Distributionally Robust Optimization with General Uncertainty Structure

Dimitris Bertsimas
Sloan School of Management, Massachusetts Institute of Technology, dbertsim@mit.edu

Judith Brugman*
Department of Econometrics and Operations Research, Tilburg University, j.m.brugman@uvt.nl

Dick den Hertog
Amsterdam Business School, University of Amsterdam, d.denhertog@uva.nl

Johan S.H. van Leeuwaarden
Department of Econometrics and Operations Research, Tilburg University, j.s.h.vanleeuwaarden@uvt.nl

Abstract

We develop an exact solution framework for a broad class of Distributionally Robust Optimization (DRO) problems with general uncertainty structure. Within the class of moment- and confidence-set-based ambiguity sets, existing exact methods are largely limited to max-of-affine functions under ambiguity sets with strictly nested confidence sets. To enlarge this scope while preserving tractability, we introduce an alternative ambiguity set based on linearly defined confidence sets that allows for weak nestedness. We then consider DRO problems whose functions are convex in the decision variables and satisfy one of the following three cases with respect to the uncertainty: (i) when the function is convex under the nested general ambiguity set, we develop a global optimization algorithm; (ii) when the function is generally nonlinear, we show that the same algorithm applies under the alternative ambiguity set; and (iii) when the function is concave, we derive an explicit conic reformulation under this alternative ambiguity set. All three cases are handled by first reformulating the DRO problem as a robust problem and then applying advanced techniques from robust optimization. By solving these problems to optimality, our framework can offer valuable insights into the conservatism and behavior of existing approximation-based DRO models. We illustrate the generality and practical relevance of the proposed framework through two applications in capital budgeting and appointment scheduling.

* Corresponding author

1 Introduction

Distributionally robust optimization (DRO) has emerged as a widely studied paradigm for decision-making under distributional uncertainty, offering a compromise between classical stochastic and worst-case robust optimization. In DRO, rather than assuming a single known probability distribution for the uncertainty, the decision-maker explicitly accounts for ambiguity in the distributional model by optimizing performance against the worst-case probability distribution within a prescribed ambiguity set. This ambiguity set incorporates partial information about the true distribution, such as moment constraints, support, or statistical distance measures, and ensures that the chosen solution remains near-optimal even under adversarial shifts in the underlying distribution.

The conceptual origins of DRO trace back to the seminal work of Scarf (1957), who studied a minimax solution to the classical newsvendor problem using only mean and variance information. After several decades of largely problem-specific research, more systematic DRO frameworks emerged, including tractable reformulations for certain linear and convex problems under moment-based and confidence-set-based ambiguity sets (e.g., Delage and Ye (2010) and Wiesemann et al. (2014)). More recently, alternative ambiguity sets have become prominent in the literature, most notably Wasserstein- and ϕ-divergence-based DRO models, for which exact formulations are available (Ben-Tal et al., 2013; Hu and Hong, 2013; Mohajerin Esfahani and Kuhn, 2018; Gao and Kleywegt, 2023). Comprehensive overviews of foundational models, methodological developments, and applications can be found in Rahimian and Mehrotra (2019), Lin et al. (2022), and Kuhn et al. (2025).

Moment- and confidence-set-based ambiguity sets are especially natural when uncertainty is described through partial structural information rather than through a large empirical sample, for example via known moments, support bounds, or probability constraints on relevant regions. Such settings arise in small-sample regimes, in expert-driven applications where only coarse distributional summaries are available, and in regulatory environments where moment restrictions must be satisfied explicitly. In these cases, a Wasserstein ball around an empirical distribution may be less transparent or less directly aligned with the available information.

However, despite recent advances, solving DRO problems to optimality remains notoriously challenging within this class of ambiguity sets. A prominent exact result is due to Wiesemann et al. (2014), hereafter WKS. They show that if the ambiguity set is defined via conic representable confidence regions which are either strictly nested or disjoint, and moment constraints with means restricted to affine manifolds, the DRO problem can be reformulated as an equivalent robust optimization problem. Under the additional assumption that the cost function can be expressed as the pointwise maximum of finitely many affine functions (the so-called "max-of-affine" structure), the worst-case expected cost admits an exact conic dual reformulation, yielding a tractable conic optimization problem. This max-of-affine structure holds in several classic models, such as the newsvendor problem, and marked a major advance in the understanding of DRO tractability.

Nevertheless, outside the max-of-affine setting, exact solution methods for nonlinear DRO problems under moment- and confidence-set-based ambiguity sets remain very limited. Many practically relevant DRO models involve nonlinear dependence on the uncertainty and therefore fall outside the scope of existing exact reformulation techniques. As a result, such problems cannot be solved to global optimality and the literature has largely relied on approximation-based approaches, such as moment-based relaxations (Delage and Ye, 2010; Goh and Sim, 2010; Zymler et al., 2013) and scenario or sample-average approximation (SAA) approaches (Shapiro, 2017; Bertsimas et al., 2018). Consequently, the accuracy of these approximation methods is difficult to assess, since the true optimal value is unknown for general nonlinear DRO problems.

In this paper, we address these limitations by extending the framework of WKS and developing an exact solution approach that substantially broadens the class of nonlinear DRO problems that can be solved to optimality. Specifically, we recover tractability for function classes beyond the max-of-affine case by reformulating the worst-case expectation as an robust optimization (RO) problem and leveraging advanced techniques from the RO literature.

RO provides a complementary paradigm to decision-making under uncertainty that does not rely on probabilistic assumptions, but enforces feasibility for all realizations of the uncertain parameters within a prescribed uncertainty set. Following early work by Soyster (1973), modern tractable RO theory was developed in Ben-Tal and Nemirovski (1999) and Ben-Tal et al. (2009). For nonlinear dependence on the uncertainty, Ben-Tal et al. (2015) derived exact tractable counterparts for constraints that are convex in the decision variables and concave in the uncertain parameters, while Bertsimas et al. (2024) introduced the Reformulation-Perspectification Technique with Branch-and-Bound (RPT-BB) as a general framework for globally solving nonlinear robust constraints when closed-form reformulations are not available. A comprehensive overview of RO theory and applications is given in the recent monograph by Bertsimas and den Hertog (2022).

Building on these advances in RO and DRO, we combine duality-based reformulations with the RPT-BB framework to solve a substantially broader class of nonlinear DRO problems to global optimality within the class of moment- and confidence-set-based ambiguity sets. Following the construction of WKS, we refer to their ambiguity set as nested general (NG). We also introduce an alternative nested linear (NL) ambiguity set, based on confidence sets defined by linear inequalities and allowing for weak nestedness. This not only restores tractability for a broader class of nonlinear DRO problems, but also accommodates naturally arising confidence sets that may share parts of their boundary, as is common when uncertainty is described through threshold, budget, or one-sided quantile-type constraints. Moreover, the DRO problems we study involve functions that are convex in the decision variables, and whose dependence on the random variables falls into one of the following three cases:

1. **Convex in uncertainty under the NG ambiguity set.** In this case, we rewrite the DRO problem as an RO problem following WKS. This problem is then solved using an iterative cutting-set algorithm that implements RPT-BB, which guarantees global optimality.

2. **General nonlinear in uncertainty under the NL ambiguity set.** In this case, we show that the DRO problem can again be reformulated as an RO problem, now with multiple robust constraints, and solved by the same global solution approach as in the convex case.

3. **Concave in uncertainty under the NL ambiguity set.** Here, we first identify a special case in which the worst-case distribution is degenerate. More generally, we leverage results on concave robust constraints to derive an exact conic reformulation of the DRO problem.

Function class in z	**Ambiguity set**	**Method**	**Reference**
Max-of-affine	NG	Conic reformulation	WKS
Convex	NG	Cutting-set algorithm	Section 3
General nonlinear	NL	Cutting-set algorithm	Section 4
Concave	NL	Conic reformulation	Section 5

Table 1: Overview of exact solution approaches for DRO problems.

These additional function classes cover many nonlinear DRO applications that were previously beyond reach for exact solution methods. Our framework therefore provides direct insight into the effect of distributional ambiguity without relying solely on approximations or conservative relaxations. Moreover, for selected nonlinear DRO instances, the exact solutions obtained here can serve as illustrative benchmarks for approximation-based methods. We demonstrate the applicability of our framework on two classical problems: the Appointment Scheduling Problem (ASP) with uncertain service times, and the Capital Budgeting Problem (CBP) with uncertain discount factors. These two case studies illustrate how our methods apply across the different function classes and ambiguity sets considered in this work, and show that exact solutions can be obtained in settings where prior DRO techniques typically rely on approximations.

This paper makes the following key contributions:

- **Exact DRO framework for a broad class of nonlinear functions:** We develop an exact solution framework for DRO problems under moment- and confidence-set-based ambiguity sets that goes beyond the classical max-of-affine assumption. In particular, our framework applies to nonlinear functions that are convex in the decision variables and covers a broad range of dependence on the uncertainty. It combines RO reformulations with global solution algorithms to recover exact solutions for the DRO problems. We illustrate this generality using the appointment scheduling problem, whose cost function is convex in the uncertain service times, and the capital budgeting problem, whose negative present-value objective is concave in the uncertain discount rates.

- **Alternative ambiguity set recovering tractability:** We introduce an alternative ambiguity set that replaces strict nestedness by weak nestedness under the additional assumption that the confidence sets are linearly defined. This both restores tractability for many nonlinear DRO problems and accommodates more natural linearly defined confidence sets. In the numerical experiments, this set allows us to model probabilistic budget constraints and component-wise event constraints.

The remainder of this paper is organized as follows. Section 2 formalizes the DRO problem, presents its reformulation as an RO problem, and introduces the alternative NL ambiguity set. Section 3 develops the cutting-set algorithm for DRO problems with convex dependence on the uncertainty. Section 4 extends this algorithm to DRO problems with uncertainty dependence that is generally nonlinear. Section 5 presents exact robust reformulations for DRO problems with concave dependence on the uncertainty. Section 6 presents numerical experiments for the appointment scheduling and capital budgeting problems. Finally, Section 7 concludes the paper and outlines directions for future research.

2 Distributionally robust framework

In DRO, we consider optimization under uncertainty in the probability distribution of the random parameters. Without loss of generality, we assume that the objective function is linear in the decision variables and the problem contains a single DRO constraint; extensions to multiple constraints are straightforward. We thus consider the generic DRO problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} && \text{(1a)} \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\big[v(\mathbf{x},\mathbf{z})\big] \le \tau, && \text{(1b)} \\ & \mathbf{x} \in \mathcal{X}. && \text{(1c)} \end{aligned}$$

The main computational challenge is the DRO constraint (1b), which is the focus of the reformulations and solution methods developed in this paper.

Here, the constraint function $v : \mathbb{R}^{N+P} \to (-\infty, +\infty]$ is proper and closed, and depends on the decision variables $\mathbf{x} \in \mathbb{R}^N$ and the random variables $\mathbf{z} \in \mathbb{R}^P$. Throughout the paper, we assume that $v(\mathbf{x}, \mathbf{z})$ is convex in the decision vector $\mathbf{x}$. With respect to the random vector $\mathbf{z}$, we distinguish three structural cases which lead to different reformulations and solution methods: convex, generally nonlinear, and concave, treated in Sections 3, 4, and 5, respectively.

The feasible set $\mathcal{X}$ is assumed to be compact and convex, and is given by $\mathcal{X} = \big\{\mathbf{x} : \mathbf{D}\mathbf{x} \le \mathbf{d},\ h_q(\mathbf{x}) \le 0\ \forall q \in \mathcal{Q}\big\}$, where $\mathcal{Q} = \{1, \dots, Q\}$, $\mathbf{D} \in \mathbb{R}^{M\times N}$, $\mathbf{d} \in \mathbb{R}^M$, and the functions $h_q : \mathbb{R}^N \mapsto (-\infty, +\infty]$ are proper, closed and convex for all $q \in \mathcal{Q}$. The random vector $\mathbf{z}$ follows a probability distribution $\mathbb{P}$ that belongs to the ambiguity set $\mathcal{P}$. The DRO constraint therefore requires that the expected value of $v(\mathbf{x}, \mathbf{z})$ does not exceed the bound $\tau \in \mathbb{R}$ for all distributions in $\mathcal{P}$, and in particular under the worst-case distribution.

Remark: The proposed framework also extends to adaptive DRO problems with a two-stage structure and affine decision rules. In such settings, the first-stage decision vector $\mathbf{x}$ is selected prior to the realization of the uncertainty, while a second-stage decision $\mathbf{y}(\mathbf{z})$ adapts affinely to $\mathbf{z}$, that is, $\mathbf{y}(\mathbf{z}) = \mathbf{D}\mathbf{z} + \mathbf{d}$. The resulting cost function $v(\mathbf{x}, \mathbf{z})$ can still be expressed as a function of $\mathbf{x}$ and $\mathbf{z}$, and remains within the function classes considered in this work. Consequently, all results developed in this paper, including the modeling assumptions, reformulations, and solution methods, extend directly to this adaptive setting.

2.1 Nested ambiguity sets

We formally introduce two types of nested ambiguity sets: the nested general (NG) ambiguity set of WKS, based on strictly nested confidence sets, and the nested linear (NL) ambiguity set, which imposes additional linear structure and allows for weak nestedness.

Nested General (NG): In the first part of this paper, we consider the NG ambiguity set $\mathcal{P}_{\text{NG}}$ introduced by WKS. In their original formulation, the ambiguity set is defined over a lifted uncertainty vector consisting of the original uncertain parameters and auxiliary variables used to encode additional distributional information. Here, we use a single random vector $\mathbf{z} \in \mathbb{R}^P$ to represent all uncertain quantities, including such auxiliary components when present. This is purely a notational simplification and does not alter the ambiguity set or its expressive power. The ambiguity set for $\mathbf{z}$ is then defined as

$$\mathcal{P}_{\text{NG}} = \left\{ \mathbb{P} : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\mathbf{z}] = \mathbf{b},\ \mathbb{P}[\mathbf{z} \in \mathcal{C}_i] \in [\underline{p}_i, \overline{p}_i]\ \forall i \in \mathcal{I} \right\}, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{K \times P}$, $\mathbf{b} \in \mathbb{R}^K$, and $\mathcal{I} = \{1, \ldots, I\}$. The confidence sets $\mathcal{C}_i \subseteq \mathbb{R}^P$ admit a conic representation of the form

$$\mathcal{C}_i = \{\mathbf{z} : \mathbf{C}_i \mathbf{z} \succeq_{\mathcal{K}_i} \mathbf{c}_i\}, \quad \forall i \in \mathcal{I},$$

where $\mathbf{C}_i \in \mathbb{R}^{L_i \times P}$, $\mathbf{c}_i \in \mathbb{R}^{L_i}$, and $\mathcal{K}_i$ is a proper cone. The parameters $\underline{p}_i, \overline{p}_i \in [0, 1]$ satisfy $\underline{p}_i \le \overline{p}_i$ for all $i \in \mathcal{I}$. We impose regularity conditions on $\mathcal{P}_{\text{NG}}$, referred to as set (A):

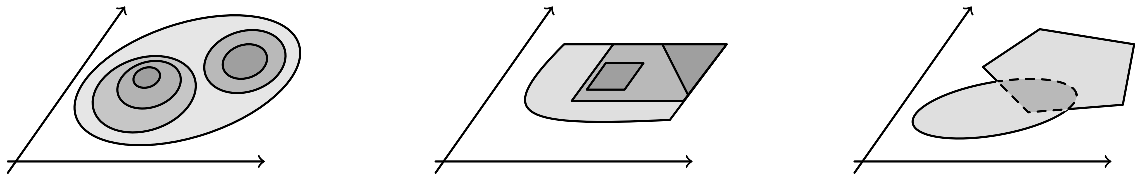
(A1) The confidence set $\mathcal{C}_I$ is bounded and has probability one, that is, $\underline{p}_I = \overline{p}_I = 1$.

