \leq 1 \\ & & \alpha, \beta \in \mathbb{R}. \end{array}$$

Since $z_n^2 - \beta z_n$ is convex in z_n, we note that

$$\alpha \geq \max_{-1\leq z_n\leq 1,\beta} \left(z_n^2 - \beta z_n\right) = \max_{\beta} \{1-\beta, 1+\beta\}.$$

Therefore,

$$\sup_{\mathbb{P}\in\mathcal{F}_n} \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right] \leq \inf \alpha = 1.$$

The strong duality holds as the above equality can be achieved by a two-point distribution that takes values in $\{-1, 1\}$ with equal probability. Consequently, we have

$$\sum_{n\in[N]} q_n^2 \sup_{\mathbb{P}\in\mathcal{F}_n} \mathbb{E}_{\mathbb{P}}\left[\tilde{z}_n^2\right] = \sum_{n\in[N]} q_n^2 = \boldsymbol{q}'\boldsymbol{q},\ \forall \boldsymbol{q} \in \mathbb{R}^N.$$

Consider any probability distribution, $\mathbb{P}$ in $\mathcal{F}_G$. The means of random variables are given by $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \nabla\phi(\boldsymbol{0}) = \boldsymbol{0}$. Likewise, their covariance is bounded by

$$\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'] \preceq \nabla^2\phi(\boldsymbol{0}) = \boldsymbol{I},$$

where $\boldsymbol{I}$ denotes the identity matrix. This is equivalent to $\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right] \leq \boldsymbol{q}'\boldsymbol{q}, \forall \boldsymbol{q} \in \mathbb{R}^N$. Therefore, $\mathbb{P} \in \mathcal{F}_C$, and more importantly, $\mathcal{F}_G \subseteq \mathcal{F}_C$. $\square$

We investigate the numerical performance of the following relaxed entropic dominance ambiguity set, $\mathcal{G}_G$

$$\mathcal{G}_G = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0} \\ \ln \mathbb{E}_{\mathbb{P}}[\exp(\boldsymbol{q}'\tilde{\boldsymbol{z}})] \leq \dfrac{1}{2}\boldsymbol{q}'\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\}$$

against the covariance dominance ambiguity set $\mathcal{F}_C$.

We initialize Algorithm 1 with $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n \mid n \in [N]\}$. In each iteration, we solve the separation problem using the trust region method with a randomly generated initial solution. If such a local minimum attains a negative objective value for the separation problem, we find a violating expectation constraint and proceed to the next iteration of Algorithm 1. Otherwise, we continue with trust region method using another randomly generated initial solution. The algorithm terminates if no violating expectation constraint is found after 100 trials.

In Table 3.3, we report the objective values for various confidence levels,

$$\epsilon \in \{0.01, 0.02, 0.05, 0.08, 0.1\}.$$

In particular, Z_C denotes the objective value obtained by the covariance dominance ambiguity set, and Z_G^i is the objective value achieved at the i-th iteration using the entropic dominance ambiguity set. The relaxed entropic dominance solutions would yield significant lower objective values than those obtained from the covariance dominance ambiguity set. We also observe that the entropic dominance approach converges reasonably well and terminates in at most ten iterations.

ϵ	0.01	0.02	0.05	0.08	0.1
Z_C	0.0674	0.0674	0.0674	0.0579	0.0430
Z_G^1	0.0674	0.0674	0.0674	0.0674	0.0674
Z_G^2	0.0674	0.0674	0.0674	0.0661	0.0661
Z_G^3	0.0474	0.0674	0.0484	0.0251	0.0569
Z_G^4	0.0443	0.0408	0.0247	0.0126	0.0517
Z_G^5	0.0442	0.0350	0.0201	0.0126	0.0486
Z_G^6	—	0.0347	0.0199	0.0106	0.0163
Z_G^7	—	—	0.0199	0.0105	0.0069
Z_G^8	—	—	—	0.0105	0.0055
Z_G^9	—	—	—	—	0.0054
Z_G^{10}	—	—	—	—	—

Tab. 3.3: Objective values of covariance dominance approach and entropic dominance approach.

We also initialize Algorithm 1 with the set $\mathcal{Q}$ contains some randomly selected $\boldsymbol{q}$'s. Particularly, we let $|\bar{\mathcal{Q}}| = N$, where components of each $\boldsymbol{q}$ are independent and identically distributed and follow a normal distribution with a zero mean. We vary the standard deviation σ_* of this normal distribution and report the objective values for various confidence levels in Table 3.4, where the superscript ∞ denotes the solution at the termination of Algorithm 1. With a random initialization, Algorithm 1 still terminates in at most ten iterations and returns the same results.

Proposition 7 shows the entropic dominance ambiguity set provides a better characterization of the independently distributed uncertainties considered in this example.

<table>
<tr><td></td><td>ϵ</td><td>0.01</td><td>0.02</td><td>0.05</td><td>0.08</td><td>0.09</td><td>0.1</td></tr>
<tr><td></td><td>Z_C</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0579</td><td>0.0499</td><td>0.0430</td></tr>
<tr><td rowspan="2">$\sigma_* = 1$</td><td>Z_G^1</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td></tr>
<tr><td>Z_G^∞</td><td>0.0442</td><td>0.0347</td><td>0.0199</td><td>0.0105</td><td>0.0079</td><td>0.0054</td></tr>
<tr><td rowspan="2">$\sigma_* = 10$</td><td>Z_G^1</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td></tr>
<tr><td>Z_G^∞</td><td>0.0442</td><td>0.0347</td><td>0.0199</td><td>0.0105</td><td>0.0079</td><td>0.0054</td></tr>
<tr><td rowspan="2">$\sigma_* = 100$</td><td>Z_G^1</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td></tr>
<tr><td>Z_G^∞</td><td>0.0442</td><td>0.0347</td><td>0.0199</td><td>0.0105</td><td>0.0079</td><td>0.0054</td></tr>
<tr><td rowspan="2">$\sigma_* = 1000$</td><td>Z_G^1</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td></tr>
<tr><td>Z_G^∞</td><td>0.0442</td><td>0.0347</td><td>0.0199</td><td>0.0105</td><td>0.0079</td><td>0.0054</td></tr>
<tr><td rowspan="2">$\sigma_* = 10000$</td><td>Z_G^1</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td><td>0.0674</td></tr>
<tr><td>Z_G^∞</td><td>0.0442</td><td>0.0347</td><td>0.0199</td><td>0.0105</td><td>0.0079</td><td>0.0054</td></tr>
</table>

Tab. 3.4: Objective values of covariance dominance approach and entropic dominance approach under different random initializations.

However practically, when we implement the relaxed entropic dominance ambiguity set instead, it may not necessarily be a better characterization than the covariance dominance ambiguity set. For example, as shown in Table 3.3 and Table 3.4, in the very first few iterations, the objective obtained from entropic dominance approach is not lower than that from covariance dominance approach. The reason is that with only a finite number of expectation constraints, the entropic dominance ambiguity set is not able to implicitly captures the covariance among uncertain components. Inspired by the concept of finite reduction, we can first implement the covariance dominance approach to obtain the finite reduction, then consider the following variant of the relaxed entropic dominance ambiguity set.

$$\bar{\mathcal{G}}_G = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^N\right) \;\middle|\; \begin{array}{ll} \tilde{\boldsymbol{z}} \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \mathbf{0} & \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{q}'\left(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}\right)\right)^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, & \forall \boldsymbol{q} \in \bar{\mathcal{Q}}^* \\ \ln \mathbb{E}_{\mathbb{P}}\left[\exp\left(\boldsymbol{q}'\tilde{\boldsymbol{z}}\right)\right] \leq \dfrac{1}{2}\boldsymbol{q}'\boldsymbol{q}, & \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 & \end{array} \right\},$$

where as shown in Theorem 6, the set $\bar{\mathcal{Q}}^*$ characterizes the finite reduction of $\mathcal{F}_C$ and $|\bar{\mathcal{Q}}^*| = N$. Following similar procedure as in Proposition 4 and Proposition 5, the resultant reformulation will be a hybrid of second order conic and exponential conic program.

We also investigate the performance of the set $\bar{\mathcal{G}}_G$. In particular, we apply Algorithm 1 with (i) $\bar{\mathcal{Q}} = \{\boldsymbol{e}_n |\ n \in [N]\}$ and (ii) a random initialization as stated previously. The results are reported in Table 3.5 and Table 3.6. As expected, now the solutions at the first iteration of the entropic dominance approach are not worse than those obtained from the covariance dominance approach. Besides, with the concept of finite reduction, the entropic dominance approach seems to converge a bit faster. Lastly, the final results obtained from different settings are identical. This shows the stability of the entropic covariance approach.

ϵ	0.01	0.02	0.05	0.08	0.09	0.1
Z_C	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
Z_G^1	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
Z_G^2	0.0674	0.0623	0.0379	0.0222	0.0181	0.0144
Z_G^3	0.0468	0.0385	0.0222	0.0120	0.0092	0.0065
Z_G^4	0.0442	0.0348	0.0199	0.0106	0.0079	0.0054
Z_G^5	—	0.0347	—	0.0105	—	—
Z_G^6	—	—	—	—	—	—

Tab. 3.5: Objective values of covariance dominance approach and entropic dominance approach (variant).

	ϵ	0.01	0.02	0.05	0.08	0.09	0.1
	Z_C	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
$\sigma_* = 1$	Z_G^1	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
	Z_G^∞	0.0442	0.0347	0.0199	0.0105	0.0079	0.0054
$\sigma_* = 10$	Z_G^1	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
	Z_G^∞	0.0442	0.0347	0.0199	0.0105	0.0079	0.0054
$\sigma_* = 100$	Z_G^1	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
	Z_G^∞	0.0442	0.0347	0.0199	0.0105	0.0079	0.0054
$\sigma_* = 250$	Z_G^1	0.0674	0.0674	0.0674	0.0579	0.0499	0.0430
	Z_G^∞	0.0442	0.0347	0.0199	0.0105	0.0079	0.0054

Tab. 3.6: Objective values of covariance dominance approach and entropic dominance approach (variant) under different random initializations.

4. ADAPTIVE DISTRIBUTIONALLY ROBUST OPTIMIZATION

4.1 Introduction

One application of robust (distributionally robust) optimization that has attracted considerable interest in recent year is adaptive robust (distributionally robust) optimization for multi-stage problems. In each stage, new *wait-and-see* decisions adaptive to revealed uncertain parameters are implemented. As such, these adaptive decisions are modeled as functions (decision rules) of the revealed uncertain parameters and lead to the computational intractability of the resultant models (see, e.g., Ben-Tal et al. 2004 and Shapiro and Nemirovski 2005). For this reason, Ben-Tal et al. (2004) suggest restricting the admissible functional forms of adaptive decisions to be affinely dependent on the random parameters, an approach that is known as linear decision rule (LDR) approximation and has been discussed in the early literature of stochastic programming (Garstka and Wets 1974). In the field of robust optimization, Ben-Tal et al. (2004) present the LDR approximation is efficient and computationally attractive and rekindle the interest in this approach.

Most of the following works in literature impose some similar restrictions and further refine adaptive decisions (i) to be piecewise-linear (for example, Chen et al. 2007, Goh and Sim 2010, and See and Sim 2010) and (ii) to be dependable on auxiliary variables that are used in describing the uncertain set or the ambiguity set (see, for instance, Chen and Zhang 2009 and Bertsimas et al. 2017). Above refinements of decision rules extend the dependence of adaptive decisions and can lead to reformulations in conic optimization formats that can ensure improved performance. In recent years, parallel to the theoretical studies in the LDR approximation, algebraic modeling packages to ease the implementations of the LDR approximation have also been developed, including YALMIP, AIMMS, ROME, JUMPER, ROC and XProg. Notably, there is also a revival of using the LDR approximation to solve multistage stochastic optimization problems (Kuhn et al. 2011 and Georghiou et al. 2015).

Certainly, there are concerns about the general sub-optimality of the LDR approximation (for example, Garstka and Wets 1974 and Bertsimas and Goyal 2012). On the other hand, it is also worth mentioning that the LDR approximation seems to perform reasonably well for some applications (e.g., Ben-Tal et al. 2005) and sometimes can even be optimal (see, for instance, Anderson and Moore 2007, Bertsimas et al. 2010b, Gounaris et al. 2013, and Iancu et al. 2013). The following quote from Shapiro and Nemirovski (2005) clarifies the rationale and trade-off for considering the LDR approximation: *"The only reason for restricting ourselves with affine decision rules stems from the desire to end up with a computationally tractable problem. We do not pretend that affine decision rules approximate well the optimal ones - whether it is so or not, it depends on the problem, and we usually have no possibility to understand how good in this respect is a particular problem we should solve. The rationale behind restricting to affine decision rules is the belief that in actual applications it is better to pose a modest and achievable goal rather than an ambitious goal which we do not know how to achieve."*

In this chapter, we adopt a framework proposed by Bertsimas et al. (2017) for solving the adaptive distributionally robust optimization problem, where the objective is to minimize the worst-case expected cost over a tractable conic representable ambiguity set. In the framework, the LDR approximation is considered to obtain tractable reformulation and is extended by being dependable on both primary random variables and auxiliary random variables that arise from the lifted ambiguity set. As such, the extended LDR approximation can be proved to improve over some more sophisticated LDR approximations developed in literature. Moreover, the adaptive distributionally robust optimization problem can be reformulated as a classical robust optimization problem with a tractable conic representable uncertainty set.

Before we proceed, we distinguish our work from that of Bertsimas et al. (2017) and summarize our contributions as follows:

1. We consider the adaptive distributionally robust optimization with infinitely constrained ambiguity sets, which would potentially allow an infinite number of expectation constraints.

2. Bertsimas et al. (2017) consider second-order conic representable ambiguity sets. We extend the consideration to ambiguity sets that are representable via tractable cones, which, besides non-negative orthants and second-order cone, refer to power cone, exponential cone, and their Cartesian product. As such, the ambiguity sets can capture the entropic dominance that provides a tighter characterization of stochastic independence than existing approach based on covariance information.

3. To solve the corresponding adaptive distributionally robust optimization problem, we apply the extended LDR approximation and propose an algorithm to obtain its solution by solving a sequence of subproblems. Each of these subproblems is associated with a relaxed finitely constrained ambiguity set and results in a tractable conic optimization problem. At each iteration, we also solve a separation problem that would lead us to a tighter relaxation of the infinitely constrained ambiguity set.

4.2 Adaptive distributionally robust optimization models

We first study a two-stage adaptive optimization problem where the first stage or here-and-now decision is a vector $\boldsymbol{x} \in \mathbb{R}^N$ chosen over the convex and compact feasible set $\mathcal{X}$. The cost incurred during the first stage in association with the decision $\boldsymbol{x}$ is deterministic and is given by $\boldsymbol{c}'\boldsymbol{x}$ for some $\boldsymbol{c} \in \mathbb{R}^N$. In progressing to the next stage, the random variable $\tilde{\boldsymbol{z}} \in \mathbb{R}^I$ with support $\mathcal{W} \subseteq \mathbb{R}^I$ is realized; thereafter, we could determine the cost incurred at the second stage. Similar to a typical stochastic programming model, for a given decision vector, $\boldsymbol{x}$ and a realization of the random variable, $\boldsymbol{z} \in \mathcal{W}$, we evaluate the second stage cost via the following linear optimization problem that involves the wait-and-see decision $\boldsymbol{y}$,

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{z}) \;=\; \min \;\; & \boldsymbol{d}'\boldsymbol{y} \\ \text{s.t.} \;\; & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\ & \boldsymbol{y} \in \mathbb{R}^L. \end{aligned} \tag{4.1}$$

Here, $\boldsymbol{A} \in \mathcal{R}^{I,M\times N}$, $\boldsymbol{b} \in \mathcal{R}^{I,M}$ are functions that map from the vector $\boldsymbol{z} \in \mathcal{W}$ to the input parameters of the linear optimization problem. As in the popular factor-based model, these functions are affinely dependent on $\boldsymbol{z}$ and are given by,

$$\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{A}_0 + \sum_{i\in[I]} \boldsymbol{A}_i z_i, \quad \boldsymbol{b}(\boldsymbol{z}) = \boldsymbol{b}_0 + \sum_{i\in[I]} \boldsymbol{b}_i z_i$$

with $\boldsymbol{A}_i \in \mathbb{R}^{M\times N}$ and $\boldsymbol{b}_i \in \mathbb{R}^M$ for any $i \in [I] \cup \{0\}$. The recourse matrix $\boldsymbol{B} \in \mathbb{R}^{M\times L}$ and the vector of cost parameters $\boldsymbol{d} \in \mathbb{R}^L$ are both constant and correspond to the stochastic programming format as fixed recourse. Note that Problem (4.1) may not always be feasible and, as in the case of complete recourse, the recourse matrix can influence the feasibility of the second stage problem.

Definition 4 (Complete recourse). The second stage problem (4.1) has complete recourse if there exists $\boldsymbol{y} \in \mathbb{R}^L$ such that $\boldsymbol{B}\boldsymbol{y} > 0$.

Complete recourse is a strong sufficient condition that guarantees the feasibility of the second stage problem for all $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{z} \in \mathbb{R}^I$. In other words, the second stage cost $f(\boldsymbol{x}, \boldsymbol{z}) < +\infty$ for any $\boldsymbol{x}$ and $\boldsymbol{z}$. Typically, a weak condition is assumed in stochastic programming to ensure that the second stage problem is feasible.

Definition 5 (Relative complete recourse). The second stage problem (4.1) has relative complete recourse if and only if the problem is feasible for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{W}$.

Beside, for the practical interest, it is reasonable and not restrictive to consider the sufficiently expensive recourse that is defined as follows.

Definition 6 (Sufficiently expensive recourse). The second stage problem (4.1) has sufficiently expensive recourse if the second stage cost $f(\boldsymbol{x}, \boldsymbol{z}) > -\infty$ for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{W}$.

The following result reveals a relation between the vector of cost parameters , $\boldsymbol{d}$ and the recourse matrix, $\boldsymbol{B}$ under the sufficiently expensive recourse condition.

Proposition 8. Under the sufficiently expensive recourse condition, the vector of cost parameters, $\boldsymbol{d}$ is a non-negative linear combination of rows of the recourse matrix, $\boldsymbol{B}$.

Proof. For any $m \in [M]$, let $\hat{\boldsymbol{b}}_m \in \mathbb{R}^{L \times 1}$ be the transpose of the m-th row of $\boldsymbol{B}$. Suppose $\boldsymbol{d}$ is not a non-negative linear combination of $\left(\hat{\boldsymbol{b}}_m\right)_{m \in [M]}$, then there exists no $\boldsymbol{p} \geq \boldsymbol{0}$ that satisfies $\sum_{m \in [M]} p_m \hat{\boldsymbol{b}}_m = \boldsymbol{d}$. This implies that the optimization problem

$$\begin{array}{ll} \max & 0 \\ \text{s.t.} & \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d} \\ & \boldsymbol{p} \geq \boldsymbol{0} \end{array}$$

is infeasible. Or equivalently, its dual problem

$$\begin{array}{ll} \min & \boldsymbol{d}'\boldsymbol{q} \\ \text{s.t.} & \boldsymbol{B}\boldsymbol{q} \geq \boldsymbol{0} \end{array}$$

is unbounded since it clearly has a feasible solution $\boldsymbol{q} = \boldsymbol{0}$. As a result, there is some feasible direction $\boldsymbol{q}$ whose second stage cost is negative, that is, $\boldsymbol{d}'\boldsymbol{q} < 0$. This contradicts the sufficiently expensive recourse that requires the second stage cost $f(\boldsymbol{x}, \boldsymbol{z})$ to be bounded from below. □

Given an ambiguity set of probability distributions, $\mathcal{F}$, the second stage cost is evaluated based on the worst-case expectation over the ambiguity set given by

$$\rho(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right]. \tag{4.2}$$

Corresponding, the here-and-now decision, $\boldsymbol{x}$ is determined by minimizing the sum of the deterministic first stage cost and the worst-case expected second stage cost given as follows:

$$\begin{aligned} Z^* = \quad & \min \quad \boldsymbol{c}'\boldsymbol{x} + \rho(\boldsymbol{x}) \\ & \text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{4.3}$$

Remark 2. The techniques developed in this chapter apply to the case of convex and piecewise affine second stage cost, i.e., the objective of Problem (4.1) can be generalized to be $\max_{k\in[k]} \boldsymbol{d}_k'\boldsymbol{y}$, which is useful in describing the utility function or risk attitude. However, for the sake of a light exposition, we only present results for the case of second stage cost being affine in $\boldsymbol{y}$.

Infinitely constrained ambiguity set

Following the stream of this thesis, we focus on a class of infinitely constrained ambiguity sets, which is defined as follows:

$$\mathcal{F}_I = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g\left(\boldsymbol{q}, \tilde{\boldsymbol{z}}\right)\right] \leq h(\boldsymbol{q}), \quad \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}\left(\tilde{\boldsymbol{z}} \in \mathcal{W}\right) = 1 \end{array} \right. \right\} \tag{4.4}$$

with parameters $\boldsymbol{G} \in \mathbb{R}^{N_1 \times I}$, $\boldsymbol{\mu} \in \mathbb{R}^{N_1}$, set $\mathcal{Q} \subseteq \mathbb{R}^{N_2}$ and functions $g : \mathcal{Q} \times \mathbb{R}^I \mapsto \mathbb{R}$ and $h : \mathcal{Q} \mapsto \mathbb{R}$. The bounded and non-empty support set $\mathcal{W}$ is a tractable conic representable set and for any given $\boldsymbol{q} \in \mathcal{Q}$, the function $g(\boldsymbol{q}, \boldsymbol{z})$ is a tractable conic representable with respect to $\boldsymbol{z}$. In addition, for any $\boldsymbol{z} \in \mathcal{W}$ such that $\boldsymbol{G}\boldsymbol{z} = \boldsymbol{\mu}$, we assume $g(\boldsymbol{q}, \boldsymbol{z}) \leq h(\boldsymbol{q}), \forall \boldsymbol{q} \in \mathcal{Q}$. As mentioned earlier in this thesis, one of the key features of an infinitely constrained ambiguity set is the potential to incorporate an infinite number of expectation constraints, which can be regarded as being parameterized by $\boldsymbol{q} \in \mathcal{Q}$. In Chapter 2, we have elucidated the generality of an infinitely constrained ambiguity set in representing a class of distributionally robust optimization problems, whose objective function is tractable conic representable in the uncertainty. The introduction of a possibly infinite number of expectation constraints also enables a tractably conic ambiguity set a greater modeling flexibility in characterizing uncertain probability distributions.

In this chapter, we investigate adaptive distributionally robust optimization problems with infinitely constrained ambiguity sets. The central question is concerned with the evaluation of the worst-case expected second stage cost over the infinitely constrained ambiguity set

$$\rho_I(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]. \tag{4.5}$$

Unfortunately, due to the possible infinitely many expectation constraints, even the static distributionally robust optimization problems with the infinitely constrained ambiguity set may not necessarily be tractable, let alone to account for adaptivity. To tackle this issue of intractability, we consider the corresponding relaxed ambiguity set that contains a finite subset of expectation constraints:

$$\mathcal{F}_R = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right]=\boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g\left(\boldsymbol{q},\tilde{\boldsymbol{z}}\right)\right]\le h(\boldsymbol{q}), \quad \forall \boldsymbol{q}\in\bar{\mathcal{Q}} \\ \mathbb{P}\left(\tilde{\boldsymbol{z}}\in\mathcal{W}\right)=1 \end{array} \right\}, \tag{4.6}$$

where $\bar{\mathcal{Q}} = \left\{\boldsymbol{q}_j : j\in[J]\right\}$ for some $\boldsymbol{q}_j\in\mathcal{Q}$, $j\in[J]$. We then have the following upper bound, $\rho_R(\boldsymbol{x})$ for the desired worst-case expected second stage cost $\rho_I(\boldsymbol{x})$:

$$\sup_{\mathbb{P}\in\mathcal{F}_I} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \rho_I(\boldsymbol{x}) \le \rho_R(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}_R} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right].$$

Lifted ambiguity set: two alternatives

Corresponding to the relaxed ambiguity set $\mathcal{F}_R$, we define the *lifted relaxed ambiguity set*, $\mathcal{G}_R$ that encompasses the primary random variable $\tilde{\boldsymbol{z}}$, and the auxiliary lifted random variable $\tilde{\boldsymbol{u}}$ as follows.

$$\mathcal{G}_R = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^J\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{u}_j\right] = h(\boldsymbol{q}_j), & \forall j \in [J] \\ \mathbb{Q}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right] = 1 & \end{array} \right\},$$

where $\bar{\mathcal{W}}$ is the *lifted support set* and is defined as the epigraph of the function $g(\cdot)$ together with the support set $\mathcal{W}$:

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^I \times \mathbb{R}^J \;\middle|\; \boldsymbol{z} \in \mathcal{W}, g\left(\boldsymbol{q}_j, \boldsymbol{z}\right) \leq u_j, \forall j \in [J] \right\}.$$

As in Chapter 3, we utilize the concept of conic representation and assume:

Assumption 2. The conic representation of the set $\left\{(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} : \boldsymbol{G}\boldsymbol{z} = \boldsymbol{\mu}\right\}$ satisfies the Slater's condition.

The above lifting technique is introduced by Wiesemann et al. (2014) in the design of a standard form of the ambiguity set. As presented in $\mathcal{G}_R$, the key features of the lifted ambiguity set include, among other things, neat expectation constraints that reside in an affine manifold. Moreover, as explored by Bertsimas et al. (2017), the inclusion of the auxiliary random variable would lead to an enhancement of the linear decision rule approximation of adaptive distributionally robust optimization problems (we will demonstrate this in details shortly). As shown in the following lifting theorem (Wiesemann et al. 2014), concerning the evaluation of the worst-case expected cost, the ambiguity sets, $\mathcal{F}_R$ and $\mathcal{G}_R$ are essentially the same.

*Theorem 7 (*Lifting Theorem*).* The ambiguity set, $\mathcal{F}_R$ is equivalent to the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$, for all $\mathbb{Q} \in \mathcal{G}_R$.

Proof. We refer the proof to Wiesemann et al. (2014). □

In the original proposal of the enhanced linear decision rule approximation, Bertsimas et al. (2017) alternatively formulate the lifted relaxed ambiguity set as below

$$\mathcal{G}_R^A = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^J\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{u}_j\right] \leq h(\boldsymbol{q}_j), & \forall j \in [J] \\ \mathbb{Q}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right] = 1 & \end{array} \right\},$$

where the expectation constraint imposed on the auxiliary variable $\tilde{\boldsymbol{u}}$ is not strictly equal. Quite surprisingly, though $\mathcal{G}_R \subseteq \mathcal{G}_R^A$, the two alternatives actually result in the same lifting theorem.

*Theorem 8 (*Lifting Theorem: an alternative*).* The ambiguity set, $\mathcal{F}_R$ is equivalent to the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$, for all $\mathbb{Q} \in \mathcal{G}_R^A$.

Proof. We refer the proof to Bertsimas et al. (2017). □

Theorem 7 and Theorem 8 reveal that the two alternatives of the lifted relaxed ambiguity set are equivalent regarding lifting theorem. In fact, we will show that the two alternatives are also equivalent regarding the enhanced linear decision rule approximation.

4.3 Linear decision rule approximation

For simplicity and clarity of the exposition, we will focus our discussion on the worst-case expected objective $\rho(\boldsymbol{x})$ in the subsequence of this chapter. We can easily incorporate the result to Problem (4.3) for obtaining the here-and-now decision $\boldsymbol{x}$.

Under the relatively complete and sufficiently expensive recourse, we can represent the second stage objective $f(\boldsymbol{x}, \boldsymbol{z})$ by leveraging the strong duality of linear programs.

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{z}) = \max \ & \boldsymbol{p}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d} \\ & \boldsymbol{p} \in \mathbb{R}^M_+ \\ = & \max_{\tau \in [P]} \left\{ \boldsymbol{p}'_\tau(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \right\}, \end{aligned} \tag{4.7}$$

where we collectively denote the extreme points of the polyhedron $\left\{\boldsymbol{p} \in \mathbb{R}^M_+ : \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d}\right\}$ by $\boldsymbol{p}_\tau, \tau \in [P]$. Consequently, the upper bound provided by considering the relaxed ambiguity set, $\rho_R(\boldsymbol{x})$ is equivalent to the following optimization problem

$$\rho_R(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_\mathbb{P} \left[\max_{\tau \in [P]} \ \left\{ \boldsymbol{p}'_\tau \left(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \right) \right\} \right], \tag{4.8}$$

which has a convex and piecewise affine objective function. Using the standard approach in distributionally robust optimization that applies duality on probability distributions (for example, Bertsimas and Popescu 2005 and Shapiro and Kleywegt 2002), Problem (4.8) can be cast as a conic optimization problem (see, e.g., Bertsimas et al. 2010a and Bertsimas et al. 2017). However, the resultant reformulation is generally intractable unless the number of extreme points is small. On the other hand, if in Problem (4.8), we consider a subset of extreme points, we would have a lower bound for $\rho_R(\boldsymbol{x})$.

Observe that we can also express Problem (4.2) as a minimization problem over a measurable decision map, $\boldsymbol{y} \in \mathcal{R}^{I,L}$ as follows:

$$\begin{aligned} \rho_R(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_\mathbb{P} \left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}}\right) \right] \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{R}^{I,L}, \end{aligned} \tag{4.9}$$

where $\mathcal{R}^{I,L}$ is the space of all measurable functions from $\mathbb{R}^I$ to $\mathbb{R}^L$. However, Problem (4.9) is computationally intractable because one is optimizing over arbitrary functions that reside in the infinite-dimensional space. Nevertheless, we can obtain an approximation from above by restricting $\boldsymbol{y}$ to a smaller class of functions. For instance, in the classical linear decision rule approximation (Garstka and Wets 1974) and Ben-Tal et al. 2004), the admissible function is restricted to one that is affinely dependent on the primary random variable $\boldsymbol{z}$, i.e., $\boldsymbol{y} \in \mathcal{L}^L$, where

$$\mathcal{L}^L = \left\{ \boldsymbol{y} \in \mathcal{R}^{I,L} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \forall i \in [I] : \\ \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{y}_{1i} z_i \end{array} \right\}.$$

Consequently, under the LDR approximation, we obtain an upper bound of $\rho_R(\boldsymbol{x})$ by solving the following problem,

$$\begin{aligned} \rho_{LDR}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P}\in\mathcal{F}_R} \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{d}' \boldsymbol{y} \left(\tilde{\boldsymbol{z}} \right) \right] \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{L}^L. \end{aligned} \tag{4.10}$$

Bertsimas et al. (2017) recently introduce an enhancement of LDR approximation, to which we refer as the extended linear decision rule (ELDR) approximation, by incorporating in the linear decision rule both $\tilde{\boldsymbol{z}}$ and the auxiliary random variable $\tilde{\boldsymbol{u}}$ that arises from the lifted ambiguity set. That is, in the ELDR approximation, we have $\boldsymbol{y} \in \bar{\mathcal{L}}^L$ such that

$$\bar{\mathcal{L}}^L = \left\{ \boldsymbol{y} \in \mathcal{R}^{I+J,L} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j}, \forall i \in [I], j \in [J] : \\ \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) = \boldsymbol{y}_0 + \sum\limits_{i\in[I]} \boldsymbol{y}_{1i} z_i + \sum\limits_{j\in[J]} \boldsymbol{y}_{2j} u_j \end{array} \right\}.$$

Under the ELDR approximation, we can work on the presentation of the lifted relaxed ambiguity set $\mathcal{G}_R$ and obtain another upper bound of $\rho_R(\boldsymbol{x})$, by solving the problem

$$\begin{aligned} \rho_{ELDR}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q}\in\mathcal{G}_R} \mathbb{E}_{\mathbb{Q}} \left[\boldsymbol{d}' \boldsymbol{y} \left(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}} \right) \right] \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^L. \end{aligned} \tag{4.11}$$

Quite notably, the ELDR approximation has been shown to lead to significant improvement over the LDR approximation.

Theorem 9. The ELDR approximation performs at least as well as the LDR approximation, i.e., $\rho_R(\boldsymbol{x}) \leq \rho_{ELDR}(\boldsymbol{x}) \leq \rho_{LDR}(\boldsymbol{x})$.