(A2) There exists a distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P}[\mathbf{z} \in \mathcal{C}_i] \in (\underline{p}_i, \overline{p}_i)$ whenever $\underline{p}_i < \overline{p}_i$, $\forall i \in \mathcal{I}$.

(A3) (Strict nestedness) For all $i \neq j$, either $\mathcal{C}_i \subseteq \text{int}(\mathcal{C}_j)$, $\mathcal{C}_j \subseteq \text{int}(\mathcal{C}_i)$, or $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$.

The strict nestedness condition is illustrated in Figure 1a. It shows that the confidence sets $\mathcal{C}_i$ form a hierarchy of nested or disjoint sets: for any two sets $\mathcal{C}_i$ and $\mathcal{C}_j$, either one is strictly contained in the other or the sets are disjoint.

This formulation of the NG ambiguity set encompasses many widely used moment-based ambiguity sets, including those incorporating information on the mean, variance (and higher-order moments), coefficient of variation, and robust statistics such as the median or mean absolute deviation; see Section 3 in WKS and Appendix A.



(a) Strictly nested general (NG) (b) Weakly nested linear (NL) (c) Non-nested (not covered)

Figure 1: Visualization of confidence sets $\mathcal{C}_i$ (based on Figure 1 in WKS).

Nested Linear (NL): In the second part of this paper, we consider the NL ambiguity set, denoted by $\mathcal{P}_{\mathrm{NL}}$. This set builds upon the NG construction, but imposes additional linear structure on the confidence sets while allowing for a relaxation of strict nestedness. Specifically, for $i \neq I$, the confidence sets $\mathcal{C}_i$ are obtained by intersecting the support $\mathcal{C}_I$ with a finite system of linear inequalities. The corresponding ambiguity set is

$$\mathcal{P}_{\mathrm{NL}} = \Big\{ \mathbb{P} : \mathbb{E}_{\mathbb{P}}[\mathbf{A}\mathbf{z}] = \mathbf{b},\ \mathbb{P}[\mathbf{z} \in \mathcal{C}_i] \in [\underline{p}_i, \overline{p}_i]\ \forall i \in \mathcal{I} \setminus \{I\},\ \mathbb{P}[\mathbf{z} \in \mathcal{C}_I] = 1 \Big\}, \tag{3}$$

with

$$\mathcal{C}_i = \Big\{ \mathbf{z} \in \mathcal{C}_I : \mathbf{G}_i \mathbf{z} \leq \mathbf{g}_i \Big\}, \qquad i \neq I,$$

for some $\mathbf{G}_i \in \mathbb{R}^{M_i \times P}$ and $\mathbf{g}_i \in \mathbb{R}^{M_i}$. Importantly, the support $\mathcal{C}_I$ is allowed to be an arbitrary convex set. Ambiguity sets defined solely through a single confidence set $\mathcal{C}_I$ are therefore naturally included as a special case. As a result, the NL ambiguity set remains highly expressive while enabling tractable results.

With respect to the regularity conditions, the NL ambiguity set allows for a slight relaxation of the nestedness condition: we replace Assumption (A3) by the following weak nestedness condition, which does not require strict containment in the sense of interior inclusion:

(A3*) (Weak nestedness) For all $i \neq j$, either $\mathcal{C}_i \subset \mathcal{C}_j$, $\mathcal{C}_j \subset \mathcal{C}_i$, or $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$.

We refer to this set of conditions as (A*). Figure 1b illustrates an example of the linearly defined confidence sets $\mathcal{C}_i$ underlying the NL ambiguity set. It also visualizes the weak nestedness condition, under which nested confidence sets may share parts of their boundary. WKS explicitly notes that such configurations cannot be handled within their framework, whereas they are accommodated by our NL ambiguity set. By comparison, the example in Figure 1c does not satisfy either the strict or weak nestedness, and cannot be modelled by any of the two ambiguity sets.

Weak nestedness is not merely a limiting approximation of strict nestedness, but gives rise to a genuinely different modeling class that captures many ambiguity sets of practical interest. In many applications, confidence sets are generated by intersecting a common support set with families of linear inequalities, for example through threshold, budget, or one-sided confidence constraints, as well as quantile-type or conditional moment restrictions. Such sets are often nested in the weak sense while sharing parts of their boundary. Hence, the NL ambiguity set is not, in general, a subset of the NG ambiguity set, but instead constitutes an alternative construction. Specific instances of these sets used in our numerical experiments are given in Section 6.

2.2 Reformulation as RO constraint

We now derive an equivalent RO representation of the DRO constraint (1b). The key ingredient is the (strict or weak) nestedness of the confidence sets, which allows us to decompose the support into disjoint regions. To this end, we define the index sets $\mathcal{A}(i) = \{\, i \,\} \cup \{\, i' \in \mathcal{I} : \mathcal{C}_i \subset \mathcal{C}_{i'} \,\}$ and $\mathcal{D}(i) = \{\, i' \in \mathcal{I} : \mathcal{C}_{i'} \subset \mathcal{C}_i \,\}$ which collect, respectively, all supersets and subsets of $\mathcal{C}_i$, $\forall i \in \mathcal{I}$. In particular, for any $k \in \mathcal{A}(i)$ and $j \in \mathcal{D}(i)$, we have $\mathcal{C}_j \subset \mathcal{C}_k$. The nested structure ensures that these collections are well defined and induces a partition of the support into disjoint regions

$$\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'}, \quad i = 1, \ldots, I, \tag{4}$$

so that each realization $\mathbf{z} \in \mathcal{C}_I$ belongs to exactly one region $\bar{\mathcal{C}}_i$.

Following the duality-based reformulation in the proof of Theorem 1 in WKS, we express the worst-case expectation as the dual of a moment problem. The resulting robust constraint involves indicator terms associated with the confidence sets; however, the partition induced by nestedness enables us to eliminate these indicators by decomposing the constraint over the regions $\bar{\mathcal{C}}_i$. This yields a finite collection of robust constraints, each corresponding to one region. For completeness, the full proof, restated from WKS, is provided in Appendix B.

Theorem 1. *Let the NG or NL ambiguity set be given and assume the corresponding regularity conditions (A) or (A*), respectively. Then, the DRO constraint* (1b) *is satisfied if and only if there exist $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}^I_+$ such that*

$$\mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \, \kappa_i - \underline{p}_i \, \lambda_i \right) \;\le\; \tau, \tag{5}$$

$$v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} \left(\kappa_{i'} - \lambda_{i'} \right) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \;\le\; 0, \quad \forall \mathbf{z} \in \bar{\mathcal{C}}_i, \; \forall i \in \mathcal{I}. \tag{6}$$

This reformulation results in an RO problem with one or more robust constraints. From this point onward, when discussing the RO reformulations, we refer to $\mathbf{z}$ as the uncertain parameters, while keeping in mind that it represents a realization of the random vector in the original DRO model. The tractability of this problem is determined by the structural properties of the function $v(\mathbf{x}, \mathbf{z})$, in combination with the assumptions on the confidence sets $\mathcal{C}_i$.

In WKS, tractability of the robust reformulation is ensured by imposing that $v(\mathbf{x}, \mathbf{z})$ can be expressed as the pointwise maximum of finitely many affine functions. This assumption plays a crucial technical role in their analysis. Due to the implied convexity of v and the strict nestedness of the sets $\mathcal{C}_i$ in the NG ambiguity set, the robust constraints defined over the disjoint regions $\bar{\mathcal{C}}_i$ can be equivalently written over the sets $\mathcal{C}_i$. The max-of-affine structure then allows each resulting robust constraint to be decomposed into a finite collection of linear inequalities (one per affine component of v), which can be reformulated using standard robust optimization tools and conic duality. While this structure yields strong tractability results, it is also restrictive and excludes many functions of practical interest; see Section 1.

In contrast, we do not rely on the max-of-affine assumption to ensure tractability of the robust reformulation. Instead, we study DRO constraint (1b) under different structural assumptions on the dependence of $v(\mathbf{x}, \mathbf{z})$ on the uncertain parameters $\mathbf{z}$. We first consider convex dependence, followed by more general nonlinear dependence, and finally concave dependence, which allows for exact reformulations. These three cases are addressed in Sections 3, 4, and 5, respectively. Table 1 gives an overview of the function classes, ambiguity sets and solution methods.

3 Convex dependence on uncertainty

This section assumes that the function $v(\mathbf{x}, \mathbf{z})$ is convex in the uncertain parameters $\mathbf{z}$ for fixed variables $\mathbf{x}$ and admits the representation

$$v(\mathbf{x}, \mathbf{z}) = h(\mathbf{T}(\mathbf{x})\mathbf{z} + \mathbf{t}(\mathbf{x})), \tag{7}$$

with $h : \mathbb{R}^N \mapsto (-\infty, +\infty]$ proper, closed and convex, and with $\mathbf{T}(\mathbf{x}) \in \mathbb{R}^{S\times P}$ and $\mathbf{t}(\mathbf{x}) \in \mathbb{R}^S$ affine in $\mathbf{x}$. Under these assumptions, v is convex in $\mathbf{x}$ as well in $\mathbf{z}$. For examples that fall within this class, see Bertsimas et al. (2023a), including quadratic optimization, piecewise linear constraints, sum-of-max-of-linear constraints, and geometric optimization.

3.1 RO reformulation

Due to the convexity of v, we can replace the disjoint sets $\bar{\mathcal{C}}_i$ in (6) by the original sets $\mathcal{C}_i$ (which is also the next step in the proof of Theorem 1 in WKS). Intuitively, convexity implies that the worst-case value of the left-hand side of the robust constraint over a nested set is attained on its boundary, and under strict nestedness the relevant boundaries of $\mathcal{C}_i$ and $\bar{\mathcal{C}}_i$ coincide. To make the structure suitable for the cutting-set algorithm, we then apply a biconjugate reformulation (Bertsimas et al., 2023a), which introduces auxiliary uncertain variables and yields an equivalent robust constraint that is affine in the decision variables, bilinear in the uncertain parameters, and posed over a convex uncertainty set. The full proof is provided in Appendix B.

Theorem 2. *Let the NG ambiguity set be given and assume that the regularity conditions (A) hold, and assume v follows* (7)*. Then, the DRO constraint* (1b) *is satisfied if and only if there exist $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}^I_+$ such that* (5) *holds and*

$$\mathbf{w}^\top \mathbf{T}(\mathbf{x})\mathbf{z} + \mathbf{w}^\top \mathbf{t}(\mathbf{x}) - w_0 - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall (w_0, \mathbf{w}, \mathbf{z}) \in \Lambda_i, \ \forall i \in \mathcal{I}, \tag{8}$$

where $\Lambda_i = \{(w_0, \mathbf{w}, \mathbf{z}) \mid \mathbf{z} \in \mathcal{C}_i,\ \mathbf{w} \in \operatorname{dom}(h^*),\ h^*(\mathbf{w}) \le w_0\}$*, and* h^* *denotes the convex conjugate of* h*, defined as* $h^*(\mathbf{w}) = \sup_{\mathbf{y}}\{\mathbf{w}^\top \mathbf{y} - h(\mathbf{y})\}$*.*

By combining Theorem 2 with the DRO problem in (1), we obtain the following equivalent robust optimization problem:

$$\begin{aligned}
\min_{\mathbf{x},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\lambda}} \quad & \mathbf{c}^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\kappa} \ge \mathbf{0}, \quad \boldsymbol{\lambda} \ge \mathbf{0}, \\
& \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i\, \kappa_i - \underline{p}_i\, \lambda_i \right) \le \tau, \\
& \mathbf{w}^\top \mathbf{T}(\mathbf{x})\mathbf{z} + \mathbf{w}^\top \mathbf{t}(\mathbf{x}) - w_0 - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \le 0,\ \forall (w_0, \mathbf{w}, \mathbf{z}) \in \Lambda_i,\ \forall i \in \mathcal{I}.
\end{aligned} \tag{9}$$

3.2 Cutting-set solution method

To solve (9), we adopt the framework of Bertsimas et al. (2024), a cutting-set algorithm that incorporates RPT-BB. Under mild additional regularity conditions on the moment conditions and probability bounds, ensuring that the extended feasible region admits a bounded reformulation, the required assumptions of this framework are satisfied: the objective function is proper, closed and convex, the constraint functions are linear or bilinear, the uncertainty sets are nonempty and convex, and the feasible region is convex; see Appendix C. The algorithm iteratively refines scenario sets that approximate the robust constraints by alternating between a master problem and subproblem evaluations. When applied to (9), the algorithm proceeds as follows:

1. *Initialization*: Initialize the scenario sets $\mathcal{S}_i$ for each $i \in \mathcal{I}$ (Section 3.2.2).

2. *Master problem*: Solve the master problem (Section 3.2.2) using the current scenario sets, yielding a candidate solution $(\mathbf{x}^*, \boldsymbol{\beta}^*, \boldsymbol{\kappa}^*, \boldsymbol{\lambda}^*)$.

3. *Subproblem*: For each $i \in \mathcal{I}$, solve the subproblem (Section 3.2.1) via RPT-BB to obtain a worst-case scenario $(w_0^*, \mathbf{w}^*, \mathbf{z}^*)$. If the optimal value is positive, add $\mathbf{z}^*$ to $\mathcal{S}_i$.

4. *Iteration*: Repeat Steps 2–3 until no subproblem yields a violating scenario. The current solution $\mathbf{x}^*$ is then globally optimal.

Under the additional assumption that $v(\mathbf{x}, \mathbf{z})$ is Lipschitz continuous in $\mathbf{x}$ for each $\mathbf{z}$, convergence of the algorithm to the global optimum of (9) follows from Theorem 3.3 of Bertsimas et al. (2024), building on the cutting-set results of Mutapcic and Boyd (2009).

3.2.1 Solving the subproblem with RPT-BB

For each constraint $i \in \mathcal{I}$, and given a fixed solution $(\mathbf{x}^*, \boldsymbol{\beta}^*, \boldsymbol{\kappa}^*, \boldsymbol{\lambda}^*)$ from the master problem, we solve the following subproblem to identify the worst-case scenario $(w_0, \mathbf{w}, \mathbf{z}) \in \Lambda_i$ that most violates the last constraint in (9):

$$\max_{(w_0,\mathbf{w},\mathbf{z})\in\Lambda_i} \quad \mathbf{w}^\top \mathbf{T}(\mathbf{x}^*)\mathbf{z} + \mathbf{w}^\top \mathbf{t}(\mathbf{x}^*) - w_0 - \sum_{i'\in\mathcal{A}(i)} (\kappa_{i'}^* - \lambda_{i'}^*) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta}^*. \tag{10}$$

If the optimal objective value of this subproblem is strictly positive (larger than a numerical tolerance of 10^{-6}), the corresponding solution $\mathbf{z}^*$ is added to the scenario set $\mathcal{S}_i$, thereby refining the master problem in the next iteration. Otherwise, the robust constraint indexed by i is satisfied. To solve this generally nonconvex maximization problem, we apply the RPT-BB framework through the following steps (further details can be found in Bertsimas et al. (2024)).