Proof. We represent a variant of the proof by Bertsimas et al. (2017) herein. Since the ELDR approximation is more flexible than the LDR approximation, it naturally follows that $\rho_{ELDR}(\boldsymbol{x}) \leq \rho_{LDR}(\boldsymbol{x})$. Let $\boldsymbol{y}^*$ be the optimal ELDR approximation in Problem (4.11). For any distribution $\mathbb{P}_* \in \mathcal{F}_R$, we construct a probability distribution $\mathbb{Q}_* \in \mathcal{G}_R$ associated with the random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathbb{R}^I \times \mathbb{R}^J$ such that $\mathbb{P}_*-$a.s.,

$$\begin{aligned}
& (\tilde{\boldsymbol{z}}, \tilde{u}_1, \ldots, \tilde{u}_J) \\
= \; & (\tilde{\boldsymbol{z}}, g(\boldsymbol{q}_1, \tilde{\boldsymbol{z}}) - \mathbb{E}_{\mathbb{P}_*}[g(\boldsymbol{q}_1, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_1), \ldots, g(\boldsymbol{q}_J, \tilde{\boldsymbol{z}}) - \mathbb{E}_{\mathbb{P}_*}[g(\boldsymbol{q}_J, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_J))
\end{aligned}$$

We consider the following linear decision rule

$$\begin{aligned}
& \boldsymbol{y}^\dagger(\boldsymbol{z}) \\
&= \boldsymbol{y}^*(\boldsymbol{z}, g(\boldsymbol{q}_1, \boldsymbol{z}) - \mathbb{E}_{\mathbb{P}_*}[g(\boldsymbol{q}_1, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_1), \ldots, g(\boldsymbol{q}_J, \boldsymbol{z}) - \mathbb{E}_{\mathbb{P}_*}[g(\boldsymbol{q}_J, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_J)) \\
&= \boldsymbol{y}^*(\boldsymbol{z}, g(\boldsymbol{q}_1, \boldsymbol{z}) - \mathbb{E}_{\mathbb{Q}_*}[g(\boldsymbol{q}_1, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_1), \ldots, g(\boldsymbol{q}_J, \boldsymbol{z}) - \mathbb{E}_{\mathbb{Q}_*}[g(\boldsymbol{q}_J, \tilde{\boldsymbol{z}})] + h(\boldsymbol{q}_J)) \\
&= \boldsymbol{y}^*(\boldsymbol{z}, u_1, \ldots, u_J),
\end{aligned}$$

which is feasible for Problem (4.9). Thus, we have $\mathbb{E}_{\mathbb{P}_*}[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})] \leq \mathbb{E}_{\mathbb{P}_*}\left[\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}})\right] = \mathbb{E}_{\mathbb{Q}_*}[\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})]$, where the inequality follows from the lifting theorem. This further implies that

$$\sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_{\mathbb{P}}[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})] \leq \mathbb{E}_{\hat{\mathbb{Q}}}[\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})]$$

for some $\hat{\mathbb{Q}} \in \mathcal{G}_R$. Therefore, we have

$$\rho_R(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_{\mathbb{P}}[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})] \leq \mathbb{E}_{\hat{\mathbb{Q}}}[\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})] \leq \sup_{\mathbb{Q} \in \mathcal{G}_R} \mathbb{E}_{\mathbb{Q}}[\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})] = \rho_{ELDR}(\boldsymbol{x}),$$

and this completes our proof. □

Instead, if we work on the alternative of the lifted relaxed ambiguity set $\mathcal{G}_R^A$, as Bertsimas et al. (2017) have done, we arrive at the same conclusion on the improvement of the ELDR approximation. Besides, Bertsimas et al. (2017) also point out that under the complete recourse assumption, the ELDR approximation can resolve the issue of infeasibility of the LDR approximation and can even obtain exact solutions for adaptive distributionally optimization problems with only one adaptive decision, i.e., $L = 1$. We summarize the advantages of the ELDR approximation in the following result.

Theorem 10. (Bertsimas et al. 2017) The ELDR approximation under $\mathcal{G}_R^A$ is given by

$$
\begin{aligned}
\rho^A_{ELDR}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q}\in\mathcal{G}_R^A} \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}}\right)\right] \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\
& \boldsymbol{y} \in \bar{\mathcal{L}}^L,
\end{aligned}
\tag{4.12}
$$

which possesses the following advantages:

(i) the ELDR approximation performs at least as well as the LDR approximation, i.e., $\rho_R(\boldsymbol{x}) \leq \rho^A_{ELDR}(\boldsymbol{x}) \leq \rho_{LDR}(\boldsymbol{x})$.

(ii) suppose Problem (4.1) has complete recourse and sufficiently expensive recourse, then for any ambiguity set, $\mathcal{F}$ such that for any $\mathbb{P} \in \mathcal{F}$, $\mathbb{E}_{\mathbb{P}}\left[|\tilde{z}_i|\right] < +\infty$, there exists a lifted ambiguity set, $\mathcal{G}$ whose corresponding ELDR approximation is feasible in Problem (4.12).

(iii) suppose Problem (4.1) has complete recourse and sufficiently expensive recourse, and has only one second stage decision variable, i.e., $L = 1$, then $\rho_R(\boldsymbol{x}) = \rho^A_{ELDR}(\boldsymbol{x})$.

Proof. We refer to the original work of these results by Bertsimas et al. (2017) for complete proofs. □

Since $\mathcal{G}_R \subseteq \mathcal{G}_R^A$, it immediately follows that $\rho_{ELDR}(\boldsymbol{x}) \leq \rho^A_{ELDR}(\boldsymbol{x})$. Notably, we next show that ELDR approximations under the two alternatives of the lifted ambiguity set are, in essence, equivalent.

Theorem 11. Suppose the second stage problem (4.1) has relative complete recourse and sufficiently expensive recourse, then $\rho_{ELDR}(\boldsymbol{x}) = \rho^A_{ELDR}(\boldsymbol{x})$. In addition, $\rho_{ELDR}(\boldsymbol{x})$ is equivalent to the following tractable conic optimization problem.

$$
\begin{array}{lll}
\inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum\limits_{j\in[J]} \gamma_j h(\boldsymbol{q}_j) & \\
\text{s.t.} & \alpha - \boldsymbol{d}'\boldsymbol{y}_0 - t_0 \geq 0 & \\
& \boldsymbol{G}'\boldsymbol{\beta} - \text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{r}_0 = \boldsymbol{0} & \\
& \boldsymbol{\gamma} - \text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{2j}\}_{j\in[J]}\right) - \boldsymbol{s}_0 = \boldsymbol{0} & \\
& \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 - t_m \geq 0, & \forall m \in [M] \\
& \text{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{r}_m = \boldsymbol{0}, & \forall m \in [M] \\
& \text{vec}\left(\left\{\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_m = \boldsymbol{0}, & \forall m \in [M] \\
& (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}, & \forall m \in [M] \cup \{0\} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{L_1}, \boldsymbol{\gamma} \in \mathbb{R}^J, \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L, & \forall i \in [I], j \in [J]
\end{array}
\tag{4.13}
$$

where $\mathcal{K}^*$ is the dual cone of

$$
\mathcal{K} = \mathbf{cl}\left\{(\boldsymbol{z}, \boldsymbol{u}, v) \in \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R} \;\middle|\; \left(\frac{\boldsymbol{z}}{v}, \frac{\boldsymbol{u}}{v}\right) \in \bar{\mathcal{W}}, v > 0\right\},
$$

$\hat{\boldsymbol{a}}_{im}, i \in [I] \cup \{0\}, m \in [M]$ is the m-th row of matrix $\boldsymbol{A}_i$, $\hat{\boldsymbol{b}}_m, m \in [M]$ is the m-th row of the recourse matrix $\hat{\boldsymbol{b}}$, and vec$(\cdot)$ is the vectorization operator such that for example, $\text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) \in \mathbb{R}^I$ is the vectorization obtained from $\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}$.

Proof. Introducing dual variables α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, we obtain the dual of $\rho_{ELDR}(\boldsymbol{x})$ as follows:

$$\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j h(\boldsymbol{q}_j) \\
\text{s.t.} \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{G}\boldsymbol{z} + \sum_{j\in[J]} \gamma_j u_j \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}), \quad \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}), \qquad\qquad \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{N_1}, \boldsymbol{\gamma} \in \mathbb{R}^J, \boldsymbol{y} \in \bar{\mathcal{L}}^L,
\end{aligned} \tag{4.14}$$

which provides an upper bound of $\rho_{ELDR}(\boldsymbol{x})$. Indeed, consider any $\mathbb{Q} \in \mathcal{G}_R$ and any feasible solution $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in the dual, the first robust counterpart in Problem (4.14) implies that

$$\mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})\right] \leq \alpha + \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{\beta}'\boldsymbol{G}\tilde{\boldsymbol{z}}\right] + \sum_{j\in[J]} \mathbb{E}_{\mathbb{P}}\left[\gamma_i \tilde{u}_j\right] \leq \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j h(\boldsymbol{q}_j).$$

Thus, weak duality follows.

Observe that each constraint in Problem (4.14) requires a classical robust counterpart to be not less than a certain threshold. For instance, we can present the first robust counterpart as

$$\begin{aligned}
&\inf\ \boldsymbol{\beta}'(\boldsymbol{G}\boldsymbol{z}) + \sum_{j\in[J]} \gamma_j u_j - \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i} z_i - \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j} u_j \\
&\text{s.t. } (\boldsymbol{z},\boldsymbol{u},1) \succeq_{\mathcal{K}} \boldsymbol{0}
\end{aligned} \quad \geq\ \boldsymbol{d}'\boldsymbol{y}_0 - \alpha,$$

where the cone $\mathcal{K}$ is a Cartesian product of tractable cones since the support set $\mathcal{W}$ and the function $g(\boldsymbol{q},\cdot)$ for any given $\boldsymbol{q} \in \mathcal{Q}$ are both tractable conic representable. The dual of the left-hand side of the above robust counterpart is given by

$$\begin{aligned}
\sup \quad & -t_0 \\
\text{s.t.} \quad & \boldsymbol{G}'\boldsymbol{\beta} - \operatorname{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{r}_0 = \boldsymbol{0} \\
& \boldsymbol{\gamma} - \operatorname{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{2j}\}_{j\in[J]}\right) - \boldsymbol{s}_0 = \boldsymbol{0} \\
& (\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) \succeq_{\mathcal{K}^*} \boldsymbol{0}.
\end{aligned}$$

Thus, we can replace the first constraint in (4.14) by a set of constraints

$$
\begin{aligned}
&\alpha - \boldsymbol{d}'\boldsymbol{y}_0 - t_0 \geq 0 \\
&\boldsymbol{G}'\boldsymbol{\beta} - \mathrm{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{r}_0 = \boldsymbol{0} \\
&\boldsymbol{\gamma} - \mathrm{vec}\left(\left\{\boldsymbol{d}'\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_0 = \boldsymbol{0} \\
&(\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) \succeq_{\mathcal{K}^*} \boldsymbol{0}.
\end{aligned}
$$

Likewise, we can replace each constraint in the second set of constraints by

$$
\begin{aligned}
&\hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 - t_m \geq 0 \\
&\mathrm{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{r}_m = \boldsymbol{0} \\
&\mathrm{vec}\left(\left\{\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_m = \boldsymbol{0} \\
&(\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}.
\end{aligned}
$$

Substituting the above dual formulations, we then have reformulation (4.13). We can apply similar analysis as in the proof of Theorem 3 in Chapter 3 to establish strong duality, which follows from the assumptions that the second stage problem (4.1) has relative complete recourse and sufficiently expensive recourse and the imposed Slater's condition.

Since $(\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}$, we have $\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j} = s_{mj} \geq 0, \forall m \in [M], j \in [J]$. Indeed, consider any $(\boldsymbol{z}, \boldsymbol{u}, v) \succeq_{\mathcal{K}} \boldsymbol{0}$ and $\boldsymbol{\delta} \in \mathbb{R}^J_+$, we have $(\boldsymbol{z}, \boldsymbol{u} + \boldsymbol{\delta}, v) \succeq_{\mathcal{K}} \boldsymbol{0}$. This implies $\boldsymbol{\delta}'\boldsymbol{s}_m \geq 0, \forall \boldsymbol{\delta} \geq \boldsymbol{0}$, and more importantly, requires $\boldsymbol{s}_m \geq \boldsymbol{0}, \forall m \in [M] \cup \{0\}$. Thus, for any $m \in [M]$ and $j \in [J]$, we have $\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j} = s_{mj} \geq 0$, and $\boldsymbol{d}'\boldsymbol{y}_{2j} \geq 0$ follows from Proposition 8. Furthermore, the constraint $\boldsymbol{\gamma} - \mathrm{vec}\left(\left\{\boldsymbol{d}'\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_0 = \boldsymbol{0}$ implies $\boldsymbol{\gamma} \geq \boldsymbol{0}$. Therefore, we can explicitly impose the constraint $\boldsymbol{\gamma} \in \mathbb{R}^J_+$ without any influence on the optimization problem (4.13). As a result, Problem (4.13) reduces to the following tractable conic optimization problem

$$
\begin{array}{lll}
\inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \displaystyle\sum_{j\in[J]} \gamma_j h(\boldsymbol{q}_j) & \\
\text{s.t.} & \alpha - \boldsymbol{d}'\boldsymbol{y}_0 - t_0 \geq 0 & \\
& \boldsymbol{G}'\boldsymbol{\beta} - \text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{r}_0 = \boldsymbol{0} & \\
& \boldsymbol{\gamma} - \text{vec}\left(\left\{\boldsymbol{d}'\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_0 = \boldsymbol{0} & \\
& \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 - t_m \geq 0, & \forall m \in [M] \\
& \text{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{r}_m = \boldsymbol{0}, & \forall m \in [M] \\
& \text{vec}\left(\left\{\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_m = \boldsymbol{0}, & \forall m \in [M] \\
& (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}, & \forall m \in [M]\cup\{0\} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{L_1}, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L, & \forall i \in [I], j \in [J]
\end{array}
$$

which is exactly the reformulation of $\rho^A_{ELDR}(\boldsymbol{x})$. In conclusion, we now can establish $\rho_{ELDR}(\boldsymbol{x}) = \rho^A_{ELDR}(\boldsymbol{x})$. □

Without any loss of generality, we will focus our discussion on $\mathcal{G}_R$ and $\rho_{ELDR}(\boldsymbol{x})$ for the rest of this chapter. We also assume the second stage problem (4.1) has relative complete recourse and sufficiently expensive recourse. Before proceeding to the next section, we show that the tractable conic optimization formulation of the ELDR approximation is representable as a class of optimization problems.

Theorem 12. Problem (4.13), the equivalent formulation of $\rho_{ELDR}(\boldsymbol{x})$, can be represented as the following tractable conic optimization problem

$$
\begin{array}{lll}
\inf & \boldsymbol{d}'\boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}\bar{\mu}_i + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}h(\boldsymbol{q}_j) & \\
\text{s.t.} & \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 - t_m \geq 0, & \forall m \in [M] \\
& \operatorname{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{r}_m = \boldsymbol{0}, & \forall m \in [M] \\
& \operatorname{vec}\left(\left\{\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_m = \boldsymbol{0}, & \forall m \in [M] \\
& (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}, & \forall m \in [M] \\
& \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L, & \forall i \in [I], j \in [J],
\end{array}
\tag{4.15}
$$

which is parameterized by some $\bar{\boldsymbol{\mu}} \in \mathcal{W}$ such that $\boldsymbol{G}\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}$.

Proof. Observe that the optimal α of Problem (4.15) takes value $\boldsymbol{d}'\boldsymbol{y}_0 + t_0$. Let $\bar{\boldsymbol{\mu}}$ be any $\bar{\boldsymbol{\mu}} \in \mathcal{W}$ such that $\boldsymbol{G}\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}$. We can plug the first three constraints into the objective and represent Problem (4.15) as follows:

$$
\begin{array}{ll}
\inf & \boldsymbol{d}'\boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}\bar{\mu}_i + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}h(\boldsymbol{q}_j) + \left(t_0 + \boldsymbol{r}_0{}'\bar{\boldsymbol{\mu}} + \sum_{j\in[J]} s_{0j}h(\boldsymbol{q}_j)\right) \\
\text{s.t.} & \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 - t_m \geq 0,\ \forall m \in [M] \\
& \operatorname{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{r}_m = \boldsymbol{0},\ \forall m \in [M] \\
& \operatorname{vec}\left(\left\{\hat{\boldsymbol{b}}'_m\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - \boldsymbol{s}_m = \boldsymbol{0},\ \forall m \in [M] \\
& (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0},\ \forall m \in [M] \cup \{0\} \\
& \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L,\ \forall i \in [I], j \in [J].
\end{array}
\tag{4.16}
$$

We can represent the fourth term in the above objective by

$$
t_0 + \boldsymbol{r}_0{}'\bar{\boldsymbol{\mu}} + \sum_{j\in[J]} s_{0j}h(\boldsymbol{q}_j) = (\boldsymbol{r}_0, \boldsymbol{s}_0, t_0)'\left(\bar{\boldsymbol{\mu}}, \operatorname{vec}\left(\left\{h(\boldsymbol{q}_j)\right\}_{j\in[J]}\right), 1\right).
$$

Recall that for any $\boldsymbol{z} \in \mathcal{W}$ such that $\boldsymbol{G}\boldsymbol{z} = \boldsymbol{\mu}$, we have assumed $g(\boldsymbol{q}, \boldsymbol{z}) \leq h(\boldsymbol{q}),\ \forall \boldsymbol{q} \in \mathcal{Q}$. Thus, we have $\left(\bar{\boldsymbol{\mu}}, \operatorname{vec}\left(\left\{h(\boldsymbol{q}_j)\right\}_{j\in[J]}\right), 1\right) \succeq_{\mathcal{K}} \boldsymbol{0}, \forall j \in [J]$. Since Problem (4.16) is a minimization problem and $(\boldsymbol{r}_0, \boldsymbol{s}_0, \boldsymbol{t}_0) \succeq_{\mathcal{K}^*} \mathbf{0}$, it follows that, at optimality of Problem (4.16), $(\boldsymbol{r}_0, \boldsymbol{s}_0, \boldsymbol{t}_0) = \mathbf{0}$. Consequently, we can conclude Problem (4.13) is equivalent to any optimization problem in the form of (4.15). □

4.4 Separation problem for tightening the relaxation

Consider two relaxed ambiguity sets, $\mathcal{F}_{R1}$ and $\mathcal{F}_{R2}$ to the infinitely constrained ambiguity set, $\mathcal{F}_I$ such that $\bar{\mathcal{Q}}_1 \subseteq \bar{\mathcal{Q}}_2$, which further implies $\mathcal{F}_I \subseteq \mathcal{F}_{R2} \subseteq \mathcal{F}_{R1}$. Combining with Theorem 9, we can have $\rho_I(\boldsymbol{x}) \leq \rho_{ELDR,2}(\boldsymbol{x}) \leq \rho_{ELDR,1}(\boldsymbol{x})$, where we obtain $\rho_{ELDR,i}(\boldsymbol{x}), i = 1, 2$ by the ELDR approximation with the corresponding lifted ambiguity set $\mathcal{G}_{Ri}$. The second inequality is attributed to the additional auxiliary random variables in $\mathcal{G}_{R2}$. Motivated by this observation, we can mitigate the conservativeness of the relaxed ambiguity set $\mathcal{F}_R$ by adding in more expectation constraints. In this section, we propose a procedure to identify effective expectation constraints to be added. These effective expectation constraints will be helpful in tightening the relaxation to the infinitely constrained ambiguity set, and ultimately, in improving the ELDR approximation to the adaptive distributionally robust optimization problems.

Observe that for a given ambiguity set $\mathcal{G}_R$, $\rho_{ELDR}(\boldsymbol{x})$ can be reformulated as Problem

(4.13), whose dual problem is given as the following conic optimization problem.

$$
\begin{aligned}
\bar{\rho}_{ELDR}(x) = \quad & \sup \quad \sum_{m\in[M]} \left\{ \phi_m(b_{0m} - \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x}) + \sum_{i\in[I]} \xi_{mi}(b_{im} - \hat{\boldsymbol{a}}'_{im}\boldsymbol{x}) \right\} \\
& \text{s.t.} \quad \phi_0 = 1 \\
& \qquad \boldsymbol{G}\boldsymbol{\xi}_0 = \boldsymbol{\mu} \\
& \qquad \zeta_{0j} = h(\boldsymbol{q}_j), \ \forall j \in [J] \\
& \qquad \phi_0 \boldsymbol{d} = \sum_{m\in[M]} \phi_m \hat{\boldsymbol{b}}_m \\
& \qquad \xi_{0i}\boldsymbol{d} = \sum_{m\in[M]} \xi_{mi}\hat{\boldsymbol{b}}_m, \ \forall i \in [I] \\
& \qquad \zeta_{0j}\boldsymbol{d} = \sum_{m\in[M]} \zeta_{mj}\hat{\boldsymbol{b}}_m, \ \forall j \in [J] \\
& \qquad (\boldsymbol{\xi}_m, \boldsymbol{\zeta}_m, \phi_m) \succeq_{\mathcal{K}} \boldsymbol{0}, \ \forall m \in [M] \cup \{0\},
\end{aligned}
\tag{4.17}
$$

where the last set of constraints is equivalent to

$$
\frac{\boldsymbol{\xi}_m}{\phi_m} \in \mathcal{W}, \ g\left(\boldsymbol{q}_j, \frac{\boldsymbol{\xi}_m}{\phi_m}\right) \leq \frac{\zeta_{mj}}{\phi_m}, \ \forall j \in [J], \forall m \in [M] \cup \{0\}.
$$

We show in the following result that $\bar{\rho}_{ELDR}(\boldsymbol{x})$ is also a valid upper bound of $\rho_R(\boldsymbol{x})$, regardless of the strong duality between Problem (4.13) and Problem (4.17).

Theorem 13. The optimal value of Problem (4.17) is also a valid upper bound of $\rho_R(\boldsymbol{x})$, i.e., $\rho_R(\boldsymbol{x}) \leq \bar{\rho}_{ELDR}(\boldsymbol{x})$.

Proof. For any realization $\boldsymbol{z} \in \mathcal{W}$, let $\boldsymbol{p}(\boldsymbol{z}) = \arg\max_{\tau\in[P]} \boldsymbol{p}'_\tau(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x})$ be the corresponding extreme point of the polyhedron $\left\{\boldsymbol{p} \in \mathbb{R}^M_+ : \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d}\right\}$ that attains the maximal objective value.

For any distribution, $\mathbb{P}_* \in \mathcal{F}_R$, we first construct a solution as follows:

$$\begin{aligned}
\phi_0 &= \int_{\boldsymbol{z}\in\mathcal{W}} \mathrm{d}\mathbb{P}_*(\boldsymbol{z}) && \\
\xi_{0i} &= \int_{\boldsymbol{z}\in\mathcal{W}} z_i \mathrm{d}\mathbb{P}_*(\boldsymbol{z}), && \forall i \in [I] \\
\zeta_{0j} &= h(\boldsymbol{q}_j), && \forall j \in [J] \\
\phi_m &= \int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z}), && \forall m \in [M] \\
\xi_{mi} &= \int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) z_i \mathrm{d}\mathbb{P}_*(\boldsymbol{z}), && \forall i \in [I], m \in [M] \\
\zeta_{mj} &= \int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \left(g(\boldsymbol{q}_j, \tilde{\boldsymbol{z}}) - \mathbb{E}_{\mathbb{P}_*}\left[g(\boldsymbol{q}_j, \tilde{\boldsymbol{z}})\right] + h(\boldsymbol{q}_j)\right) \mathrm{d}\mathbb{P}_*(\boldsymbol{z}), && \forall j \in [J], m \in [M].
\end{aligned} \tag{4.18}$$

We next check the feasibility of this solution in Problem (4.17). Note that we only check its feasibility in the set of conic constraints $(\boldsymbol{\xi}_m, \boldsymbol{\zeta}_m, \phi_m) \succeq_{\mathcal{K}} \mathbf{0}, \forall m \in [M] \cup \{0\}$, because its feasibility in other sets of constraints are relatively straightforward.

- For $m = 0$, note that $\boldsymbol{\xi}_0/\phi_0 = \boldsymbol{\xi}_0 = \mathbb{E}_{\mathbb{P}_*}[\tilde{\boldsymbol{z}}]$. Thus $\boldsymbol{\xi}_0/\phi_0 \in \mathcal{W}$ since the support set is convex. In addition, for any $j \in [J]$, we have

$$\begin{aligned}
g\left(\boldsymbol{q}_j, \frac{\boldsymbol{\xi}_0}{\phi_0}\right) &= g\left(\boldsymbol{q}_j, \int_{\boldsymbol{z}\in\mathcal{W}} \boldsymbol{z} \mathrm{d}\mathbb{P}_*(\boldsymbol{z})\right) \\
&\leq \int_{\boldsymbol{z}\in\mathcal{W}} g\left(\boldsymbol{q}_j, \boldsymbol{z}\right) \mathrm{d}\mathbb{P}_*(\boldsymbol{z}) \\
&= \mathbb{E}_{\mathbb{P}_*}\left[g\left(\boldsymbol{q}_j, \tilde{\boldsymbol{z}}\right)\right] \\
&\leq h\left(\boldsymbol{q}_j\right) = \zeta_{0j},
\end{aligned}$$

where the first inequality follows from the fact that $g(\boldsymbol{q}_j, \boldsymbol{z})$ is a convex function with respect to $\boldsymbol{z}$ for any given $\boldsymbol{q}_j$.

- For $m \in [M]$, we have

$$\frac{\boldsymbol{\xi}_m}{\phi_m} = \frac{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \boldsymbol{z} \mathrm{d}\mathbb{P}_*(\boldsymbol{z})}{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z})} = \int_{\boldsymbol{z}\in\mathcal{W}} \frac{p_m(\boldsymbol{z})}{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z})} \boldsymbol{z} \mathrm{d}\mathbb{P}_*(\boldsymbol{z}) \in \mathcal{W}.$$

In addition, for any $j \in [J]$, we have

$$
\begin{aligned}
& g\left(\boldsymbol{q}_j, \frac{\boldsymbol{\xi}_m}{\phi_m}\right) \\
= & g\left(\boldsymbol{q}_j, \int_{\boldsymbol{z}\in\mathcal{W}} \frac{p_m(\boldsymbol{z})}{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z})} \boldsymbol{z} \mathrm{d}\mathbb{P}_*(\boldsymbol{z})\right) \\
\leq & \int_{\boldsymbol{z}\in\mathcal{W}} \frac{p_m(\boldsymbol{z})}{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z})} g(\boldsymbol{q}_j, \boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z}) \\
\leq & \int_{\boldsymbol{z}\in\mathcal{W}} \frac{p_m(\boldsymbol{z})}{\int_{\boldsymbol{z}\in\mathcal{W}} p_m(\boldsymbol{z}) \mathrm{d}\mathbb{P}_*(\boldsymbol{z})} \left(g(\boldsymbol{q}_j, \boldsymbol{z}) - \mathbb{E}_{\mathbb{P}_*}\left[g(\boldsymbol{q}_j, \tilde{\boldsymbol{z}})\right] + h(\boldsymbol{q}_j)\right) \mathrm{d}\mathbb{P}_*(\boldsymbol{z}) \\
= & \frac{\zeta_{mj}}{\phi_m},
\end{aligned}
$$

where the second inequality follows from the fact that $p_m(\boldsymbol{z}) \geq 0$ and $h(\boldsymbol{q}_j) - \mathbb{E}_{\mathbb{P}_*}\left[g(\boldsymbol{q}_j, \tilde{\boldsymbol{z}})\right] \geq 0$.

Finally, we observe that

$$
\begin{aligned}
& \mathbb{E}_{\mathbb{P}_*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \\
= & \mathbb{E}_{\mathbb{P}_*}\left[\boldsymbol{p}(\tilde{\boldsymbol{z}})'(\boldsymbol{b}(\tilde{\boldsymbol{z}}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x})\right] \\
= & \mathbb{E}_{\mathbb{P}_*}\left[\sum_{m\in[M]} p_m(\tilde{\boldsymbol{z}}) \left(b_{0m} + \sum_{i\in[I]} b_{im}\tilde{z}_i - \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} - \sum_{i\in[I]} \hat{\boldsymbol{a}}_{im}'\boldsymbol{x}\tilde{z}_i\right)\right] \\
= & \mathbb{E}_{\mathbb{P}_*}\left[\sum_{m\in[M]} p_m(\tilde{\boldsymbol{z}}) \left(b_{0m} - \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x}\right) + \sum_{m\in[M]} p_m(\tilde{\boldsymbol{z}}) \sum_{i\in[I]} \left(b_{im} - \hat{\boldsymbol{a}}_{im}'\boldsymbol{x}\right)\tilde{z}_i\right] \\
= & \sum_{m\in[M]} \left\{\phi_m(b_{0m} - \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x}) + \sum_{i\in[I]} \xi_{mi}(b_{im} - \hat{\boldsymbol{a}}_{im}'\boldsymbol{x})\right\},
\end{aligned}
$$

which is identical to the objective value of Problem (4.17) of the constructive feasible solution (4.18). This implies $\mathbb{E}_{\mathbb{P}_*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \bar{\rho}_{ELDR}(\boldsymbol{x}), \forall \mathbb{P}_* \in \mathcal{F}_R$, and concludes our claim. □

As has been discussed in the proof of Theorem 11, when strong duality holds, i.e., $\bar{\rho}_{ELDR}(\boldsymbol{x}) = \rho_{ELDR}(\boldsymbol{x})$, it is already straightforward that $\bar{\rho}_{ELDR}(\boldsymbol{x})$ is a valid upper bound for $\rho(\boldsymbol{x})$. Nevertheless, Theorem 13 generally serves as the first step to derive the extremal distribution (e.g., Bertsimas et al. 2010a and Hanasusanto et al. 2015a) or the worst-case distribution (see Hanasusanto et al. 2016b) that attains the optimal value $\rho_R(\boldsymbol{x})$. In Chapter 3, we also utilize a variant of Problem (4.17) for identifying a violating distribution so as to tighten the relaxation of the relaxed ambiguity set for a static infinitely constrained distributionally robust optimization problem. For our interested case of the infinitely constrained adaptive distributionally robust optimization problem, we discuss how such an idea can be extended to two-stage (multi-stage) problems.

Observe that in Problem (4.17), ambiguity sets with different sets of expectation constraints would contribute to different variables $\boldsymbol{\zeta}_m, m \in [M]$. Inspired by this observation, given the optimal solution $(\phi_m^*, \boldsymbol{\xi}_m^*)_{m\in[M]\cup\{0\}}$ of Problem (4.17), we can identify a violating expectation constraint such that $\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})\right] > h(\boldsymbol{q})$, if for the corresponding $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$, the following system

$$\begin{cases} \zeta_0 = h(\boldsymbol{q}) \\ \zeta_0 \boldsymbol{d} = \displaystyle\sum_{m\in[M]} \zeta_m \hat{\boldsymbol{b}}_m \\ \phi_m^* g\left(\boldsymbol{q}, \dfrac{\boldsymbol{\xi}_m^*}{\phi_m^*}\right) \leq \zeta_m, \ \forall m \in [M] \cup \{0\} \end{cases}$$

is infeasible. That is to say, we can identify a violating expectation constraint if the following optimization problem is infeasible for the particular $\boldsymbol{q}$.

$$\begin{aligned} \min \quad & 0 \\ \text{s.t.} \quad & \zeta_0 = h(\boldsymbol{q}) \\ & \zeta_0 \boldsymbol{d} = \sum_{m\in[M]} \zeta_m \hat{\boldsymbol{b}}_m \\ & \zeta_m \geq \theta_m(\boldsymbol{q}), \qquad \forall m \in [M] \cup \{0\}, \end{aligned} \tag{4.19}$$

where for $m \in [M] \cup \{0\}$ and the given $(\phi_m, \boldsymbol{\xi}_m)$, we denote $\theta_m(\boldsymbol{q}) = \phi_m^* g\left(\boldsymbol{q}, \boldsymbol{\xi}_m^*/\phi_m^*\right)$.

We refer to the following dual of Problem (4.19) as the separation problem.

$$
\begin{aligned}
\max \quad & \sum_{m \in [M] \cup \{0\}} \theta_m(\boldsymbol{q})\psi_m - h(\boldsymbol{q})\tau \\
\text{s.t.} \quad & \boldsymbol{d}'\boldsymbol{\eta} + \psi_0 = \tau \\
& \hat{\boldsymbol{b}}_m'\boldsymbol{\eta} = \psi_m, && \forall m \in [M] \\
& \boldsymbol{\eta} \in \mathbb{R}^L, \tau \in \mathbb{R}, \psi_m \in \mathbb{R}_+, && \forall m \in [M] \cup \{0\}.
\end{aligned} \tag{4.20}
$$

Observe that the separation problem is always feasible, thus its objective goes to positive infinity whenever Problem (4.19) is infeasible.

Theorem 14. Let $\operatorname{recc}(\boldsymbol{B})$ be the recession cone generated by the recourse matrix $\boldsymbol{B}$, i.e., $\operatorname{recc}(\boldsymbol{B}) = \left\{\boldsymbol{\eta} \in \mathbb{R}^L : \boldsymbol{B}\boldsymbol{\eta} \geq \boldsymbol{0}\right\}$. For the particular $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$, the separation problem (4.20) is unbounded if and only if some extreme ray $\boldsymbol{\eta}$ of $\operatorname{recc}(\boldsymbol{B})$ satisfies

$$
\sum_{m \in [M]} \theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) > 0 \tag{4.21}
$$

Proof. Observe that Problem (4.20) can be represented as

$$
\begin{aligned}
\max \quad & \sum_{m \in [M]} \theta_m(\boldsymbol{q})\hat{\boldsymbol{b}}_m'\boldsymbol{\eta} + (\theta_0(\boldsymbol{q}) - h(\boldsymbol{q}))\psi_0 - h(\boldsymbol{q})\boldsymbol{d}'\boldsymbol{\eta} \\
\text{s.t.} \quad & \hat{\boldsymbol{b}}_m'\boldsymbol{\eta} \geq 0, && \forall m \in [M] \\
& \boldsymbol{\eta} \in \mathbb{R}^L, \psi_0 \in \mathbb{R}_+.
\end{aligned} \tag{4.22}
$$

We notice that $\theta_0(\boldsymbol{q}) = \phi_0^* g\left(\boldsymbol{q}, \boldsymbol{\xi}_0^*/\phi_0^*\right) = g(\boldsymbol{q}, \boldsymbol{\xi}_0^*)$, where $\boldsymbol{\xi}_0^*$ satisfies $\boldsymbol{\xi} \in \mathcal{W}$ and $\boldsymbol{G}\boldsymbol{\xi}_0^* = \boldsymbol{\mu}$. Recall our assumption, we must have $\theta_0(\boldsymbol{q}) \leq h(\boldsymbol{q}), \forall \boldsymbol{q} \in \mathcal{Q}$. Thus, the optimal solution $(\boldsymbol{\eta}^*, \psi_0^*)$ of Problem (4.22) satisfies $\psi_0^* = 0$. As a result, Problem (4.22) is equivalent to

$$
\begin{aligned}
\max \quad & \sum_{m \in [M]} \theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) \\
\text{s.t.} \quad & \hat{\boldsymbol{b}}_m'\boldsymbol{\eta} \geq 0, && \forall m \in [M] \\
& \boldsymbol{\eta} \in \mathbb{R}^L.
\end{aligned} \tag{4.23}
$$

Now we can identify a violating expectation constraint by equivalently verifying whether the separation problem (4.23) is unbounded for a particular $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$. Since the separation problem (4.23) is a linear program over the recession cone $\mathrm{recc}\,(\boldsymbol{B})$, it is unbounded if and only if for some extreme ray $\boldsymbol{\eta}$ of $\mathrm{recc}\,(\boldsymbol{B})$, it satisfies $\sum_{m\in[M]} \theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) > 0$. □

Quite surprisingly, it turns out that for any extreme ray $\boldsymbol{\eta}$ of $\mathrm{recc}(\boldsymbol{B})$, there is one distribution in the ambiguity set $\mathcal{F}_R$ that corresponds to the objective of Problem (4.21). Moreover, such a distribution can be interpreted as the worst-case distribution in the ambiguity set $\mathcal{F}_R$ because the value of $\rho_{ELDR}(\boldsymbol{x})$ is essentially the expectation of the optimal ELDR approximation with respect to this distribution. We formally summarize these interpretive results in terms of the worst-case distribution as below.

Theorem 15. Let $(\phi_m^*, \boldsymbol{\xi}_m^*, \boldsymbol{\zeta}_m^*)_{m\in[M]\cup\{0\}}$ be the optimal solutions of Problem (4.17). We consider a probability distribution $\mathbb{Q}_v \in \mathcal{G}_R$ defined as follows:

$$\mathbb{Q}_v\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) = \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*}, \frac{\boldsymbol{\zeta}_m^*}{\phi_m^*}\right)\right] = \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^*, \ \forall m \in [M],$$

and we denote by $\mathbb{P}_v$ the marginal distribution of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}_v$. We have:

(i) for any extreme ray $\boldsymbol{\eta}$ of $\mathrm{recc}(\boldsymbol{B})$, the objective of the separation problem (4.21) can be represented as $\boldsymbol{d}'\boldsymbol{\eta}\,(\mathbb{E}_{\mathbb{P}_v}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] - h(\boldsymbol{q}))$,

(ii) the expected second stage cost under $\mathbb{P}_v$ is bounded from above by $\rho_{ELDR}(x)$. That is, $\mathbb{E}_{\mathbb{P}_v}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \leq \rho_{ELDR}(\boldsymbol{x})$,

(iii) consider the equivalent reformulation of Problem (4.13), Problem (4.15) with $\bar{\boldsymbol{\mu}} = \boldsymbol{\xi}_0^*$ and let $\boldsymbol{y}^\dagger$ be the corresponding optimal ELDR approximation. The value of $\mathbb{E}_{\mathbb{P}_v}\left[\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})\right]$, is given by $\boldsymbol{d}'\boldsymbol{y}_0^\dagger + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}^\dagger \xi_{0i}^* + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}^\dagger h(\boldsymbol{q}_j)$, which is identical to $\rho_{ELDR}(\boldsymbol{x})$.

Proof. Recall that $\boldsymbol{d}$ is a non-negative linear combination of rows of $\boldsymbol{B}$. Under the sufficiently expensive recourse assumption, we only need to focus on the case where every extreme ray $\boldsymbol{\eta}$ of $\mathrm{recc}(\boldsymbol{B})$ satisfies $\boldsymbol{d}'\boldsymbol{\eta} > 0$. In fact, the assumption of sufficiently expensive recourse will be violated if $\boldsymbol{d}'\boldsymbol{\eta} < 0$ for some extreme ray $\boldsymbol{\eta}$. Besides, if $\boldsymbol{d}'\boldsymbol{\eta} = 0$, we then have $\hat{\boldsymbol{b}}_m'\boldsymbol{\eta} = 0$ for every m-th rows of $\boldsymbol{B}$, which follows from the fact that $\boldsymbol{B}\boldsymbol{\eta} \geq \boldsymbol{0}$. Such a case is trivial since the objective of the separation problem (4.23) is always zero for any $\boldsymbol{q} \in \mathcal{Q}$. We next verify $\mathbb{Q}_v \in \mathcal{G}_R$ as below.

- The support constraint:

$$\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} \in \mathcal{W}, \ \forall m \in [M].$$

- The probability constraint:

$$\sum_{m\in[M]} \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^* = \frac{\sum_{m\in[M]}\phi_m^*\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}} = \frac{\phi_0^*\boldsymbol{d}'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}} = \phi_0^* = 1.$$

- The expectation constraint:

$$\mathbb{E}_{\mathbb{P}_v}[\boldsymbol{G}\tilde{\boldsymbol{z}}] = \sum_{m\in[M]} \boldsymbol{G}\frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^*\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} = \boldsymbol{G}\sum_{m\in[M]}\frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\boldsymbol{\xi}_m^* = \boldsymbol{G}\boldsymbol{\xi}_0^* = \boldsymbol{\mu}.$$

- The relaxed finitely many expectation constraint:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_v}\left[g(\boldsymbol{q}_j,\tilde{\boldsymbol{z}})\right] &= \sum_{m\in[M]}\frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^* g\left(\boldsymbol{q}_j,\frac{\boldsymbol{\xi}_m^*}{\phi_m^*}\right) \\
&\leq \sum_{m\in[M]}\frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\zeta_{mj}^* \\
&= \zeta_{0j}^* = h(\boldsymbol{q}_j), \ \forall j \in [J].
\end{aligned}$$

For the first argument, we have

$$\begin{aligned}
\sum_{m\in[M]}\theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) &= \boldsymbol{d}'\boldsymbol{\eta}\left(\sum_{m\in[M]}\frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^* g\left(\boldsymbol{q},\frac{\boldsymbol{\xi}_m^*}{\phi_m^*}\right) - h(\boldsymbol{q})\right) \\
&= \boldsymbol{d}'\boldsymbol{\eta}\left(\mathbb{E}_{\mathbb{P}_v}[g(\boldsymbol{q},\tilde{\boldsymbol{z}})] - h(\boldsymbol{q})\right).
\end{aligned}$$

For the second argument, since $\mathbb{Q}_v \in \mathcal{G}_R$, it follows that $\mathbb{P}_v \in \mathcal{F}_R$, which implies $\mathbb{E}_{\mathbb{P}_v}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] \leq \rho_R(\boldsymbol{x}) \leq \rho_{ELDR}(\boldsymbol{x})$.

Lastly, for the third argument, we compute the value of $\mathbb{E}_{\mathbb{Q}_v}\left[\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})\right]$ as follows:

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_v}\left[\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})\right] \\
&= \sum_{m\in[M]} \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^* \left(\boldsymbol{d}'\boldsymbol{y}_0^\dagger + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}^\dagger \frac{\xi_{mi}^*}{\phi_m^*} + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}^\dagger \frac{\zeta_{mj}^*}{\phi_m^*} \right) \\
&= \boldsymbol{d}'\boldsymbol{y}_0^\dagger + \sum_{m\in[M]} \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\phi_m^* \left(\sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}^\dagger \frac{\xi_{mi}^*}{\phi_m^*} + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}^\dagger \frac{\zeta_{mj}^*}{\phi_{mi}^*} \right) \\
&= \boldsymbol{d}'\boldsymbol{y}_0^\dagger + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}^\dagger \left(\sum_{m\in[M]} \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\xi_{mi}^* \right) + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}^\dagger \left(\sum_{m\in[M]} \frac{\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}}{\boldsymbol{d}'\boldsymbol{\eta}}\zeta_{mj}^* \right) \\
&= \boldsymbol{d}'\boldsymbol{y}_0^\dagger + \sum_{i\in[I]} \boldsymbol{d}'\boldsymbol{y}_{1i}^\dagger \xi_{0i}^* + \sum_{j\in[J]} \boldsymbol{d}'\boldsymbol{y}_{2j}^\dagger h(\boldsymbol{q}_j),
\end{aligned}$$

which is identical to the objective value of Problem (4.15) with $\bar{\boldsymbol{\mu}} = \boldsymbol{\xi}_0^*$, and more importantly, is identical to $\bar{\rho}_{ELDR}(\boldsymbol{x})$. □

To conclude this section, we propose an algorithm that leverages the unboundedness of the separation problem for tightening the upper bound of the desired $\rho_I(\boldsymbol{x})$, by shrinking the relaxed ambiguity set and improving the ELDR approximation.

Algorithm 2. Adaptive Distributionally Robust Optimization

Input: An initial finite subset $\bar{\mathcal{Q}} \subseteq \mathcal{Q}$.

1. Use the ELDR approximation to solve Problem (4.3) with the relaxed ambiguity set and obtain an optimal solution $\boldsymbol{x}$.

2. Solve Problem (4.17) for the optimal solution $\boldsymbol{x}$.

3. Check the unboundedness of the separation problem (4.20). If the separation problem is unbounded for a particular $\boldsymbol{q}$, update $\bar{\mathcal{Q}} = \bar{\mathcal{Q}} \cup \{\boldsymbol{q}\}$ and Go to Step 1. If the separation problem is bounded for all $\boldsymbol{q} \in \mathcal{Q}$, then STOP.

Output: Solution $\boldsymbol{x}$.

4.5 Generalization to multistage problems

Though two-stage and multi-stage adaptive problems appear to be similar, their computational complexity differs considerably (Georghiou et al. 2015). For instance, obtaining the exact solution of linear two-stage adaptive problems is #P-hard, while multi-stage adaptive problems are believed to be *"computationally intractable already when medium-accuracy solutions are sought"* (Shapiro and Nemirovski 2005). The essentials of the multistage problems, among others, include the non-anticipativity constraints that are necessary to capture the nature of multistage decisions, in which information is revealed sequentially in stages. Fortunately, our results that iteratively improve the ELDR approximation can be extended to multistage problems.

Let us consider a $(T+1)$-stage problem. For every $t \in [T]$, in processing from stage t to stage $(t+1)$, the uncertain components, $z_i, i \in [I_t] \setminus [I_{t-1}]$ of the overall uncertainty $\boldsymbol{z}$ reveal, where $0 = I_0 < I_t < \cdots < I_T = I$. We consider the following infinitely constrained ambiguity set

$$\mathcal{F}_I = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g\left(\boldsymbol{q}, \tilde{\boldsymbol{z}}\right)\right] \leq h(\boldsymbol{q}), \quad \forall \boldsymbol{q} \in \mathcal{Q}_t, t \in [T] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}.$$

In particular, the sets $\mathcal{Q}_t$, $t \in [T]$ satisfy $\mathcal{F}_t = \left\{\boldsymbol{q} \in \mathbb{R}^{N_2} : q_v = 0, \forall v \in [N_2] \setminus [N_{2t}]\right\}$, where $\{N_{2t}\}_{t \in [T]}$ is a sequence such that $0 < N_{21} < N_{22} < \cdots < N_{2T} = N_2$. Let $\bar{\mathcal{Q}}_t$, $t \in [T]$ be the sets such that $\bar{\mathcal{Q}}_t = \left\{\boldsymbol{q}_{tj} \in \mathcal{Q}_t : j \in [J_t]\right\} \subseteq \mathcal{Q}_t$. We then have a relaxed ambiguity set as follows:

$$\mathcal{F}_R = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \left| \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[g\left(\boldsymbol{q}_{tj}, \tilde{\boldsymbol{z}}\right)\right] \leq h(\boldsymbol{q}_{tj}), \quad \forall t \in [T], j \in [J_t] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right. \right\}.$$

Correspondingly, we have the following lifted relaxed ambiguity set, $\mathcal{G}_R$ that encompasses the auxiliary lifted random variable $\boldsymbol{u} \in \mathbb{R}^J$:

$$\mathcal{G}_R = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^J\right) \left| \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{u}_{tj}\right] \leq h(\boldsymbol{q}_{tj}), & \forall t \in [T], j \in [J_t] \\ \mathbb{Q}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right] = 1 & \end{array} \right. \right\}$$

with $J = \sum_{t\in[T]} J_t$ and

$$\mathcal{W} = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^I \times \mathbb{R}^J \mid \boldsymbol{z} \in \mathcal{W}, g(\boldsymbol{q}_{tj}, \boldsymbol{z}) \leq u_{tj},\ \forall t \in [T], j \in [J_t]\right\}.$$

Given the subsets $\mathcal{I}_l \subseteq [I], l \in [L]$ that reflect the information dependency of the adaptive decisions, y_l, we consider the generalization of Problem (4.9) as follows:

$$\begin{array}{rll} \rho(\boldsymbol{x}) = & \min & \sup\limits_{\mathbb{P}\in\mathcal{F}_R} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}}\right)\right] \\ & \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & & y_l \in \mathcal{R}^I(\mathcal{I}_l), \qquad\qquad\qquad\quad \forall l \in [L], \end{array} \tag{4.24}$$

where we define the space of restricted measurable functions as

$$\mathcal{R}^I(\mathcal{I}) = \left\{ y \in \mathcal{R}^{I,1} \;\middle|\; y(\boldsymbol{v}) = y(\boldsymbol{w}), \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^I : v_k = w_k, \forall k \in \mathcal{I} \right\}.$$

Problem (4.24) solves for the optimal decision rule $\boldsymbol{y} \in \mathcal{R}^{I,L}$ that takes into account of the information dependence requirement, i.e., the non-anticipativity requirement. Note that without loss of generality, we can assume $\emptyset \neq \mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \cdots \subseteq \mathcal{I}_L = [I]$. Under the

ELDR approximation, we consider the following upper bound of $\rho(\boldsymbol{x})$:

$$\begin{aligned} \rho_{ELDR}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P} \in \mathcal{F}_R} \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{d}' \boldsymbol{y}\left(\tilde{\boldsymbol{z}}\right)\right] \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & y_l \in \bar{\mathcal{L}}(\mathcal{I}_l, \mathcal{J}_l), \quad \forall l \in [L], \end{aligned}$$

where

$$\bar{\mathcal{L}}(\mathcal{I}, \mathcal{J}) = \left\{ y \in \mathcal{R}^{I+J,1} \;\middle|\; \begin{array}{l} \exists y_0, y_{1i}, y_{2j} \in \mathbb{R}, \forall i \in \mathcal{I}, j \in \mathcal{J} : \\ y(\boldsymbol{z}, \boldsymbol{u}) = y_0 + \sum_{i \in \mathcal{I}} y_{1i} z_i + \sum_{j \in \mathcal{J}} y_{2j} u_j \end{array} \right\}$$

and the subsets $\mathcal{J}_l \subseteq [J], l \in [L]$ are consistent with the information restriction imposed by $\mathcal{I}_l \subseteq [I]$.

Theorem 16. Under the assumption of relative complete recourse and sufficiently expensive recourse, the upper bound $\rho_{ELDR}(\boldsymbol{x})$ is equivalent to the following problem.

$$\begin{aligned} \inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j \in [J]} \gamma_j h(\boldsymbol{q}_j) \\ \text{s.t.} \quad & \alpha - \boldsymbol{d}'\boldsymbol{y}_0 - t_0 \geq 0 \\ & \boldsymbol{G}'\boldsymbol{\beta} - \text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i \in [I]} \right) - \boldsymbol{r}_0 = \boldsymbol{0} \\ & \boldsymbol{\gamma} - \text{vec}\left(\left\{\boldsymbol{d}'\boldsymbol{y}_{2j}\right\}_{j \in [J]} \right) - \boldsymbol{s}_0 = \boldsymbol{0} \\ & \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}_m'\boldsymbol{y}_0 - t_m \geq 0, && \forall m \in [M] \\ & \text{vec}\left(\left\{ \hat{\boldsymbol{a}}_{im}'\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}_m'\boldsymbol{y}_{i1} \right\}_{i \in [I]} \right) - \boldsymbol{r}_m = \boldsymbol{0}, && \forall m \in [M] \\ & \text{vec}\left(\left\{ \hat{\boldsymbol{b}}_m'\boldsymbol{y}_{2j} \right\}_{j \in [J]} \right) - \boldsymbol{s}_m = \boldsymbol{0}, && \forall m \in [M] \\ & (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0}, && \forall m \in [M] \cup \{0\} \\ & y_{1il} = 0, && \forall i \in [I], l \in [L], i \notin \mathcal{I}_l \\ & y_{2jl} = 0, && \forall j \in [J], l \in [L], j \notin \mathcal{J}_l \\ & \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{N_1}, \boldsymbol{\gamma} \in \mathbb{R}^J, \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L, && \forall i \in [I], j \in [J]. \end{aligned}$$

Proof. We omit the proof as it follows from Theorem 11 straightforwardly. □

Taking dual of $\rho_{ELDR}(\boldsymbol{x})$, we obtain

$$\begin{array}{lll}
\sup & \displaystyle\sum_{m\in[M]}\left\{\phi_m(b_{0m}-\hat{\boldsymbol{a}}'_{0m}\boldsymbol{x})+\sum_{i\in[I]}\xi_{mi}(b_{im}-\hat{\boldsymbol{a}}'_{im}\boldsymbol{x})\right\} & \\
\text{s.t.} & \phi_0=1 & \\
& \boldsymbol{G}\boldsymbol{\xi}_0=\boldsymbol{\mu} & \\
& \zeta_{0j}=h(\boldsymbol{q}_j), & \forall j\in[J] \\
& \phi_0\boldsymbol{d}=\displaystyle\sum_{m\in[M]}\phi_m\hat{\boldsymbol{b}}_m & \\
& \xi_{0i}d_l=\displaystyle\sum_{m\in[M]}\xi_{mi}w_{ml}, & \forall i\in[I],l\in[L],i\in\mathcal{I}_l \\
& \zeta_{0j}d_l=\displaystyle\sum_{m\in[M]}\zeta_{mj}w_{ml}, & \forall j\in[J],l\in[L],j\in\mathcal{T}_l \\
& (\boldsymbol{\xi}_m,\boldsymbol{\zeta}_m,\phi_m)\succeq_{\mathcal{K}}\boldsymbol{0}, & \forall m\in[M]\cup\{0\}.
\end{array}$$

Given the optimal solution $(\phi_m^*,\boldsymbol{\xi}_m^*)_{m\in[M]\cup\{0\}}$, we can identify a violating expectation constraint if the following optimization problem is infeasible for the particular $\boldsymbol{q}^*\in\mathcal{Q}_t$ for some $t\in[T]$.

$$\begin{array}{lll}
\min & 0 & \\
\text{s.t.} & \zeta_0\leq h(\boldsymbol{q}^*) & \\
& \zeta_m\geq\theta_m(\boldsymbol{q}^*), & \forall m\in[M]\cup\{0\} \\
& \zeta_0d_l-\displaystyle\sum_{m\in[M]}\zeta_m w_{ml}=0, & \forall l\in[L],\kappa\left(\boldsymbol{q}^*\right)_l=1,
\end{array}\tag{4.25}$$

where $\boldsymbol{\kappa}:\mathcal{Q}\mapsto\{0,1\}^L$ is an indicator function that determines whether the adaptive decisions, y_l are dependable on the specific auxiliary random variable u^* associated with the expectation constraint induced by $\boldsymbol{q}^*$. That is, the identity $\kappa_l\left(\boldsymbol{q}^*\right)=1$ means that the adaptive decision, y_l is dependable on u^*, which is associated with the expectation constraint $\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{q}^*,\tilde{\boldsymbol{z}})\right]\leq h(\boldsymbol{q}^*)$. According to aforementioned set-ups, for any $t\in[T]$, a certain $\boldsymbol{q}^* \in \mathcal{Q}_t$ implies

$$\kappa_l(\boldsymbol{q}^*) = \begin{cases} 1, & [I_t] \subseteq \mathcal{I}_l \\ 0, & \text{otherwise.} \end{cases}$$

The dual of Problem (4.25) is

$$\begin{aligned} \max \quad & \sum_{m \in [M] \cup \{0\}} \theta_m(\boldsymbol{q}^*)\psi_m - h(\boldsymbol{q}^*)\tau \\ \text{s.t.} \quad & \sum_{l \in [L]:\ \kappa_l(\boldsymbol{f}^*)=1} d_l \eta_l + \psi_0 = \tau \\ & \sum_{l \in [L]:\ \kappa_l(\boldsymbol{f}^*)=1} w_{ml}\eta_l = \psi_m, \forall m \in [M] \\ & \tau \in \mathbb{R}_+, \psi_m \in \mathbb{R}_+, \forall m \in [M] \cup \{0\}, \eta_l \in \mathbb{R}, \forall l \in [L] : \kappa_l\left(\boldsymbol{q}^*\right) = 1, \end{aligned} \tag{4.26}$$

which is naturally feasible and whose objective goes to positive infinity whenever Problem (4.25) is infeasible.

Let L^* be the cardinality of the set $\{l \in [L] : \kappa_l(\boldsymbol{q}^*) = 1\}$, $\boldsymbol{B}^* \in \mathbb{R}^{M,L^*}$ be the sub-matrix of $\boldsymbol{B}$ whose columns correspond to those non-zero columns in $\boldsymbol{B}\text{diag}\left(\boldsymbol{\kappa}\left(\boldsymbol{q}^*\right)\right)$, and $\boldsymbol{d}^*$ be the sub-vector of $\boldsymbol{d}$ whose elements correspond to those non-zero components in $\text{diag}\left(\boldsymbol{\kappa}\left(\boldsymbol{q}^*\right)\right)\boldsymbol{d}$. Notwithstanding L^*, $\boldsymbol{B}^*$ and $\boldsymbol{d}^*$ all depend on $\boldsymbol{q}^*$, we omit the dependency herein for a light exposition. A natural corollary of Proposition 8 concludes that under the sufficiently expensive recourse condition, $\boldsymbol{d}^*$ defined above is a non-negative linear combination of rows of $\boldsymbol{B}^*$. Thus, it follows from Theorem 14 that the separation problem is in a simpler form of Problem (4.26) as below:

$$\begin{aligned} \max \quad & \sum_{m \in [M]} \theta_m(\boldsymbol{q}^*)(\hat{\boldsymbol{b}}_m^{*\,\prime}\boldsymbol{\eta}^*) - h(\boldsymbol{q}^*)(\boldsymbol{d}^{*\prime}\boldsymbol{\eta}^*) \\ \text{s.t.} \quad & \hat{\boldsymbol{b}}_m^{*\,\prime}\boldsymbol{\eta}^* \geq 0, \forall m \in [M] \\ & \boldsymbol{\eta}^* \in \mathbb{R}^{L^*}. \end{aligned}$$

Therefore, we can identify a violating expectation constraint by equivalently verifying for the particular $\boldsymbol{q}^* \in \mathcal{F}_t$, whether there is some extreme ray $\boldsymbol{\eta}^*$ of $\text{recc}\left(\boldsymbol{B}^*\right)$, the recession cone generated by $\boldsymbol{B}^*$, that satisfies

$$\sum_{m\in[M]} \theta_m(\boldsymbol{q}^*)(\hat{\boldsymbol{b}}_m^{*\,\prime}\boldsymbol{\eta}^*) - h(\boldsymbol{q}^*)(\boldsymbol{d}^{*\prime}\boldsymbol{\eta}^*) > 0.$$

4.6 Linear decision rule approximation with covariance dominance ambiguity set

In this section, we focus on the commonly encountered covariance dominance ambiguity set that captures support, mean and an upper bound on covariance of the random variable $\tilde{\boldsymbol{z}}$.

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})'\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}.$$

We assume the support set is a polytope given by

$$\mathcal{W} = \left\{ \boldsymbol{z} \in \mathbb{R}^I \mid \exists \boldsymbol{v} \in \mathbb{R}^{N_3} : \boldsymbol{D}\boldsymbol{z} + \boldsymbol{E}\boldsymbol{v} \geq \boldsymbol{q} \right\},$$

and we refer to Chapter 3 for a similar treatment for the general tractable conic representable support set. The associated lifted covariance dominance ambiguity set is given by

$$\mathcal{G}_C = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^{I\times I}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{\boldsymbol{U}}\right] \preceq \boldsymbol{\Sigma} \\ \mathbb{Q}\left[\left(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{U}}\right) \in \bar{\mathcal{W}}_C\right] = 1 \end{array} \right\},$$

where $\tilde{\boldsymbol{U}} \in \mathbb{R}^{I\times I}$ is an auxiliary lifted random variable in the lifted support set

$$\bar{\mathcal{W}}_C = \left\{ (\boldsymbol{z}, \boldsymbol{U}) \in \mathbb{R}^I \times \mathbb{R}^{I\times I} \;\mid \boldsymbol{z} \in \mathcal{W},\ \boldsymbol{U} \succeq (\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})' \right\}.$$

Under the covariance dominance ambiguity set, the worst-case second stage cost is

$$\begin{aligned}
\rho(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}}\right)\right] \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W} \\
& \boldsymbol{y} \in \mathcal{R}^{I,L}.
\end{aligned}$$

Or equivalently, the worst-case second stage cost can be given by

$$\begin{aligned}
\rho(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q}\in\mathcal{G}_C} \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}}\right)\right] \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{U}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall (\boldsymbol{z},\boldsymbol{U}) \in \bar{\mathcal{W}}_C \\
& \boldsymbol{y} \in \mathcal{R}^{I,L},
\end{aligned}$$

which considers the associated lifted covariance dominance ambiguity set, $\mathcal{G}_C$. Using the ELDR approximation, we obtain an upper bound of $\rho(\boldsymbol{x})$ as follows:

$$\begin{aligned}
\bar{\rho}_C(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q}\in\mathcal{G}_C} \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{U}}\right)\right] \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{U}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \left(\boldsymbol{z},\boldsymbol{U}\right) \in \bar{\mathcal{W}}_C \\
& \boldsymbol{y} \in \bar{\mathcal{L}}_C^L,
\end{aligned}$$

where

$$\bar{\mathcal{L}}_C^L = \left\{ \boldsymbol{y} \in \mathcal{R}^{I+I\times I,L} \;\middle|\; \begin{array}{l} \forall l \in [L], \exists y_{0l}, y_{1il}, i \in [I], \boldsymbol{Y}_l \in \mathcal{S}^I : \\ y_l(\boldsymbol{z},\boldsymbol{U}) = y_{0l} + \sum_{i\in[I]} y_{1il} z_i + \langle \boldsymbol{Y}_l, \boldsymbol{U} \rangle \end{array} \right\}.$$

On the other hand, we can also cast the covariance dominance ambiguity set as the following infinitely constrained ambiguity set

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \mathbb{R}^I, \|\boldsymbol{q}\|_2 \leq 1 \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\}. \tag{4.27}$$

The corresponding relaxed covariance dominance ambiguity set is given by

$$\mathcal{F}_R = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},$$

where $\bar{\mathcal{Q}} = \left\{\boldsymbol{q}_j : j \in [J]\right\}$ for some $\boldsymbol{q}_j \in \mathbb{R}^I, \|\boldsymbol{q}_j\|_2 \leq 1, j \in [J]$, and the lifted relaxed covariance dominance ambiguity set, $\mathcal{G}_R$ is given by

$$\mathcal{G}_R = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^J\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} & \\ \mathbb{Q}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}_R\right] = 1 & \\ \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}\left[\tilde{u}_j\right] \leq \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j, & \forall j \in [J] \end{array} \right\}$$

with a lifted support set

$$\bar{\mathcal{W}}_R = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^I \times \mathbb{R}^J \;\middle|\; \boldsymbol{z} \in \mathcal{W},\ u_j \geq (\boldsymbol{q}_j'(\boldsymbol{z} - \boldsymbol{\mu}))^2,\ \forall j \in [J] \right\}.$$

With the lifted relaxed covariance dominance ambiguity set and the ELDR approximation, we can obtain another upper bound of $\rho(\boldsymbol{x})$, shown as below:

$$\begin{array}{rll} \bar{\rho}_R(\boldsymbol{x}) = & \min & \sup\limits_{\mathbb{Q} \in \mathcal{G}_R} \mathbb{E}_{\mathbb{Q}}\left[\boldsymbol{d}'\boldsymbol{y}\left(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}\right)\right] \\ & \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall\, (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}_R \\ & & \boldsymbol{y} \in \bar{\mathcal{L}}^L. \end{array}$$

The relations between different evaluations of the second stage cost generally hold as $\rho(\boldsymbol{x}) \leq \bar{\rho}_C(\boldsymbol{x})$ and $\rho(\boldsymbol{x}) \leq \bar{\rho}_R(\boldsymbol{x})$, where the former follows from the ELDR approximation, while the latter follows from both $\mathcal{F}_C \subseteq \mathcal{F}_R$ and the ELDR approximation. We next show that in general, we have $\bar{\rho}_C(\boldsymbol{x}) \leq \bar{\rho}_R(\boldsymbol{x})$.

Theorem 17. The ELDR approximation with the lifted covariance dominance ambiguity set should perform not worse than that with the relaxed counterpart, i.e., $\bar{\rho}_C(\boldsymbol{x}) \leq \bar{\rho}_R(\boldsymbol{x})$.