First, we apply RPT to the subproblem (10). Introducing the lifted matrix variable $\mathbf{V} = \mathbf{w}\mathbf{z}^\top$ to linearize the bilinear terms, we obtain

$$\max_{(w_0,\mathbf{w},\mathbf{z},\mathbf{V})\in\Theta_i} \quad \mathrm{Tr}\Big(\mathbf{T}(\mathbf{x}^*)^\top \mathbf{V}\Big) + \mathbf{w}^\top \mathbf{t}(\mathbf{x}^*) - w_0 - \sum_{i'\in\mathcal{A}(i)} (\kappa_{i'}^* - \lambda_{i'}^*) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta}^*, \tag{11}$$

where $\Theta_i = \big\{(w_0,\mathbf{w},\mathbf{z},\mathbf{V}) \mid (w_0,\mathbf{w},\mathbf{z}) \in \Lambda_i,\ \mathbf{V} = \mathbf{w}\mathbf{z}^\top\big\}$. The constraint $\mathbf{V} = \mathbf{w}\mathbf{z}^\top$ renders the problem nonconvex. Following the RPT-BB framework, we replace Θ_i by a convex outer approximation $\hat{\Theta}_i$, obtained by generating valid inequalities through multiplications of constraints in Λ_i and substituting the resulting product terms with $\mathbf{V}$ (Bertsimas et al., 2023b; Bertsimas et al., 2026). Solving (11) over $\hat{\Theta}_i$ yields a convex relaxation of the original subproblem and therefore provides an upper bound on the optimal value of (10).

Second, a corresponding lower bound is obtained by evaluating the original constraint violation in (6) on a set of candidate scenarios derived from the solution of the RPT relaxation. Specifically, we define the scenario set

$$\mathcal{V} = \left\{\mathbf{z}^*, \frac{\mathbf{V}_1^*}{w_1^*}, \ldots, \frac{\mathbf{V}_S^*}{w_S^*}\right\}, \tag{12}$$

where $\mathbf{V}_s^*$ denotes the s-th row of $\mathbf{V}^*$ and w_s^* is the corresponding component of $\mathbf{w}^*$. When $w_s^* = 0$, the ratio is omitted, and note $\mathcal{V} \neq \emptyset$. These candidates are heuristically motivated by the rank-one relation $\mathbf{V} = \mathbf{w}\mathbf{z}^\top$ satisfied in the original problem. Optionally, local improvement procedures such as hill climbing may be applied to further refine the scenarios. The maximum value of the original violation function over $\mathcal{V}$ provides a valid lower bound for (10).

Lastly, in case the two bounds coincide within a prescribed tolerance ϵ, the subproblem is solved to global optimality. Otherwise, a branch-and-bound process is initiated to close the optimality gap. At each iteration, a branching hyperplane is introduced to partition the feasible region. The branching direction is determined using the eigenvector $\mathbf{f}$ corresponding to the largest eigenvalue of the matrix $(\mathbf{V}^* - \mathbf{w}^*(\mathbf{z}^*)^\top)^\top(\mathbf{V}^* - \mathbf{w}^*(\mathbf{z}^*)^\top)$, which measures the deviation from the rank-one structure. This yields the branching decision $\mathbf{f}^\top\mathbf{z} \le l$ and $\mathbf{f}^\top\mathbf{z} \ge l$, where $l = \mathbf{f}^\top\mathbf{z}^*$. The subproblem is then re-solved on each branch by repeating the bounding procedure. At each iteration, the active node with the highest upper bound is selected for further exploration.

3.2.2 Initializing and solving the master problem

We initialize the scenario set $\mathcal{S}_i$ for each $i \in \mathcal{I}$ by solving an RPT-based relaxation of (9). Specifically, each robust constraint is replaced by its RPT relaxation, introducing the lifted variable $\mathbf{V} = \mathbf{w}\mathbf{z}^\top$ and replacing the uncertainty set by a convex outer approximation $\hat{\Theta}_i$. This yields the constraint

$$\operatorname{Tr}\Big(\mathbf{T}(\mathbf{x})^\top \mathbf{V}\Big) + \mathbf{w}^\top \mathbf{t}(\mathbf{x}) - w_0 - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \quad \forall (w_0, \mathbf{w}, \mathbf{z}, \mathbf{V}) \in \hat{\Theta}_i.$$

The resulting relaxation is linear with a convex uncertainty set and can be solved efficiently, for instance via support function evaluations (Ben-Tal et al., 2015). Given the solution $(\mathbf{x}^*, \boldsymbol{\beta}^*, \boldsymbol{\kappa}^*, \boldsymbol{\lambda}^*)$, we solve the subproblem (10) for each $i \in \mathcal{I}$ to obtain candidate worst-case scenarios $(w_0^*, \mathbf{w}^*, \mathbf{z}^*, \mathbf{V}^*)$. These scenarios are then used to initialize the sets $\mathcal{S}_i$ according to (12).

Given scenario sets $\mathcal{S}_i$, $i \in \mathcal{I}$, the following master problem is solved at each iteration to obtain an optimal solution $(\mathbf{x}^*, \boldsymbol{\beta}^*, \boldsymbol{\kappa}^*, \boldsymbol{\lambda}^*)$, which is subsequently used in the subproblem:

$$\begin{aligned}
\min_{\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\lambda}} \quad & \mathbf{c}^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\kappa} \geq 0, \quad \boldsymbol{\lambda} \geq 0, \\
& \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\overline{p}_i \kappa_i - \underline{p}_i \lambda_i] \leq \tau, \\
& v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall \mathbf{z} \in \mathcal{S}_i, \ \forall i \in \mathcal{I}.
\end{aligned} \tag{13}$$

4 General nonlinear dependence on uncertainty

This section assumes that the function $v(\mathbf{x}, \mathbf{z})$ is convex in the variables $\mathbf{x}$. Moreover, for each fixed $\mathbf{x}$, we assume that $v(\mathbf{x}, \mathbf{z})$ is sum-of-linear times concave (SLC) in the uncertainty $\mathbf{z}$, meaning that it can be written as

$$v(\mathbf{x}, \mathbf{z}) = v_0(\mathbf{x}, \mathbf{z}) + \sum_{l \in \mathcal{L}} (q_l(\mathbf{x}) - \mathbf{r}_l(\mathbf{x})^\top \mathbf{z}) v_l(\mathbf{x}, \mathbf{z}), \tag{14}$$

where each component function $v_l(\mathbf{x}, \mathbf{z}) : \mathbb{R}^{N+P} \to (-\infty, +\infty]$ is concave in $\mathbf{z}$ and finite-valued on an open set containing $\mathcal{C}_I$, for every fixed $\mathbf{x}$, and for all $l \in \mathcal{L} = \{0, \ldots, L\}$. Additionally, for each term where $v_l(\mathbf{x}, \mathbf{z})$ is nonlinear in $\mathbf{z}$, we impose $q_l(\mathbf{x}) - \mathbf{r}_l(\mathbf{x})^\top \mathbf{z} \geq 0$ for all $\mathbf{z} \in \mathcal{C}_I$. The SLC class is expressive and includes, as special cases, max-of-affine and robust convex constraints after suitable reformulations; see, e.g., examples in Bertsimas et al. (2023b) and Bertsimas et al. (2024).

Remark: In above works, the SLC representation is stated in a form where the linear terms are independent of $\mathbf{x}$ (i.e., q_l and $\mathbf{r}_l$ are constants). In contrast, we only require an SLC representation in the uncertainty for each fixed $\mathbf{x}$, allowing $q_l(\mathbf{x})$ and $\mathbf{r}_l(\mathbf{x})$ to depend on the decision vector.

This is less restrictive for modeling. For example, the quadratic form $v(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top \mathbf{T}(\mathbf{z})^\top \mathbf{T}(\mathbf{z})\mathbf{x}$, where $\mathbf{T}(\mathbf{z})$ is affine in $\mathbf{z}$, is permitted under our definition. Importantly, this generalized definition remains fully compatible with the cutting-set framework combined with RPT-BB.

4.1 Decomposition of confidence sets

We consider the NL ambiguity set as defined in (3). While $\mathcal{C}_I$ is allowed to be any convex set, the confidence sets $\mathcal{C}_i$ are defined by linear inequalities for $i \neq I$, i.e.,

$$\mathcal{C}_i = \mathcal{C}_I \cap P_i, \quad P_i = \{\mathbf{z} : \mathbf{G}_i \mathbf{z} \leq \mathbf{g}_i\} = \bigcap_{r=1}^{R_i} \{\mathbf{z} : \mathbf{G}_{ir}^\top \mathbf{z} \leq g_{ir}\}, \quad \forall i \neq I,$$

where $\mathbf{G}_i \in \mathbb{R}^{R_i \times P}$, $\mathbf{g}_i \in \mathbb{R}^{R_i}$, $\mathbf{G}_{ir}^\top$ denotes the r-th row of $\mathbf{G}_i$, and g_{ir} the r-th component of $\mathbf{g}_i$. Hence, each set $\mathcal{C}_i$ is obtained by intersecting the support set $\mathcal{C}_I$ with a polyhedral set.

To exploit the nested structure of the confidence sets, we introduce for each $i \in \mathcal{I}$ the set of children (i.e., maximal proper subsets) of $\mathcal{C}_i$:

$$\mathcal{G}(i) = \Big\{k \in \mathcal{I} : \mathcal{C}_k \subset \mathcal{C}_i, \ \nexists l \in \mathcal{I} \text{ such that } \mathcal{C}_k \subset \mathcal{C}_l \subset \mathcal{C}_i\Big\}.$$

In Figure 1b, the outermost support set $\mathcal{C}_I$ has one child, namely the large square contained in it. This set in turn has two children, the two smaller sets, which themselves have no children.

Using this definition, the disjoint sets $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{k \in \mathcal{D}(i)} \mathcal{C}_k$ that appear in the robust constraint (6) can equivalently be written as $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{k \in \mathcal{G}(i)} \mathcal{C}_k$. These sets are in general not convex. However, due to the linear structure of the confidence sets, each region $\bar{\mathcal{C}}_i$ can be decomposed into a finite union of convex subsets $\bar{\mathcal{C}}_{ij}$, which are convex but generally not disjoint. The following lemma formalizes this decomposition; its proof is given in Appendix B.

Lemma 1. *Assume the NL ambiguity set* (3) *and weak nestedness (A3*). Then each set $\bar{\mathcal{C}}_i$ can be written as a finite union of (not necessarily disjoint) convex sets $\bar{\mathcal{C}}_{ij}$, each obtained as the intersection of $\mathcal{C}_i$ with a polyhedron Q_{ij} induced by the linear boundaries of the sets $\mathcal{C}_k$, $k \in \mathcal{G}(i)$. That is,*

$$\bar{\mathcal{C}}_i = \bigcup_{j \in \mathcal{J}_i} \bar{\mathcal{C}}_{ij}, \qquad \bar{\mathcal{C}}_{ij} := \mathcal{C}_i \cap Q_{ij} = \mathcal{C}_i \cap \bigcap_{k \in \mathcal{G}(i)} \{\mathbf{z} : \mathbf{G}_{kr_k}^\top \mathbf{z} > g_{kr_k}\}.$$

In particular, the number of such convex sets satisfies $|\mathcal{J}_i| \leq \prod_{k \in \mathcal{G}(i)} R_k$.

The lemma shows that $\bar{\mathcal{C}}_i$ can be decomposed into convex regions by selecting, for each child $k \in \mathcal{G}(i)$, exactly one violated inequality from the defining system of $\mathcal{C}_k$ (which explains the use of strict inequalities in the definition of Q_{ij}). Each such choice induces a region obtained by intersecting the corresponding halfspaces with $\mathcal{C}_i$. Since each child k contributes R_k possible inequalities, the total number of regions is at most $\prod_{k \in \mathcal{G}(i)} R_k$. In practice, this number may be smaller, as some regions can be empty or redundant. This construction is illustrated in Figure 2.

In practice, it is desirable to keep the number of regions $\bar{\mathcal{C}}_{ij}$ as small as possible, since each region induces a separate robust constraint in the reformulation (see Section 4.2). Redundant regions or regions that can be merged may therefore be eliminated using standard techniques from convex optimization (Boyd and Vandenberghe, 2004). Alternatively, the regions can be generated lazily: starting from a small subset of regions, the corresponding master problem is solved and additional regions are introduced only when their associated robust constraints are violated. Such a region-generation scheme may substantially reduce the computational burden in practice.

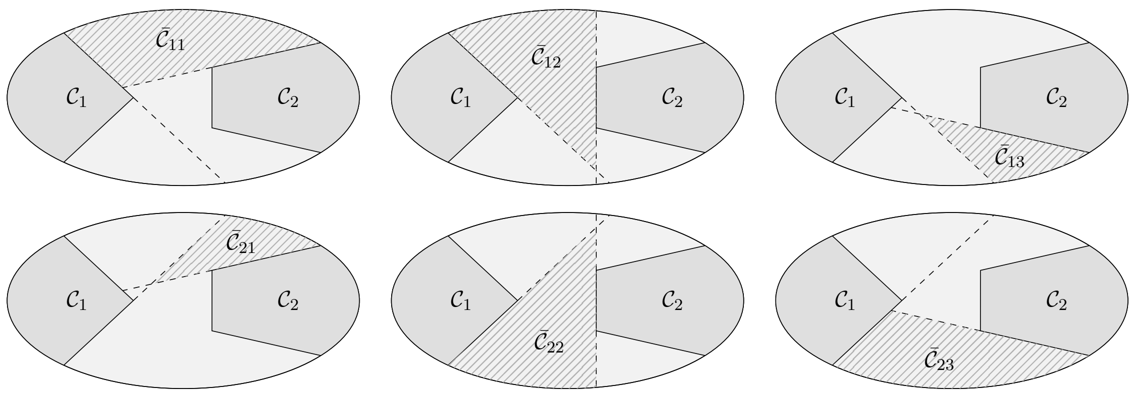


Figure 2: Example of the decomposition induced by two child sets with $R_1 = 2$ and $R_2 = 3$, resulting in six candidate regions, of which two are redundant.

4.2 RO reformulation

Since we construct $\bar{\mathcal{C}}_i = \bigcup_{j \in \mathcal{J}_i} \bar{\mathcal{C}}_{ij}$, the i-th robust constraint (6) can be reformulated by imposing the constraint on each region $\bar{\mathcal{C}}_{ij}$ separately. Moreover, this can be replaced by its closure

$$\tilde{\mathcal{C}}_{ij} := \mathrm{cl}(\bar{\mathcal{C}}_{ij}),$$

without changing feasibility, because for fixed $\mathbf{x}$ the constraint function is continuous in $\mathbf{z}$. The set $\tilde{\mathcal{C}}_{ij}$ is tractable, as it is convex and obtained by intersecting the support with finitely many weak linear inequalities. Using closures is also required by the cutting-set method, which works with closed uncertainty sets. The full proof is given in Appendix B.

Theorem 3. *Let the NL ambiguity set be given and assume that the regularity conditions (A*) hold, and suppose that v is SLC in the uncertainty $\mathbf{z}$. Then, the DRO constraint* (1b) *is satisfied if and only if there exist $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}^I_+$ such that* (5) *holds and*

$$v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \le 0, \quad \forall \mathbf{z} \in \tilde{\mathcal{C}}_{ij},\ \forall i \in \mathcal{I},\ j \in \mathcal{J}_i. \tag{15}$$

Theorem 3 provides an equivalent reformulation of the DRO constraint for SLC functions. Substituting this result into the DRO problem (1) yields the following robust optimization problem:

$$
\begin{aligned}
\min_{\mathbf{x},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\lambda}} \quad & \mathbf{c}^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\kappa} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \\
& \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \, \kappa_i - \underline{p}_i \, \lambda_i \right) \leq \tau, \\
& v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall \mathbf{z} \in \tilde{\mathcal{C}}_{ij}, \ \forall i \in \mathcal{I}, \ j \in \mathcal{J}_i.
\end{aligned} \tag{16}
$$

4.3 Cutting-set solution method

We solve (16) using the same cutting-set framework as introduced in Section 3.2. This algorithm, based on the work of Bertsimas et al. (2024), is originally developed for optimization problems with SLC constraints; the convex case discussed previously can be viewed as a special case. The required assumptions are satisfied here under mild regularity conditions; see Appendix C. Convergence follows under the same Lipschitz continuity assumption on $v(\mathbf{x}, \mathbf{z})$.

The master problem remains identical to (13), and its initialization proceeds as before by generating scenario sets $\mathcal{S}_{ij}$ for all $i \in \mathcal{I}$ and $j \in \mathcal{J}_i$, in the same way as the sets $\mathcal{S}_i$ in Section 3.2. The only difference lies in the structure of the subproblem. For fixed $(\mathbf{x}^*, \boldsymbol{\beta}^*, \boldsymbol{\kappa}^*, \boldsymbol{\lambda}^*)$ and for each region $\bar{\mathcal{C}}_{ij}$, the subproblem becomes

$$
\max_{\mathbf{z} \in \tilde{\mathcal{C}}_{ij}} \ v_0(\mathbf{x}^*, \mathbf{z}) + \sum_{l \in \mathcal{L}} (q_l(\mathbf{x}^*) - \mathbf{r}_l(\mathbf{x}^*)^\top \mathbf{z}) v_l(\mathbf{x}^*, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} (\kappa^*_{i'} - \lambda^*_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta}^*.
$$

Following Bertsimas et al. (2024), we apply a perspective reformulation to the SLC terms and introduce the lifted variable $\mathbf{Z} = \mathbf{z}\mathbf{z}^\top$. This yields the lifted formulation

$$
\max_{(\mathbf{z},\mathbf{Z}) \in \Theta_{ij}} \ v_0(\mathbf{x}^*, \mathbf{z}) + \sum_{l \in \mathcal{L}} (q_l(\mathbf{x}^*) - \mathbf{r}_l(\mathbf{x}^*)^\top \mathbf{z}) v_l \left(\mathbf{x}^*, \frac{q_l(\mathbf{x}^*)\mathbf{z} - \mathbf{Z}\mathbf{r}_l(\mathbf{x}^*)}{q_l(\mathbf{x}^*) - \mathbf{r}_l(\mathbf{x}^*)^\top \mathbf{z}} \right) - \sum_{i' \in \mathcal{A}(i)} (\kappa^*_{i'} - \lambda^*_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta}^*,
$$

where $\Theta_{ij} = \{(\mathbf{z}, \mathbf{Z}) \mid \mathbf{z} \in \tilde{\mathcal{C}}_{ij}, \ \mathbf{Z} = \mathbf{z}\mathbf{z}^\top\}$. We replace Θ_{ij} by a convex outer approximation $\hat{\Theta}_{ij}$ obtained via RPT. In particular, we pairwise multiply inequalities defining $\tilde{\mathcal{C}}_{ij}$ and linearize the resulting product terms using $\mathbf{Z} = \mathbf{z}\mathbf{z}^\top$.

5 Concave dependence on uncertainty

In this section, we consider functions $v(\mathbf{x}, \mathbf{z})$ that are convex in the decision variables $\mathbf{x}$ and concave in the uncertainty $\mathbf{z}$. Typical examples include functions that are affine in $\mathbf{z}$, as well as separable functions of the form $v(\mathbf{x}, \mathbf{z}) = \sum_i x_i \, \phi(z_i)$, where ϕ is concave (e.g., logarithmic or square-root functions).

5.1 Exact result for degenerate distribution

An important observation is that, when the uncertain parameters have a fixed mean and the degenerate distribution at that mean belongs to the ambiguity set, the worst-case expectation in (1b) reduces to its deterministic counterpart whenever v is concave in the uncertainty. In that case, the worst-case distribution places all its probability mass at the mean. This follows directly from Jensen's inequality; see Appendix B for a proof.

Theorem 4. *Consider a function $v(\mathbf{x}, \mathbf{z})$ such that, for every fixed $\mathbf{x} \in \mathcal{X}$, it is concave in $\mathbf{z}$. Suppose that the ambiguity set imposes a fixed mean $\boldsymbol{\mu}$ and that the degenerate distribution concentrated at $\mathbf{z} = \boldsymbol{\mu}$ belongs to $\mathcal{P}$. Then, for every $\mathbf{x} \in \mathcal{X}$,*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})] = v(\mathbf{x}, \boldsymbol{\mu}), \tag{17}$$

and the supremum is attained by the degenerate distribution at $\boldsymbol{\mu}$.

For the NG and NL ambiguity sets, the conditions of Theorem 4 are satisfied whenever the moment constraint $\mathbb{E}_{\mathbb{P}}[\mathbf{z}] = \boldsymbol{\mu}$ is imposed and the confidence-set probability bounds are such that the degenerate distribution at $\boldsymbol{\mu}$ is feasible. In particular, for every $i \in \mathcal{I}$, this requires that $\overline{p}_i = 1$ whenever $\boldsymbol{\mu} \in \mathcal{C}_i$, and $\underline{p}_i = 0$ whenever $\boldsymbol{\mu} \notin \mathcal{C}_i$. The simplest special case is $I = 1$ with $\boldsymbol{\mu} \in \mathcal{C}_1$, since by assumption $\underline{p}_1 = \overline{p}_1 = 1$.

Under these conditions, the DRO problem (1) reduces to the deterministic problem

$$\min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[v(\mathbf{x}, \mathbf{z})\right] = \min_{\mathbf{x} \in \mathcal{X}} v(\mathbf{x}, \boldsymbol{\mu}).$$

Remark: If the ambiguity set does not impose a fixed mean, we define the set of feasible means by $\mathcal{M} := \{\boldsymbol{\mu} : \exists\, \mathbb{P} \in \mathcal{P} \text{ such that } \mathbb{E}_{\mathbb{P}}[\mathbf{z}] = \boldsymbol{\mu}\}$. A similar result holds if, for all $\boldsymbol{\mu} \in \mathcal{M}$, the degenerate distribution concentrated at $\mathbf{z} = \boldsymbol{\mu}$ belongs to $\mathcal{P}$. In that case, the right-hand side of (17) can be replaced by $\sup_{\boldsymbol{\mu} \in \mathcal{M}} v(\mathbf{x}, \boldsymbol{\mu})$.

5.2 RO reformulation

When the conditions of Theorem 4 are not satisfied, the DRO problem no longer reduces to a deterministic counterpart and requires an explicit reformulation of the robust constraint. Under the NL ambiguity set, Theorem 3 implies that the robust constraint must be enforced separately on the regions $\tilde{\mathcal{C}}_{ij}$ for uncertainty $\mathbf{z}$. Since the resulting constraint involves the maximization of a concave function in $\mathbf{z}$, we can apply Fenchel duality to derive a tractable reformulation (Ben-Tal et al., 2015). This reformulation expresses the constraint in terms of two key components: the support function of the uncertainty set and the concave conjugate of v.

The support function of a set $\mathcal{C}$ is defined as

$$\delta^*(\mathbf{y} \mid \mathcal{C}) = \sup_{\mathbf{z}\in\mathcal{C}} \mathbf{y}^\top \mathbf{z},$$

which is convex in $\mathbf{y}$. Expressions for common uncertainty sets are summarized in Table 2. Further useful composition rules, such as for intersections and Minkowski sums of sets, are also provided in Section 3 of Ben-Tal et al. (2015), which allows for the derivation of support functions for more complex uncertainty sets.

Uncertainty	Description $\mathcal{C}$	Support Function $\delta^*(\mathbf{y} \mid \mathcal{C})$
Box	$\{\mathbf{z} \mid \\|\mathbf{z}\\|_\infty \leq \rho\}$	$\rho\\|\mathbf{y}\\|_1$
Ball	$\{\mathbf{z} \mid \\|\mathbf{z}\\|_2 \leq \rho\}$	$\rho\\|\mathbf{y}\\|_2$
Budget	$\{\mathbf{z} \mid \\|\mathbf{z}\\|_\infty \leq \rho, \\|\mathbf{z}\\|_1 \leq \Gamma\}$	$\min_{\mathbf{w}^1,\mathbf{w}^\infty}\{\rho\\|\mathbf{w}^1\\|_1 + \Gamma\\|\mathbf{w}^\infty\\|_\infty \mid \mathbf{w}^1 + \mathbf{w}^\infty = \mathbf{y}\}$
Polyhedral	$\{\mathbf{z} \mid \mathbf{D}\mathbf{z} \leq \mathbf{d}\}$	$\min_{\mathbf{w}}\{\mathbf{d}^\top\mathbf{w} \mid \mathbf{D}^\top\mathbf{w} = \mathbf{y}, \mathbf{w} \geq \mathbf{0}\}$
Convex cone	$\{\mathbf{z} \mid \mathbf{d} - \mathbf{D}\mathbf{z} \in \mathcal{K}\}$	$\min_{\mathbf{w}}\{\mathbf{d}^\top\mathbf{w} \mid \mathbf{D}^\top\mathbf{w} = \mathbf{y}, \mathbf{w} \in \mathcal{K}^*\}$

Table 2: Common uncertainty sets $\mathcal{C}$ and their support functions $\delta^*(\mathbf{y} \mid \mathcal{C})$.

The concave conjugate of $v(\mathbf{x}, \mathbf{z})$ with respect to $\mathbf{z}$ is defined as

$$v_*(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{z}}\{\mathbf{y}^\top\mathbf{z} - v(\mathbf{x}, \mathbf{z})\}.$$

An explicit closed-form expression for v_* is not required. Rather, it suffices that v_* admits a tractable representation, possibly involving auxiliary variables and constraints, or a supremum, that can be incorporated into the resulting optimization model. Note that if z_i is an auxiliary uncertain parameter that does not appear in v, then finiteness of v_* requires $y_i = 0$. Table 3 summarizes several useful structural cases. For further details and examples on computing concave conjugates, see Section 4 of Ben-Tal et al. (2015).

Structure of $v(\mathbf{x}, \mathbf{z})$	Concave conjugate $v_*(\mathbf{x}, \mathbf{y})$
Linear in $\mathbf{z}$: $v(\mathbf{x}, \mathbf{z}) = h(\mathbf{x})^\top\mathbf{z} + c(\mathbf{x})$	$v_*(\mathbf{x}, \mathbf{y}) = -c(\mathbf{x})$ if $\mathbf{y} = h(\mathbf{x})$, and $-\infty$ otherwise
Sum of functions: $v(\mathbf{x}, \mathbf{z}) = \sum_{l\in\mathcal{L}} v_l(\mathbf{x}, \mathbf{z})$	$v_*(\mathbf{x}, \mathbf{y}) = \sup\left\{\sum_{l\in\mathcal{L}}(v_l)_*(\mathbf{x}, \mathbf{y}^l) \mid \sum_{l\in\mathcal{L}}\mathbf{y}^l = \mathbf{y}\right\}$
Separable in $\mathbf{z}$: $v(\mathbf{x}, \mathbf{z}) = \sum_{n\in\mathcal{N}} v_n(\mathbf{x}, z_n)$	$v_*(\mathbf{x}, \mathbf{y}) = \sum_{n\in\mathcal{N}}(v_n)_*(\mathbf{x}, y_n)$

Table 3: Examples of structural cases for computing concave conjugates.

Using these definitions, the DRO constraint (1b) admits the following equivalent convex reformulation via Fenchel duality, by dualizing the inner maximization over $\mathbf{z}$ Ben-Tal et al. (2015). The full proof is provided in Appendix B.

Theorem 5. *Let the NL ambiguity set be given and assume that the regularity conditions (A*) hold. If v is concave in* $\mathbf{z}$*, then the DRO constraint* (1b) *is satisfied if and only if the following condition is satisfied: for every $i \in \mathcal{I}$ and $j \in \mathcal{J}_i$, there exists a vector $\mathbf{y}^{ij} \in \mathbb{R}^P$ such that*

$$\delta^*(\mathbf{y}^{ij} \mid \tilde{\mathcal{C}}_{ij}) - v_*(\mathbf{x}, \mathbf{y}^{ij} + \mathbf{A}^\top \boldsymbol{\beta}) - \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] \leq 0.$$

Theorem 5 provides the robust counterpart of the constraint and yields the following robust optimization reformulation of (1):

$$\begin{aligned}
\min_{\substack{\mathbf{x},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\lambda},\\ \{\mathbf{y}^{ij}\}_{i,j}}} \quad & \mathbf{c}^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\kappa} \geq 0, \quad \boldsymbol{\lambda} \geq 0, \\
& \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \kappa_i - \underline{p}_i \lambda_i \right) \leq \tau, \\
& \delta^*(\mathbf{y}^{ij} \mid \tilde{\mathcal{C}}_{ij}) - v_*(\mathbf{x}, \mathbf{y}^{ij} + \mathbf{A}^\top \boldsymbol{\beta}) - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \leq 0, \qquad \forall i \in \mathcal{I}, j \in \mathcal{J}_i.
\end{aligned} \tag{18}$$

This is a deterministic convex optimization problem. If the uncertainty set is polyhedral or conic representable, then its support function admits a linear or conic representation, respectively. If, in addition, the structure of v yields a compatible tractable representation of v_*, then the reformulation reduces to a linear or conic optimization problem. In such cases, the model can be solved efficiently using standard convex optimization solvers.

A key advantage of this reformulation is that the computations involving confidence sets $\tilde{\mathcal{C}}_{ij}$ are completely independent of those involving the function v. Thus, once a particular structure of v has been analyzed, modifying the ambiguity set does not require re-deriving the conjugate of v. Moreover, known support functions of commonly used confidence sets can be implemented directly, making the framework modular and easily extensible.

6 Numerical experiments

We conduct all numerical experiments in this section using Julia 1.12.2 together with the solvers Gurobi 12.0.3 and MOSEK 11.0.30. All experiments are performed on a Windows machine equipped with an Intel(R) Core(TM) i7-1165G7 CPU running at 2.80 GHz and 16 GB of RAM. The Julia code is available upon request.