Proof. Applying duality on probability distributions, $\bar{\rho}_C(\boldsymbol{x})$ is equivalent to the following positive semidefinite conic program:

$$
\begin{aligned}
\inf \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}_0, \boldsymbol{\Sigma} \rangle \\
\text{s.t.} \quad & \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \langle \boldsymbol{\Gamma}_0, \boldsymbol{U} \rangle \geq \sum_{l \in [L]} d_l \left(y_{0l} + \sum_{i \in [I]} y_{1il} z_i + \langle \boldsymbol{Y}_l, \boldsymbol{U} \rangle \right), \forall (\boldsymbol{z}, \boldsymbol{U}) \in \bar{\mathcal{W}}_C \\
& \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} + \sum_{i \in [I]} \hat{\boldsymbol{a}}_{im}'\boldsymbol{x} z_i + \sum_{l \in [L]} w_{ml} \left(y_{0l} + \sum_{i \in [I]} y_{1il} z_i + \langle \boldsymbol{Y}_l, \boldsymbol{U} \rangle \right) \\
& \qquad \geq b_{0m} + \sum_{i \in [I]} b_{im} z_i, \forall (\boldsymbol{z}, \boldsymbol{U}) \in \bar{\mathcal{W}}_C, \forall m \in [M] \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{N_1}, \boldsymbol{\Gamma}_0 \in \mathcal{S}_+^I, \boldsymbol{y}_0 \in \mathbb{R}^L, \boldsymbol{y}_{i1} \in \mathbb{R}^L, \forall i \in [I], \boldsymbol{Y}_l \in \mathcal{S}^I, \forall l \in [L],
\end{aligned}
\tag{4.28}
$$

where $\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}_0$ are the dual variables corresponding to the constraints in $\mathcal{G}_C$ and $\boldsymbol{y}_0, \boldsymbol{y}_{i1}, \boldsymbol{Y}_l$ account for the ELDR approximation. The first robust counterpart can be represented as

$$
\begin{aligned}
\inf \ & \boldsymbol{\beta}'\boldsymbol{z} + \langle \boldsymbol{\Gamma}_0, \boldsymbol{U} \rangle - \sum_{i \in [I]} \boldsymbol{d}'\boldsymbol{y}_{i1} z_i - \sum_{l \in [L]} d_l \langle \boldsymbol{Y}_l, \boldsymbol{U} \rangle \\
\text{s.t.} \ & \boldsymbol{D}\boldsymbol{z} + \boldsymbol{E}\boldsymbol{v} \geq \boldsymbol{q} \\
& \begin{pmatrix} 1 & (\boldsymbol{z} - \boldsymbol{\mu})' \\ (\boldsymbol{z} - \boldsymbol{\mu}) & \boldsymbol{U} \end{pmatrix} \in \mathcal{S}_+^{I+1}
\end{aligned}
\quad \geq \ \boldsymbol{d}'\boldsymbol{y}_0 - \alpha.
$$

Taking dual of the left-hand side of the robust counterpart, we can replace the first constraint by a set of constraints as follows:

$$
\begin{aligned}
& \alpha - \boldsymbol{d}'\boldsymbol{y}_0 + \boldsymbol{\pi}_0'\boldsymbol{q} - \delta_0 + 2\boldsymbol{\varphi}_0'\boldsymbol{\mu} \geq 0 \\
& \boldsymbol{\beta} - \operatorname{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i \in [I]} \right) - \boldsymbol{D}'\boldsymbol{\pi}_0 - 2\boldsymbol{\varphi}_0 = \boldsymbol{0} \\
& \boldsymbol{\Gamma}_0 - \sum_{l \in [L]} d_l \boldsymbol{Y}_l - \boldsymbol{\Phi}_0 = \boldsymbol{0} \\
& \boldsymbol{E}'\boldsymbol{\pi}_0 = \boldsymbol{0}, \ \boldsymbol{\pi}_0 \geq \boldsymbol{0}, \ \begin{pmatrix} \delta_0 & \boldsymbol{\varphi}_0' \\ \boldsymbol{\varphi}_0 & \boldsymbol{\Phi}_0 \end{pmatrix} \in \mathcal{S}_+^{I+1}.
\end{aligned}
$$

Likewise, we can replace each m-th ($m \in [M]$) robust counterpart in the second set of constraints by

$$
\begin{aligned}
&\hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 + \boldsymbol{\pi}'_m\boldsymbol{q} - \delta_m + 2\boldsymbol{\varphi}'_m\boldsymbol{\mu} \geq 0\\
&\text{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{D}'\boldsymbol{\pi}_m - 2\boldsymbol{\varphi}_m = \boldsymbol{0}\\
&\sum_{l\in[L]} w_{ml}\boldsymbol{Y}_l - \boldsymbol{\Phi}_m = \boldsymbol{0}\\
&\boldsymbol{E}'\boldsymbol{\pi}^m = \boldsymbol{0},\ \boldsymbol{\pi}^m \geq \boldsymbol{0},\ \begin{pmatrix} \delta_m & \boldsymbol{\varphi}'_m \\ \boldsymbol{\varphi}_m & \boldsymbol{\Phi}_m \end{pmatrix} \in \mathcal{S}_+^{I+1}.
\end{aligned}
$$

Thus, the complete explicit positive semidefinite conic reformulation of $\bar{\rho}_C(\boldsymbol{x})$ is:

$$
\begin{aligned}
\inf\quad & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}_0, \boldsymbol{\Sigma}\rangle \\
\text{s.t.}\quad & \alpha - \boldsymbol{d}'\boldsymbol{y}_0 + \boldsymbol{\pi}'_0\boldsymbol{q} - \delta_0 + 2\boldsymbol{\varphi}'_0\boldsymbol{\mu} \geq 0 \\
& \boldsymbol{\beta} - \text{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{D}'\boldsymbol{\pi}_0 - 2\boldsymbol{\varphi}_0 = \boldsymbol{0} \\
& \boldsymbol{\Gamma}_0 - \sum_{l\in[L]} d_l\boldsymbol{\Gamma}_l - \boldsymbol{\Phi}_0 = \boldsymbol{0} \\
& \hat{\boldsymbol{a}}'_{0m}\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_0 + \boldsymbol{\pi}'_m\boldsymbol{q} - \delta_m + 2\boldsymbol{\varphi}'_m\boldsymbol{\mu} \geq 0, && \forall m \in [M] \\
& \text{vec}\left(\left\{\hat{\boldsymbol{a}}'_{im}\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}'_m\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{D}'\boldsymbol{\pi}_m - 2\boldsymbol{\varphi}_m = \boldsymbol{0}, && \forall m \in [M] \\
& \sum_{l\in[L]} w_{ml}\boldsymbol{\Gamma}_l - \boldsymbol{\Phi}_m = \boldsymbol{0}, && \forall m \in [M] \\
& \boldsymbol{E}'\boldsymbol{\pi}_m = \boldsymbol{0},\ \boldsymbol{\pi}_m \geq 0,\ \begin{pmatrix} \delta^m & \boldsymbol{\varphi}'_m \\ \boldsymbol{\varphi}_m & \boldsymbol{\Phi}_m \end{pmatrix} \in \mathcal{S}_+^{I+1}, && \forall m \in [M] \cup \{0\} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{N_1}, \boldsymbol{\Gamma}_0 \in \mathcal{S}_+^I, \boldsymbol{y}_0, \boldsymbol{y}_{1i} \in \mathbb{R}^L, \boldsymbol{Y}_l \in \mathcal{S}^I, && \forall i \in [I], l \in [L].
\end{aligned}
\tag{4.29}
$$

By Theorem 11, $\bar{\rho}_R(\boldsymbol{x})$ is equivalent to

$$
\begin{aligned}
\inf\ & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j \\
\text{s.t.}\ & \alpha - \boldsymbol{d}'\boldsymbol{y}_0 + \boldsymbol{\pi}_0'\boldsymbol{q} + \sum_{j\in[J]} (s_{0j} - r_{0j} + 2t_{0j}\boldsymbol{q}_j'\boldsymbol{\mu}) \geq 0 \\
& \boldsymbol{\beta} - \mathrm{vec}\left(\{\boldsymbol{d}'\boldsymbol{y}_{1i}\}_{i\in[I]}\right) - \boldsymbol{D}'\boldsymbol{\pi}_0 - \sum_{j\in[J]} 2t_{0j}\boldsymbol{q}_j = \boldsymbol{0} \\
& \boldsymbol{\gamma} - \mathrm{vec}\left(\left\{\boldsymbol{d}'\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - (\boldsymbol{r}_0 + \boldsymbol{s}_0) = \boldsymbol{0} \\
& \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} - b_{0m} + \hat{\boldsymbol{b}}_m'\boldsymbol{y}_0 + \boldsymbol{\pi}_m'\boldsymbol{q} + \sum_{j\in[J]} (s_{mj} - r_{mj} + 2t_{mj}\boldsymbol{q}_j'\boldsymbol{\mu}) \geq 0,\ \forall m \in [M] \\
& \mathrm{vec}\left(\left\{\hat{\boldsymbol{a}}_{im}'\boldsymbol{x} - b_{im} + \hat{\boldsymbol{b}}_m'\boldsymbol{y}_{1i}\right\}_{i\in[I]}\right) - \boldsymbol{D}'\boldsymbol{\pi}_m - \sum_{j\in[J]} 2t_{mj}\boldsymbol{q}_j = \boldsymbol{0},\ \forall m \in [M] \\
& \mathrm{vec}\left(\left\{\hat{\boldsymbol{b}}_m'\boldsymbol{y}_{2j}\right\}_{j\in[J]}\right) - (\boldsymbol{r}_m + \boldsymbol{s}_m) = \boldsymbol{0},\ \forall m \in [M] \\
& \boldsymbol{E}'\boldsymbol{\pi}_m = \boldsymbol{0},\ \boldsymbol{\pi}_m \geq \boldsymbol{0},\ \forall m \in [M]\cup\{0\} \\
& (r_{mj}, s_{mj}, t_{mj}) \in \mathcal{K}_{soc},\ \forall j \in [J], m \in [M]\cup\{0\} \\
& \alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{L_1}, \boldsymbol{\gamma} \in \mathbb{R}^J_+, \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L,\ \forall i \in [I], j \in [J].
\end{aligned}
\tag{4.30}
$$

Let $\left(\alpha^\dagger, \boldsymbol{\beta}^\dagger, \boldsymbol{\gamma}^\dagger, \boldsymbol{y}_0^\dagger, \boldsymbol{y}_{1i}^\dagger, \boldsymbol{y}_{2j}^\dagger, \boldsymbol{\pi}_m^\dagger, \boldsymbol{r}_m^\dagger, \boldsymbol{s}_m^\dagger, \boldsymbol{t}_m^\dagger\right)$ be any feasible solution for Problem (4.30). We next show that there is a corresponding feasible solution for Problem (4.29), which we construct as below:

$$
\begin{array}{rcll}
\alpha &=& \alpha^\dagger & \\
\boldsymbol{\beta} &=& \boldsymbol{\beta}^\dagger & \\
\boldsymbol{\Gamma}_0 &=& \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j \boldsymbol{q}_j' & \\
\boldsymbol{y}_0 &=& \boldsymbol{y}_0^\dagger & \\
\boldsymbol{y}_{1i} &=& \boldsymbol{y}_{1i}^\dagger, & \forall i \in [I] \\
\boldsymbol{\pi}_m &=& \boldsymbol{\pi}_m^\dagger, & \forall m \in [M]\cup\{0\} \\
\boldsymbol{Y}_l &=& \sum_{j\in[J]} y_{2jl}^\dagger \boldsymbol{q}_j \boldsymbol{q}_j', & \forall l \in [L] \\
\delta_m &=& \sum_{j\in[J]} (r_{mj}^\dagger - s_{mj}^\dagger), & \forall m \in [M]\cup\{0\} \\
\boldsymbol{\varphi}_m &=& \sum_{j\in[J]} t_{mj}^\dagger \boldsymbol{q}_j, & \forall m \in [M]\cup\{0\} \\
\boldsymbol{\Phi}_m &=& \sum_{j\in[J]} (r_{mj}^\dagger + s_{mj}^\dagger)\boldsymbol{q}_j \boldsymbol{q}_j', & \forall m \in [M]\cup\{0\}.
\end{array}
\tag{4.31}
$$

Observe that

$$
\langle \boldsymbol{\Gamma}_0, \boldsymbol{\Sigma}\rangle = \left\langle \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j \boldsymbol{q}_j', \boldsymbol{\Sigma} \right\rangle = \sum_{j\in[J]} \langle \gamma_j \boldsymbol{q}_j \boldsymbol{q}_j', \boldsymbol{\Sigma}\rangle = \sum_{j\in[J]} \gamma_j \boldsymbol{q}_j' \boldsymbol{\Sigma} \boldsymbol{q}_j,
$$

which implies the feasible solution (4.31) yields the same objective as that of the solution $\left(\alpha^\dagger, \boldsymbol{\beta}^\dagger, \boldsymbol{\gamma}^\dagger, \boldsymbol{y}_0^\dagger, \boldsymbol{y}_{1i}^\dagger, \boldsymbol{y}_{2j}^\dagger, \boldsymbol{\pi}_m^\dagger, \boldsymbol{r}_m^\dagger, \boldsymbol{s}_m^\dagger, \boldsymbol{t}_m^\dagger\right)$ for Problem (4.30). Thus it is clear that $\bar{\rho}_C(\boldsymbol{x}) \leq \bar{\rho}_R(\boldsymbol{x})$. □

Theorem 17 shows that with the covariance dominance ambiguity set, the ELDR approximation, in general, performs not worse off than that with the relaxed ambiguity set. However, the resultant optimization problem is a positive semidefinite conic program. In contrast, under the relaxed covariance dominance ambiguity set, it is a second order conic program, which attractively possesses the computational tractability and is well supported by the state-of-the-art commercial solvers such as CPLEX, Gurobi, and Mosek.

Quite surprisingly, the concept of finite reduction we have presented in Chapter 3 is still valid. That is, for some relaxed covariance dominance ambiguity set, we have $\bar{\rho}_C(\boldsymbol{x}) = \bar{\rho}_R(\boldsymbol{x})$.

Theorem 18. There exists one relaxed covariance dominance ambiguity set such that $\bar{\rho}_C(\boldsymbol{x}) = \bar{\rho}_R(\boldsymbol{x})$.

Proof. In Theorem 17, we note that $\bar{\rho}_C(\boldsymbol{x})$ is equivalent to the implicit positive semidefinite conic optimization problem (4.28). Let $(\alpha^*, \boldsymbol{\beta}^*, \boldsymbol{y}_0^*, \boldsymbol{y}_i^*, \boldsymbol{\Gamma}^*, \boldsymbol{Y}_l^*)$ be the optimal solution of Problem (4.28). By the eigenvalue decomposition of a symmetric matrix, we can represented $\boldsymbol{\Gamma}^*$ and $\boldsymbol{Y}_l^*$ as

$$\boldsymbol{\Gamma}_0^* = \sum_{i \in [I]} \chi_{0i}^* \boldsymbol{\lambda}_{0i}^* \boldsymbol{\lambda}_{0i}^{*\prime}, \quad \boldsymbol{Y}_l^* = \sum_{i \in [I]} \chi_{li}^* \boldsymbol{\lambda}_{li}^* \boldsymbol{\lambda}_{li}^{*\prime}, \ \forall l \in [L],$$

where χ_l^*, $\boldsymbol{\lambda}_l^*$ are corresponding eigenvalues and eigenvectors. Note that $\boldsymbol{\chi}_0^* \geq \boldsymbol{0}$ since $\boldsymbol{\Gamma}_0$ is positive semidefinite. We collectively denote all these eigenvectors in a set Λ, that is, $\Lambda = \{\boldsymbol{\lambda}_{li} : l \in [L] \cup \{0\}, i \in [I]\}$. The first constraint in Problem (4.28) can be represented as

$$\alpha^* + \boldsymbol{\beta}^{*\prime}\boldsymbol{z} + \sum_{i \in [I]} \chi_{0i}^* \boldsymbol{\lambda}_{0i}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{0i}^* \geq \sum_{l \in [L]} d_l \left(y_{0l}^* + \sum_{i \in [I]} y_{1il}^* z_i + \sum_{i \in [I]} \chi_{li}^* \boldsymbol{\lambda}_{li}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{li}^* \right)$$

for all $(\boldsymbol{z}, \boldsymbol{U}) \in \bar{\mathcal{W}}_C$, which implies

$$\alpha^* + \boldsymbol{\beta}^{*\prime}\boldsymbol{z} + \sum_{i \in [I]} \chi_{0i}^* \boldsymbol{\lambda}_{0i}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{0i}^* \geq \sum_{l \in [L]} d_l \left(y_{0l}^* + \sum_{i \in [I]} y_{1il}^* z_i + \sum_{i \in [I]} \chi_{li}^* \boldsymbol{\lambda}_{li}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{li}^* \right)$$

for all $\boldsymbol{z} \in \mathcal{W}$, $\boldsymbol{q} \in \Lambda$, and $\boldsymbol{q}'\boldsymbol{U}\boldsymbol{q} \geq (\boldsymbol{q}'(\boldsymbol{z} - \boldsymbol{\mu}))^2$. Likewise, each constraint in the second set of constraints in Problem (4.28) implies that for all $\boldsymbol{z} \in \mathcal{W}$, $\boldsymbol{q} \in \Lambda$, and

$\boldsymbol{q}'\boldsymbol{U}\boldsymbol{q} \geq (\boldsymbol{q}'(\boldsymbol{z}-\boldsymbol{\mu}))^2$, the term

$$\hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} + \sum_{i\in[I]} \hat{\boldsymbol{a}}_{im}'\boldsymbol{x}z_i + \sum_{l\in[L]} w_{ml} \left(y_{0l} + \sum_{i\in[I]} y_{1il}z_i + \sum_{i\in[I]} \chi_{li}^* \boldsymbol{\lambda}_{li}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{li}^* \right)$$

is not less than $b_{0m} + \sum_{i\in[I]} b_{im}z_i$.

We next consider $\bar{\rho}_R(\boldsymbol{x})$ with one particular relaxed covariance dominance ambiguity set such that $\bar{\mathcal{Q}} = \Lambda$. We can formulate $\bar{\rho}_R(\boldsymbol{x})$ as

$$\begin{array}{rl}
\inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \displaystyle\sum_{v\in[\hat{L}]}\sum_{i\in[I]} \gamma_{vi} \boldsymbol{\lambda}_{vi}^{*\prime} \boldsymbol{\Sigma} \boldsymbol{\lambda}_{vi}^* \\
\text{s.t.} & \begin{array}{l} \alpha + \boldsymbol{\beta}'\boldsymbol{z} + \displaystyle\sum_{v\in[\hat{L}]}\sum_{i\in[I]} \gamma_{vi} u_{vi} \\
\geq \displaystyle\sum_{l\in[L]} d_l \left(y_{0l} + \sum_{i\in[I]} y_{1il} z_i + \sum_{v\in[\hat{L}]}\sum_{i\in[I]} y_{2l}^{vi} u_{vi} \right), \ \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}}_R \\
\hat{\boldsymbol{a}}_{0m}'\boldsymbol{x} + \displaystyle\sum_{i\in[I]} \hat{\boldsymbol{a}}_{im}'\boldsymbol{x}z_i + \sum_{l\in[L]} w_{ml} \left(y_{0l} + \sum_{i\in[I]} y_{1il} z_i + \sum_{v\in[\hat{L}]}\sum_{i\in[I]} y_{2l}^{vi} u_{vi} \right) \\
\geq b_{0m} + \displaystyle\sum_{i\in[I]} b_{im} z_i, \ \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}}_R, m \in [M] \\
\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^{N_1}, \boldsymbol{\gamma} \in \mathbb{R}^J \\
\boldsymbol{y}_0 \in \mathbb{R}^L, \boldsymbol{y}_{1i} \in \mathbb{R}^L, \forall i \in [I], \boldsymbol{y}_2^{vi} \in \mathbb{R}^L, \forall i \in [I], v \in [\hat{L}],
\end{array}
\end{array} \tag{4.32}$$

where $[\hat{L}] = [L] \cup \{0\}$.

We now show that for the optimal solution of Problem (4.28), there correspondingly exists the following feasible solution for Problem (4.32) with the same objective.

$$
\begin{aligned}
\alpha &= \alpha^* \\
\boldsymbol{\beta} &= \boldsymbol{\beta}^* \\
\boldsymbol{y}_0 &= \boldsymbol{y}_0^* \\
\boldsymbol{y}_{1i} &= \boldsymbol{y}_{1i}^*, && \forall i \in [I] \\
\boldsymbol{y}_{2l}^{0i} &= 0, && \forall i \in [I], l \in [L] \\
y_{2l}^{vi} &= \begin{cases} \chi_{li}^*, & v = l \\ 0, & v \neq l \end{cases} && \forall l \in [L] \\
\gamma_{vi} &= \begin{cases} \chi_{0i}^*, & v = 0 \\ 0, & v \neq 0 \end{cases} && \forall i \in [I].
\end{aligned}
$$

Observe that the objective of Problem (4.30) becomes

$$
\begin{aligned}
\alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \sum_{v \in [\hat{L}]} \sum_{i \in [I]} \gamma_{vi} \boldsymbol{\lambda}_{vi}^{*\prime} \boldsymbol{\Sigma} \boldsymbol{\lambda}_{vi}^* &= \alpha^* + \boldsymbol{\beta}^{*\prime}\boldsymbol{\mu} + \sum_{i \in [I]} \chi_{0i}^* \boldsymbol{\lambda}_{0i}^{*\prime} \boldsymbol{\Sigma} \boldsymbol{\lambda}_{0i}^* \\
&= \alpha^* + \boldsymbol{\beta}^{*\prime}\boldsymbol{\mu} + \langle \boldsymbol{\Gamma}^*, \boldsymbol{\Sigma} \rangle ,
\end{aligned}
$$

which coincides with the optimal objective of Problem (4.28). We verify the feasibility of the first constraint in Problem (4.32). The left-hand side of the first constraint becomes

$$
\alpha + \boldsymbol{\beta}'\boldsymbol{z} + \sum_{v \in [\hat{L}]} \sum_{i \in [I]} \gamma_{vi} u_{vi} = \alpha^* + \boldsymbol{\beta}^{*\prime}\boldsymbol{z} + \sum_{i \in [I]} \chi_{0i}^* u_{0i},
$$

while the right-hand side becomes

$$
\sum_{l \in [L]} d_l \left(y_{0l} + \sum_{i \in [I]} y_{1il} z_i + \sum_{v \in [\hat{L}]} \sum_{i \in [I]} y_{2l}^{vi} u_{li} \right) = \sum_{l \in [L]} d_l \left(y_{0l}^* + \sum_{i \in [I]} y_{1il}^* z_i + \sum_{i \in [I]} \chi_{li}^* u_{li} \right).
$$

If for all $v \in [\hat{L}], i \in [I]$, we replace u_{vi} by $\boldsymbol{\lambda}_{vi}^{*\prime} \boldsymbol{U} \boldsymbol{\lambda}_{vi}^*$, we then see the first constraint is feasible. Similarly, we can verify the feasibility of the second set of constraints in

Problem (4.32). Therefore, for this particular relaxed dominance ambiguity set, we have $\bar{\rho}_R(\boldsymbol{x}) \leq \bar{\rho}_C(\boldsymbol{x})$. With Theorem 17, we can conclude that $\bar{\rho}_C(\boldsymbol{x}) = \bar{\rho}_R(\boldsymbol{x})$ for this particular relaxed dominance ambiguity set. □

In the case of relaxed covariance dominance ambiguity set $\mathcal{G}_R$, the dual of $\bar{\rho}_R(\boldsymbol{x})$ is given by the following second order conic optimization problem.

$$
\begin{array}{lll}
\sup & \displaystyle\sum_{m\in[M]} \left\{ \phi_m(b_{0m} - \hat{\boldsymbol{a}}_{0m}'\boldsymbol{x}) + \sum_{i\in[I]} \xi_{mi}(b_{im} - \hat{\boldsymbol{a}}_{im}'\boldsymbol{x}) \right\} & \\
\text{s.t.} & \phi_0 = 1 & \\
& \boldsymbol{\xi}_0 = \boldsymbol{\mu} & \\
& \zeta_{0j} \leq \boldsymbol{q}_j'\boldsymbol{\Sigma}\boldsymbol{q}_j, & \forall j \in [J] \\
& \phi_0 \boldsymbol{d} = \displaystyle\sum_{m\in[M]} \phi_m \hat{\boldsymbol{b}}_m & \\
& \xi_{0i}\boldsymbol{d} = \displaystyle\sum_{m\in[M]} \xi_{mi}\hat{\boldsymbol{b}}_m, & \forall i \in [I] \\
& \zeta_{0j}\boldsymbol{d} = \displaystyle\sum_{m\in[M]} \zeta_{mj}\hat{\boldsymbol{b}}_m, & \forall j \in [J] \\
& \boldsymbol{D}\boldsymbol{\xi}_m + \boldsymbol{E}\boldsymbol{\epsilon}_m \geq \phi_m\boldsymbol{q}, & \forall m \in [M]\cup\{0\} \\
& \big(\zeta_{mj} + \phi_m, \zeta_{mj} - \phi_m, 2\boldsymbol{q}_j'(\boldsymbol{\xi}_m - \phi_m\boldsymbol{\mu})\big) \in \mathcal{K}_{soc}, & \forall j \in [J], m \in [M]\cup\{0\},
\end{array}
$$

whose optimal solution we denote by $(\phi_m^*, \boldsymbol{\xi}_m^*, \boldsymbol{\zeta}_m^*)_{m\in[M]\cup\{0\}}$. By Theorem 14, for any extreme ray $\boldsymbol{\eta}$ of $\text{recc}(\boldsymbol{B})$, we would identify a violating expectation constraint with the particular $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$ if it satisfies that

$$
\begin{aligned}
& \sum_{m\in[M]} \theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m'\boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) \\
= & \sum_{m\in[M]} \hat{\boldsymbol{b}}_m'\boldsymbol{\eta}\phi_m^* \left(\boldsymbol{q}' \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} - \boldsymbol{\mu} \right) \right)^2 - \boldsymbol{d}'\boldsymbol{\eta}\boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q} \\
= & \, \boldsymbol{q}' \left(\sum_{m\in[M]} \phi_m^* \hat{\boldsymbol{b}}_m'\boldsymbol{\eta} \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} - \boldsymbol{\mu} \right) \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} - \boldsymbol{\mu} \right)' - \boldsymbol{d}'\boldsymbol{\eta}\boldsymbol{\Sigma} \right) \boldsymbol{q} \\
\geq & \, 0.
\end{aligned}
$$

Or equivalently, we can study the following classical minimal eigenvalue problem

$$\min_{\boldsymbol{q}:\ \|\boldsymbol{q}\|_2 \le 1} \boldsymbol{q}' \left(\boldsymbol{d}'\boldsymbol{\eta}\boldsymbol{\Sigma} - \sum_{m \in [M]} \phi_m^* \hat{\boldsymbol{b}}_m' \boldsymbol{\eta} \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} - \boldsymbol{\mu} \right) \left(\frac{\boldsymbol{\xi}_m^*}{\phi_m^*} - \boldsymbol{\mu} \right)' \right) \boldsymbol{q},$$

which can be solved efficiently using numerical techniques. If the minimal eigenvalue is negative, we then add the corresponding eigenvector into $\bar{\mathcal{Q}}$.

4.7 Numerical examples

4.7.1 Distributionally robust multi-item newsvendor problem

In the first numerical example, we study the multi-item newsvendor problem, one of the most fundamental inventory problems, in which the newsvendor sells different perishable items under uncertain demand (Hadley and Whitin 1963). For each item i, $i \in [I]$, the unit selling price is v_i, the unit ordering cost is c_i, and the unit salvage cost and the unit stock-out cost at the end of the sale season are g_i and b_i, respectively. We assume $c_i < v_i$ and $g_i < v_i$ for any $i \in [I]$, which is the prerequisite for a profitable business without arbitrage opportunity. For each item i, we denote the order quantity as x_i and the realized demand as z_i, hence the corresponding sale is $\min\{x_i, z_i\}$. The total cost for the newsvendor is given as:

$$\begin{aligned}
& f(\boldsymbol{x}, \boldsymbol{z}) \\
= & \sum_{i \in [I]} \{c_i x_i - v_i \min\{x_i, z_i\} - g_i (x_i - \min\{x_i, z_i\}) + b_i (z_i - \min\{x_i, z_i\})\} \\
= & \sum_{i \in [I]} \{(c_i - v_i - b_i) x_i + b_i z_i + (v_i + b_i - g_i) \max\{x_i - z_i, 0\}\} \\
= & (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})'\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{z} + (\boldsymbol{v} + \boldsymbol{b} - \boldsymbol{g})'(\boldsymbol{x} - \boldsymbol{z})^+,
\end{aligned}$$

where $(\boldsymbol{x} - \boldsymbol{z})^+ = \max\{\boldsymbol{x} - \boldsymbol{z}, 0\}$ and the 'max' denotes component-wise maximization. Given a known demand distribution $\mathbb{P} \in \mathbb{R}^I$, the multi-item newsvendor problem is formulated as

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{x}})\right] = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{(\boldsymbol{c}-\boldsymbol{v}-\boldsymbol{b})'\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{\mu} + \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{v}+\boldsymbol{b}-\boldsymbol{g})'(\boldsymbol{x}-\boldsymbol{z})^{+}\right]\right\},$$

where $\boldsymbol{\mu}$ is the means of demands and the order decision $\boldsymbol{x}$ is subjected to a feasible budget set $\mathcal{X}$.

In the case of single-item newsvendor problem without budget constraint, i.e., $I = 1$ and $\mathcal{X} = \mathbb{R}_{+}$, the optimal order quantity is known to be determined by a celebrated critical quantile of the demand distribution that relates to the cost parameters. However, in the case of multi-item newsvendor problem, the possible correlation among items prohibits the extension of this result. For a real world multi-item newsvendor problem, estimating the joint demand distribution of items is also statistically challenging. Even the joint demand distribution is known, evaluating the expected positive part involves multidimensional integration and is computationally prohibitive in general. Therefore, it is often of interest to investigate an alternative approach to solve multi-item newsvendor problem.

Distributionally robust multi-item newsvendor problem that can account for the ambiguity of uncertain demands has been extensively discussed (see, for instance, Hanasusanto et al. 2015a and Natarajan and Teo 2017). Instead of unrealistically assuming the complete knowledge of demand distribution, the distributionally robust model assumes that the demand distribution belongs to an ambiguity set $\mathcal{F}$ and solve the following optimization problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \left\{(\boldsymbol{c}-\boldsymbol{v}-\boldsymbol{b})'\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{\mu} + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{v}+\boldsymbol{b}-\boldsymbol{g})'(\boldsymbol{x}-\boldsymbol{z})^{+}\right]\right\}. \tag{4.33}$$

Hanasusanto et al. (2015a) study the distributionally robust multi-item newsvendor problem under CVaR risk measure, and they consider ambiguous demand distribution to be a known mixture of distinct distributions that resides in ambiguity sets with known ellipsoid support, known mean and known covariance. The authors show the distributionally robust optimization problem, in this case, is $\mathcal{NF}$-hard, even in the absence of support constraints and even the number of mixed distributions is one. Consequently, they provide an approximate reformulation in the positive semidefinite optimization format. Natarajan and Teo (2017) investigate the ambiguity set with unbounded support, known mean and known covariance, and they provide a positive semidefinite relaxation. They presents the solution improves significantly upon solution obtained from only uses marginal moments. Inspired by both models, we consider the following covariance dominance ambiguity set

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 \end{array} \right\},$$

where $\mathcal{Q} = \left\{\boldsymbol{q} \in \mathbb{R}^I \middle| \ \|\boldsymbol{q}\|_2 \leq 1\right\}$, and which specifies a tractable conic representable support set $\mathcal{W}$ and an upper bound $\boldsymbol{\Sigma}$ on covariance. We reformulate the multi-item newsvendor problem (4.33) as the following two-stage problem

$$\begin{aligned} \min \ & \left\{ (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})'\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{\mu} + \sup_{\mathbb{P} \in \mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{v} + \boldsymbol{b} - \boldsymbol{g})'\boldsymbol{y}(\boldsymbol{z})\right] \right\} \\ \text{s.t. } & \boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{x} - \boldsymbol{z}, \ \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{0}, \ \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{y} \in \mathcal{R}^{I,L}. \end{aligned} \tag{4.34}$$

We then instead tackle the adaptive distributionally robust optimization problem by the ELDR approximation in correspondence to some relaxed covariance ambiguity sets parameterized by certain $\bar{\mathcal{Q}} \subseteq \mathcal{Q}$. Quite notably, the recourse matrix $\boldsymbol{B}$ herein possesses a property that is stronger than complete recourse, which is formally defined as simple recourse.

Definition 7 (Simple recourse). The second stage problem (4.1) has simple recourse, if and only if it has complete recourse and each row of the recourse matrix $\boldsymbol{B}$ is some standard basis vector.

The simple recourse requires that $\hat{\boldsymbol{b}}_m = \boldsymbol{e}_{v(m)}$ for some $v(m)$-th standard basis vector $\boldsymbol{e}_{v(m)}$, where $v(\cdot)$ is a mapping from the set $[M]$ to the set $[L]$. In the case of simple recourse, the extreme rays of $\text{recc}(\boldsymbol{B})$ are standard basis vectors and the number of extreme rays equals to the number of adaptive decisions $\boldsymbol{y}$. Consequently, the simple recourse would lead to significant simplification of the separation problem (4.20).