6.1 Appointment scheduling problem

The ASP considers a set of N customers that must be scheduled in consecutive order over a planning horizon of length T. Each customer $n \in \mathcal{N} = \{1, \ldots, N\}$ requires an uncertain service time z_n, and we denote the service time vector by $\mathbf{z} = (z_1, \ldots, z_N)$. The decision variables $\mathbf{x} = (x_1, \ldots, x_N)$

represent the appointment slot lengths assigned to each customer and must satisfy $\mathcal{X} = \{\mathbf{x} \mid \mathbf{x} \geq 0,\ \mathbf{1}^\top \mathbf{x} = T\}$. The cost function penalizes both total customer waiting time and overtime. For a given schedule $\mathbf{x}$ and realization $\mathbf{z}$, the total cost is

$$v(\mathbf{x}, \mathbf{z}) = \sum_{n=2}^{N} w_n + \gamma w_{N+1},$$

where waiting times are recursively defined as $w_n = (w_{n-1} + z_{n-1} - x_{n-1})^+$ for $n = 2, \ldots, N+1$, and $w_1 = 0$, and overtime is modeled as the waiting time of an artificial customer $N+1$ and is weighted by a penalty parameter $\gamma > 0$. By induction, this recursion admits the equivalent closed-form representation $w_n = \max_{1 \leq l \leq n-1} \left\{0, \sum_{j=n-l}^{n-1} (z_j - x_j)\right\}$.

The ambiguity set $\mathcal{P}$ encodes partial distributional information about the uncertain service times $\mathbf{z}$. The resulting DRO problem seeks a schedule $\mathbf{x} \in \mathcal{X}$ that minimizes the worst-case expected cost over all distributions $\mathbb{P} \in \mathcal{P}$:

$$\min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})]. \tag{19}$$

Here, we directly minimize the worst-case expected cost rather than imposing it through a bound τ as in (1), so the expectation appears in the objective and the parameter τ can be omitted.

The ASP cost function $v(\mathbf{x}, \mathbf{z})$ can be written in the form $h(\mathbf{T}\mathbf{z} + \mathbf{t}(\mathbf{x}))$, where $\mathbf{T}$ is constant and h is a sum of maxima of linear function. Hence, h is convex and Lipschitz continuous on its domain, and its convex conjugate is the indicator function of a polyhedron given by

$$h^*(\mathbf{w}) = \begin{cases} 0 & \text{if } \mathbf{w} \in \operatorname{dom}(h^*) = \left\{\mathbf{w} \geq 0 \ : \ \mathbf{H}^\top \mathbf{w} = \mathbf{1}\right\}, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathbf{H}$ is a binary matrix encoding which linear functions belong to each max operator. Details of this derivation, including the construction of $\mathbf{H}$ and $\mathbf{T}$, as well as the derivation of $\mathbf{t}(\mathbf{x}) = -\mathbf{T}\mathbf{x}$, are provided in Brugman et al. (2024).

Remark: Although the cost function $v(\mathbf{x}, \mathbf{z})$ can in principle be written as the maximum of finitely many affine functions, this representation quickly becomes computationally intractable. Indeed, the recursive max-structure underlying the ASP gives rise to an exponential number of affine pieces. In the special case $\gamma = 1$, this counting appears to follow the Catalan number $C_{N+1} = \frac{1}{N+2}\binom{2N+2}{N+1}$, which grows exponentially in N. For example, when $N = 12$, this yields $C_{13} = 742,900$ affine terms. For general $\gamma \neq 1$, the number of distinct affine terms may be even larger. A reformulation based on the WKS framework would therefore involve a very large number of robust constraints and becomes computationally prohibitive for realistic problem sizes.

6.1.1 NG ambiguity set

Since the ASP cost function is convex in the uncertainty, the DRO problem under the NG ambiguity set admits the robust optimization reformulation (9). In this formulation, the uncertainty is given by the pair $(\mathbf{w}, \mathbf{z})$. Moreover, as $h^*(\mathbf{w})$ is an indicator function, the variable w_0 can be omitted.

Corollary 1. *Under the NG ambiguity set, the appointment scheduling problem* (19) *admits the following equivalent robust optimization reformulation:*

$$\begin{aligned}
\min_{\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\lambda}} \quad & \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \, \kappa_i - \underline{p}_i \, \lambda_i \right) \\
s.t. \quad & \mathbf{x} \geq 0, \quad \mathbf{1}^\top \mathbf{x} = T, \\
& \boldsymbol{\kappa} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \\
& \mathbf{w}^\top \mathbf{T} \mathbf{z} - \mathbf{w}^\top \mathbf{T} \mathbf{x} - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall (\mathbf{w}, \mathbf{z}) \in \Lambda_i, \ \forall i \in \mathcal{I},
\end{aligned} \tag{20}$$

where $\Lambda_i = \left\{ (\mathbf{w}, \mathbf{z}) \mid \mathbf{z} \in \mathcal{C}_i, \ \mathbf{w} \geq \mathbf{0}, \ \mathbf{H}^\top \mathbf{w} = \mathbf{1} \right\}$.

Because this formulation falls within the convex class introduced in Section 3 and satisfies the required assumptions, the RO problem (20) is solved using the cutting-set algorithm that iteratively solves the master and subproblem.

Subproblem: (11) can be further specified depending on the cones defining the confidence sets (see Appendix A). The RPT relaxation yields the following convex subproblem:

$$\max_{(\mathbf{w}, \mathbf{z}, \mathbf{V}) \in \hat{\Theta}_i} \quad \mathrm{Tr}\Big(\mathbf{T}^\top \mathbf{V} \Big) - \mathbf{w}^\top \mathbf{T} \mathbf{x}^* - \sum_{i' \in \mathcal{A}(i)} (\kappa^*_{i'} - \lambda^*_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta}^*. \tag{21}$$

Here, $\hat{\Theta}_i$ denotes the convex outer approximation obtained via RPT by pairwise multiplying constraints involving $\mathbf{w}$ and $\mathbf{z}$. When $\mathcal{C}_i$ contains only linear constraints, that is $\mathcal{C}_i = \{\mathbf{z} : \mathbf{C}_i \mathbf{z} \leq \mathbf{c}_i\}$, as in the mean-support ambiguity set, the relaxed uncertainty set becomes

$$\hat{\Theta}_i \ = \ \Big\{ (\mathbf{w}, \mathbf{z}, \mathbf{V}) \mid \mathbf{C}_i \mathbf{z} \leq \mathbf{c}_i, \ \mathbf{w} \geq 0, \ \mathbf{H}^\top \mathbf{w} = \mathbf{1}, \ \mathbf{H}^\top \mathbf{V} = \mathbf{1} \mathbf{z}^\top, \ \mathbf{V} \mathbf{C}_i^\top - \mathbf{w} \mathbf{c}_i^\top \leq \mathbf{0} \Big\}.$$

When $\mathcal{C}_i$ additionally contains quadratic constraints of the form $(z_n - \mu_n)^2 \leq z_{N+n}$, as in the mean-variance and mean-variance-PCM ambiguity sets introduced later, $\hat{\Theta}_i$ is extended with rotated second-order cone (SOC) constraints (Bertsimas et al., 2026). Specifically, for each component n and for each bilinear term (j, n) arising from the lifting, we impose

$$\left\| \left(z_n - \mu_n, \ \frac{z_{N+n} - 1}{2} \right) \right\| \leq \frac{z_{N+n} + 1}{2}, \qquad \left\| \left(V_{j,n} - \mu_n w_j, \ \frac{V_{j,N+n} - w_j}{2} \right) \right\| \leq \frac{V_{j,N+n} + w_j}{2}.$$

Master problem: (13) can be applied directly to the ASP. To obtain initial scenario sets $\mathcal{S}_i$, we solve the RPT relaxation of (20), replacing the i-th robust constraint with

$$\operatorname{Tr}\Big(\mathbf{T}^\top \mathbf{V}\Big) - \mathbf{w}^\top \mathbf{T}\mathbf{x} - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \;\leq\; 0. \quad \forall (\mathbf{w}, \mathbf{z}, \mathbf{V}) \in \hat{\Theta}_i.$$

The full RPT relaxation is provided in Appendix D.1, together with its dual formulation. Furthermore, we provide a simplified version of the dual for the case of a single confidence set, which can be solved directly.

Under the NG ambiguity set, we consider the same numerical experiments of Bertsimas et al. (2019) (hereafter BSZ). BSZ considers several ambiguity sets, including the mean-support and mean-marginal-moment sets from Mak et al. (2015) and their own partial cross-moment (PCM) ambiguity set with correlation coefficient α. Their approach is based on an adaptive DRO reformulation of the ASP, in which the adjustable variables $\mathbf{y}(\mathbf{z})$ for the waiting and overtime are approximated using linear decision rules (LDRs). Consequently, their reported values are conservative upper bounds on the true optimum.

Table 4 reports the results for instances 3 and 4 (given in Table 8 in Appendix E.1). For each ambiguity set, we report the upper bound from Table 1 of BSZ, our optimal objective value, the relative deviation of the BSZ upper bound from our objective value, computation time, the number of scenarios added to the master problem, and the total number of branches generated across all iterations. All instances are solved to the prescribed tolerance $\epsilon = 0.01$ in the RPT-BB procedure. For both instances, the optimal values under the Mean and Mean-MM ambiguity sets coincide with the values reported by BSZ, confirming the tightness of their bounds. For Mean-PCM-α with $0 < \alpha < 1$, our optimal objective values are slightly higher than the bounds reported in BSZ (deviating more than 1%), which suggests that the bounds may be numerically inaccurate.[1] For Mean-PCM-0, our optimal values fall below the reported bounds, indicating that the LDR approximation can be relatively conservative, which is in line with the observations in Zhen et al. (2018). Regarding computational performance, the Mean and Mean-MM instances are solved within one minute. The Mean-PCM instances are substantially more challenging and require a larger number of branches. Our method requires up to 12 minutes for the most difficult instance, while most cases require less than 10 minutes.

Table 5 reports a scalability experiment for the Mean-MM ambiguity set as the number of customers N increases (instances given in Table 9 in Appendix E.1). The optimal values are nearly identical to the upper bounds reported by BSZ in Table 3 (which goes up to $N = 30$), again confirming the tightness of their approximation. As expected, computational effort increases with the problem size. Solution times grow from less than one minute for $N = 8$ to just over one hour for $N = 20$, primarily due to the increasing size of the subproblem. The number of scenarios added to the

[1] The precise source of this discrepancy remains unclear. While we obtained code from the authors for a different ambiguity set and used it to recover the values of instances 3 and 4, we did not have access to their implementation for the PCM ambiguity set.

Inst.	Ambiguity	UB BSZ	Obj. value	BSZ gap(%)	Time(s)	#Scen.	#Br.
3	Mean	1987.66	1987.66	0.00	5	1	0
	Mean-MM	37.79	37.79	0.00	53	250	164
	Mean-PCM-0.75	37.76	37.78	-0.05	477	313	1933
	Mean-PCM-0.5	37.21	37.68	-1.25	402	235	2345
	Mean-PCM-0.25	36.27	36.63	-0.98	231	258	1652
	Mean-PCM-0	34.51	32.33	6.74	417	257	3243
4	Mean	2089.22	2089.22	0.00	6	0	0
	Mean-MM	66.88	66.88	0.00	43	264	104
	Mean-PCM-0.75	66.82	66.87	-0.07	296	291	1330
	Mean-PCM-0.5	66.01	66.75	-1.11	415	283	2219
	Mean-PCM-0.25	64.13	65.13	-1.54	710	250	5313
	Mean-PCM-0	61.00	59.36	2.76	359	253	3242

Table 4: Results for appointment scheduling instances 3 and 4 under NG ambiguity ($\epsilon = 0.01$).

master problem increases from 146 to 885, while the number of branches grows from 70 to 646. Despite this growth, the results indicate that the proposed exact approach remains computationally tractable for problem sizes of practical interest.

Size	UB BSZ	Obj. value	BSZ gap(%)	Time(s)	#Scen.	#Br.
$N = 8$	-	124.34	-	23	146	70
$N = 10$	-	158.42	-	121	419	215
$N = 12$	219.65	218.99	0.30	193	364	228
$N = 14$	288.48	288.00	0.17	313	450	131
$N = 16$	374.71	374.71	0.00	784	560	288
$N = 18$	380.50	380.49	0.00	2013	696	500
$N = 20$	538.52	538.52	0.00	4139	885	646

Table 5: Results for appointment scheduling with varying N under NG ambiguity ($\epsilon = 0.01$).

6.1.2 NL ambiguity set

After the bilinear reformulation, the ASP cost function satisfies the SLC structure in (14) when $(\mathbf{w}, \mathbf{z})$ is treated jointly as the uncertainty, and therefore can be treated under the NL ambiguity set as well. Using (16) implies that the robust constraint must be enforced separately on the corresponding uncertainty sets built from the regions $\tilde{\mathcal{C}}_{ij}$. This leads to the following equivalent robust optimization reformulation.

Corollary 2. *Under the NL ambiguity set, the appointment scheduling problem* (19) *is equivalent to* (20)*, with the uncertainty sets* Λ_i *replaced by*

$$\Lambda_{ij} := \left\{ (\mathbf{w}, \mathbf{z}) \mid \mathbf{z} \in \tilde{\mathcal{C}}_{ij},\ \mathbf{w} \geq \mathbf{0},\ \mathbf{H}^\top \mathbf{w} = \mathbf{1} \right\},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{J}_i.$$

For the numerical experiments, the goal is to illustrate the modeling flexibility enabled by the NL ambiguity set, rather than to provide a full scalability study. For this, we again consider ASP instances 3 and 4, now using an NL ambiguity set. All parameters are identical to those used for the NG ambiguity set. In addition, we define a budget threshold $\phi = \sum_{i=1}^{N}(\mu_i + \sigma_i)$, which is assumed to hold with probability at least 0.9 (see Table 8 in Appendix E.1 for a full overview). This leads to the confidence sets

$$\mathcal{C}_2 = \{\boldsymbol{z} \mid \boldsymbol{a} \leq \boldsymbol{z} \leq \boldsymbol{b}\}, \qquad \mathcal{C}_1 = \{\boldsymbol{z} \in \mathcal{C}_2 \mid \mathbf{1}^\top \mathbf{z} \leq \phi\},$$

with $\mathbb{P}(\boldsymbol{z} \in \mathcal{C}_1) \in [0.9, 1]$ and, by assumption, $\mathbb{P}(\boldsymbol{z} \in \mathcal{C}_2) = 1$. Thus, at least 90% of the probability mass satisfies the budget constraint on the total service time. This natural restriction satisfies weak nestedness, but not the strict nestedness required by the NG ambiguity set.

To apply our method, we consider the disjoint set $\bar{\mathcal{C}}_2 = \mathcal{C}_2 \setminus \mathcal{C}_1$ and its closure $\tilde{\mathcal{C}}_2 = \{\boldsymbol{z} \in \mathcal{C}_2 \mid \mathbf{1}^\top \boldsymbol{z} \geq \phi\}$. Then, Corollary 2 requires the robust constraint to be enforced on the two corresponding uncertainty sets built from $\mathcal{C}_1$ and $\tilde{\mathcal{C}}_2$. Since both sets are already convex, no further decomposition into smaller sets is required.

Table 6 compares the NL ambiguity set with two ambiguity sets with $I = 1$: (i) a support-only set (S) with confidence set $\mathcal{C}_2$, and (ii) a support-and-budget set (SB) in which both constraints hold with probability one. The table reports the optimal objective value, computation time, the number of generated scenarios, and the total number of branches across all iterations of the subproblem. For the NL ambiguity set, the reported numbers correspond to the combined totals of the two subproblems. As expected, the objective value obtained with the NL ambiguity set lies between those for the two $I = 1$ ambiguity sets. The S-set is the least restrictive and therefore yields the most conservative solution, whereas the SB-set imposes the strongest restrictions and yields the smallest objective value. The NL ambiguity set enforces the budget constraint only probabilistically and thus represents an intermediate level of conservatism. For these instances, the NL ambiguity set behaves similarly to the SB set in terms of computation time. Although two subproblems must be solved in each iteration, the total computation times remain comparable. This is also reflected in the number of generated scenarios and branches.

Instance	**Ambiguity**	**Obj. value**	**Time (s)**	**#Scenarios**	**#Branches**
	S	660.87	4.5	10	0
3	NL	609.51	6.6	3	27
	SB	584.42	8.3	2	27
	S	697.92	4.4	8	0
4	NL	662.07	7.4	2	35
	SB	599.77	6.8	8	29

Table 6: Results for appointment scheduling instances 3 and 4 using three ambiguity sets ($\epsilon = 0.01$).

6.2 Capital budgeting problem with NL ambiguity

Consider a capital budgeting problem (Alfandari and García, 2018; Bertsimas and den Hertog, 2022; Wang et al., 2025), in which a decision maker must select a subset of N available projects. Each project $n \in \mathcal{N} = \{1, \dots, N\}$ has an associated cost $q_n > 0$, and the total available budget is $\theta > 0$. The objective is to select a portfolio of projects that maximizes the total net present value (NPV), subject to the budget constraint. Let $x_n \in \{0, 1\}$ be a binary decision variable indicating whether project n is selected ($x_n = 1$) or not ($x_n = 0$). The NPV of a project depends on its future cash flows, discounted using an uncertain discount rate. Let F_{n0} denote the initial cash flow of project n, and let F_{nt} denote the forecasted cash flow for project n at time $t \in \mathcal{T} := \{1, \dots, T\}$. Let z_n denote the uncertain discount rate associated with project n. We collect these in the vector $\mathbf{z} = (z_1, \dots, z_N)$. To formulate the problem as a minimization problem, we consider the negative present value of project n, given by $f_n(z_n) = -\left(F_{n0} + \sum_{t\in\mathcal{T}} \frac{F_{nt}}{(1+z_n)^t}\right)$. The total cost function is then

$$v(\mathbf{x}, \mathbf{z}) = -\sum_{n\in\mathcal{N}} x_n \left(F_{n0} + \sum_{t\in\mathcal{T}} \frac{F_{nt}}{(1+z_n)^t}\right),$$

which is linear in $\mathbf{x}$ and concave in $\mathbf{z}$, assuming nonnegative cash flows and discount rates satisfying $z_n > -1$. The resulting DRO problem seeks a project portfolio $\mathbf{x}$ from the feasible set $\mathcal{X} = \{\mathbf{x} \mid \mathbf{x} \in \{0, 1\}^N, \mathbf{q}^\top \mathbf{x} \le \theta\}$ that minimizes the worst-case expected cost over all distributions $\mathbb{P} \in \mathcal{P}$:

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})]. \tag{22}$$

Again, we minimize the worst-case expected cost directly instead of imposing a bound τ.

Since v is concave in $\mathbf{z}$, we apply Section 5 and study it under the NL ambiguity set. In this numerical example, we assume that the support $\mathcal{C}_I$ is polyhedral. This implies that the convex sets $\tilde{\mathcal{C}}_{ij}$ are polyhedral as well for all $i \in \mathcal{I}, j \in \mathcal{J}_i$, so we write $\tilde{\mathcal{C}}_{ij} = \left\{\mathbf{z} : \mathbf{G}_{ij}\mathbf{z} \le \mathbf{g}_{ij}\right\}$. The support function δ^* of these polyhedral sets is given by

$$\delta^*(\mathbf{y} \mid \tilde{\mathcal{C}}_{ij}) = \min_{\mathbf{w}} \left\{\mathbf{g}_{ij}^\top \mathbf{w} \;\middle|\; \mathbf{G}_{ij}^\top \mathbf{w} = \mathbf{y},\ \mathbf{w} \ge \mathbf{0}\right\}.$$

Let $\alpha_{nt} := (t+1)t^{-\frac{t}{t+1}} F_{nt}^{\frac{1}{t+1}}$ for all $n \in \mathcal{N}$ and $t \in \mathcal{T}$. The concave conjugate of $v(\mathbf{x}, \mathbf{z})$ with respect to $\mathbf{z}$, derived in Appendix D, is given by

$$v_*(\mathbf{x}, \mathbf{y}) = \sum_{n\in\mathcal{N}} x_n F_{n0} + \sup_{\mathbf{r}} \left\{\sum_{n\in\mathcal{N}} \sum_{t\in\mathcal{T}} \left(\alpha_{nt} x_n^{\frac{1}{t+1}} r_{nt}^{\frac{t}{t+1}} - r_{nt}\right) \mid \sum_{t\in\mathcal{T}} r_{nt} \le y_n\ \forall n,\ r_{nt} \ge 0\ \forall n, t\right\}.$$

The nonlinear terms in this expression admit an equivalent representation using power cone constraints. For each n and t, introduce an auxiliary variable $u_{nt} \ge 0$ such that

$$u_{nt} \le x_n^{\frac{1}{t+1}} r_{nt}^{\frac{t}{t+1}} \quad \Longleftrightarrow \quad (x_n, r_{nt}, u_{nt}) \in \mathcal{P}_3^{\frac{1}{t+1}, \frac{t}{t+1}},$$

where the equivalence uses $x_n \geq 0$, $r_{nt} \geq 0$, and $u_{nt} \geq 0$. Then the term $\alpha_{nt} x_n^{\frac{1}{t+1}} r_{nt}^{\frac{t}{t+1}} - r_{nt}$ is represented by $\alpha_{nt} u_{nt} - r_{nt}$. Therefore, the nonlinear terms in the reformulation of $v_*(\mathbf{x}, \mathbf{y})$ are power-cone representable.

We now substitute the expressions for δ^* and v_* into (18). Since the support function appears with a positive sign and the concave conjugate with a negative sign in a constraint of the form "≤ 0", the inner minimum in δ^* and the supremum in v_* can be dropped, as feasibility only requires the existence of variables satisfying the resulting inequalities. Replacing the remaining nonlinear terms in v_* by their power cone representation yields the following result.

Corollary 3. *Under the NL ambiguity set with $\mathcal{C}_I$ polyhedral, the capital budgeting problem* (22) *is equivalent to the following mixed-integer conic optimization problem:*

$$
\begin{aligned}
\min_{\substack{\mathbf{x},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\lambda} \\ \{\mathbf{y}^{ij},\mathbf{w}^{ij},\mathbf{r}^{ij},\mathbf{u}^{ij}\}_{i,j}}} \quad & \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \kappa_i - \underline{p}_i \lambda_i \right) \\
\text{s.t.} \quad & \boldsymbol{\kappa} \geq 0, \quad \boldsymbol{\lambda} \geq 0, \\
& \mathbf{x} \in \{0,1\}^N, \quad \mathbf{q}^\top \mathbf{x} \leq \theta, \\
& \mathbf{g}_{ij}^\top \mathbf{w}^{ij} - \sum_{n \in \mathcal{N}} x_n F_{n0} - \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} \left(\alpha_{nt} u_{nt}^{ij} - r_{nt}^{ij} \right) \\
& \qquad - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \leq 0, && \forall i \in \mathcal{I},\ j \in \mathcal{J}_i, \\
& \mathbf{G}_{ij}^\top \mathbf{w}^{ij} = \mathbf{y}^{ij}, \quad \mathbf{w}^{ij} \geq 0, && \forall i \in \mathcal{I},\ j \in \mathcal{J}_i, \\
& \sum_{t \in \mathcal{T}} r_{nt}^{ij} \leq y_n^{ij} + (\mathbf{A}^\top \boldsymbol{\beta})_n, && \forall n \in \mathcal{N},\ i \in \mathcal{I},\ j \in \mathcal{J}_i, \\
& (x_n, r_{nt}^{ij}, u_{nt}^{ij}) \in \mathcal{P}_3^{\frac{1}{t+1}, \frac{t}{t+1}},\ u_{nt}^{ij} \geq 0 && \forall n \in \mathcal{N},\ t \in \mathcal{T},\ i \in \mathcal{I},\ j \in \mathcal{J}_i.
\end{aligned}
\tag{23}
$$

For the capital budgeting experiments, the intention is to illustrate how the NL ambiguity set can incorporate additional linear event information, and to provide a first indication of the computational effect of increasing the number of induced regions. For this, we adopt the numerical setup from the code of Wang et al. (2025) [2]. We consider a fixed instance with $N = 12$ projects and $T = 5$ periods, and use the same budget parameter θ, cash flow matrix $\mathbf{F}$, and project cost vector $\mathbf{q}$. The full parameter settings are reported in Appendix E.2. For the project-specific discount rates, we follow the structure of their data generation process. In particular, for project n, we set the support bounds to $a_n = 0.005(n-1)$ and $b_n = 0.025n$, and define the mean as $\mu_n = \frac{a_n + b_n}{2}$. This yields a mean-support ambiguity set with support $[\mathbf{a}, \mathbf{b}]$ and mean $\boldsymbol{\mu}$.

To enrich this ambiguity description, we incorporate additional linear event information on selected

[2] Although our data are based on their setup, their study considers Wasserstein ambiguity sets and is therefore not directly comparable to the ambiguity sets studied here. Moreover, to the best of our knowledge, no directly comparable results for this capital budgeting problem under such ambiguity sets are available in the literature.

components of the discount-rate vector. For an index set $\mathcal{M} \subseteq \mathcal{N}$, we consider the confidence sets

$$\mathcal{C}_2 = \{\boldsymbol{z} \mid \boldsymbol{a} \leq \boldsymbol{z} \leq \boldsymbol{b}\}, \qquad \mathcal{C}_1 = \{\mathbf{z} \mid z_n \leq \tau_n \text{ for all } n \in \mathcal{M}\},$$

where the thresholds are chosen as $\tau_n = \frac{\mu_n + a_n}{2}$, corresponding to relatively low discount rates. The probability of this event is restricted to lie in the interval $[\underline{p}_1, \overline{p}_1] = [0.65, 0.75]$. Then, we model

$$\bar{\mathcal{C}}_2 = \{\boldsymbol{z} \in \mathcal{C}_2 \mid z_n > \tau_n \text{ for at least one } n \in \mathcal{M}\}$$

This set is generally nonconvex. We therefore decompose it into the union of convex regions

$$\bar{\mathcal{C}}_2 = \bigcup_{n \in \mathcal{M}} \bar{\mathcal{C}}_{2n}, \qquad \bar{\mathcal{C}}_{2n} = \{\boldsymbol{z} \in \mathcal{C}_2 \mid z_n > \tau_n\}.$$

The robust constraints are then introduced on the closure of these sets.

Table 7 reports results starting from the mean-support case without event constraints and gradually imposing the linear event on an increasing number of components, following the sequence

$$\mathcal{M} \in \{\emptyset,\ \{1\},\ \{1,6\},\ \{1,6,7\},\ \{1,6,7,8\},\ \{1,6,7,8,9\},\ \{1,6,7,8,9,11\}\}.$$

The selected indices correspond to projects that are chosen in all solutions considered, so that the imposed constraints directly affect the objective value. The table shows that, as the number of components increases, the objective value (the negative NPV) decreases monotonically from -10.5347 to -10.6366, while computation times remain below 1.2 seconds in all cases. This indicates that adding linear event information makes the ambiguity set more informative and leads to a less conservative worst-case value for this instance. In all cases considered here, the optimal project selection remains unchanged and is given by $\{1, 6, 7, 8, 9, 11\}$.

Case	**Index set**	**Obj. value**	**Time (s)**
0	$\emptyset$	-10.5347	0.22
1	$\{1\}$	-10.5353	0.48
2	$\{1,6\}$	-10.5477	0.41
3	$\{1,6,7\}$	-10.5653	0.55
4	$\{1,6,7,8\}$	-10.5833	0.54
5	$\{1,6,7,8,9\}$	-10.6070	0.94
6	$\{1,6,7,8,9,11\}$	-10.6366	1.17

Table 7: Results for the capital budgeting instance under increasing event index sets.

7 Conclusions and outlook

This paper develops a general and tractable framework for exactly solving DRO problems with broad nonlinear functions under moment- and confidence-set-based ambiguity sets. Building on the foundational work of Wiesemann et al. (2014), we reformulate the DRO problem as an RO problem and introduce an alternative ambiguity set that is particularly convenient for deriving exact solution methods. Our approach thus bridges recent advances in RO with DRO.

Beyond the max-of-linear structure, we distinguish three structural cases for the dependence on the uncertainty. For convex dependence, we tailor a cutting-set algorithm based on RPT-BB to globally solve the resulting robust problem, and for SLC dependence, we slightly extend this algorithm. For concave dependence, we derive exact robust reformulations based on support functions and concave conjugates. Together, these results cover a wide range of function classes that are difficult to handle with existing DRO approaches. Through the appointment scheduling and capital budgeting problems, we demonstrate both the modeling flexibility and computational potential of the proposed framework. In particular, the framework yields exact solutions in settings where existing methods provide only safe approximations.

Several limitations and directions for further research remain. First, scalability is limited by the size of the lifted RPT reformulations, which grow quadratically in the uncertainty dimension and can become computationally demanding for larger instances or many confidence or decomposition regions. Developing more compact reformulations, stronger relaxations, or decomposition-based solution approaches is therefore an important direction for future work. Second, the NL ambiguity set relies on linearly defined confidence sets. While this structure is useful for modeling threshold, budget, and one-sided event information, it excludes nonlinear nested confidence regions such as ellipsoidal sets unless they can be approximated or reformulated in a compatible way. Finally, the exact solutions obtained here can serve as reference values for more systematic benchmarking of approximation-based DRO methods, such as LDR approximations, SAA-based approaches, and SDP relaxations. Extensions to multistage settings and applications with richer nonlinear structure are also promising directions for future research.

Acknowledgements

This work is supported by an NWO Mathematics Clusters grant and an NWO VICI grant. We thank Wolfram Wiesemann, Danique de Moor and Jianzhe Zhen for their valuable comments on earlier drafts.

References

Alfandari, Laurent and Juan-Carlos Espinoza García (2018). “Robust optimization for non-linear impact of data variation”. In: *Computers & Operations Research* 99, pp. 38–47.

Beck, Amir and Aharon Ben-Tal (2009). "Duality in robust optimization: primal worst equals dual best". In: *Operations Research Letters* 37.1, pp. 1–6.

Ben-Tal, Aharon, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen (2013). "Robust solutions of optimization problems affected by uncertain probabilities". In: *Management Science* 59.2, pp. 341–357.

Ben-Tal, Aharon, Dick den Hertog, and Jean-Philippe Vial (2015). "Deriving robust counterparts of nonlinear uncertain inequalities". In: *Mathematical Programming* 149.1, pp. 265–299.

Ben-Tal, Aharon and Arkadi Nemirovski (1999). "Robust solutions of uncertain linear programs". In: *Operations Research Letters* 25.1, pp. 1–13.

Ben-Tal, Aharon, Arkadi Nemirovski, and Laurent El Ghaoui (2009). *Robust optimization.* Princeton university press.

Bertsimas, Dimitris and Dick den Hertog (2022). *Robust and adaptive optimization.* Belmont: Dynamic Ideas.

Bertsimas, Dimitris, Vishal Gupta, and Nathan Kallus (2018). "Robust sample average approximation". In: *Mathematical Programming* 171.1, pp. 217–282.