Theorem 19. Suppose the second stage problem (4.1) has simple recourse, then the separation problem (4.20) is unbounded for a particular $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$, if and only if for some $l \in [L]$, the following inequality holds $\sum_{m\in\mathcal{M}_l} \theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l > 0$, where for any $l \in [L]$, the set $\mathcal{M}_l$ is defined as $\mathcal{M}_l = \{m \in [M] : v(m) = l\}$.

Proof. Note that under the simple recourse condition, any extreme ray $\boldsymbol{\eta} \geq \boldsymbol{0}$. In addition, for each $m \in [M]$, we have $\hat{\boldsymbol{b}}_m' \boldsymbol{\eta} = \boldsymbol{e}_{v(m)}' \boldsymbol{\eta} = \eta_{v(m)}$ for some $v(m) \in [L]$. The objective of the separation problem (4.20) can be represented as below:

$$\begin{aligned}\sum_{m\in[M]} \theta_m(\boldsymbol{q})(\hat{\boldsymbol{b}}_m' \boldsymbol{\eta}) - h(\boldsymbol{q})(\boldsymbol{d}'\boldsymbol{\eta}) &= \sum_{m\in[M]} \theta_m(\boldsymbol{q})\eta_{v(m)} - \sum_{l\in[L]} h(\boldsymbol{q})d_l\eta_l \\ &= \sum_{l\in[L]} \eta_l \left(\sum_{m\in\mathcal{M}_l} \theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l \right),\end{aligned}$$

which is additive and positive homogeneous in $\boldsymbol{\eta}$. Thus, the objective goes to positive infinity if and only if for some $l \in [L]$, the following inequality holds

$$\sum_{m\in\mathcal{M}_l} \theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l > 0,$$

and this completes our proof. □

To test the performance of the ELDR approximation to the formulation (4.34), we consider a numerical experiment inspired by the set-ups in both Hanasusanto et al.

(2015a) and Natarajan and Teo (2017). For a fixed number of items, we generate 100 random instances as follows. We sample the unit selling price $\boldsymbol{v}$ uniformly from $[5, 10]^I$. We set the unit salvage cost and unit stock-out cost to 10% and 25% of the unit selling price, and we sample the unit ordering cost uniformly from 50% to 60% of the unit selling price. The mean demand $\boldsymbol{\mu}$ is sampled uniformly from $[5, 100]^I$, while the standard deviation $\boldsymbol{\sigma}$ is sampled from independent uniform distributions on $[\boldsymbol{\mu}, 5\boldsymbol{\mu}]$. The correlations among different demands are generated by first sampling a random matrix $\boldsymbol{\Upsilon} \in \mathbb{R}^{I\times I}$ with independent elements uniformly distributed in $[\Delta, 1]$, and then setting the correlation matrix to $\text{diag}(\boldsymbol{w})\boldsymbol{V}\text{diag}(\boldsymbol{w})$, where $\boldsymbol{V} = \boldsymbol{\Upsilon}'\boldsymbol{\Upsilon}$ and $\boldsymbol{w}$ is a vector whose i-th element is defined as $w_i = 1/\sqrt{V_{ii}}$. Note that in the above process, we set the parameter Δ to be non-negative, hence the correlations among demands are assumed to be positive and are stronger with larger value of Δ. In particular, we vary the parameter Δ from 0 to 0.75 in steps of 0.25. We consider a box-typed support set such that the lower bound of is $\mathbf{0}$, and the upper bound is $\boldsymbol{\mu} + 3\boldsymbol{\sigma}$ that scales with both mean and standard deviation. Lastly, we set the feasible budget set $\mathcal{X} = \left\{\boldsymbol{x} \in \mathcal{R}^I \;\middle|\; \boldsymbol{x} \geq \mathbf{0},\ \sum_{i\in[I]} x_i \leq \boldsymbol{c}'(\boldsymbol{\mu} + \boldsymbol{\sigma})\right\}$. In particular, we consider different ELDR approximations as below.

- (**Approx CD**): we consider the ELDR approximation (4.27) in association with the auxiliary random matrix in the lifted covariance dominance ambiguity set.

- (**Approx MM**): we consider the ELDR approximation with the marginal moment ambiguity set, i.e., we consider $\mathcal{F}_{MM} \subseteq \mathcal{F}_C$ with $\bar{\mathcal{Q}} = \{\boldsymbol{e}_t |\ t \in [T]\}$.

- **Approx ALG**): we start with the marginal moment ambiguity set and utilize the Algorithm 2 with 50 iterations to progressively incorporate the advantage of covariance information.

We benchmark against the exact solution of Problem (4.33) with the covariance dominance ambiguity set $\mathcal{F}_C$. Note that we can represent Problem (4.33) as the following problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \left\{ (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})'\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{\mu} + \sup_{\mathbb{P}\in\mathcal{F}_C} \mathbb{E}_{\mathbb{P}} \left[\max_{\mathcal{S}:\mathcal{S}\subseteq[I]} \sum_{i\in\mathcal{S}} (v_i + b_i - g_i)(x_i - \tilde{z}_i) \right] \right\},$$

which is a special case of Problem (4.8) and would result in a positive semidefinite optimization format with the representation of ambiguity set $\mathcal{F}_C$. Since the number of subsets $\mathcal{S}$ equals to 2^I, we will study the case of 5, 8, and 10 items so that it is computationally variable for comparison.

We present the results of different cases in Table 4.1, Table 4.2, and Table 4.3. In particular, the tables represents the median (minimal, average, maximal) relative gaps among all instances of different ELDR approximations in the objective value to that of the exact solution, under various values of Δ. In this example, the conservativeness of considering only marginal moments information grows as the positive correlations among demands get stronger. By iteratively incorporating the covariance information, the Algorithm 2 can yield solutions that significantly mitigate the conservativeness. This reveals the importance and benefits of considering covariance information in adaptive distributionally robust optimization problems. Observe that the maximal gap suggests that there are some extreme instances where the ELDR approximation performs poorly. If we compare the median relative gap with the average relative gap, we then see such extreme instances drag down the average performance of the ELDR approximation. Consequently, the ELDR approximation should perform even better in a more reasonably 'expected' sense where we are allowed exclude these outliers.

Δ	Approx CD	Approx MM	Approx ALG
0	0.3 $(0, 0.6, 9.5)$	0.5 $(0, 2.1, 26.6)$	0.3 $(0, 0.6, 9.5)$
0.25	0.5 $(0, 2.9, 97.4)$	2.4 $(0.1, 22.6, 439.5)$	0.6 $(0, 3.0, 97.5)$
0.50	0.5 $(0, 3.3, 89.5)$	16.2 $(0.2, 44.9, 690.1)$	0.6 $(0.1, 3.4, 89.5)$
0.75	0.7 $(0, 3.4, 40.0)$	48.2 $(0.5, 166.2, 2270.2)$	0.1 $(0.8, 3.5, 40.0)$

Tab. 4.1: Median (minimal, average, maximal) relative gap (%) in the objective to the exact solution (5 items).

Δ	Approx CD	Approx MM	Approx ALG
0	0.9 $(0.1, 3.4, 89.4)$	4.1 $(0.1, 44.3, 890.4)$	1.0 $(0.1, 4.4, 92.6)$
0.25	0.9 $(0.0, 3.3, 74.1)$	33.9 $(0.2, 81.3, 1082.0)$	1.4 $(0.2, 3.8, 74.1)$
0.50	0.7 $(0, 3.1, 55.2)$	74.7 $(0.6, 357.7, 14238.0)$	1.2 $(0.1, 4.6, 91.1)$
0.75	0.2 $(0, 1.5, 70.8)$	128.4 $(6.2, 336.3, 10864.3)$	0.4 $(0.1, 2.8, 91.0)$

Tab. 4.2: Median (minimal, average, maximal) relative gap (%) in the objective to the exact solution (8 items).

Δ	Approx CD	Approx MM	Approx ALG
0	0.8 $(0, 3.9, 162.5)$	8.0 $(0, 31.6, 942.0)$	1.5 $(0, 6.4, 292.5)$
0.25	0.8 $(0.1, 3.4, 49.1)$	74.7 $(0.4, 133.2, 2270.9)$	3.3 $(0.4, 7.9, 70.4)$
0.50	0.7 $(0.1, 1.6, 32.7)$	149.3 $(2.9, 384.3, 11342.1)$	2.8 $(0.6, 7.7, 132.1)$
0.75	0.2 $(0, 0.8, 12.9)$	176.6 $(36.5, 415.5, 6903.4)$	0.9 $(0.2, 3.8, 106.4)$

Tab. 4.3: Median (minimal, average, maximal) relative gap (%) in the objective to the exact solution (10 items).

4.7.2 Distributionally robust multi-period inventory control problem

For the second numerical example, we study a distributionally robust multi-period single item inventory control problem to showcase our approach to the multi-stage problem. We consider a finite T-period, where the uncertain demand at the t period is $\tilde{d}_t, t \in [T]$. At the beginning of period t, the order quantity is $x_t \in [0, \bar{x}_t]$, which is assumed to arrive immediately to replenish the stock before the demand is realized. The unit ordering cost is c_t, the unit holding cost of excessive inventory is h_t, and the unit backlogged cost is b_t. We consider the following demand process motivated by Graves (1999) and See and Sim (2010):

$$\tilde{d}_t = d_t(\boldsymbol{z}_t) = \tilde{z}_t + \alpha\tilde{z}_{t-1} + \cdots + \alpha\tilde{z}_1 + \mu,$$

where the uncertain factors z_t are realized periodically and are identically distributed in $[-\bar{z}, \bar{z}]$ with zero mean. For any $t \in [T]$, let $\boldsymbol{z}_t = (z_1, \ldots, z_t)$ for ease of exposition.

We consider the following covariance dominance ambiguity set that encompasses the distributions of the uncertain factor $\boldsymbol{z}_T$.

$$\mathcal{F}_C = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^T\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{q}'\tilde{\boldsymbol{z}})^2\right] \leq \boldsymbol{q}'\boldsymbol{\Sigma}\boldsymbol{q}, \quad \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in [-\bar{z}, \bar{z}]^T\right] = 1 \end{array} \right\},$$

where $\mathcal{Q} = \left\{\boldsymbol{q} \in \mathbb{R}^I \middle| \, \|\boldsymbol{q}\|_2 \leq 1\right\}$ and the upper bound on covariance matrix $\boldsymbol{\Sigma}$ is a diagonal matrix such that $\Sigma_{tt} = \bar{z}^2/3$ for any $t \in [T]$. Given an ambiguity set $\mathcal{F}_C$, our objective is to minimize the worst-case expected total cost over the entire horizon:

$$\begin{array}{lll} \min & \sup\limits_{\mathbb{P} \in \mathcal{F}_C} \mathbb{E}_{\mathbb{P}}\left[\sum\limits_{t=1}^{T}\left(c_t x_t(\tilde{\boldsymbol{z}}_{t-1}) + y_t(\tilde{\boldsymbol{z}}_t)\right)\right] & \\ \text{s.t.} & y_t(\boldsymbol{z}_t) \geq b_t\left(\sum\limits_{v=1}^{t}(d_v(\boldsymbol{z}_v) - x_v(\boldsymbol{z}_{v-1}))\right), & \forall \boldsymbol{z} \in \mathcal{W}, t \in [T] \\ & y_t(\boldsymbol{z}_t) \geq h_t\left(\sum\limits_{v=1}^{t}(x_v(\boldsymbol{z}_{v-1}) - d_v(\boldsymbol{z}_v))\right), & \forall \boldsymbol{z} \in \mathcal{W}, t \in [T] \\ & 0 \leq x_t(\boldsymbol{z}_{t-1}) \leq \bar{x}_t, & \forall \boldsymbol{z} \in \mathcal{W}, t \in [T] \\ & x_t \in \mathcal{R}^{t-1,1}, y_t \in \mathcal{R}^{t,1}, & \forall t \in [T]. \end{array} \tag{4.35}$$

Quite notably, the number of extreme rays of the recession cone generated by the recourse matrix in Problem (4.35) is identical to the number of periods.

Theorem 20. Consider a finite horizon, T-period inventory control problem, the recession cone generated by the recourse matrix in Problem (4.35) has T extreme rays.

Proof. Observe that the first stage decision in Problem (4.35) is x_1, and the adaptive decisions are $x_t, t = 2, \ldots, T$ and $y_t, t = 1, \ldots, T$. We can represent the constraints in the following abstract form: $\hat{\boldsymbol{a}}(\boldsymbol{z})x_1 + \boldsymbol{B}(x_2, \ldots, x_T, y_1, \ldots, y_T)' \geq \boldsymbol{b}(\boldsymbol{z})$. Specifically, we have

$$\hat{\boldsymbol{a}}(\boldsymbol{z}) = (0, 0, \ldots, 0, 0, b_1, -h_1, \ldots, b_T, -h_T) \in \mathbb{R}_{(4T-2)\times 1},$$

the matrix $\boldsymbol{b}(\boldsymbol{z}) \in \mathbb{R}^{(4T-2)\times 1}$ is given by

$$\left(0, -\bar{x}_2, \ldots, 0, -\bar{x}_T, b_1 d_1(z_1), -h_1 d_1(z_1), \ldots, b_T \sum_{v=1}^{T} d_v(\boldsymbol{z}_v), -h_T \sum_{v=1}^{T} d_v(\boldsymbol{z}_v)\right),$$

and the recourse matrix $\boldsymbol{B} \in \mathbb{R}^{(4T-2)\times(2T-1)}$ takes the form

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{O} \\ \boldsymbol{B}_2 & \boldsymbol{B}_3 \end{pmatrix},$$

where

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & -1 \end{pmatrix} \in \mathbb{R}^{2(T-1)\times(T-1)},$$

$$\boldsymbol{B}_2 = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ b_2 & 0 & 0 & \cdots & 0 \\ -h_2 & 0 & 0 & \cdots & 0 \\ b_3 & b_3 & 0 & \cdots & 0 \\ -h_3 & -h_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_T & b_T & \cdots & \cdots & b_T \\ -h_T & -h_T & \cdots & \cdots & -h_T \end{pmatrix} \in \mathbb{R}^{2T\times(T-1)},$$

and

$$\boldsymbol{B}_3 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2T \times T},$$

and the zero matrix $\boldsymbol{O} \in \mathbb{R}^{2(T-1)\times T}$.

By the definition of an extreme ray, in the linear system $\boldsymbol{B\eta} \geq \boldsymbol{0}$, there are $2T-1$ linearly independent constraints active at the extreme ray $\boldsymbol{\eta}$. It is clear that for all $i \in [T-1]$, $\eta_i = 0$. As a consequence, the extreme ray $\boldsymbol{\eta}$ must further have $T-1$ zeros among its last T components. In particular, the extreme ray $\boldsymbol{\eta}$ takes the form $\boldsymbol{\eta} = (\boldsymbol{0}, \boldsymbol{e}_i)$ for some i-th standard unit basis in $\mathbb{R}^T$. Therefore, there are T extreme rays in total. □

Theorem 20 reveals that the recession cone generated by the recourse matrix has a number of extreme rays that scales linearly with the number of periods. A corollary further implies that so does the recession cone generated by any sub-matrix of the recourse matrix that corresponds to a particular $\boldsymbol{q}^*$. This is an attractive feature for us to consider the ELDR approximation and implement Algorithm 1 to improve its performance. Specifically, we consider the ELDR approximation among different relaxed covariance dominance ambiguity sets as follows.

• (**Approx MM**): we consider the marginal moment ambiguity set $\mathcal{F}_{MM} \subseteq \mathcal{F}_C$ such that in $\mathcal{F}_{MM}$, we have $\bar{\mathcal{Q}} = \{\boldsymbol{e}_t |\ t \in [T]\}$.

• (**Approx PCM**): motivated by Bertsimas et al. (2017), we consider the partial cross moment ambiguity set $\mathcal{F}_{PCM}$, in which we specify an upper bound on the variance of $\sum_{r=s}^{t} \tilde{z}_r$ for any $s \leq t, t \in [T]$, that is, we have $\bar{\mathcal{Q}} = \{\boldsymbol{q}_{st} |\ t \in [T], s \leq t\}$ such that

$$\boldsymbol{q}_{st} = \Bigg(\underbrace{0,\ldots,0}_{s-1},\underbrace{1,\ldots,1}_{t-s+1},\underbrace{0,\ldots,0}_{T-t}\Bigg).$$

• (**Approx ALG**): we start with the above partial cross moment ambiguity set and utilize the Algorithm 1 with 15 iterations.

We study instances with $T = 5, 10, 20$, and we report, respectively, the objective values in Table 4.4, Table 4.5, and Table 4.6. We set $\bar{x}_t = 260, c_t = 0.1, h_t = 0.02$ for all $t \in [T]$, $b_t = h_t b/h$ for all $t \in [T-1]$ and $b_T = 10b_{T-1}$. Because unfulfilled demands at the last period are lost, we set the corresponding unit backlogged cost relatively high. For the case of $T = 5$, we set $\mu = 200$ and $\bar{z} = 40$, while for $T = 10$, we set $\mu = 200$ and $\bar{z} = 20$; for $T = 20$, we set $\mu = 240$ and $\bar{z} = 12$. The symbol '$-$' denotes the Algorithm 2 does not improve over the approach that considers the partial cross moment ambiguity set. We observe that when the problem size becomes larger and when the parameter α increases, the differences in objective values among different approaches become larger. As in the previous numerical example, we show the significance of incorporating covariance information in adaptive distributionally robust optimization problems.

$T = 5$			
b/h	Approx MM	Approx PCM	Approx ALG
$\alpha = 0$			
10	108.0	108.0	$-$
30	108.0	108.0	$-$
50	108.0	108.0	$-$
$\alpha = 0.25$			
10	109.2	109.2	$-$
30	109.2	109.2	$-$
50	109.2	109.2	$-$
$\alpha = 0.50$			
10	160.3	124.9	123.9
30	265.4	152.7	152.3
50	369.7	179.5	179.2
$\alpha = 0.75$			
10	219.9	145.2	144.0
30	435.1	208.3	204.5
50	648.6	268.9	262.5
$\alpha = 1$			
10	280.1	170.1	166.9
30	605.5	276.1	264.7
50	928.4	379.0	358.6

Tab. 4.4: Performance of the ELDR approximation under different relaxed covariance dominance ambiguity sets ($T = 5$).

$T = 10$			
b/h	Approx MM	Approx PCM	Approx ALG
$\alpha = 0$			
10	206.0	206.0	–
30	206.0	206.0	–
50	206.0	206.0	–
$\alpha = 0.25$			
10	206.1	206.1	–
30	206.1	206.1	–
50	206.1	206.1	–
$\alpha = 0.50$			
10	237.4	217.6	217.2
30	287.9	225.0	224.9
50	338.0	231.9	231.8
$\alpha = 0.75$			
10	376.2	245.3	244.5
30	686.0	296.2	293.7
50	993.4	343.1	338.8
$\alpha = 1$			
10	527.9	283.8	281.0
30	1114.4	390.7	382.2
50	1696.1	491.5	476.3

Tab. 4.5: Performance of the ELDR approximation under different relaxed covariance dominance ambiguity sets ($T = 10$).

$T = 20$			
b/h	Approx MM	Approx PCM	Approx ALG
$\alpha = 0$			
10	486.0	486.0	–
30	486.0	486.0	–
50	486.0	486.0	–
$\alpha = 0.25$			
10	827.6	539.1	537.9
30	1442.0	604.0	600.9
50	2050.4	661.8	655.6
$\alpha = 0.50$			
10	1347.7	642.5	637.0
30	2890.5	849.0	828.0
50	4419.5	1044.3	1005.3
$\alpha = 0.75$			
10	1872.8	762.2	746.9
30	4367.0	1156.5	1095.7
50	6840.8	1539.2	1425.5
$\alpha = 1$			
10	2402.0	893.6	862.7
30	5875.3	1512.4	1404.6
50	9340.5	2120.3	1931.8

Tab. 4.6: Performance of the ELDR approximation under different relaxed covariance dominance ambiguity sets ($T = 20$).

4.8 Conclusion

We study the adaptive distributionally robust optimization with infinitely constrained ambiguity sets. To obtain approximate solutions, we consider the relaxed ambiguity set and adopt the ELDR approximation that includes the auxiliary random variables originating from the corresponding lifted ambiguity set. We show the benefits of such an approach and propose a solution procedure that iteratively returns a refined approximation that follows the aforementioned fashion. However, a limitation in our solution procedure is that it does not guarantee to refine the approximation towards optimality. Recently, Zhen et al. (2016) demonstrate how an adaptive (distributionally) robust optimization problem can be transferred to a static robust optimization problem via Fourier-Motzkin elimination (FME). This approach would generally create exponential number constraints. Nevertheless, a hybrid of FME and ELDR approximation, that is, partially performing FME and then applying ELDR approximation, is a promising and attractive approach to obtain the best approximation within the available computational recourses.

5. TRACTABLE DISTRIBUTIONALLY ROBUST OPTIMIZATION WITH DATA

5.1 Introduction

In Chapter 4, we focus on adaptive optimization problems involving only continuous recourse decisions. There are also important works on extending to discrete recourse decisions (Bertsimas et al. 2011b, Bertsimas and Goyal 2012, and Bertsimas and Georghiou 2015, 2017). Meanwhile, there is another approach for addressing optimization problems with recourse, which is called finite adaptability. This notion was first introduced by Bertsimas and Caramanis (2010) and was further explored in Hanasusanto et al. (2015b) and Hanasusanto et al. (2016b). Unlike the LDR, instead of being fully adaptive to the realization of uncertainty, finite adaptability pre-determines a set of recourse decisions and implements the best one of them once the uncertain parameters are observed. Such a key feature of finite adaptability is also favorable in situations that involve discrete recourse variables.

In this chapter, we discuss our treatment on discrete recourse decisions. It is worth mentioning that motivated by this initial aim, we arrive at a proposal on the design of a new framework for distributionally robust optimization. We summarize the contributions of this chapter as follows:

1. Inspired by the format for specifying ambiguity sets proposed by Wiesemann et al. (2014), we propose a new format for tractable ambiguity sets that is more intuitive in their representation and could encompass a richer family of distributional ambiguity without compromising on computational tractability.

2. We allow for partitioning the support set into a collection of disjoint confidence sets that may not necessarily be convex and closed. We demonstrate its potential for specifying conditional expectation, disjoint confidence sets with uncertain probabilities, and ϕ-divergence on uncertain probabilities, as well as handling discrete recourse variables for problems with recourse.

3. We show that the Wasserstein-based ambiguity set has an equivalent formulation via our proposed ambiguity set and this has ramification in our ability to tractably approximate a Wasserstein-based distributionally robust optimization problem with recourse

4. We propose a tractable adaptive recourse scheme (TARS), which enables us to obtain the *here-and-now* solution of a distributionally robust optimization problem with recourse by reformulating it as a tractable robust optimization problem. The TARS leverages the features from several classical models for addressing optimization under uncertainty, such as linear decision rules and finite adaptability, in a unified fashion.

5.2 A unified format for tractable ambiguity sets

We consider a minimization problem where the objective function is uncertain and involves a biconvex and piecewise affine objective function $f(\boldsymbol{x}, \boldsymbol{z}) : \mathbb{R}^{N_1} \times \mathbb{R}^{I_1} \mapsto \mathbb{R}$. Here, $\boldsymbol{x} \in \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}^{N_1}$ denotes the decision variable and $\boldsymbol{z} \in \mathbb{R}^{I_1}$ represents a realization of the uncertain random variable, $\tilde{\boldsymbol{z}}$ with probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_1})$, which is unknown but is an element of the ambiguity set, $\mathcal{F} \subseteq \mathcal{P}_0(\mathbb{R}^{I_1})$. The decision maker is ambiguity averse (see, Gilboa and Schmeidler 1989) and evaluates the uncertain objective function by taking the worst-case expectation as below

$$F(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right].$$

While the objective function, $F(\boldsymbol{x})$ remains convex in $\boldsymbol{x} \in \mathcal{X}$, evaluating the expectation over a probability distribution is a multidimensional integration problem, which is generally intractable. Therefore, the ambiguity set $\mathcal{F}$ plays a crucial role in our ability to evaluate the function $F(\boldsymbol{x})$ efficiently, which has to be addressed before we can proceed to the next step of optimizing over $\boldsymbol{x} \in \mathcal{X}$. For notational convenience, we suppress the decision variable $\boldsymbol{x}$ and focus on evaluating $\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$ for a convex and piecewise affine function, $f(\boldsymbol{z}) : \mathbb{R}^{I_1} \mapsto \mathbb{R}$ as follows:

$$f(\boldsymbol{z}) = \max_{k \in [K]} \left\{\boldsymbol{a}_k' \boldsymbol{z} + b_k\right\} \tag{5.1}$$

for some parameters $\boldsymbol{a}_k \in \mathbb{R}^{I_1}, b_k \in \mathbb{R}, \forall k \in [K]$.

Unfortunately, many ambiguity sets are based on expectation constraints on moments, which may not necessarily lead to tractable evaluation of the worst-case expectation of the objective function (see, for instance, Bertsimas and Popescu 2005). Wiesemann et al. (2014) propose a framework for characterizing ambiguity sets that would enable reformulation of the worst-case expectation in the form of a tractable minimization problem. The "standard form" of the WKS ambiguity set has an input format given by

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{G}\tilde{\boldsymbol{z}} + \boldsymbol{H}\tilde{\boldsymbol{u}}\right] = \boldsymbol{\mu} \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_j\right] = [\underline{p}_j, \overline{p}_j], \quad \forall j \in [J] \end{array} \right\}$$

for some given parameters $\boldsymbol{G} \in \mathbb{R}^{L \times I_1}$, $\boldsymbol{H} \in \mathbb{R}^{L \times I_2}$, $\boldsymbol{\mu} \in \mathbb{R}^L$ and lifted confidence sets $\mathcal{C}_j$, $j \in [J]$ for the joint probability distribution of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$. The confidence sets $\mathcal{C}_j$ are closed, convex and *tractable*.

Definition 8 (tractable functions and sets). We say a convex function $f : \mathbb{R}^I \mapsto \mathbb{R}$ is tractable if and only if it can be evaluated in polynomial time and its epigraph epi$f :=$ $\left\{(\boldsymbol{x}, t) \in \mathbb{R}^I \times \mathbb{R} \mid f(\boldsymbol{x}) \leq t\right\}$ can be compactly represented as a polynomial sized convex set. We say a set $\mathcal{C} \in \mathbb{R}^I$, which may not be convex or closed, is tractable if and only if its support function, $h_{\mathcal{C}}(\boldsymbol{x}) := \sup_{\boldsymbol{z}\in\mathcal{C}} \boldsymbol{x}'\boldsymbol{z}$ is tractable.

The distinctive features of the WKS ambiguity set are the introduction of the auxiliary random variable, $\tilde{\boldsymbol{u}}$ that does not explicitly appear in the objective function f, the joint random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$ that resides within the lifted confidence sets, $\mathcal{C}_j$ with potentially unknown probabilities in $[\underline{p}_j, \overline{p}_j]$, $j \in [J]$, and the expectation of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$ that is constrained within an affine manifold.

The inclusion of the auxiliary random variable, $\tilde{\boldsymbol{u}}$ allows the ambiguity set to model a richer variety of distributional information pertaining to the random variable $\tilde{\boldsymbol{z}}$. As an illustration, we consider the following ambiguity set

$$\mathcal{F}_1 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) \;\middle|\; \begin{array}{ll} \tilde{\boldsymbol{z}} \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} & \\ \boldsymbol{\Sigma} - \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'\right] \in \mathbb{S}_+^{I_1} & \\ \mathbb{E}_{\mathbb{P}}\left[g_i(\tilde{\boldsymbol{z}})\right] \leq \sigma_i, & \forall i \in [I] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1 & \end{array} \right\},$$

where $\mathbb{S}_+^{I_1}$ denotes the cone of $I_1 \times I_1$ symmetric positive semidefinite matrices. The first expectation constraint specifies the mean of $\tilde{\boldsymbol{z}}$, the second expectation constraint imposes the covariance of $\tilde{\boldsymbol{z}}$ so that $\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{v}'\tilde{\boldsymbol{z}})^2\right] \leq \boldsymbol{v}'\boldsymbol{\Sigma}\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{R}^{I_1}$, and the third collection of constraints involving tractable convex functions $g_i, i \in [I]$ that can provide useful and interesting characterizations of probability distributions beyond moments including, among others, expected absolute deviation, semi-variance and so forth. Based on the lifting theorem of Wiesemann et al. (2014), we have

$$\mathcal{F}_1 = \Pi_{\tilde{\boldsymbol{z}}}\bar{\mathcal{F}}_1,$$

where $\Pi_{\tilde{\boldsymbol{z}}}\bar{\mathcal{F}}_1$ is defined as the set of marginal probability distributions of $\tilde{\boldsymbol{z}}$ under any

$\mathbb{P} \in \bar{\mathcal{F}}_1$,

$$\bar{\mathcal{F}}_1 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{S}_+^{I_1} \times \mathbb{R}^I\right) \left| \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{U}}\right] = \boldsymbol{\Sigma} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}\right] = \boldsymbol{\sigma} \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}\right] = 1 \end{array} \right. \right\},$$

and

$$\mathcal{C} = \left\{ (\boldsymbol{z}, \boldsymbol{U}, \boldsymbol{u}) \in \mathcal{W} \times \mathbb{S}_+^{I_1} \times \mathbb{R}^I \;\middle|\; \boldsymbol{U} - \boldsymbol{z}\boldsymbol{z}' \in \mathbb{S}_+^{I_1}, g_i(\tilde{\boldsymbol{z}}) \leq u_i, \forall i \in [I] \right\}.$$

Hence,

$$\sup_{\mathbb{P} \in \mathcal{F}_1} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P} \in \bar{\mathcal{F}}_1} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right].$$

Observe that in the expectation constraints, the auxiliary random variables $\tilde{\boldsymbol{U}}$ and $\tilde{u}_i, i \in [I]$ are respectively associated with $\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'$ and $g_i(\tilde{\boldsymbol{z}}), i \in [I]$, which are nonlinear functions of the nominal random variable, $\tilde{\boldsymbol{z}}$. Accordingly, the lifted support set, $\mathcal{C}$ incorporates the support set of the random variable $\tilde{\boldsymbol{z}}$ and the "conic epigraphs" of $\boldsymbol{z}\boldsymbol{z}'$ and $g_i(\boldsymbol{z}), i \in [I]$. By Schur complement,

$$\boldsymbol{U} - \boldsymbol{z}\boldsymbol{z}' \in \mathbb{S}_+^{I_1} \Leftrightarrow \begin{bmatrix} 1 & \boldsymbol{z}' \\ \boldsymbol{z} & \boldsymbol{U} \end{bmatrix} \in \mathbb{S}_+^{I_1+1},$$

hence, the lifted support set, $\mathcal{C}$ is a tractable set in the form of linear matrix inequalities (Ben-Tal and Nemirovski 2001). However, this is not the case when we impose in the ambiguity set $\mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}'\right] = \boldsymbol{\Sigma}$, because the corresponding lifted support set,

$$\mathcal{C} = \left\{ (\boldsymbol{z}, \boldsymbol{U}, \boldsymbol{u}) \in \mathcal{W} \times \mathbb{S}_+^{I_1} \times \mathbb{R}^I \;\middle|\; \boldsymbol{U} = \boldsymbol{z}\boldsymbol{z}', g_i(\tilde{\boldsymbol{z}}) \leq u_i, \forall i \in [I] \right\},$$

is intractable (Bertsimas and Popescu 2005). Therefore, the ambiguity set is designed in the way to ensure that the worst-case expectation can be obtained by solving a tractable optimization problem using the state-of-the-art solvers such as CPLEX and Gurobi.