Bertsimas, Dimitris, Dick den Hertog, Jean Pauphilet, and Jianzhe Zhen (2023a). "Robust convex optimization: A new perspective that unifies and extends". In: *Mathematical Programming* 200.2, pp. 877–918.

Bertsimas, Dimitris, Danique de Moor, Dick den Hertog, Thodoris Koukouvinos, and Jianzhe Zhen (2024). "An exact method for a class of robust nonlinear optimization problems". In: *Optimization Online*. URL: https://optimization-online.org/?p=27202.

Bertsimas, Dimitris, Danique de Moor, Dick den Hertog, Thodoris Koukouvinos, and Jianzhe Zhen (2026). "Cone product reformulation for global optimization". In: *INFORMS Journal on Computing*.

Bertsimas, Dimitris, Danique de Moor, Dick den Hertog, Thodoris Koukouvinos, and Jianzhe Zhen (2023b). "Reformulation-Perspectification Technique for nonconvex optimization problems". In: *Optimization Online*. URL: https://optimization-online.org/?p=23399.

Bertsimas, Dimitris, Melvyn Sim, and Meilin Zhang (2019). "Adaptive distributionally robust optimization". In: *Management Science* 65.2, pp. 604–618.

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex optimization.* Cambridge university press.

Brugman, Judith, Johan van Leeuwaarden, and Dick den Hertog (2024). "Robust appointment scheduling for general convex uncertainty sets". In: *Optimization Online*. URL: https://optimization-online.org/?p=27791.

Delage, Erick and Yinyu Ye (2010). "Distributionally robust optimization under moment uncertainty with application to data-driven problems". In: *Operations Research* 58.3, pp. 595–612.

Gao, Rui and Anton Kleywegt (2023). "Distributionally robust stochastic optimization with Wasserstein distance". In: *Mathematics of Operations Research* 48.2, pp. 603–655.

Goh, Joel and Melvyn Sim (2010). "Distributionally robust optimization and its tractable approximations". In: *Operations Research* 58.4-part-1, pp. 902–917.

Gorissen, Bram L, Dick den Hertog, and Meike Reusken (2025). "Hidden convexity in a class of optimization problems with bilinear terms". In: *Operations Research*.

Hu, Zhaolin and L Jeff Hong (2013). "Kullback-Leibler divergence constrained distributionally robust optimization". In: *Optimization Online*. URL: https://optimization-online.org/?p=12225.

Kuhn, Daniel, Soroosh Shafiee, and Wolfram Wiesemann (2025). "Distributionally robust optimization". In: *Acta Numerica* 34, pp. 579–804.

Lin, Fengming, Xiaolei Fang, and Zheming Gao (2022). "Distributionally robust optimization: A review on theory and applications". In: *Numerical Algebra, Control and Optimization* 12.1, pp. 159–212.

Mak, Ho-Yin, Ying Rong, and Jiawei Zhang (2015). "Appointment scheduling with limited distributional information". In: *Management Science* 61.2, pp. 316–334.
Mohajerin Esfahani, Peyman and Daniel Kuhn (2018). "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations". In: *Mathematical Programming* 171.1, pp. 115–166.
Mutapcic, Almir and Stephen Boyd (2009). "Cutting-set methods for robust convex optimization with pessimizing oracles". In: *Optimization Methods & Software* 24.3, pp. 381–406.
Rahimian, Hamed and Sanjay Mehrotra (2019). "Distributionally robust optimization: A review". In: *arXiv preprint arXiv:1908.05659*.
Scarf, Herbert E (1957). *A min-max solution of an inventory problem.* Rand Corporation Santa Monica.
Shapiro, Alexander (2017). "Distributionally robust stochastic programming". In: *SIAM Journal on Optimization* 27.4, pp. 2258–2275.
Shapiro, Alexander (2001). "On duality theory of conic linear problems". In: *Nonconvex Optimization and its Applications* 57, pp. 135–155.
Soyster, Allen L (1973). "Convex programming with set-inclusive constraints and applications to inexact linear programming". In: *Operations Research* 21.5, pp. 1154–1157.
Wang, Irina, Cole Becker, Bart Van Parys, and Bartolomeo Stellato (2025). "Mean robust optimization: I. Wang et al." In: *Mathematical Programming* 213.1, pp. 1235–1277.
Wiesemann, Wolfram, Daniel Kuhn, and Melvyn Sim (2014). "Distributionally robust convex optimization". In: *Operations Research* 62.6, pp. 1358–1376.
Zhen, Jianzhe, Dick den Hertog, and Melvyn Sim (2018). "Adjustable robust optimization via Fourier–Motzkin elimination". In: *Operations Research* 66.4, pp. 1086–1100.
Zymler, Steve, Daniel Kuhn, and Berç Rustem (2013). "Distributionally robust joint chance constraints with second-order moment information". In: *Mathematical Programming* 137.1, pp. 167–198.

Appendix

A Ambiguity sets

Here we demonstrate that the ambiguity sets considered in this paper conform to the NG ambiguity set defined in (2). Each ambiguity set is obtained by specifying appropriate moment constraints $(\mathbf{A}, \mathbf{b})$ and a single confidence set $\mathcal{C}$ (so $I = 1$ and $\underline{p} = \bar{p} = 1$).

The mean-support ambiguity set with mean $\boldsymbol{\mu}$ and support $[\mathbf{a}, \mathbf{b}]$ is defined as

$$\mathcal{P}_{(\boldsymbol{\mu})} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\mathbf{z}] = \boldsymbol{\mu},\ \mathbb{P}[\mathbf{a} \leq \mathbf{z} \leq \mathbf{b}] = 1\}.$$

This corresponds to $\mathbf{A} = \mathbf{I}_P$, $\mathbf{b} = \boldsymbol{\mu}$, and polyhedral set $\mathcal{C} = \{\mathbf{z} : \mathbf{a} \leq \mathbf{z} \leq \mathbf{b}\}$.

The mean-MM (marginal moment) ambiguity set with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\sigma}^2$ and support $[\mathbf{a}, \mathbf{b}]$ uses auxiliary variables $\mathbf{u}$ to model the marginal moment, and is defined as

$$\mathcal{P}_{(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[[\mathbf{z}, \mathbf{u}]] = [\boldsymbol{\mu}, \boldsymbol{\sigma}^2], \mathbb{P}\left(\mathbf{u} \geq (\mathbf{z} - \boldsymbol{\mu})^\top \mathbf{I}_p (\mathbf{z} - \boldsymbol{\mu}), \mathbf{a} \leq \mathbf{z} \leq \mathbf{b}\right) = 1\}.$$

This corresponds to $\mathbf{A} = \mathbf{I}_{2P}$, $\mathbf{b} = [\boldsymbol{\mu}; \boldsymbol{\sigma}^2]$, and set $\mathcal{C}$ with linear constraints $a_j \leq z_j \leq b_j$ and second order cone constraints $(\frac{u_j+1}{2}, z_j - \mu_j, \frac{u_j-1}{2}) \in \mathcal{Q}$ for all $j = 1, \ldots, P$.

The mean-PCM (partial cross moment) ambiguity set with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\sigma}^2$, covariance matrix $\boldsymbol{\Sigma}$ which defines partial cross moment $\phi = \mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}$, and support $[\mathbf{a}, \mathbf{b}]$ gives:

$$\begin{aligned}\mathcal{P}_{(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \phi)} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^{P+1}) : \mathbb{E}_{\mathbb{P}}[[\mathbf{z}, \mathbf{u}, u_{N+1}]] = [\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \phi], \\ \mathbb{P}\left(\mathbf{u} \geq (\mathbf{z} - \boldsymbol{\mu})^2, u_{N+1} \geq \left(\sum_j (z_j - \mu_j)\right)^2, \mathbf{a} \leq \mathbf{z} \leq \mathbf{b}\right) = 1\}.\end{aligned}$$

This corresponds to $\mathbf{A} = \mathbf{I}_{2P+1}$, $\mathbf{b} = [\boldsymbol{\mu}; \boldsymbol{\sigma}^2; \phi]$, and set $\mathcal{C}$ can be written as linear constraints $a_j \leq z_j \leq b_j$ and second order cone constraints $(\frac{u_j+1}{2}, z_j - \mu_j, \frac{u_j-1}{2}) \in \mathcal{Q}$ for all $j = 1, \ldots, P$, and $(\frac{u_{N+1}+1}{2}, \sum_j (z_j - \mu_j), \frac{u_{N+1}-1}{2}) \in \mathcal{Q}$.

B Proofs

Proof of Theorem 1 (copy of Proof of Theorem 1 in WKS). The left-hand side of the distributionally robust constraint (1b) coincides with the optimal value of the following moment problem:

$$\begin{array}{ll} \text{maximize} & \displaystyle\int_{\mathcal{C}_I} v(\mathbf{x}, \mathbf{z}) \, \mathrm{d}\mu(\mathbf{z}, \mathbf{u}) \\ \text{subject to} & \mu \in \mathcal{M}_+\big(\mathbb{R}^P \times \mathbb{R}^Q\big), \\ & \displaystyle\int_{\mathcal{C}_I} [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}] \, \mathrm{d}\mu(\mathbf{z}, \mathbf{u}) = \mathbf{b}, \\ & \left.\begin{array}{l} \displaystyle\int_{\mathcal{C}_I} \mathbf{1}_{\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i\}} \, \mathrm{d}\mu(\mathbf{z}, \mathbf{u}) \ \geq \ \underline{p}_i \\ \displaystyle\int_{\mathcal{C}_I} \mathbf{1}_{\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i\}} \, \mathrm{d}\mu(\mathbf{z}, \mathbf{u}) \ \leq \ \overline{p}_i \end{array}\right\} \quad \forall i \in \mathcal{I}. \end{array}$$

By assumption, we have $\underline{p}_I = \bar{p}_I = 1$. Hence, every feasible measure μ in this problem is naturally identified with a probability measure $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ that is supported on $\mathcal{C}_I$. The dual of the moment problem is given by

$$\begin{array}{ll} \text{minimize} & \mathbf{b}^\top \boldsymbol{\beta} + \displaystyle\sum_{i \in \mathcal{I}} \left[\overline{p}_i \kappa_i - \underline{p}_i \lambda_i\right] \\ \text{subject to} & \boldsymbol{\beta} \in \mathbb{R}^K, \quad \boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^{\mathcal{I}}, \\ & [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \displaystyle\sum_{i \in \mathcal{I}} \mathbf{1}_{\{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i\}} [\kappa_i - \lambda_i] \ \geq \ v(\mathbf{x}, \mathbf{z}), \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_I. \end{array}$$

Strong duality is guaranteed by Proposition 3.4 in Shapiro (2001), which is applicable due to condition (A2). Next, the nesting condition (either (A3) or (A3*)) allows us to partition the support $\mathcal{C}_I$ into I nonempty and disjoint sets $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'},\ i = 1, \ldots, I$, where $\mathcal{D}(i)$ denotes the index set of strict subsets of $\mathcal{C}_i$. The constraint in the dual problem is therefore equivalent to the constraint set

$$[\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] \ \geq \ v(\mathbf{x}, \mathbf{z}), \qquad \forall (\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i,\ \forall i \in \mathcal{I}.$$

We can then reformulate the i-th constraint as

$$\max_{(\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i} \left\{ v(\mathbf{x}, \mathbf{z}) - [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} - \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] \right\} \ \leq \ 0.$$

Thus the robust expectation constraint in (1b) is satisfied if and only if

$$\begin{aligned} &\mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left[\bar{p}_i \kappa_i - p_i \lambda_i \right] \ \leq \ \tau, \\ &v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] - [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} \ \leq \ 0, \qquad \forall (\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i,\ \forall i \in \mathcal{I}, \end{aligned}$$

is satisfied by some $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{I}}_+$. □

Proof of Theorem 2. The left-hand side of the robust constraint (6) inherits convexity from v and is therefore maximized on the boundary of $\bar{\mathcal{C}}_i$. Due to the strict nesting condition (A3), the relevant boundary of $\bar{\mathcal{C}}_i$ coincides with the boundary of $\mathcal{C}_i$. This argument follows the final step in the proof of Theorem 1 in WKS. Consequently, the robust expectation constraint (1b) is satisfied if and only if there exists $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{I}}_+$ such that

$$\begin{aligned} &\mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left[\bar{p}_i \kappa_i - \underline{p}_i \lambda_i \right] \ \leq \ \tau, \\ &v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \ \leq \ 0, \qquad \forall \mathbf{z} \in \mathcal{C}_i,\ \forall i \in \mathcal{I}. \end{aligned}$$

We now build upon this reformulation and apply a biconjugate reformulation to the robust constraint, following Bertsimas et al. (2023a). This step is not present in WKS and allows us to express as an equivalent bilinear constraint, specifically:

$$\mathbf{w}^\top \mathbf{T}(\mathbf{x}) \mathbf{z} + \mathbf{t}(\mathbf{x})^\top \mathbf{w} - w_0 - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall (w_0, \mathbf{w}, \mathbf{z}) \in \Lambda_i,\ \forall i \in \mathcal{I},$$

where $\Lambda_i = \{(w_0, \mathbf{w}, \mathbf{z}) \mid \mathbf{z} \in \mathcal{C}_i,\ \mathbf{w} \in \mathrm{dom}(h^*),\ h^*(\mathbf{w}) \leq w_0\}$ and h^* denotes the convex conjugate of h, defined as $h^*(\mathbf{w}) = \sup_{\mathbf{y}} \{\mathbf{w}^\top \mathbf{y} - h(\mathbf{y})\}$. □

Proof of Lemma 1. We define the confidence sets in the NL ambiguity set as

$$\mathcal{C}_i = \mathcal{C}_I \cap P_i, \qquad P_i = \{\mathbf{z} : \mathbf{G}_i \mathbf{z} \le \mathbf{g}_i\} = \bigcap_{r=1}^{R_i} \{\mathbf{z} : \mathbf{G}_{ir}^\top \mathbf{z} \le g_{ir}\},$$

where $\mathbf{G}_i \in \mathbb{R}^{R_i \times P}$, $\mathbf{g}_i \in \mathbb{R}^{R_i}$, $\mathbf{G}_{ir}^\top$ denotes row r of $\mathbf{G}_i$, and g_{ir} component r of $\mathbf{g}_i$.

Fix $i \in \mathcal{I}$. By nestedness, removing all subsets of $\mathcal{C}_i$ is equivalent to removing only its maximal subsets collected in $\mathcal{G}(i)$, i.e., $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{k \in \mathcal{D}(i)} \mathcal{C}_k = \mathcal{C}_i \setminus \bigcup_{k \in \mathcal{G}(i)} \mathcal{C}_k$. Using the identity $A \setminus \bigcup_k B_k = \bigcap_k (A \setminus B_k)$, we obtain $\bar{\mathcal{C}}_i = \bigcap_{k \in \mathcal{G}(i)} (\mathcal{C}_i \setminus \mathcal{C}_k)$.