More recently, Bertsimas et al. (2017) show that the auxiliary random variable $\tilde{\boldsymbol{u}}$ can be incorporated in linear decision rules, which can significantly improve the approximation of distributionally robust optimization problems with recourse (see an application to a multi-stage vehicle repositioning problem in He et al. 2017). To further improve the representation of distributional ambiguity and facilitate a more intuitive modeling framework, we propose a new format for defining tractable ambiguity sets as follows:

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) \left| \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})\right] = \boldsymbol{v} \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{D}_j\right] = p_j, \quad \forall j \in [J] \\ \text{for some } (\boldsymbol{v}, \boldsymbol{p}) \in \mathcal{V} \end{array} \right. \right\}, \tag{5.2}$$

where $\mathcal{V} \subseteq \mathbb{R}^{I_1+I_2+J}$ is a tractable convex set such that $\Pi_{\boldsymbol{p}}\mathcal{V} \subseteq \{\boldsymbol{p} \in \mathbb{R}^J_+ \mid \sum_{j\in[J]} p_j = 1\}$, and $\mathcal{D}_j, j \in [J]$ are tractable disjoint confidence sets that do not overlap and constitute a partition of the support set of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$. To obtain a tractable reformation, we utilize the concept of conic representation and assume the following Slater's condition:

Assumption 3. The conic representation of the following system

$$\begin{aligned} & \sum_{j\in[J]} \boldsymbol{z}_j = \boldsymbol{v}_1 \\ & \sum_{j\in[J]} \boldsymbol{u}_j = \boldsymbol{v}_2 \\ & \boldsymbol{w} = \boldsymbol{p} \\ & (\boldsymbol{z}_j, \boldsymbol{u}_j, w_j) \in \mathcal{K}(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))), \quad \forall j \in [J] \\ & (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{p}) \in \mathcal{V} \end{aligned} \tag{5.3}$$

satisfies the Slater's condition (see, Theorem 1.4.2 in Ben-Tal and Nemirovski 2001) and has a solution such that $\boldsymbol{p} > 0$.

We next describe some new features of our proposed format for tractable ambiguity sets.

Convex expectation constraints

Despite the generality of the WKS ambiguity set, the requirement for the expectation of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$ to reside within an affine manifold may lead to cumbersome representation of ambiguity sets. For instance, to specify the boundary of the means of random variables such as $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}]$ and $\|\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] - \boldsymbol{\mu}\|_2 \leq \epsilon$, we will require to introduce new auxiliary random variables that are not directly related to the realization of $\tilde{\boldsymbol{z}}$. As an illustration, we can express the following ambiguity set,

$$\mathcal{F}_2 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})] \in \mathcal{V} \\ \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{D}] = 1 \end{array} \right\}$$

as an WKS ambiguity set. Indeed, observing that $(\mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})], 1) \in \mathcal{K}(\mathcal{V})$ and using the lifting theorem, the following WKS ambiguity set,

$$\bar{\mathcal{F}}_2 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_2, \tilde{s}_0) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_2, \tilde{s}_0)] = (\boldsymbol{0}, \boldsymbol{0}, 1) \\ \mathbb{P}\left[\begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{D} \\ (\tilde{\boldsymbol{s}}_1 + \tilde{\boldsymbol{z}}, \tilde{\boldsymbol{s}}_2 + \tilde{\boldsymbol{u}}, \tilde{s}_0) \in \mathcal{K}(\mathcal{V}) \end{array} \right] = 1 \end{array} \right\}$$

satisfies $\mathcal{F}_2 = \Pi_{(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})} \bar{\mathcal{F}}_2$, where $\Pi_{(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})} \bar{\mathcal{F}}_2$ is the set of marginal probability distribution of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$ under any $\mathbb{P} \in \bar{\mathcal{F}}_2$. Nevertheless, despite the equivalence, from the modeling perspective, $\bar{\mathcal{F}}_2$ would have been less intuitive.

Disjoint confidence sets

Our proposed ambiguity set requires the collection of confidence sets $\mathcal{D}_j, j \in [J]$ to be disjoint, which is not mandatory in the WKS ambiguity set. Moreover, although we require the sets $\mathcal{D}_j, j \in [J]$ to be tractable, in contrast to the WKS ambiguity set, it is not necessary for them to be convex and closed. It should be noted that the lack of convexity in the confidence sets $\mathcal{D}_j, j \in [J]$ does not necessarily imply intractability. For instance, although the set $\mathcal{D} = \{\boldsymbol{z} \mid 1 < \|\boldsymbol{z}\|_2 < 2\}$ is neither closed nor convex, its support function $h_{\mathcal{D}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_2/2$ is a tractable function.

Our proposed ambiguity set always leads to tractable formation if the cost function $f(\boldsymbol{z})$ is convex, piecewise affine and with moderate number of segments. In contrast, the tractability of the WKS ambiguity set depends on how its confidence sets are overlapped. Nevertheless, Wiesemann et al. (2014) propose a "nesting condition" that will ensure the tractability of the WKS ambiguity set, which, as we will demonstrate in the later section, we use to construct a laminar family of confidence sets that are disjoint and tractable. We will also elucidate the benefits of our proposed format when addressing distributionally robust optimization problems with recourse, especially in the presence of discrete constraints on the recourse variables.

Incorporating conditional expectation

Another new feature of our proposed ambiguity set is having the joint uncertainty set for characterizing the expectations and the probabilities of confidence sets, which will provide additional flexibility such as, among other things, conditional expectation of the random variable arising from the realization of a random scenario. For instance, let $\tilde{w}$ denote a random scenario taking the value $j \in [J]$ with probability p_j, which, conditioning on its realization, will affect the means and support of the random variable $\tilde{\boldsymbol{z}}$. Consider the following ambiguity set

$$\mathcal{F}_3 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times [J]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{w}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}} \mid \tilde{w} = j\right] \in [\underline{\mu}_j, \overline{\mu}_j], & \forall j \in [J] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{W}_j, \tilde{w} = j\right] = p_j, & \forall j \in [J] \\ \text{for some } \boldsymbol{p} \in \mathcal{S} & \end{array} \right\},$$

in which $\mathcal{S} = \{\boldsymbol{p} \in \mathbb{R}^J_+ \mid \sum_{j\in[J]} p_j = 1\}$ and the conditional support sets $\mathcal{W}_j,\ j \in [J]$ are not required to be disjoint. We can equivalently represent this ambiguity set as a projection of the following ambiguity set

$$\mathcal{G}_3 = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_1\times J} \times [J]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}_1, \dots, \tilde{\boldsymbol{u}}_J, \tilde{w}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}_j\right] \in [\underline{\mu}_j p_j, \overline{\mu}_j p_j], & \forall j \in [J] \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}_1, \dots, \tilde{\boldsymbol{u}}_J, \tilde{w}) \in \mathcal{D}_j\right] = p_j, & \forall j \in [J] \\ \text{for some } \boldsymbol{p} \in \mathcal{S} & \end{array} \right\},$$

where $\mathcal{D}_j = \left\{(\boldsymbol{z}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_J, w) \in \mathcal{W}_j \times \mathbb{R}^{I_1\times J} \times [J] \mid w = j, \boldsymbol{u}_j = \boldsymbol{z}, \boldsymbol{u}_k = \boldsymbol{0}, \forall k \in [J] \setminus \{j\}\right\}$. Observe that with the inclusion of the random scenario in the confidence sets, the sets $\mathcal{D}_j$ are now disjoint.

Reformation as a robust optimization problem

Under our proposed ambiguity set, we next show how we can determine the worst-case expectation by solving a classical robust optimization problem. Given the nature of our proposed ambiguity set for which the confidence sets may not be convex, it is not obvious how we can apply the strong duality results for moment problems of Isii (1962), Shapiro (2001), and Bertsimas and Popescu (2005). Hence, we provide a standalone proof that will enable us to elicit the conditions for strong duality as we have presented in Assumption 1.

Theorem 21. Given a convex and piecewise affine function, $f(\cdot)$ in (5.1) and an ambiguity set, $\mathcal{G}$ in (5.2), the worst-case expectation,

$$\sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] \tag{5.4}$$

is the same as the optimal value of the following robust optimization problem:

$$\begin{aligned}
\inf \quad & s_0 \\
\text{s.t.} \quad & r_j + \boldsymbol{s}_1'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq f(\boldsymbol{z}), && \forall\, (\boldsymbol{z}, \boldsymbol{u}) \in \mathrm{cl}(\mathrm{CH}(\mathcal{D}_j)),\ \forall j \in [J] \\
& s_0 \geq \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2, && \forall (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{p}) \in \mathcal{V} \\
& \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2}.
\end{aligned} \tag{5.5}$$

Moreover, Problem (5.5) is solvable, i.e., its optimal value is attainable.

Proof. Observe that the worst-case expectation (5.4) can be re-expressed as:

$$\lambda^* = \sup_{(\boldsymbol{v},\boldsymbol{p}) \in \mathcal{V}} \lambda(\boldsymbol{v}, \boldsymbol{p}),$$

where for any $(\boldsymbol{v}, \boldsymbol{p}) \in \mathcal{V}$, we define

$$\mathcal{G}(\boldsymbol{v}, \boldsymbol{p}) = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})\right] = \boldsymbol{v} \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{D}_j\right] = p_j, \quad \forall j \in [J] \end{array} \right\}$$

and

$$\lambda(\boldsymbol{v}, \boldsymbol{p}) = \sup_{\mathbb{P} \in \mathcal{G}(\boldsymbol{v},\boldsymbol{p})} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right].$$

The dual of $\lambda(\boldsymbol{v}, \boldsymbol{p})$ is given by

$$\begin{aligned}
\lambda_1(\boldsymbol{v}, \boldsymbol{p}) = \quad \inf \quad & \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2 \\
\text{s.t.} \quad & r_j + \boldsymbol{s}_1'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq f(\boldsymbol{z}), \qquad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{D}_j,\ \forall j \in [J] \\
& \boldsymbol{r} \in \mathbb{R}^J, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2},
\end{aligned}$$

which provides an upper bound of $\lambda(\boldsymbol{v}, \boldsymbol{p})$. Indeed, consider any $\mathbb{P} \in \mathcal{G}(\boldsymbol{v}, \boldsymbol{p})$ and any feasible solution $(\boldsymbol{r}, \boldsymbol{s}_1, \boldsymbol{s}_2)$ in the dual, the set of robust counterparts in the dual implies that

$$\mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] \leq \sum_{j\in[J]} r_j p_j + \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{s}_1'\tilde{\boldsymbol{z}} + \boldsymbol{s}_2'\tilde{\boldsymbol{u}}\right] = \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2.$$

Hence, weak duality holds between $\lambda(\boldsymbol{v},\boldsymbol{p})$ and $\lambda_1(\boldsymbol{v},\boldsymbol{p})$, i.e., $\lambda(\boldsymbol{v},\boldsymbol{p}) \leq \lambda_1(\boldsymbol{v},\boldsymbol{p})$. By the general min-max theorem, we observe that

$$\begin{array}{rll} \lambda_1^* = \sup\limits_{(\boldsymbol{v},\boldsymbol{p})\in\mathcal{V}} \lambda_1(\boldsymbol{v},\boldsymbol{p}) \leq \lambda_2^* = & \inf & s_0 \\ & \text{s.t.} & r_j + \boldsymbol{s}_1'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq f(\boldsymbol{z}), \qquad \forall(\boldsymbol{z},\boldsymbol{u}) \in \mathcal{D}_j,\ \forall j \in [J] \\ & & s_0 \geq \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2, \qquad \forall(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{p}) \in \mathcal{V} \\ & & \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2}. \end{array}$$

We next proceed to establish $\lambda^* = \lambda_1^* = \lambda_2^*$ and show that Problem (5.5) is solvable.

Since f is a convex and piecewise affine function, the collection of robust counterparts becomes

$$r_j - b_k + (\boldsymbol{s}_1 - \boldsymbol{a}_k)'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq 0,\ \forall(\boldsymbol{z},\boldsymbol{u}) \in \mathcal{D}_j,\ \forall j \in [J],\ \forall k \in [K],$$

where we can replace $\mathcal{D}_j$ by $\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))$ because the left-hand side is affine in $(\boldsymbol{z},\boldsymbol{u})$. Thus, we have

$$\begin{array}{rll} \lambda_2^* = & \inf & s_0 \\ & \text{s.t.} & r_j - b_k + (\boldsymbol{s}_1 - \boldsymbol{a}_k)'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq 0, \qquad \forall(\boldsymbol{z},\boldsymbol{u}) \in \mathrm{cl}(\mathrm{CH}(\mathcal{D}_j)),\ \forall j \in [J],\ \forall k \in [K] \\ & & s_0 \geq \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2, \qquad \forall(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{p}) \in \mathcal{V} \\ & & \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2}. \end{array}$$

By the definition of dual cone, the collection of robust counterparts is equivalent to

$$(\boldsymbol{s}_1 - \boldsymbol{a}_k, \boldsymbol{s}_2, r_j - b_k) \in \mathcal{K}^*(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))),\ \forall j \in [J],\ \forall k \in [K].$$

Similarly, the second constraint can be re-expressed as $(-\boldsymbol{s}_1, -\boldsymbol{s}_2, -\boldsymbol{r}, s_0) \in \mathcal{K}^*(\mathcal{V})$. With these conic inequalities, we obtain

$$
\begin{aligned}
\lambda_2^* = \quad & \inf \quad s_0 \\
& \text{s.t.} \quad (\boldsymbol{s}_1, \boldsymbol{s}_2, r_j) - (\boldsymbol{a}_k, \boldsymbol{0}, b_k) \in \mathcal{K}^*(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))), \ \forall j \in [J], \ \forall k \in [K] \\
& \qquad (-\boldsymbol{s}_1, -\boldsymbol{s}_2, -\boldsymbol{r}, s_0) \in \mathcal{K}^*(\mathcal{V}) \\
& \qquad \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2}.
\end{aligned}
\tag{5.6}
$$

By conic duality, the dual of Problem (5.6) is given by

$$
\begin{aligned}
\lambda_3^* = \quad \sup \quad & \sum_{j \in [J]} \sum_{k \in [K]} \left(\boldsymbol{a}_k' \boldsymbol{\xi}_{jk} + b_k \eta_{jk} \right) \\
\text{s.t.} \quad & \sum_{k \in [K]} \sum_{j \in [J]} \boldsymbol{\xi}_{jk} = \boldsymbol{\alpha} \\
& \sum_{k \in [K]} \sum_{j \in [J]} \boldsymbol{\zeta}_{jk} = \boldsymbol{\beta} \\
& \sum_{k \in [K]} \eta_{jk} = \gamma_j, && \forall j \in [J] \\
& (\boldsymbol{\xi}_{jk}, \boldsymbol{\zeta}_{jk}, \eta_{jk}) \in \mathcal{K}(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))), && \forall j \in [J], \ \forall k \in [K] \\
& (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, 1) \in \mathcal{K}(\mathcal{V}) \\
& \boldsymbol{\xi}_{jk} \in \mathbb{R}^{I_1}, \boldsymbol{\zeta}_{jk} \in \mathbb{R}^{I_2}, \eta_{jk} \in \mathbb{R}, && \forall j \in [J], \ \forall k \in [K] \\
& \boldsymbol{\alpha} \in \mathbb{R}^{I_1}, \boldsymbol{\beta} \in \mathbb{R}^{I_2}, \boldsymbol{\gamma} \in \mathbb{R}^J,
\end{aligned}
\tag{5.7}
$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\xi}_{jk}, \boldsymbol{\zeta}_{jk}, \eta_{jk}, \forall j \in [J], \forall k \in [K]$ are the dual variables associated with the specified constraints respectively. Note that $\mathcal{K}^{**}(\mathcal{V}) = \mathcal{K}(\mathcal{V})$ and $\mathcal{K}^{**}(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))) = \mathcal{K}(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))), \forall j \in [J]$.

Under Assumption 1, Slater's condition holds for Problem (5.7), hence strong duality holds between λ_2^* and λ_3^*, i.e., $\lambda_2^* = \lambda_3^*$. In fact, there exists a sequence of feasible solutions in Problem (5.7)

$$
\left\{ \left(\bar{\boldsymbol{\alpha}}^l, \bar{\boldsymbol{\beta}}^l, \bar{\boldsymbol{\gamma}}^l, \left(\bar{\boldsymbol{\xi}}_{jk}^l, \bar{\boldsymbol{\zeta}}_{jk}^l, \bar{\eta}_{jk}^l \right)_{j \in [J], k \in [K]} \right) \right\}_{l \geq 0}
$$

such that

$$
(\bar{\boldsymbol{\alpha}}^l, \bar{\boldsymbol{\beta}}^l, \bar{\boldsymbol{\gamma}}^l) \in \mathrm{ri}(\mathcal{V}), \ \bar{\boldsymbol{\gamma}}^l > \boldsymbol{0}, \ (\bar{\boldsymbol{\xi}}_{jk}^l / \bar{\eta}_{jk}^l, \bar{\boldsymbol{\zeta}}_{jk}^l / \bar{\eta}_{jk}^l) \in \mathrm{ri}(\mathcal{D}_j), \ \bar{\eta}_{jk}^l > 0, \ \forall j \in [J], k \in [K],
$$

and

$$\lim_{l\to\infty}\sum_{j\in[J]}\sum_{k\in[K]}\left(\boldsymbol{a}_k'\bar{\boldsymbol{\xi}}_{jk}^l + b_k\bar{\eta}_{jk}^l\right) = \lambda_3^*.$$

Observe that for all $l \geq 0$, $(\bar{\boldsymbol{\xi}}_{jk}^l/\bar{\eta}_{jk}^l, \bar{\boldsymbol{\zeta}}_{jk}^l/\bar{\eta}_{jk}^l) \in \mathrm{ri}(\mathrm{cl}(\mathrm{CH}(\mathcal{D}_j))), \forall j \in [J], k \in [K]$. However, since $\mathcal{D}_j$ may not be convex, it may not necessarily imply $(\bar{\boldsymbol{\xi}}_{jk}^l/\bar{\eta}_{jk}^l, \bar{\boldsymbol{\zeta}}_{jk}^l/\bar{\eta}_{jk}^l) \in \mathrm{ri}(\mathcal{D}_j)$. Nevertheless, by Carathéodory's theorem, there exists $(\bar{\boldsymbol{z}}_{jk}^{lm}, \bar{\boldsymbol{u}}_{jk}^{lm}) \in \mathcal{D}_j$, $\nu_{jk}^{lm} \geq 0$, $m \in [M]$, and $M = I_1 + I_2 + 1$ satisfying

$$(\bar{\boldsymbol{\xi}}_{jk}^l/\bar{\eta}_{jk}^l, \bar{\boldsymbol{\zeta}}_{jk}^l/\bar{\eta}_{jk}^l) = \sum_{m\in[M]}\nu_{jk}^{lm}(\bar{\boldsymbol{z}}_{jk}^{lm}, \bar{\boldsymbol{u}}_{jk}^{lm})$$

and $\sum_{m\in[M]}\nu_{jk}^{lm} = 1$. We can thus correspondingly construct a sequence of discrete probability distributions $\left\{\mathbb{P}_l \in \mathcal{P}_0(\mathbb{R}^{I_1}\times\mathbb{R}^{I_2})\right\}_{l\geq 0}$ on random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathbb{R}^{I_1}\times\mathbb{R}^{I_2}$ in the following way:

$$\mathbb{P}_l\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) = \left(\bar{\boldsymbol{z}}_{jk}^{lm}, \bar{\boldsymbol{u}}_{jk}^{lm}\right)\right] = \nu_{jk}^{lm}\bar{\eta}_{jk}^l,\ \forall j \in [J],\ \forall k \in [K],\ \forall m \in [M].$$

Since for all $l \geq 0$, $\mathbb{E}_{\mathbb{P}_l}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})\right] = (\bar{\boldsymbol{\alpha}}^l, \bar{\boldsymbol{\beta}}^l)$ and $\mathbb{P}_l\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{D}_j\right] = \bar{\gamma}_j^l, \forall j \in [J]$, we have $\mathbb{P}_l \in \mathcal{G}$. Moreover, we have

$$\begin{aligned}
\lambda_3^* = & \lim_{l\to\infty}\sum_{j\in[J]}\sum_{k\in[K]}\left(\boldsymbol{a}_k'\bar{\boldsymbol{\xi}}_{jk}^l + b_k\bar{\eta}_{jk}^l\right) \\
= & \lim_{l\to\infty}\sum_{j\in[J]}\sum_{k\in[K]}\bar{\eta}_{jk}^l\left(\boldsymbol{a}_k'\left(\sum_{m\in[M]}\nu_{jk}^{lm}\bar{\boldsymbol{z}}_{jk}^{lm}\right) + b_k\left(\sum_{m\in[M]}\nu_{jk}^{lm}\right)\right) \\
= & \lim_{l\to\infty}\sum_{j\in[J]}\sum_{k\in[K]}\sum_{m\in[M]}\nu_{jk}^{lm}\bar{\eta}_{jk}^l\left(\boldsymbol{a}_k'\bar{\boldsymbol{z}}_{jk}^{lm} + b_k\right) \\
\leq & \lim_{l\to\infty}\sum_{j\in[J]}\sum_{k\in[K]}\sum_{m\in[M]}\nu_{jk}^{lm}\bar{\eta}_{jk}^l\left(\max_{n\in[K]}\left\{\boldsymbol{a}_n'\bar{\boldsymbol{z}}_{jn}^{lm} + b_n\right\}\right) \\
= & \lim_{l\to\infty}\mathbb{E}_{\mathbb{P}_l}\left[\max_{n\in[K]}\left\{\boldsymbol{a}_n'\tilde{\boldsymbol{z}} + b_n\right\}\right] \\
\leq & \sup_{\mathbb{P}\in\mathcal{G}}\mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] \\
= & \lambda^*.
\end{aligned}$$

Recall that $\lambda^* \le \lambda_1^* \le \lambda_3^*$, we now arrive at $\lambda^* \le \lambda_1^* \le \lambda_2^* = \lambda_3^* \le \lambda^*$. By Theorem 1.4.2 in Ben-Tal and Nemirovski (2001), Problem (5.5), the dual of Problem (5.7), is solvable. □

5.3 Ambiguity set with Wasserstein distance

Wasserstein-based distributionally robust optimization problems have attracted considerable interest recently from stochastic programming and robust optimization communities. The Wasserstein ambiguity set is characterized by limiting the statistical distance of its probability distributions to the reference probability distribution. Differing from ambiguity sets that are prescribed by ϕ-divergences, the Wasserstein ambiguity set provides a higher confidence in containing the true probability distribution that generates the observable past realizations and offers better out-of-sample performance guarantee (see, e.g., Esfahani and Kuhn 2015).

We consider a similar setting as in Zhao and Guan (2015) and Esfahani and Kuhn (2015) on the design of a Wasserstein ambiguity set. Particularly, as the reference probability distribution, we consider the random variable $\tilde{\boldsymbol{z}}^\dagger$ with realizations in $\{\hat{\boldsymbol{z}}_1, \dots, \hat{\boldsymbol{z}}_N\}$ and the corresponding empirical probability distribution $\mathbb{P}^\dagger$ as follows:

$$\mathbb{P}^\dagger\left[\tilde{\boldsymbol{z}}^\dagger = \hat{\boldsymbol{z}}_n\right] = \frac{1}{N},\ \forall n \in [N].$$

Given a tractable distance metric $\rho : \mathbb{R}^{I_1} \times \mathbb{R}^{I_1} \mapsto [0, +\infty)$ such that $\rho(\boldsymbol{z}, \boldsymbol{z}^\dagger) = 0$ if and only if $\boldsymbol{z} = \boldsymbol{z}^\dagger$, the Wasserstein distance $d_W : \mathcal{P}_0(\mathcal{W}) \times \mathcal{P}_0(\mathcal{W}) \mapsto \mathbb{R}$ is defined via

$$\begin{aligned} d_W(\mathbb{P}, \mathbb{P}^\dagger) := \quad \inf \quad & \mathbb{E}_{\bar{\mathbb{P}}}\left[\rho(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^\dagger)\right] \\ \text{s.t.} \quad & (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^\dagger) \sim \bar{\mathbb{P}} \\ & \Pi_{\tilde{\boldsymbol{z}}}\bar{\mathbb{P}} = \mathbb{P} \\ & \Pi_{\tilde{\boldsymbol{z}}^\dagger}\bar{\mathbb{P}} = \mathbb{P}^\dagger \\ & \bar{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^\dagger) \in \mathcal{W} \times \mathcal{W}\right] = 1. \end{aligned}$$

Correspondingly, the statistical-distance-based Wasserstein ambiguity set is given by

$$\mathcal{F}_\theta = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{W}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P}, \tilde{\boldsymbol{z}}^\dagger \sim \mathbb{P}^\dagger \\ d_W(\mathbb{P}, \mathbb{P}^\dagger) \leq \theta \end{array} \right\}, \tag{5.8}$$

which can be viewed as the Wasserstein ball of radius θ centered at the empirical probability distribution $\mathbb{P}^\dagger$.

It is well-known that the worst-case expectation taking on a statistical-distance-based Wasserstein ambiguity set can be evaluated through a finite-dimensional optimization problem. For the purpose of this section, we derive a reformation herein that re-expresses the worst-case expectation as a classical robust optimization problem.

Theorem 22. Given a convex and bounded support set $\mathcal{W}$, the worst-case expectation

$$\sup_{\mathbb{P} \in \mathcal{F}_\theta} \mathbb{E}_\mathbb{P}\left[f(\tilde{\boldsymbol{z}})\right] \tag{5.9}$$

is the same as the optimal value of the following robust optimization problem:

$$\begin{aligned} \inf \quad & r\theta + \frac{1}{N}\sum_{n\in[N]} s_n \\ \text{s.t.} \quad & ru + \sum_{n\in[N]} s_n w_n \geq f(\boldsymbol{z}), \quad \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\ & r \in \mathbb{R}_+, \boldsymbol{s} \in \mathbb{R}^N, \end{aligned} \tag{5.10}$$

where

$$\mathcal{D} = \mathrm{cl}\left\{ (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathbb{R}^{I_1} \times \mathbb{R}_+ \times \mathbb{R}^N_+ \;\middle|\; \begin{array}{l} \exists\ \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N \in \mathbb{R}^{I_1}, u_1, \ldots, u_N \in \mathbb{R} : \\ \displaystyle\sum_{n\in[N]} \boldsymbol{z}_n = \boldsymbol{z}, \ \sum_{n\in[N]} u_n = u, \ \sum_{n\in[N]} w_n = 1 \\ \displaystyle\frac{\boldsymbol{z}_n}{w_n} \in \mathcal{W}, \ u_n \geq w_n \rho\left(\frac{\boldsymbol{z}_n}{w_n}, \hat{\boldsymbol{z}}_n\right), \ w_n > 0, \ \forall n \in [N] \end{array} \right\}. \tag{5.11}$$

Proof. By the definition of the Wasserstein distance, the worst-case expectation (5.9) can be re-expressed as follows:

$$\begin{array}{rl} \sup & \mathbb{E}_{\bar{\mathbb{P}}}\left[f(\tilde{\boldsymbol{z}})\right] \\ \text{s.t.} & \mathbb{E}_{\bar{\mathbb{P}}}\left[\rho(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\dagger})\right] \leq \theta \\ & \Pi_{\tilde{\boldsymbol{z}}^{\dagger}}\bar{\mathbb{P}} = \mathbb{P}^{\dagger} \\ & \bar{\mathbb{P}}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\dagger}) \in \mathcal{W} \times \mathcal{W})\right] = 1. \end{array}$$

Using the law of total probability, we can represent the joint probability distribution $\bar{\mathbb{P}}$ of $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{z}}^{\dagger}$ as the marginal probability distribution $\mathbb{P}^{\dagger}$ of $\tilde{\boldsymbol{z}}^{\dagger}$ supported on $\{\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_N\}$ and the conditional probability distributions $\mathbb{P}_n, n \in [N]$ given $\tilde{\boldsymbol{z}}^{\dagger} = \hat{\boldsymbol{z}}_n$. Consequently, we can reformulate the above problem as

$$\begin{array}{rl} \sup & \displaystyle\frac{1}{N}\sum_{n\in[N]} \mathbb{E}_{\mathbb{P}_n}\left[f(\tilde{\boldsymbol{z}})\right] \\ \text{s.t.} & \displaystyle\frac{1}{N}\sum_{n\in[N]} \mathbb{E}_{\mathbb{P}_n}\left[\rho(\tilde{\boldsymbol{z}}, \hat{\boldsymbol{z}}_n)\right] \leq \theta \\ & \mathbb{P}_n\left[\tilde{\boldsymbol{z}} \in \mathcal{W}\right] = 1, \ \forall n \in [N]. \end{array}$$

Under the condition stated in the theorem, this problem is well defined and strong duality holds (see, e.g., Esfahani and Kuhn 2015), and its dual problem is given by

$$\begin{array}{rl} \inf & \displaystyle r\theta + \sum_{n\in[N]} s_n \\ \text{s.t.} & \displaystyle\frac{r}{N}\rho(\boldsymbol{z}, \hat{\boldsymbol{z}}_n) + s_n \geq \frac{1}{N}f(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{W}, \forall n \in [N] \\ & r \in \mathbb{R}_+, \boldsymbol{s} \in \mathbb{R}^N. \end{array}$$

Using the change of variable from $\boldsymbol{s}$ to $\boldsymbol{s}/N$ and introducing the auxiliary variable $\boldsymbol{u} \in \mathbb{R}^N$, the dual becomes:

$$\begin{aligned}
\inf \quad & r\theta + \frac{1}{N}\sum_{n\in[N]} s_n \\
\text{s.t.} \quad & ru_n + s_n - f(\boldsymbol{z}) \geq 0,\ \forall \boldsymbol{z} \in \mathcal{W}, u_n \geq \rho(\boldsymbol{z}, \hat{\boldsymbol{z}}_n), \quad \forall n \in [N] \\
& r \in \mathbb{R}_+, \boldsymbol{s} \in \mathbb{R}^N, \boldsymbol{u} \in \mathbb{R}^N.
\end{aligned} \tag{5.12}$$

Using the observation that,

$$\min_{n\in[N]} a_n \geq 0 \quad \Leftrightarrow \quad a_n \geq 0,\ \forall n \in [N] \quad \Leftrightarrow \quad \min_{\boldsymbol{w}\in\mathcal{S}} \left\{ \sum_{n\in[N]} w_n \alpha_n \right\} \geq 0,$$

where $\mathcal{S} = \{\boldsymbol{w} \in \mathbb{R}_+^N \mid \sum_{n\in[N]} w_n = 1\}$, we can represent the set of robust counterparts in (5.12) as

$$\min_{n\in[N]} \inf_{(\boldsymbol{z},\boldsymbol{u})\in\mathcal{U}_n} \{ru + s - f(\boldsymbol{z})\} \geq 0$$

with $\mathcal{U}_n = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}_+ \mid \boldsymbol{z} \in \mathcal{W}, u_n \geq \rho(\boldsymbol{z}, \hat{\boldsymbol{z}}_n)\right\}, \forall n \in [N]$, which is equivalent to

$$\inf_{(\boldsymbol{z}_n,\boldsymbol{u}_n)\in\mathcal{U}_n, n\in[N]} \min_{n\in[N]} \{ru_n + s_n - f(\boldsymbol{z}_n)\} \geq 0.$$

Implementing the observation once more, we arrive at

$$\inf_{(\boldsymbol{z}_n,\boldsymbol{u}_n)\in\mathcal{U}_n, n\in[N]} \min_{\boldsymbol{w}\in\mathcal{S}} \left\{ r\left(\sum_{n\in[N]} w_n u_n\right) + \sum_{n\in[N]} s_n w_n - \sum_{n\in[N]} w_n f(\boldsymbol{z}_n) \right\} \geq 0. \tag{5.13}$$

Since f is convex, we have $\sum_{n\in[N]} w_n f(\boldsymbol{z}_n) \geq f\left(\sum_{n\in[N]} w_n \boldsymbol{z}_n\right), \forall \boldsymbol{w} \in \mathcal{S}$. Thus the inequality (5.13) implies

$$\inf_{(\boldsymbol{z}_n,\boldsymbol{u}_n)\in\mathcal{U}_n, n\in[N]} \min_{\boldsymbol{w}\in\mathcal{S}} \left\{ r\left(\sum_{n\in[N]} w_n u_n\right) + \sum_{n\in[N]} s_n w_n - f\left(\sum_{n\in[N]} w_n \boldsymbol{z}_n\right) \right\} \geq 0. \tag{5.14}$$

In addition, for any $(\boldsymbol{z}_n, \boldsymbol{u}_n) \in \mathcal{U}_n, n \in [N]$, the optimization problem

$$\min_{\boldsymbol{w} \in \mathcal{S}} \left\{ r\left(\sum_{n \in [N]} w_n u_n \right) + \sum_{n \in [N]} s_n w_n - \sum_{n \in [N]} w_n f(\boldsymbol{z}_n) \right\}$$

is linear in $\boldsymbol{w}$. Therefore, it has an optimal solution $\boldsymbol{w}^*$ that satisfies $w^*_{n_0} = 1$ and $w^*_n = 0, \forall n \in [N] \backslash \{n_0\}$ for some $n_0 \in [N]$. Consequently, there exists an optimal solution $((\boldsymbol{z}^*_n, \boldsymbol{u}^*_n)_{n \in [N]}, \boldsymbol{w}^*)$ to the optimization problem in (5.13) that satisfies $\sum_{n \in [N]} w^*_n f(\boldsymbol{z}^*_n) = f\left(\sum_{n \in [N]} w^*_n \boldsymbol{z}^*_n\right)$. This establishes the equivalence between (5.13) and (5.14).

Using the change of variables from $w_n \boldsymbol{z}_n$ to $\boldsymbol{z}_n$ and $w_n u_n$ to u_n, for all $n \in [N]$ and introducing the variable $\boldsymbol{z}$ and u such that $\boldsymbol{z} = \sum_{n \in [N]} \boldsymbol{z}_n$ and $u = \sum_{n \in [N]} u_n$, respectively, we can rewrite (5.14) as

$$\inf_{(\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D}} \left\{ ru + \sum_{n \in [N]} s_n w_n - f(\boldsymbol{z}) \right\} \geq 0,$$

which reformulates Problem (5.12) to Problem (5.10). We remark that $\boldsymbol{z}$, as a convex combination of $\boldsymbol{z}_n / w_n \in \mathcal{W}$ for $w_n \geq 0$, $n \in [N]$, also lies in the support set, i.e., $\boldsymbol{z} \in \mathcal{W}$. □

We next provide a new representation of the Wasserstein ambiguity set that is based on our proposed ambiguity set as follows:

$$\mathcal{G}_\theta = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R} \times \mathbb{R}^N\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{w}}] = \dfrac{\mathbf{1}}{N} \\ \mathbb{E}_{\mathbb{P}}[\tilde{u}] \leq \theta \\ \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}}) \in \mathcal{D}] = 1 \end{array} \right\} \tag{5.15}$$

where $(u, \boldsymbol{w})$ is the auxiliary random variable and $\mathcal{D}$ is defined as in (5.11). Correspondingly, we investigate the worst-case expectation

$$\sup_{\mathbb{P} \in \mathcal{G}_\theta} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})],$$

which we show to be equal to the worst-case expectation (5.9).

Theorem 23. Given a convex function, $f : \mathbb{R}^{I_1} \mapsto \mathbb{R}$,

$$\sup_{\mathbb{P} \in \mathcal{F}_\theta} \mathbb{E}_\mathbb{P}\left[f(\tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P} \in \mathcal{G}_\theta} \mathbb{E}_\mathbb{P}\left[f(\tilde{\boldsymbol{z}})\right].$$

Proof. We show that $\sup_{\mathbb{P} \in \mathcal{G}_\theta} \mathbb{E}_\mathbb{P}\left[f(\tilde{\boldsymbol{z}})\right]$ equals to the optimal value of Problem (5.10). Indeed, its value equals to the optimal value of its dual program, which is given as follows:

$$\begin{aligned} \inf \quad & t + r\theta + \frac{1}{N} \sum_{n \in [N]} s_n \\ \text{s.t.} \quad & t + ru + \sum_{n \in [N]} s_n w_n \geq f(\boldsymbol{z}), \quad \forall\, (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\ & t \in \mathbb{R}, r \in \mathbb{R}_+, \boldsymbol{s} \in \mathbb{R}^N. \end{aligned} \tag{5.16}$$

The robust counterpart can be represented as

$$ru + \sum_{n \in [N]} (t + s_n) w_n \geq f(\boldsymbol{z}), \; \forall\, (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D}$$

which follows from the fact that any $(\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D}$ satisfies $\sum_{n \in [N]} w_n = 1$. The objective can be re-expressed it as

$$t + r\theta + \frac{1}{N} \sum_{n \in [N]} s_n = \sum_{n \in [N]} \frac{t}{N} + r\theta + \frac{1}{N} \sum_{n \in [N]} s_n = r\theta + \sum_{n \in [N]} \frac{1}{N}(t + s_n).$$

Using the change of variables from $t + s_n$ to s_n for all $n \in [N]$, we then see that Problem (5.16) is exactly Problem (5.10). This completes our proof. $\square$

Speaking intuitively, the auxiliary random variable, $\tilde{\boldsymbol{w}}$ in the Wasserstein ambiguity set corresponds to the reference random variable, $\tilde{\boldsymbol{z}}^\dagger$ such that $\boldsymbol{w} = \boldsymbol{e}_n$ if and only if $\boldsymbol{z}^\dagger = \hat{\boldsymbol{z}}_n$. Consequently, the auxiliary random variable $\tilde{u}$ is related to the epigraphical representation of the distance metric such that $\tilde{u} \geq \rho(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^\dagger)$. The key advantage of representing the Wasserstein ambiguity set as a special case of our proposed ambiguity set is the potential for incorporating other forms of distributional information such as moments and confidence sets. Moreover, based on the recent work of Bertsimas et al. (2017), the representation of the Wasserstein ambiguity set via our proposed ambiguity set could provide a direct approach for improving the approximations of distributionally robust optimization problems with recourse. We will discuss this in the subsequent section.

5.4 Handling large data via confidence sets

The provision of confidence sets with uncertain probabilities enables us to handle empirical probability distributions. However, for a given large collection of data $\{\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_N\}$, it may not be computationally viable to incorporate the entire data set in the optimization model. Apart from specifying the expectation constraints, another practical approach is to partition the support set into J disjoint confidence sets $\mathcal{D}_j,\ j \in [J]$, for which their probabilities $p_j = \mathbb{P}(\tilde{\boldsymbol{z}}_j \in \mathcal{D}_j)$ can be estimated from data, i.e.,

$$q_j = \sum_{n \in [N]} \mathbb{1}(\hat{\boldsymbol{z}}_n \in \mathcal{D}_j),$$

where $\mathbb{1}(\hat{\boldsymbol{z}}_n \in \mathcal{D}_j)$ denotes an indicator function that returns one if $\hat{\boldsymbol{z}}_n \in \mathcal{D}_j$ and zero otherwise. We assume that it is relatively easy to check the feasibility of $\mathcal{D}_j, j \in [J]$. As we will study in the next section, having the disjoint confidence sets allows us to better approximate distributaionlly robust optimization problems with discrete recourse variables as functions of the realization of $\tilde{\boldsymbol{z}}$ in the confidence sets.

As the true probability $\boldsymbol{p}$ is uncertain, we adopt the approach of Ben-Tal et al. (2013) by restricting it to be within the ϕ-divergence (for example, chi-squared, Hellinger, and Kullback-Leibler) ball of radius θ centered at probability $\boldsymbol{q}$ as follows:

$$\mathcal{V}(\theta) = \left\{ \boldsymbol{p} \in \mathbb{R}^J_+ \;\middle|\; \sum_{j \in [J]} p_j = 1, \sum_{j \in [J]} q_j \phi\left(\frac{q_j}{p_j}\right) \leq \theta \right\},$$

where the divergence function $\phi(t)$ is convex for $t \geq 0$. The ϕ-divergence ambiguity set is

$$\mathcal{G}(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) \;\middle|\; \begin{array}{ll} \tilde{\boldsymbol{z}} \sim \mathbb{P} & \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{D}_j\right] = p_j, & \forall j \in [J] \\ \text{for some } \boldsymbol{p} \in \mathcal{V}(\theta). & \end{array} \right\}.$$

Applying Theorem 21 and Corollary 4.2 of Ben-Tal et al. (2013), the worst-case expectation, $\sup_{\mathbb{P} \in \mathcal{G}(\theta)} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right]$ equals to

$$\begin{array}{ll} \inf & s_0 \\ \text{s.t.} & r_j \geq f(\boldsymbol{z}), \;\forall \boldsymbol{z} \in \text{cl}(\text{CH}(\mathcal{D}_j)), \;\forall j \in [J] \\ & s_0 \geq \eta + \theta\lambda + \lambda \displaystyle\sum_{j \in [J]} q_j \phi^* \left(\frac{r_j - \eta}{\lambda}\right) \\ & \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \lambda \in \mathbb{R}_+, \eta \in \mathbb{R}, \end{array}$$

where $\phi^*(s)$ is the conjugate function of the divergence function, $\phi(t)$ and we refer interested readers to Table 2 in Ben-Tal et al. (2013) for some common ϕ-divergence functions and their conjugates.

Constructing disjoint confidence sets

We now discuss how to construct disjoint confidence sets by adopting the "nesting condition" of Wiesemann et al. (2014) to construct a tractable laminar family (Schrijver 2003) of sets.

Definition 9 (tractable laminar family). A tractable laminar family of sets $\mathcal{C}_j, j \in [J]$ satisfies:

1. For any $i, j \in [J]$, either $\mathcal{C}_i \subseteq \mathcal{C}_j$, or $\mathcal{C}_j \subseteq \mathcal{C}_i$, or $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$.

2. For all $j \in [J]$, $\mathcal{C}_j$ is tractable but may not necessarily be convex or closed.

3. For all $j \in [J]$, $\text{cl}(\mathcal{C}_j) = \text{cl}\left(\text{CH}\left(\mathcal{C}_j \setminus \cup_{i \in \mathcal{S}(j)} \mathcal{C}_i\right)\right)$, where $\mathcal{S}(j) = \{i \in [J] \setminus \{j\} \mid \mathcal{C}_i \subseteq \mathcal{C}_j\}$.

Given the tractable laminar family of sets, we can obtain the disjoint confidence sets as follows:

$$\mathcal{D}_j = \mathcal{C}_j \setminus \cup_{i \in \mathcal{S}(j)} \mathcal{C}_i, \ \forall j \in [J].$$

We note that the "nesting condition" of Wiesemann et al. (2014) mandates that two confidence sets should be either disjoint, or otherwise, one set should strictly be in the interior of the other. Although we do not impose this condition, we require the third condition in Definition 9 to ensure that a set would preserve the same convex hull after its "inner" nested sets are removed. We motivate some examples of the tractable laminar family of disjoint confidence sets.

Norm-based laminar family

Given a norm $\|\cdot\|$ and $\theta_j, j \in [J]$ such that $0 = \theta_1 < \theta_2 < \cdots < \theta_J = 1$, we can construct a norm-based laminar family as follows:

$$\mathcal{D}_j = \begin{cases} \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \mid \theta_j < \|\boldsymbol{z}\| < \theta_{j+1} \right\} & \text{if } j \in [J-1] \\ \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \mid \|\boldsymbol{z}\| \leq 1 \right\} \setminus \cup_{i \in [J-1]} \mathcal{D}_i & \text{otherwise.} \end{cases}$$

The norm-based laminar family partitions the sets into regions that depend on the magnitude of the deviation from the nominal position via the norm distance measure. Observe that the last set has zero volume and may be removed if $\tilde{\boldsymbol{z}}$ is almost surely infeasible in the set.

Inspired by the *budgeted uncertainty* of Bertsimas and Sim (2004), we can also construct the following laminar sets that are intuitively associated with the number of uncertain parameters that have deviated from their nominal values as follows:

$$\mathcal{D}_j = \begin{cases} \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \mid \|\boldsymbol{z}\|_\infty < 1, \Gamma_j < \|\boldsymbol{z}\|_1 < \Gamma_{j+1} \right\} & \text{if } j \in [J-1] \\ \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \mid \|\boldsymbol{z}\|_\infty \leq 1 \right\} \setminus \cup_{i \in [J-1]} \mathcal{D}_i & \text{otherwise} \end{cases}$$

for some $\Gamma_j, j \in [J]$ such that $0 = \Gamma_1 < \Gamma_2 < \cdots < \Gamma_J \leq I_1$.

Direction-based laminar family

Observe that the norm-based laminar family is insensitive to the direction of the deviation associated with the random variable, $\tilde{\boldsymbol{z}}$. For instance, under a regular norm, the realizations $\boldsymbol{e}_j/2$ and $-\boldsymbol{e}_j/2$, $j \in [J]$ would be contained in the same disjoint confidence set. To address this issue, we propose the direction-based laminar family of disjoint confidence sets as follows:

$$\mathcal{D}_j^+ = \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \;\middle|\; \begin{array}{l} 0 < z_j < 1 \\ |z_i| < z_j, \ \forall i \in [J] \setminus \{j\} \end{array} \right\},$$

$$\mathcal{D}_j^- = \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \;\middle|\; \begin{array}{l} 0 < -z_j < 1 \\ |z_i| < -z_j, \ \forall i \in [J] \setminus \{j\} \end{array} \right\},$$

and

$$\mathcal{D}^0 = \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \mid \|\boldsymbol{z}\|_\infty \leq 1 \right\} \setminus \cup_{i \in [J-1]} (\mathcal{D}_i^+ \cup \mathcal{D}_i^-).$$

Note that these sets are pairwise disjoint and thus collectively form a laminar family. However, they are not closed and the closures of each of them are not necessarily pairwise disjoint.

We remark that it is also possible to combine the direction-based family with the norm-based family to yield a new laminar family that captures both the magnitude and direction of the deviations.

5.5 Distributionally robust optimization with recourse

Many real-world optimization problems involve recourse decisions that are adaptive to the uncertain outcomes as they unfold in stages. We focus on the two-stage linear optimization problem, the building block for optimization problems with recourse. The *here-and now* decision $\boldsymbol{x} \in \mathcal{X}$ is chosen before the realization of the uncertainty $\tilde{\boldsymbol{z}} \in \mathcal{W}$, and the *wait-and-see* recourse decision $\boldsymbol{y} \in \mathbb{R}^{N_2}$ is determined after the value of $\tilde{\boldsymbol{z}}$ is revealed. Similar to a typical stochastic programming approach, for a given decision, $\boldsymbol{x}$ and a realization of the uncertainty, $\boldsymbol{z}$, we evaluate the recourse objective value through a linear optimization problem as follows:

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{z}) = \quad \min \quad & \boldsymbol{d}'\boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\ & \boldsymbol{y} \in \mathbb{R}^{N_2}. \end{aligned} \tag{5.17}$$

Here, $\boldsymbol{A} : \mathbb{R}^{I_1} \mapsto \mathbb{R}^{M\times N_1}$ and $\boldsymbol{b} : \mathbb{R}^{I_1} \mapsto \mathbb{R}^{M}$ are functions that map from $\boldsymbol{z}$ to the input parameters of the linear optimization problem. We adopt the popular factor-based model such that these functions are affinely dependent on $\boldsymbol{z}$ and are given by,

$$\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{A}^0 + \sum_{i\in[I_1]} \boldsymbol{A}^i z_i, \quad \boldsymbol{b}(\boldsymbol{z}) = \boldsymbol{b}^0 + \sum_{i\in[I_1]} \boldsymbol{b}^i z_i$$

with $\boldsymbol{A}^i \in \mathbb{R}^{M,N_1}, \boldsymbol{b}^i \in \mathbb{R}^M$ for all $i \in [I_1] \cup \{0\}$. The recourse matrix $\boldsymbol{B} \in \mathbb{R}^{M\times N_2}$ and the vector of cost parameters $\boldsymbol{d} \in \mathbb{R}^{N_2}$ are both constant. In the terminology of stochastic programming, if there exists $\boldsymbol{y} \in \mathbb{R}^{N_2}$ such that $\boldsymbol{B}\boldsymbol{y} > \boldsymbol{0}$, then Problem (5.17) is said to have *complete recourse*. Given our proposed ambiguity set, $\mathcal{G}$ in (5.2), the worst-case expected recourse objective value is evaluated by

$$\lambda(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right].$$

Correspondingly, the optimal *here-and-now* decision $\boldsymbol{x}$ is determined by considering the sum of the deterministic first stage objective value, $\boldsymbol{c}'\boldsymbol{x}$ and the worse-case expected recourse objective value, $\lambda(\boldsymbol{x})$. The goal is to obtain the *here-and-now* decision $\boldsymbol{x}$ by solving a tractable deterministic optimization problem. Thereafter, we can obtain the optimal *wait-and-see* solution by solving Problem (5.17).

Under the factor-based model, the recourse objective value $f(\boldsymbol{x}, \boldsymbol{z})$ is convex and piecewise affine in $\boldsymbol{z}$. Applying Theorem 21, we have

$$\begin{aligned}
\lambda(\boldsymbol{x}) = \quad \inf \quad & s_0 \\
\text{s.t.} \quad & r_j + \boldsymbol{s}_1'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq f(\boldsymbol{x}, \boldsymbol{z}), && \forall(\boldsymbol{z}, \boldsymbol{u}) \in \mathrm{cl}(\mathrm{CH}(\mathcal{D}_j)), \forall j \in [J] \\
& s_0 \geq \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2, && \forall(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{p}) \in \mathcal{V} \\
& \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2}.
\end{aligned} \tag{5.18}$$

However, Problem (5.18) is generally intractable because for each $\boldsymbol{x} \in \mathcal{X}$, the recourse variable $\boldsymbol{y}$ for evaluating $f(\boldsymbol{x}, \boldsymbol{z})$ depends on the realizations of $\tilde{\boldsymbol{z}}$, which is usually exponential in its dimension or potentially infinite. Hence, to approach an upper bound of Problem (5.18) via a computationally tractable approximation, it is common to restrict the recourse variable $\boldsymbol{y}$ to a smaller class of affine function maps or the so-called linear decision rules (LDR).

Typically, a decision rule may be interpreted as an implementable "policy" as we can express the recourse variable as a function of the nominal random variable $\tilde{\boldsymbol{z}}$ or a subset of auxiliary random variable $\tilde{\boldsymbol{u}}$ that is associated with the realization of $\tilde{\boldsymbol{z}}$ (see, Bertsimas et al. 2017). However, Bertsimas et al. (2017) argue against using the LDR as a "policy" even in the situations where it could provide the exact worst-case expected recourse objective values because of the inferior *wait-and-see* solutions. Moreover, as in the case of the Wasserstein ambiguity set, the auxiliary random variable $\tilde{\boldsymbol{u}}$ may not naturally be associated with the realization of $\tilde{\boldsymbol{z}}$. Therefore, to avoid being misconstrued as a "decision rule", we use the term *tractable adaptive recourse scheme* (TARS) as a family of functions of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})$ that imposes the structure of the recourse variable, which will enable us to obtain a tractable approximation of Problem (5.18). Specifically, to obtain upper bounds of $\lambda(\boldsymbol{x})$, we propose to consider the following problem based on our proposed TARS:

$$
\begin{array}{rlll}
\bar{\lambda}(\boldsymbol{x}) = & \inf & s_0 & \\
& \text{s.t.} & r_j + \boldsymbol{s}_1'\boldsymbol{z} + \boldsymbol{s}_2'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}), & \forall\,(\boldsymbol{z},\boldsymbol{u}) \in \mathrm{cl}(\mathrm{CH}(\mathcal{D}_j)),\ \forall j \in [J] \\
& & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}), & \forall(\boldsymbol{z},\boldsymbol{u}) \in \mathrm{cl}(\mathrm{CH}(\mathcal{D}_j)),\ \forall j \in [J] \\
& & s_0 \geq \boldsymbol{r}'\boldsymbol{p} + \boldsymbol{s}_1'\boldsymbol{v}_1 + \boldsymbol{s}_2'\boldsymbol{v}_2, & \forall(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{p}) \in \mathcal{V} \\
& & y_n \in \mathcal{T}(\mathcal{Y}^n, \mathcal{I}_0^n, \mathcal{I}_1^n, \mathcal{I}_2^n), & \forall n \in [N_2] \\
& & \boldsymbol{r} \in \mathbb{R}^J, s_0 \in \mathbb{R}, \boldsymbol{s}_1 \in \mathbb{R}^{I_1}, \boldsymbol{s}_2 \in \mathbb{R}^{I_2} &
\end{array}
\tag{5.19}
$$

for some $\mathcal{Y}^n \subseteq \mathbb{R}, \mathcal{I}_0^n \in [J], \mathcal{I}_1^n \subseteq [I_1], \mathcal{I}_2^n \subseteq [I_2], \forall n \in [N_2]$, where the TARS is the family of functions defined as

$$
\begin{aligned}
&\mathcal{T}(\mathcal{Y}, \mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2) \\
&= \left\{ y(\boldsymbol{z},\boldsymbol{u}) : \mathbb{R}^{I_1+I_2} \mapsto \mathbb{R} \;\middle|\; \begin{array}{l} \exists y_j^0, y_i^1, y_k^2 \in \mathcal{Y}, \forall j \in \mathcal{I}_0, \forall i \in \mathcal{I}_1, \forall k \in \mathcal{I}_2 : \\ y(\boldsymbol{z},\boldsymbol{u}) := \displaystyle\sum_{j \in \mathcal{I}_0} y_j^0 \mathbb{1}((\boldsymbol{z},\boldsymbol{u}) \in \mathcal{D}_j) + \sum_{i \in \mathcal{I}_1} y_i^1 z_i + \sum_{k \in \mathcal{I}_2} y_k^2 u_k \end{array} \right\}.
\end{aligned}
$$

Proposition 9. Given $\boldsymbol{x} \in \mathcal{X}$, and $\mathcal{Y}^n \subseteq \mathbb{R}$, $\mathcal{I}_0^n \subseteq [J]$, $\mathcal{I}_1^n \subseteq [I_1]$, and $\mathcal{I}_2^n \subseteq [I_2]$ for all $n \in [N_2]$, we have $\lambda(\boldsymbol{x}) \leq \bar{\lambda}(\boldsymbol{x})$.

Proof. Let $\boldsymbol{y}^*$ be the optimal TARS solution in Problem (5.19). Recall that $f(\boldsymbol{x},\boldsymbol{z}) = \min\{\boldsymbol{d}'\boldsymbol{y} \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z})\}$. For any fixed $\boldsymbol{z}$ of interest, $\boldsymbol{y}^*(\boldsymbol{z},\boldsymbol{u})$ is feasible in $f(\boldsymbol{x},\boldsymbol{z})$. Thus, $\boldsymbol{d}'\boldsymbol{y}^*(\boldsymbol{z},\boldsymbol{u}) \geq f(\boldsymbol{x},\boldsymbol{z})$. Replacing $f(\boldsymbol{x},\boldsymbol{z})$ by $\boldsymbol{d}'\boldsymbol{y}^*(\boldsymbol{z},\boldsymbol{u})$ in Problem (5.18), we obtain a problem that is more restrictive than Problem (5.18) but less restrictive than Problem (5.19). Thus, $\lambda(\boldsymbol{x}) \leq \bar{\lambda}(\boldsymbol{x})$. □

Our proposed TARS incorporates several key features arising from models that have been proposed for addressing optimization under uncertainty, including:

- **Robust and affinely adjustable robust optimization:** In the classical robust optimization approach, there is no constraint on expectation in the ambiguity set. Setting $\mathcal{T}(\mathbb{R}, \{1\}, \emptyset, \emptyset)$, the TARS recovers the classical robust optimization approach. Setting $\mathcal{T}(\mathbb{R}, \{1\}, [I_1], \emptyset)$, the TARS captures the affinely adjustable robust solutions in Ben-Tal et al. (2004). Choosing $\mathcal{T}(\mathbb{R}, \{1\}, [I_1], [I_2])$, our TARS can incorporate the extended affinely adjustable robust solutions in Chen and Zhang (2009).

- **Stochastic programming with discrete probability distributions:** Given a discrete probability distribution $\mathbb{P}^\dagger$ such that $\mathbb{P}^\dagger(\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}_j) = p_j$ for all possible realizations $\hat{\boldsymbol{z}}_j, j \in [J]$, we can consider $\mathcal{D}_j = \{\hat{\boldsymbol{z}}_j\}$ and $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{D}_j) = p_j$ in the description of the ambiguity set, $\mathcal{G}$ in (5.2). Setting $\mathcal{T}(\mathbb{R}, [J], \emptyset, \emptyset)$ in the TARS, we obtain solutions as in stochastic programming approach.

- **Finite adaptability and discrete recourse variables:** We include the confidence set indicator function in the TARS to encompass the key notion of finite adaptability rule proposed by Bertsimas and Caramanis (2010), which would enable us to enforce discrete constraints on the recourse variables. For instance, if a subset of recourse variables $y_n, n \in \mathcal{N} \subseteq [N_2]$ is required to be binary, we can specify the TARS as follows: $y_n \in \mathcal{T}(\{0,1\}, [J], \emptyset, \emptyset), \forall n \in \mathcal{N}$. We note that the TARS differs from finite adaptability because its disjoint confidence sets are fixed a priori, while finite adaptability can further account for determining disjoint confidence sets by the optimization model. It is also possible to adopt the splitting heuristics arising from finite adaptability (see, e.g., Postek and Hertog 2016 and Bertsimas and Dunning 2016) to obtain the disjoint confidence sets.

- **Nonanticipativity constraints:** With different choices of subsets, $\mathcal{I}_0^n \subset [J], \mathcal{I}_1^n \subseteq [I_1], \mathcal{I}_2^n \subseteq [I_2], \forall n \in [N_2]$, we can model the information restriction associated with each recourse variable, y_n, $n \in [N_2]$. Hence, the TARS can take into account of the nonanticipativity constraints arising from multi-stage optimization models.

- **Adaptive distributionally robust optimization of Bertsimas et al. (2017):** Bertsimas et al. (2017) propose to use the LDR approximation for addressing distributionally robust optimization problems with recourse, which is the same as setting $\mathcal{T}(\mathbb{R}, \{1\}, [I_1], [I_2])$ in the TARS. In this case, the random variable $\tilde{\boldsymbol{u}}$ is associated with the conic epigraph of $\tilde{\boldsymbol{z}}$ and hence, the decision rule is directly related to the realization of the nominal random variable, $\tilde{\boldsymbol{z}}$. Bertsimas et al. (2017) show that the inclusion of $\tilde{\boldsymbol{u}}$ in the LDR has the benefits of (i) resolving the infeasibility in complete recourse problems, (ii) obtaining exact solutions for complete recourse problems when $N_2 = 1$, and (iii) improving over more complex piecewise linear decision rules of Chen et al. (2008) and Goh and Sim (2010).

It is also worth mentioning that Zhen et al. (2016) recently propose a new strategy by eliminating a subset of the recourse variables via Fourier-Motzkin elimination (FME), which can obtain optimal solutions when all the recourse variables have been eliminated. Although FME would generally lead to an exponential number of constraints, to improve the sequence of approximations, we can perform partial FME and implement the TARS to the remaining recourse variables.

Case of Wasserstein ambiguity set

Although the worst-case expectation taking on a Wasserstein ambiguity set can be reformulated into a finite-dimensional optimization problem, the corresponding distributionally robust optimization problems with recourse may not necessarily be tractable (see, e.g., Hanasusanto and Kuhn 2016). With the observation in Theorem 23 and noting that the recourse objective value $f(\boldsymbol{x}, \boldsymbol{z})$ is convex and piecewise affine in $\boldsymbol{z}$, we can apply the TARS to obtain tractable reformulations. Using Theorem 22 and implementing the TARS, we can obtain an upper bound of the worst-case expected recourse objective value with the Wasserstein ambiguity set as follows:

$$
\begin{array}{rll}
\bar{\lambda}_\theta(\boldsymbol{x}) = \inf & r\theta + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} s_n & \\
\text{s.t.} & ru + \displaystyle\sum_{n\in[N]} s_n w_n \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z}, u, \boldsymbol{w}), & \forall\, (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, u, \boldsymbol{w}) \geq \boldsymbol{b}(\boldsymbol{z}), & \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\
& y_n \in \mathcal{T}(\mathbb{R}, \{1\}, [I_1], [N+1]), & \forall n \in [N_2] \\
& r \in \mathbb{R}_+, s \in \mathbb{R}^N. &
\end{array}
\tag{5.20}
$$

Interestingly, when $\theta = 0$ and the problem has complete recourse, only the empirical probability distribution would be feasible in the Wasserstein ambiguity set, and we can determine its optimal solution as follows:

$$\begin{aligned} \lambda_0(\boldsymbol{x}) = \quad \min \quad & \frac{1}{N}\sum_{n\in[N]} \boldsymbol{d}'\boldsymbol{y}_n \\ \text{s.t.} \quad & \boldsymbol{A}(\hat{\boldsymbol{z}}_n)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_n \geq \boldsymbol{b}(\hat{\boldsymbol{z}}_n), \quad \forall n \in [N] \\ & \boldsymbol{y}_n \in \mathbb{R}^{N_2}, \qquad\qquad\qquad \forall n \in [N]. \end{aligned} \tag{5.21}$$

Theorem 24. Suppose Problem (5.17) has complete recourse, then $\bar{\lambda}_0(\boldsymbol{x}) = \lambda_0(\boldsymbol{x})$.

Proof. Under complete recourse, there exists $\bar{\boldsymbol{y}}$ such that $\boldsymbol{B}\bar{\boldsymbol{y}} \geq \boldsymbol{1}$. Let $\boldsymbol{y}_n^\dagger, n \in [N]$ be the optimal solutions to Problem (5.21) and consider the following sequence of solutions involving a TARS solution, $\boldsymbol{y}_\kappa^*$:

$$\boldsymbol{y}_\kappa^*(\boldsymbol{z}, u, \boldsymbol{w}) = \kappa u \bar{\boldsymbol{y}} + \sum_{n\in[N]} \boldsymbol{y}_n^\dagger w_n, \; r_\kappa = \kappa \max\{\boldsymbol{d}'\bar{\boldsymbol{y}}, 0\}, \; s_n = \boldsymbol{d}'\boldsymbol{y}_n^\dagger, \; \forall n \in [N] \tag{5.22}$$

as $\kappa \to \infty$. We next show that this solution is feasible in Problem (5.20) as κ approaches to infinity. When $\theta = 0$, we note that Assumption 1 may no longer hold and hence, we may not expect the optimal value of Problem (5.20) to be attainable.

For the first constraint in Problem (5.20), we note that $r_\kappa u \geq \boldsymbol{d}'\bar{\boldsymbol{y}}\kappa u, \forall u \geq 0$, which clearly holds according to (5.22). For the second constraint in Problem (5.20), it is sufficient to show

$$\lim_{\kappa\to\infty}\left([\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_m + \kappa u + \sum_{n\in[N]} \left[\boldsymbol{B}\boldsymbol{y}_n^\dagger\right]_m w_n - [\boldsymbol{b}(\boldsymbol{z})]_m \right) \geq 0, \; \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D}, \; \forall m \in [M], \tag{5.23}$$

where we use $[\cdot]_m$ to denote the m-th component of a vector. Indeed, whenever $u > 0$, the inequality (5.23) holds because

$$[\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_m + \sum_{n\in[N]} \left[\boldsymbol{B}\boldsymbol{y}_n^\dagger\right]_m w_n - [\boldsymbol{b}(\boldsymbol{z})]_m$$

is finite, which follows the fact that it is affine in $(\boldsymbol{z}, \boldsymbol{w})$ and $\Pi_{(\boldsymbol{z},\boldsymbol{w})}\mathcal{D}$ is compact.

Since the distance metric ρ satisfies $\rho(\boldsymbol{z}, \boldsymbol{z}^\dagger) = 0$ if and only if $\boldsymbol{z} = \boldsymbol{z}^\dagger$, it implies that whenever $u = 0$, we have $\boldsymbol{z} = \sum_{n\in[N]} w_n \hat{\boldsymbol{z}}_n$, $\sum_{n\in[N]} w_n = 1, w_n \geq 0, \forall n \in [N]$. Recall that for all $n \in [N]$, we have $\boldsymbol{A}(\hat{\boldsymbol{z}}_n)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_n^\dagger \geq \boldsymbol{b}(\hat{\boldsymbol{z}}_n)$, which implies that the inequality (5.23) still holds whenever $u = 0$.

Hence, as κ approaches to infinity, the feasible solution defined in (5.22) achieves an objective value of $\sum_{n\in[N]} s_n = \sum_{n\in[N]} \boldsymbol{d}'\boldsymbol{y}_n^\dagger$ for Problem (5.20), which coincides with the optimal value of Problem (5.21). This completes our proof. □

Theorem 24 indicates that if θ is small, the TARS would provide good approximations to Wasserstein-based distributionally robust optimization problems with complete recourse.

5.6 Numerical experiments

In our experiment, we consider a multi-item newsvendor problem with I different items. For each item $i, i \in [I]$, its selling price and ordering cost are denoted by p_i and c_i, respectively. Under a total budget of Γ, the manager decides the ordering quantity x_i of each item before the random demand, $\tilde{z}_i$ is observed. Once the demand realizes, the selling quantity of each item is decided as $\min\{x_i, z_i\}$. The manager is ambiguity-averse and maximizes the worst-case expected operating revenue by solving the following distributionally robust optimization problem:

$$
\begin{aligned}
\max \quad & \inf_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[I]} p_i \min\{x_i, \tilde{z}_i\}\right] \\
\text{s.t.} \quad & \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\
& \boldsymbol{x} \in \mathbb{R}^I_+,
\end{aligned}
$$

which, to be consistent with the previous framework, can be recast as a minimization problem,

$$
\begin{aligned}
\min \quad & -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[I]} p_i(x_i - \tilde{z}_i)^+\right] \\
\text{s.t.} \quad & \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\
& \boldsymbol{x} \in \mathbb{R}^I_+.
\end{aligned}
$$

The random demand has support $\mathcal{W} = [\mathbf{0}, \bar{\boldsymbol{z}}]$ and a collection of past observations $\{\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_N\}$ is available. This experimental set-up is similar to Gotoh et al. (2015) and Hanasusanto and Kuhn (2016), where the authors have demonstrated that the statistical-distance-based distributionally robust approach could yield solutions with low out-of-sample variability.

In this section, we focus on demonstrating the modeling power of our proposed ambiguity set and showing that our proposed TARS is easy to implement and can provide high-quality approximation at modest computational effort. Using the Euclidean norm $\|\cdot\|_2$ as the distance metric, we construct the Wasserstein ambiguity set in the form of (5.15) as follows:

$$
\mathcal{G}_\theta = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R} \times \mathbb{R}^N\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{w}}\right] = \dfrac{\mathbf{1}}{N} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{u}\right] \leq \theta \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}}) \in \mathcal{D}\right] = 1 \end{array} \right\},
$$

where the set $\mathcal{D}$ is given by

$$
\mathcal{D} = \left\{ (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathbb{R}^{I_1} \times \mathbb{R}_+ \times \mathbb{R}^{N+1}_+ \;\middle|\; \begin{array}{l} \exists\ \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N \in \mathbb{R}^{I_1}, u_1, \ldots, u_N \in \mathbb{R} : \\ \displaystyle\sum_{n\in[N]} \boldsymbol{z}_n = \boldsymbol{z},\ \sum_{n\in[N]} u_n = u,\ \sum_{n\in[N]} w_n = 1 \\ \boldsymbol{z}_n \in [\mathbf{0}, w_n\bar{\boldsymbol{z}}],\ u_n \geq \|\boldsymbol{z}_n - w_n\hat{\boldsymbol{z}}_n\|_2,\ w_n \geq 0,\ \forall n \in [N] \end{array} \right\}.
$$

Observe that $\sum_{i\in[I]} p_i(x_i - z_i)^+ = \max_{\mathcal{S}\subseteq[I]} \left\{\sum_{j\in[S]} p_j(x_j - z_j)\right\}$, i.e., it can be expressed as a convex and piecewise affine function involving 2^I pieces. Thus, the distributionally robut multi-item newsvendor problem over the Wasserstein ambiguity set, $\mathcal{G}_\theta$ can be exactly solved by considering the following problem,

$$
\begin{aligned}
\min \quad & -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{G}_\theta} \mathbb{E}_{\mathbb{P}}\left[y(\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}})\right] \\
\text{s.t.} \quad & y(\boldsymbol{z}, u, \boldsymbol{w}) \geq \max_{\mathcal{S}\subseteq[I]} \left\{ \sum_{j\in\mathcal{S}} p_j(x_j - z_j) \right\}, \quad \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\
& \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\
& \boldsymbol{x} \in \mathbb{R}^I_+, \\
& y \in \mathcal{T}(\mathbb{R}, \{1\}, [I], [N+1]),
\end{aligned}
\tag{5.24}
$$

where we introduce a recourse variable y and specify it in the form of the TARS. This follows from the useful result that our TARS is optimal for a complete recourse problem with only one recourse variable (see Theorem 4 of Bertsimas et al. 2017). Note that the size of the above exact approach increases exponentially in the number of items. Alternatively, we can use the TARS to obtain an upper bound by solving the following optimization problem,

$$
\begin{aligned}
\min \quad & -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{G}_\theta} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{p}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{u}, \tilde{\boldsymbol{w}})\right] \\
\text{s.t.} \quad & \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\
& \boldsymbol{y}(\boldsymbol{z}, u, \boldsymbol{w}) \geq \boldsymbol{0}, && \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\
& \boldsymbol{y}(\boldsymbol{z}, u, \boldsymbol{w}) \geq \boldsymbol{x} - \boldsymbol{z}, && \forall (\boldsymbol{z}, u, \boldsymbol{w}) \in \mathcal{D} \\
& \boldsymbol{x} \in \mathbb{R}^I_+ \\
& y_i \in \mathcal{T}(\mathbb{R}, \{1\}, [I], \mathcal{I}_2), && \forall i \in [I],
\end{aligned}
\tag{5.25}
$$

where, by varying the set $\mathcal{I}_2$, we can control the dependency of the recourse variables $y_i, i \in [I]$ on a combination of $\boldsymbol{z}$, u, and $\boldsymbol{w}$.

We consider the I-item newsvendor problem with $I = 5$ and 7, and $N = 50$. In each instance of our experiment, we randomly generate the upper bound of the demand, $\bar{\boldsymbol{z}}$ from a uniform distribution on $[0, 100]^I$. Subsequently, the N reference demands are randomly generated from a uniform distribution on $[\mathbf{0}, \bar{\boldsymbol{z}}]$. We set the cost to $\boldsymbol{c} = \mathbf{1}$ and the total budget to $\Gamma = 50I$, and we generate the price, $\boldsymbol{p}$ from a uniform distribution on $[0, 5]^I$.

For different choices of θ, we run 100 instances and compare the performance of the exact solution against different cases of the TARS implementation as follows:

- *Case 1:* $\boldsymbol{y}$ depends on $(\boldsymbol{z}, u, \boldsymbol{w})$.
- *Case 2:* $\boldsymbol{y}$ depends on $(u, \boldsymbol{w})$.
- *Case 3:* $\boldsymbol{y}$ depends on $(\boldsymbol{z}, u)$.
- *Case 4:* $\boldsymbol{y}$ depends on $(\boldsymbol{z}, \boldsymbol{w})$.
- *Case 5:* $\boldsymbol{y}$ depends on $\boldsymbol{z}$.

For each case, we report the relative gap between the TARS objective value and the exact optimal objective value given by $\dfrac{\lambda - \bar{\lambda}}{\lambda} \times 100\%$. The results for $I = 5$ items and $I = 7$ items are summarized in Table 1 and Table 2, respectively. Notably, with the inclusion of auxiliary random variable $(u, \boldsymbol{w})$, the TARS can provide high-quality approximation to the exact solution. Meanwhile, excluding either u or $\boldsymbol{w}$ may lead to more conservative approximation. As shown in Theorem 24, the TARS with recourse variable $\boldsymbol{y}$ being dependable only on $(u, \boldsymbol{w})$ can obtain exact solutions when $\theta = 0$. On the other hand, excluding $\boldsymbol{z}$ would general lead to relatively inferior approximations when θ is not equal to zero.

We evaluate the scalability of both the exact and the TARS approaches, by comparing their computation times in Table 3. For the exact approach, the computer runs out of memory when the number of items exceeds 10. In contrast, we are able to obtain high-quality solution via the TARS with modest computational effort. Note that all numerical experiments are implemented in XProg (see Appendix) and solved using Gurobi 6.0.5 with its default settings on an Intel Core (TM) @ 3.40 GHz with 8GB RAM.

θ	0	1	2	5	10	15	20
Case 1	0% (0%, 0%)	0.1% (0%, 0.3%)	0.1% (0%, 0.6%)	0.1% (0%, 1.2%)	0.3% (0%, 3.2%)	0.5% (0%, 5.4%)	0.6% (0%, 5.3%)
Case 2	0% (0%, 0%)	2.1% (1.1%, 4.8%)	4.3% (1.8%, 11.1%)	10.6% (4.8%, 22.0%)	21.5% (8.1%, 41.3%)	33.4% (16.9%, 97.1%)	41.4% (9.4%, 83.4%)
Case 3	1.5% (0%, 15.4%)	1.9% (0%, 13.0%)	2.1% (0%, 15.2%)	1.7% (0%, 12.5%)	2.3% (0%, 13.9%)	2.7% (0%, 20.3%)	2.6% (0%, 15.7%)
Case 4	2.0% (0%, 19.5%)	2.1% (0%, 13.6%)	2.2% (0%, 15.2%)	1.7% (0%, 12.5%)	2.3% (0%, 13.9%)	2.7% (0%, 20.3%)	2.6% (0%, 15.7%)
Case 5	2.0% (0%, 19.5%)	2.1% (0%, 13.6%)	2.2% (0%, 15.2%)	1.7% (0%, 12.5%)	2.3% (0%, 13.9%)	2.7% (0%, 20.3%)	2.6% (0%, 15.7%)

Tab. 5.1: Average (minimum, maximum) relative gap of TARS for $I = 5$.

θ	0	1	2	5	10	15	20
Case 1	0% (0%, 0%)	0.1% (0%, 0.5%)	0.1% (0%, 0.8%)	0.2% (0%, 1.8%)	0.3% (0%, 3.8%)	0.5% (0%, 3.6%)	0.6% (0%, 6.7%)
Case 2	0% (0%, 0%)	2.6% (1.3%, 7.2%)	4.6% (2.8%, 8.3%)	12.4% (6.7%, 39.2%)	24.0% (12.3%, 43.4%)	37.4% (19.1%, 68.4%)	47.9% (23.6%, 80.6%)
Case 3	0.8% (0%, 9.0%)	1.5% (0%, 10.2%)	1.8% (0%, 10.2%)	1.7% (0%, 12.7%)	1.6% (0%, 13.0%)	1.9% (0%, 13.3%)	1.9% (0%, 14.9%)
Case 4	1.2% (0%, 10.6%)	1.6% (0%, 11.2%)	1.9% (0%, 11.1%)	1.7% (0%, 12.7%)	2.2% (0%, 9.7%)	1.9% (0%, 13.3%)	1.9% (0%, 14.9%)
Case 5	1.2% (0%, 10.6%)	1.6% (0%, 11.2%)	1.9% (0%, 11.1%)	1.7% (0%, 12.7%)	1.6% (0%, 13.0%)	1.9% (0%, 13.3%)	1.9% (0%, 14.9%)

Tab. 5.2: Average (minimum, maximum) relative gap of TARS for $I = 7$.

# item	1	2	3	4	5	6	7	8	9	10	11
Exact approach	0.7s	0.7s	0.9s	1.6s	3.1s	9.5s	15.6s	53.2s	130.1s	366.7s	—
TARS	0.7s	0.7s	0.8s	1.0s	1.0s	1.6s	2.2s	4.1s	5.9s	7.6s	25.5s

Tab. 5.3: Computation time comparison ("—" denotes "out of memory").

5.7 Conclusion

In this section, we propose a new format for tractable ambiguity sets, which could encompass a rich family of statistical distributional information that has been investigated in the literature. By introducing the disjoint confidence sets, the new format can further specify conditional expectation and uncertain probabilities associated with these confidence sets. We also provide a new perspective that shows the Wasserstein ambiguity set has an equivalent reformation in the format of our proposed ambiguity set.

We investigate the distributionally robust optimization problem with recourse and introduce the tractable adaptive recourse scheme (TARS), which enables us to obtain an upper bound of the worst-case expected recourse objective value by solving a tractable deterministic optimization problem. The scheme leverages key features from the classical linear decision rule and finite adaptability, and can be applied in situations with discrete recourse variables. In the numerical experiment, we demonstrate the effectiveness of the TARS in our computational study on a multi-item newsvendor problem.

Code segment

To formulate our models in the numerical experiments, we use XProg, a Matlab-based modeling toolbox that allows us to efficiently specify the distributionally robust optimization problem in an intuitive way. We provide a sample code in XProg for Problems (5.24) and (5.25) and we refer interested readers to `http://xprog.weebly.com` for the users' guide.

Implementing the TARS

```
% I: number of items
% N: number of past observations,
% para_theta: upper bound of distance,
% cost: cost parameters,
% price: price parameters,
% Gamma: total budget,
% zbar: upper bound of demand.
% Zs = (z_1, ..., z_N): past realizations
% define model
TARS = xprog('TARS_Wasserstein');
x = TARS.decision(I);          % define here-and-now decision
y = TARS.recourse(I);          % define recourse decision
z = TARS.random(I);            % define random variable z
u = TARS.random(1);            % define random variable u
w = TARS.random(N);            % define random variable w
Un = TARS.random(N);           % define random variable Un
```

```
Zn = TARS.random(I,N);          % define random variable Zn
y.depend(z);                    % define dependency of y on z
y.depend(u);                    % define dependency of y on u
y.depend(w);                    % define dependency of y on w
% define expectation constraints
TARS.uncertain(expect(u) <= para_theta);
TARS.uncertain(expect(w) == ones(N,1)/N);
% define support set
TARS.uncertain(z' == sum(Zn'));
TARS.uncertain(u == sum(Un));
TARS.uncertain(w >= 0);
TARS.uncertain(sum(w) == 1);
for n = 1:N
TARS.uncertain(Zn(:,n) >= 0);
TARS.uncertain(Zn(:,n) <= ones(I,1)*w(n).*zbar);
TARS.uncertain(Un(n) >= norm(Zn(:,n)-ones(I,1)*w(n).*Zs(:,n))
   );
end
% define objective
TARS.min(-price'*x + expect(price'*y));
% define constraints
TARS.add(cost'*x <= Gamma);
TARS.add(x >= 0);
TARS.add(y >= x - z);
```

```
TARS.add(y >= 0);

TARS.solve; % solve model
```

Implementing the exact approach

```
EXACT = xprog('EXACT_Wasserstein');

x = EXACT.decision(I);          % here-and-now decision

y = EXACT.recourse(1);          % define recourse decision

z = EXACT.random(I);            % define random variable z

u = EXACT.random(1);            % define random variable u

w = EXACT.random(N);            % define random variable w

Un = EXACT.random(N);           % define random variable Un

Zn = EXACT.random(I,N);         % define random variable Zn

y.depend(z);                    % define dependency of y on z

y.depend(u);                    % define dependency of y on u

y.depend(w);                    % define dependency of y on w

% define expectation constraints

EXACT.uncertain(expect(u) <= para_theta);

EXACT.uncertain(expect(w) == ones(N,1)/N);

% define support set

EXACT.uncertain(z' == sum(Zn'));

EXACT.uncertain(u == sum(Un));

EXACT.uncertain(w >= 0);

EXACT.uncertain(sum(w) == 1);

for n = 1:N
```

```
EXACT.uncertain(Zn(:,n) >= 0);
EXACT.uncertain(Zn(:,n) <= ones(I,1)*w(n).*zbar);
EXACT.uncertain(Un(n) >= norm(Zn(:,n)-ones(I,1)*w(n).*Zs(:,n)
   ));
end
% define objective
EXACT.min(-price'*x + expect(y));
% define constraints
EXACT.add(cost'*x <= Gamma);
EXACT.add(x >= 0);
P = double(dec2bin(0:2^I-1,I) == '1'); % Power set matrix
EXACT.add(ones(2^I,1)*y >= P*diag(price)*(x-z));
EXACT.solve; % solve model
```

6. CONCLUSIONS

Moments-based ambiguity sets and statistical-distance-based ambiguity sets have received considerable attentions in the field of distributionally robust optimization. In these types of ambiguity sets, the number of expectation constraints is finite. To the best of our knowledge, no ambiguity set with possibly an infinite number of expectation constraints has been well investigated yet. In this thesis, we motivate and introduce the class of infinitely constrained ambiguity sets, in which the number of expectation constraints can be infinite. We showcase the generality and modeling potential of this class of ambiguity sets. Besides, we study distributionally robust optimization with these infinitely constrained ambiguity sets, in both static settings and adaptive settings. We design solution procedures, respectively, for the resulting optimization problems. Numerical experiments show the efficiency of our solutions procedures, and more importantly, present the advantages of infinitely constrained ambiguity sets.

On the other hand, moments-based ambiguity sets and statistical-distance-based ambiguity sets are usually considered separably in the literature. In this thesis, we study the both the ϕ-divergence ambiguity set and the Wasserstein ambiguity set, which are notable and popular instances in the class of statistical-distance-based ambiguity sets. We show that the Wasserstein ambiguity set has an equivalent reformation via our proposed ambiguity set, which encompasses the moments-based ambiguity sets. Therefore, leveraging such an equivalent reformation, we can incorporate moments information into the

Wasserstein ambiguity set. We believe our result in this regard provides a good perspective to link the moments-based ambiguity sets and statistical-distance-based ambiguity sets. More importantly, the equivalent reformation enables us to handle distributionally robust optimization with recourse in the consideration of the Wasserstein ambiguity set.

We propose a unified and tractable framework for ambiguity set. To address a distributional robust optimization problem with recourse, we introduce the *tractable adaptive recourse scheme* (TARS), which captures key features from several classical models that have been proposed for optimization problems under uncertainty and can be applied in situations where the recourse decisions are discrete.

Certainly, there are several opportunities for future exploration.

- **Finite reduction of the infinitely constrained ambiguity set:** We show that the distributionally robust optimization problem with the covariance dominance ambiguity set admits a finite reduction such that it is equivalent to a distributionally robust optimization problem with a relaxed covariance dominance ambiguity set. This result may be useful in situations where the decision variable $\boldsymbol{x}$ is discrete. It would be interesting to explore whether the finite reduction result exists for a general infinitely constrained ambiguity set.

- **The ELDR approximation for generalized conic ambiguity sets and its sub-optimality:** While we are focusing on tractable conic ambiguity set in this thesis, some ambiguity sets we have excluded would still be 'tractable' at least in practice. For example, at the operational level, some problems do not need to be scalable to a large size. As such, we are still able to consider the ambiguity set with semidefinite expectation constraints. The numerical experiment on multi-item newsvendor problem in Chapter 4, our results show that the ELDR approximation associated with the lifted covariance dominance ambiguity set performs very well and may even provide the exactly optimal solutions in some cases. It is worthwhile to investigate the ELDR approximation for generalized conic ambiguity set and its sub-optimality.

- **Distributionally robust chance constrained models:** Significantly different from the objective maximization (minimization) approach, the satisficing approach provides an alternative perspective on optimization problems that are subject to uncertainty. The chance constrained models, to some extent, lie in the domain of satisficing approach. It would be of great interest to look at distributionally robust chance constrained models (especially joint chance constrained models) with infinitely constrained ambiguity sets.

- **Specialized solver for exponential conic programs:** Exponential conic representable ambiguity sets often lead distributionally robust optimization models to reformulations of exponential conic programs. For small problems, one can make use of the software CVX that implements the successive approximation method to solve a sequence of second-order conic programs. However, for large-scale problems, it is not advisable to attempt this approach (Grant et al. 2009). Ideally, it is best to solve exponential conic programs by an interior-point method that makes use of analytical properties of the exponential cone. Unfortunately, to the best of our knowledge, there is no public release of an efficient solver for exponential conic programs yet. It will be a significant benefit to the field of distributionally robust optimization if a specialized solver for exponential conic programs is published.

- **Algebraic modeling toolbox:** Moving forward, we intend to develop an algebraic modeling toolbox for our proposed framework, which will help us to develop prototypes quickly and to investigate the effectiveness of various distributionally robust optimization models.

ENDNOTES

1. JUMPER homepage: http://jumper.readthedocs.org. See also Dunning et al. (2017).
2. We refer to http://www.meilinzhang.com/software for more details about ROC. See also Bertsimas et al. (2017).
3. ROME homepage: http://robustopt.com. See also Goh and Sim (2011).
4. SDPNAL+ homepage: http://www.math.nus.edu.sg/ mattohkc/SDPNALplus.html.
5. SDPT3 homepage: http://www.math.nus.edu.sg/ mattohkc/sdpt3.html. See also Toh et al. (1999) and Tütüncü et al. (2003).
6. SeDumi homepage: http://sedumi.ie.lehigh.edu/.
7. We refer to http://xprog.weebly.com/ for more information on XProg.
8. YALMIP homepage: http://users.isy.liu.se/johanl/yalmip/. See also Löfberg (2012).
9. Numerical experiments in Chapter 3 are implemented in CVX and solved using SDPT3 with their default settings on an Intel Core (TM) @ 3.40 GHz with 8GB RAM.
10. Numerical experiments in Chapter 4 are implemented in CVX and solved using MOSEK 7.1.0 with their default settings on an Intel Core (TM) @ 3.40 GHz with 8GB RAM.

11. Numerical experiments in Chapter 5 are implemented in XProg and solved using Gurobi 6.0.5 with its default settings on an Intel Core (TM) @ 3.40 GHz with 8GB RAM.

BIBLIOGRAPHY

Anderson, Brian DO, John B Moore. 2007. *Optimal control: linear quadratic methods*. Courier Corporation.

Ben-Tal, Aharon, Dimitris Bertsimas, David B Brown. 2010. A soft robust model for optimization under ambiguity. *Operations research* **58**(4-part-2) 1220–1234.

Ben-Tal, Aharon, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, Gijs Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* **59**(2) 341–357.

Ben-Tal, Aharon, Dick Den Hertog, Jean-Philippe Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming* **149**(1-2) 265–299.

Ben-Tal, Aharon, Boaz Golany, Arkadi Nemirovski, Jean-Philippe Vial. 2005. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management* **7**(3) 248–271.

Ben-Tal, Aharon, Alexander Goryashko, Elana Guslitzer, Arkadi Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2) 351–376.

Ben-Tal, Aharon, Arkadi Nemirovski. 1998. Robust convex optimization. *Mathematics of operations research* **23**(4) 769–805.

Ben-Tal, Aharon, Arkadi Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations research letters* **25**(1) 1–13.

Ben-Tal, Aharon, Arkadi Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming* **88**(3) 411–424.

Ben-Tal, Aharon, Arkadi Nemirovski. 2001. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM.

Bertsimas, Dimitris, David B Brown, Constantine Caramanis. 2011a. Theory and applications of robust optimization. *SIAM review* **53**(3) 464–501.

Bertsimas, Dimitris, Constantine Caramanis. 2010. Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control* **55**(12) 2751–2766.

Bertsimas, Dimitris, Xuan Vinh Doan, Karthik Natarajan, Chung-Piaw Teo. 2010a. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580–602.

Bertsimas, Dimitris, Iain Dunning. 2016. Multistage robust mixed-integer optimization with adaptive partitions. *Operations Research* **64**(4) 980–998.

Bertsimas, Dimitris, Angelos Georghiou. 2015. Design of near optimal decision rules in multi-stage adaptive mixed-integer optimization. *Operations Research* **63**(3) 610–627.

Bertsimas, Dimitris, Angelos Georghiou. 2017. Binary decision rules for multistage adaptive mixed-integer optimization. *Mathematical Programming* 1–39.

Bertsimas, Dimitris, Vineet Goyal. 2012. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical programming* **134**(2) 491–531.

Bertsimas, Dimitris, Vineet Goyal, Xu Andy Sun. 2011b. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research* **36**(1) 24–54.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2013. Data-driven robust optimization. *arXiv preprint arXiv:1401.0212* .

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

Bertsimas, Dimitris, Karthik Natarajan, Chung-Piaw Teo. 2006. Persistence in discrete optimization under data uncertainty. *Mathematical programming* **108**(2) 251–274.

Bertsimas, Dimitris, Ioana Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations research* **52**(1) 35–53.

Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2017. A practically efficient approach for solving adaptive distributionally robust linear optimization problems. URL `http://www.optimization-online.org/DB_FILE/2016/03/5353.pdf`.

Birge, John R, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Breton, Michèle, Saeb El Hachem. 1995. Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications* **4**(4) 317–345.

Brown, David B, Melvyn Sim. 2009. Satisficing measures for analysis of risky positions. *Management Science* **55**(1) 71–84.

Calafiore, Giuseppe Carlo, Laurent El Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* **130**(1) 1–22.

Chares, Robert. 2009. Cones and interior-point algorithms for structured convex optimization involving powers and exponentials.

Chen, Wenqing, Melvyn Sim. 2009. Goal-driven optimization. *Operations Research* **57**(2) 342–357.

Chen, Wenqing, Melvyn Sim, Jie Sun, Chung-Piaw Teo. 2010. From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations research* **58**(2) 470–485.

Chen, Xin, Melvyn Sim, Peng Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Chen, Xin, Melvyn Sim, Peng Sun, Jiawei Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

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

Cheng, Jianqiang, Erick Delage, Abdel Lisser. 2014. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization* **24**(3) 1485–1506.

Conn, Andrew R, Nicholas IM Gould, Philippe L Toint. 2000. *Trust region methods*. SIAM.

Delage, Erick, Shie Mannor. 2010. Percentile optimization for markov decision processes with parameter uncertainty. *Operations research* **58**(1) 203–213.

Delage, Erick, Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* **58**(3) 595–612.

Dentcheva, Darinka, Andrzej Ruszczynski. 2003. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* **14**(2) 548–566.

Dentcheva, Darinka, Andrzej Ruszczyński. 2006. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance* **30**(2) 433–451.

Dunning, Iain, Joey Huchette, Miles Lubin. 2017. Jump: A modeling language for mathematical optimization. *SIAM Review* **59**(2) 295–320.

Dupačová, Jitka. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes* **20**(1) 73–88.

El Ghaoui, Laurent, Hervé Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications* **18**(4) 1035–1064.

El Ghaoui, Laurent, Maksim Oks, Francois Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research* **51**(4) 543–556.

El Ghaoui, Laurent, Francois Oustry, Hervé Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization* **9**(1) 33–52.

Esfahani, Peyman Mohajerin, Daniel Kuhn. 2015. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116* .

Gao, Rui, Anton J Kleywegt. 2016. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199* .

Garstka, Stanley J, Roger J-B Wets. 1974. On decision rules in stochastic programming. *Mathematical Programming* **7**(1) 117–143.

Georghiou, Angelos, Wolfram Wiesemann, Daniel Kuhn. 2015. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* **152**(1-2) 301–338.

Gilboa, Itzhak, David Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of mathematical economics* **18**(2) 141–153.

Goh, Joel, Melvyn Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations research* **58**(4-part-1) 902–917.

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

Gotoh, Jun-ya, Michael Jong Kim, Andrew Lim. 2015. Robust empirical optimization is almost the same as mean-variance optimization .

Gounaris, Chrysanthos E, Wolfram Wiesemann, Christodoulos A Floudas. 2013. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research* **61**(3) 677–693.

Grant, Michael, Stephen Boyd, Yinyu Ye. 2008. Cvx: Matlab software for disciplined convex programming.

Grant, Michael, Stephen Boyd, Yinyu Ye. 2009. cvx users guide. *Technical report, Technical Report Build 711, Citeseer* .

Grant, Michael C, Stephen P Boyd. 2008. Graph implementations for nonsmooth convex programs. *Recent advances in learning and control*. Springer, 95–110.

Graves, Stephen C. 1999. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* **1**(1) 50–61.

Hadley, George, Thomson M Whitin. 1963. *Analysis of inventory systems*. Prentice Hall.

Hanasusanto, Grani A, Daniel Kuhn. 2016. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *arXiv preprint arXiv:1609.07505* .

Hanasusanto, Grani A, Daniel Kuhn, Stein W Wallace, Steve Zymler. 2015a. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* **152**(1-2) 1–32.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2015b. K-adaptability in two-stage robust binary programming. *Operations Research* **63**(4) 877–891.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2016a. A comment on computational complexity of stochastic programming problems. *Mathematical Programming* **159**(1-2) 557–569.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2016b. K-adaptability in two-stage distributionally robust binary programming. *Operations Research Letters* **44**(1) 6–11.

Hanasusanto, Grani A, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2015c. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151**(1) 35–62.

He, Long, Zhenyu Hu, Meilin Zhang. 2017. Robust repositioning for vehicle sharing. URL `https://ssrn.com/abstract=2973739orhttp://dx.doi.org/10.2139/ssrn.2973739.`

Hu, Zhaolin, L Jeff Hong. 2013. Kullback-leibler divergence constrained distributionally robust optimization. *Available at Optimization Online* .

Iancu, Dan A, Mayank Sharma, Maxim Sviridenko. 2013. Supermodularity and affine policies in dynamic robust optimization. *Operations Research* **61**(4) 941–956.

Isii, Keiiti. 1962. On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics* **14**(1) 185–197.

Jiang, Ruiwei, Yongpei Guan. 2016. Data-driven chance constrained stochastic program. *Mathematical Programming* **158**(1-2) 291–327.

Kaas, Rob, Marc Goovaerts, Jan Dhaene, Michel Denuit. 2008. *Modern actuarial risk theory: using R*, vol. 128. Springer Science & Business Media.

Kim, Michael Jong, Andrew EB Lim. 2015. Robust multiarmed bandit problems. *Management Science* **62**(1) 264–285.

Kong, Qingxia, Chung-Yee Lee, Chung-Piaw Teo, Zhichao Zheng. 2013. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations research* **61**(3) 711–726.

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

Li, Xiaobo, Karthik Natarajan, Chung-Piaw Teo, Zhichao Zheng. 2014. Distributionally robust mixed integer linear programs: Persistency models with applications. *European Journal of Operational Research* **233**(3) 459–473.

Löfberg, Johan. 2012. Automatic robust convex programming. *Optimization methods and software* **27**(1) 115–129.

Mak, Ho-Yin, Ying Rong, Jiawei Zhang. 2014. Appointment scheduling with limited distributional information. *Management Science* **61**(2) 316–334.

Mathworks. 2017. Unconstrained nonlinear optimization algorithms. URL `www.mathworks.com/help/optim/ug/unconstrained-nonlinear-optimization-algorithms.html.`

Mehrotra, Sanjay, Dávid Papp. 2014. A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization* **24**(4) 1670–1697.

Natarajan, Karthik, Dessislava Pachamanova, Melvyn Sim. 2009. Constructing risk measures from uncertainty sets. *Operations research* **57**(5) 1129–1141.

Natarajan, Karthik, Chung-Piaw Teo. 2017. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming* **161**(1-2) 487–518.

Natarajan, Karthik, Chung Piaw Teo, Zhichao Zheng. 2011. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations research* **59**(3) 713–728.

Nemirovski, Arkadi, Alexander Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

Popescu, Ioana. 2005. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research* **30**(3) 632–657.

Postek, Krzysztof, Dick den Hertog, Bertrand Melenberg. 2016. Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review* **58**(4) 603–650.

Postek, Krzysztof, Dick den Hertog. 2016. Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing* **28**(3) 553–574.

Qi, Jin. 2016. Mitigating delays and unfairness in appointment systems. *Management Science* .

Rockafellar, R Tyrrell, Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of banking & finance* **26**(7) 1443–1471.

Ruszczynski, Andrzej P, Alexander Shapiro. 2003. *Stochastic programming*, vol. 10. Elsevier Amsterdam.

Scarf, Herbert. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* .

Schrijver, Alexander. 2003. *Combinatorial optimization: polyhedra and efficiency*, vol. 24. Springer Science & Business Media.

See, Chuen-Teck, Melvyn Sim. 2010. Robust approximation to multiperiod inventory management. *Operations research* **58**(3) 583–594.

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shapiro, Alexander, Anton Kleywegt. 2002. Minimax analysis of stochastic problems. *Optimization Methods and Software* **17**(3) 523–542.

Shapiro, Alexander, Arkadi Nemirovski. 2005. On complexity of stochastic programming problems. *Continuous optimization*. Springer, 111–146.

Sion, Maurice, et al. 1958. On general minimax theorems. *Pacific J. Math* **8**(1) 171–176.

Skajaa, Anders, Yinyu Ye. 2015. A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. *Mathematical Programming* **150**(2) 391–422.

Soyster, Allen L. 1973. Technical noteconvex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* **21**(5) 1154–1157.

Toh, Kim-Chuan, Michael J Todd, Reha H Tütüncü. 1999. Sdpt3a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software* **11**(1-4) 545–581.

Tütüncü, Reha H, Kim-Chuan Toh, Michael J Todd. 2003. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical programming* **95**(2) 189–217.

Von Neumann, John, Oskar Morgenstern. 1947. *Theory of games and economic behavior*. Princeton university press.

Wainwright, JM. 2015. High-dimensional statistics: A non-asymptotic viewpoint. *preparation. University of California, Berkeley* .

Wiesemann, Wolfram, Daniel Kuhn, Melvyn Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.

Xu, Huan, Shie Mannor. 2012. Distributionally robust markov decision processes. *Mathematics of Operations Research* **37**(2) 288–300.

Zhao, C, Y Guan. 2015. Data-driven risk-averse stochastic optimization with wasserstein metric. *Available on Optimization Online* .

Zhen, Jianzhe, Dick Den Hertog, Melvyn Sim. 2016. Adjustable robust optimization via fourier-motzkin elimination. URL http://www.optimization-online.org/DB_HTML/2016/07/5564.html.

Zymler, Steve, Daniel Kuhn, Berç Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* 1–32.