Next, fix $k \in \mathcal{G}(i)$. Since $\mathcal{C}_k = \mathcal{C}_I \cap P_k$ and $\mathcal{C}_i \subseteq \mathcal{C}_I$, it follows that

$$\mathcal{C}_i \setminus \mathcal{C}_k = \mathcal{C}_i \setminus (\mathcal{C}_I \cap P_k) = \mathcal{C}_i \setminus P_k = \mathcal{C}_i \cap P_k^c \quad \Rightarrow \quad \bar{\mathcal{C}}_i = \mathcal{C}_i \cap \bigcap_{k \in \mathcal{G}(i)} P_k^c,$$

where P_k^c denotes the complement of P_k. For each P_k^c, we have by De Morgan's law the following expression:

$$P_k^c = \left(\bigcap_{r=1}^{R_k} \{\mathbf{z} : \mathbf{G}_{kr}^\top \mathbf{z} \le g_{kr}\} \right)^c = \bigcup_{r=1}^{R_k} \{\mathbf{z} : \mathbf{G}_{kr}^\top \mathbf{z} \le g_{kr}\}^c = \bigcup_{r=1}^{R_k} \{\mathbf{z} : \mathbf{G}_{kr}^\top \mathbf{z} > g_{kr}\}.$$

Substituting this expression yields

$$\bar{\mathcal{C}}_i = \mathcal{C}_i \cap \bigcap_{k \in \mathcal{G}(i)} \left(\bigcup_{r=1}^{R_k} \{\mathbf{z} : \mathbf{G}_{kr}^\top \mathbf{z} > g_{kr}\} \right).$$

Distributing intersections over unions produces a finite union indexed by selections $j = (r_k)_{k \in \mathcal{G}(i)}$ with $r_k \in \{1, \ldots, R_k\}$

$$\bar{\mathcal{C}}_i = \bigcup_{j \in \mathcal{J}_i} \left(\mathcal{C}_i \cap \bigcap_{k \in \mathcal{G}(i)} \{\mathbf{z} : \mathbf{G}_{kr_k}^\top \mathbf{z} > g_{kr_k}\} \right) = \bigcup_{j \in \mathcal{J}_i} \bar{\mathcal{C}}_{ij}.$$

Each region $\bar{\mathcal{C}}_{ij}$ is convex, being the intersection of the convex set $\mathcal{C}_i$ with finitely many open halfspaces. Since each region corresponds to one choice $r_k \in \{1, \ldots, R_k\}$ per child polyhedron, we obtain

$$|\mathcal{J}_i| \le \prod_{k \in \mathcal{G}(i)} R_k.$$

□

Proof of Theorem 3. Since $\bar{\mathcal{C}}_i = \bigcup_{j \in \mathcal{J}_i} \bar{\mathcal{C}}_{ij}$, the i-th robust constraint (6) holds over $\bar{\mathcal{C}}_i$ if and only if it holds over each region $\bar{\mathcal{C}}_{ij}$ separately. Moreover, since each component function $v_l(\mathbf{x}, \cdot)$ is concave and finite-valued on an open set containing $\mathcal{C}_I$, it is continuous on $\mathcal{C}_I$. The affine coefficients $q_l(\mathbf{x}) - \mathbf{r}_l(\mathbf{x})^\top \mathbf{z}$ are continuous in $\mathbf{z}$. Hence, $v(\mathbf{x}, \cdot)$, and therefore the left-hand side of the robust constraint (6), is continuous on $\mathcal{C}_I$. Thus the constraint holds over $\bar{\mathcal{C}}_{ij}$ if and only if it holds over its closure $\tilde{\mathcal{C}}_{ij}$. Consequently, the robust expectation constraint (1b) is satisfied if and only if there exists $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^{\mathcal{I}}$ such that

$$\begin{aligned} &\mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left[\bar{p}_i \kappa_i - p_i \lambda_i\right] \ \leq \ \tau, \\ &v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] - [\mathbf{A}\mathbf{z}]^\top \boldsymbol{\beta} \ \leq \ 0, \quad \forall \mathbf{z} \in \tilde{\mathcal{C}}_{ij}, \ \forall i \in \mathcal{I}, \ \forall j \in \mathcal{J}_i. \end{aligned}$$

□

Proof of Theorem 4. Fix $\mathbf{x} \in \mathcal{X}$. By assumption, $\mathbb{E}_{\mathbb{P}}[\mathbf{z}] = \boldsymbol{\mu}$. Since $v(\mathbf{x}, \mathbf{z})$ is concave in $\mathbf{z}$, Jensen's inequality yields

$$\mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})] \leq v(\mathbf{x}, \mathbb{E}_{\mathbb{P}}[\mathbf{z}]) = v(\mathbf{x}, \boldsymbol{\mu}).$$

As this holds for every $\mathbb{P} \in \mathcal{P}$, we obtain

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})] \leq v(\mathbf{x}, \boldsymbol{\mu}).$$

To show tightness, consider the degenerate distribution at $\mathbf{z} = \boldsymbol{\mu}$, which belongs to $\mathcal{P}$ by assumption. Under this distribution, $\mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \mathbf{z})] = v(\mathbf{x}, \boldsymbol{\mu})$, which establishes the result. □

Proof of Theorem 5. Under NL ambiguity, the DRO constraint can be written as a collection of robust constraints over the sets $\tilde{\mathcal{C}}_{ij}$ by Theorem 3. For each $i \in \mathcal{I}, j \in \mathcal{J}_i$, this yields

$$f_i(\mathbf{x}, \mathbf{z}) := v(\mathbf{x}, \mathbf{z}) - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall \mathbf{z} \in .$$

Since $f_i(\mathbf{x}, \mathbf{z})$ is concave in $\mathbf{z}$, we can apply Theorem 2 of Ben-Tal et al. (2015), which yields that this constraint holds if and only if there exists a vector $\mathbf{y}^{ij}$ such that

$$\delta^*(\mathbf{y}^{ij} \mid) - (f_i)_*(\mathbf{x}, \mathbf{y}^{ij}) \leq 0. \tag{24}$$

Using the definition of the concave conjugate, we first obtain

$$\begin{aligned} (f_i)_*(\mathbf{x}, \mathbf{y}) &= \inf_{\mathbf{z} \in \mathrm{dom}(f_i(\mathbf{x}, \cdot))} \left\{\mathbf{y}^\top \mathbf{z} - f_i(\mathbf{x}, \mathbf{z})\right\} \\ &= \inf_{\mathbf{z} \in \mathrm{dom}(v)} \left\{\mathbf{z}^\top (\mathbf{y} + \mathbf{A}^\top \boldsymbol{\beta}) - v(\mathbf{x}, \mathbf{z})\right\} + \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \\ &= v_*\big(\mathbf{x}, \, \mathbf{y} + \mathbf{A}^\top \boldsymbol{\beta}\big) + \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}). \end{aligned}$$

Substituting this expression into (24) yields this result. Since δ^* is convex in $\mathbf{y}$ and v_* is concave in $(\mathbf{x}, \mathbf{y})$, the resulting constraint is convex in $(\mathbf{x}, \mathbf{y})$. □

C Convergence assumptions

This appendix verifies the assumptions required for convergence of the framework of Bertsimas et al. (2024) for both convex functions in Section 3 and SLC functions in Section 4.

For convex functions, we obtained the robust reformulation in (9):

$$\min_{\mathbf{x},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\lambda}} \quad \mathbf{c}^\top \mathbf{x} \tag{25a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X}, \quad \boldsymbol{\kappa} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \tag{25b}$$

$$\mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \left(\overline{p}_i \, \kappa_i - \underline{p}_i \, \lambda_i \right) \leq \tau, \tag{25c}$$

$$\mathbf{w}^\top \mathbf{T}(\mathbf{x})\mathbf{z} + \mathbf{w}^\top \mathbf{t}(\mathbf{x}) - w_0 - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - \mathbf{z}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \quad \forall (w_0, \mathbf{w}, \mathbf{z}) \in \Lambda_i, \ \forall i \in \mathcal{I}. \tag{25d}$$

- The objective function (25a) is proper, closed and convex, while each constraint function is linear or bilinear.
- The uncertainty sets Λ_i are nonempty and convex. Convexity follows from the convexity of $\mathcal{C}_i$, the convexity of $\mathrm{dom}(h^*)$, and the fact that $\{h^*(\mathbf{w}) \leq w_0\}$ is the epigraph of the convex function h^*. Although Λ_i is not compact due to the absence of an upper bound on w_0, this does not affect the robust constraint as its worst-case realization always selects the minimal feasible value $w_0 = h^*(\mathbf{w})$. Moreover, under a boundedness assumption on $\mathrm{dom}(h^*)$ (e.g., when h is Lipschitz), the relevant uncertainty $(\mathbf{w}, \mathbf{z})$ ranges over a bounded set.
- The feasible region in $(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\lambda})$ is convex and admits a bounded formulation. For the boundedness argument below, we use the following sufficient regularity conditions:

  (B1) $\mathbf{A}$ is nonsingular and $\mathbb{E}[\mathbf{z}] = \mathbf{A}^{-1}\mathbf{b} \in \mathrm{int}(\mathcal{C}_I)$.

  (B2) The probability bounds are nondegenerate on the nested structure, i.e., there exists $\boldsymbol{\pi} \in [\underline{\mathbf{p}}, \overline{\mathbf{p}}]$ such that

  $$\pi_m - \sum_{\ell \in \mathcal{G}(m)} \pi_\ell > 0, \qquad \forall m \in \mathcal{I},$$

  where $\mathcal{G}(m)$ denotes the direct children of m in the nested sets.

  - Since $\mathcal{X}$ is compact, $\mathbf{x}$ is bounded.
  - To show that β_i is bounded above, we maximize β_i over the constraints that involve $\boldsymbol{\beta}$. If this auxiliary maximization problem has a finite optimal value, then β_i is bounded. The duality step below relies on the result of Beck and Ben-Tal (2009); the required regularity conditions are satisfied here because $\mathcal{C}_I$ is compact and condition (B1) provides the required Slater-type condition. We verify finiteness through the following primal-dual pair:

$$\begin{aligned} \max_{\boldsymbol{\beta}} \quad & \mathbf{e}_i^\top \boldsymbol{\beta} \\ \text{s.t.} \quad & \mathbf{b}^\top \boldsymbol{\beta} \le U, \\ & -(\mathbf{A}\mathbf{z})^\top \boldsymbol{\beta} \le V \quad \forall \mathbf{z} \in \mathcal{C}_I. \end{aligned} \qquad \begin{aligned} \min_{\alpha,\nu,\mathbf{z}} \quad & U\alpha + V\nu \\ \text{s.t.} \quad & \alpha \mathbf{b} - \nu \mathbf{A}\mathbf{z} = \mathbf{e}_i, \\ & \alpha \ge 0, \quad \nu \ge 0, \quad \mathbf{z} \in \mathcal{C}_I. \end{aligned}$$

Hence, β_i is bounded above whenever the dual problem is feasible. By condition (B1), $\mathbf{A}^{-1}\mathbf{b} \in \operatorname{int}(\mathcal{C}_I)$. Therefore, for each i, there exists an $\varepsilon > 0$ such that

$$\mathbf{z}^- := \mathbf{A}^{-1}\mathbf{b} - \varepsilon \mathbf{A}^{-1}\mathbf{e}_i \in \mathcal{C}_I.$$

Choose $\alpha = \frac{1}{\varepsilon}, \nu = \frac{1}{\varepsilon}$. Then

$$\alpha \mathbf{b} - \nu \mathbf{A}\mathbf{z}^- = \frac{1}{\varepsilon}\mathbf{b} - \frac{1}{\varepsilon}\mathbf{A}(\mathbf{A}^{-1}\mathbf{b} - \varepsilon \mathbf{A}^{-1}\mathbf{e}_i) = \mathbf{e}_i,$$

so the dual is feasible and the primal objective is finite, so β_i is bounded above.

Boundedness from below follows analogously by considering $\max_{\boldsymbol{\beta}}(-\mathbf{e}_i^\top \boldsymbol{\beta})$.

– Finally, consider $(\boldsymbol{\kappa}, \boldsymbol{\lambda})$. In the special case $\underline{p}_i = \overline{p}_i =: p_i$, these variables enter both (25c) and (25d) only through the differences $\delta_i := \kappa_i - \lambda_i$. Without loss of generality, we may instead work with the variables $\boldsymbol{\delta}$ and choose the representative

$$\kappa_i = \delta_i^+, \qquad \lambda_i = (-\delta_i)^+.$$

In this equality case, condition (B2) reduces to

$$q_m := p_m - \sum_{\ell \in \mathcal{G}(m)} p_\ell > 0, \qquad \forall m \in \mathcal{I}.$$

To bound δ_j from above, consider the following primal-dual pair:

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \delta_j \\ \text{s.t.} \quad & \sum_{i \in \mathcal{I}} p_i \delta_i \le U, \\ & -\sum_{i' \in \mathcal{A}(i)} \delta_{i'} \le R_i, \quad \forall i \in \mathcal{I}, \end{aligned} \qquad \begin{aligned} \min_{\alpha,\boldsymbol{\eta}} \quad & U\alpha + \sum_{i \in \mathcal{I}} R_i \eta_i \\ \text{s.t.} \quad & p_m \alpha - \sum_{i \in \mathcal{D}(m) \cup \{m\}} \eta_i = \mathbf{1}_{\{m=j\}}, \quad \forall m \in \mathcal{I}, \\ & \alpha \ge 0, \quad \boldsymbol{\eta} \ge \mathbf{0}. \end{aligned}$$

Here, $\mathcal{A}(i) = \{\, i \,\} \cup \{\, i' \in \mathcal{I} : \mathcal{C}_i \subset \mathcal{C}_{i'} \,\}$ and $\mathcal{D}(i) = \{\, i' \in \mathcal{I} : \mathcal{C}_{i'} \subset \mathcal{C}_i \,\}$ collect, respectively, all supersets and strict subsets of $\mathcal{C}_i$.

Choose $\alpha \ge \frac{1}{q_j}$ and set $\eta_m = q_m \alpha - \mathbf{1}_{\{m=j\}} + \mathbf{1}_{\{m=pa(j)\}}$, where $pa(j)$ denotes the parent of j; if j has no parent, the last indicator is taken to be zero. Then $\boldsymbol{\eta} \geq \mathbf{0}$: for $m = j$ this follows from $\alpha \geq 1/q_j$ (and $q_j > 0$ by assumption (ii)), for $m = pa(j)$ we have $\eta_m = q_m\alpha + 1$, and for all other m, we have $\eta_m = q_m\alpha$. Hence,

$$\sum_{i \in \mathcal{D}(m) \cup \{m\}} \eta_i = \alpha p_m - \mathbf{1}_{\{j \in \mathcal{D}(m) \cup \{m\}\}} + \mathbf{1}_{\{pa(j) \in \mathcal{D}(m) \cup \{m\}\}}.$$

Since $pa(j)$ is the parent of j, the two indicator terms cancel except for $m = j$. Therefore,

$$p_m\alpha - \sum_{i \in \mathcal{D}(m) \cup \{m\}} \eta_i = \mathbf{1}_{\{m=j\}}, \qquad \forall m \in \mathcal{I}.$$

Thus the dual is feasible and hence δ_j is bounded above.

For the general case $\underline{p}_i \leq \overline{p}_i$, define again $\delta_i := \kappa_i - \lambda_i$. For fixed δ_i, choosing

$$\kappa_i = \delta_i^+, \qquad \lambda_i = (-\delta_i)^+$$

preserves all robust constraints and weakly decreases the left-hand side of (25c). Hence, this representative can be chosen without loss of generality.

Let $\pi_i \in [\underline{p}_i, \overline{p}_i]$ be chosen as in condition (B2), and define

$$q_m^\pi := \pi_m - \sum_{\ell \in \mathcal{G}(m)} \pi_\ell > 0, \qquad \forall m \in \mathcal{I}.$$

For every $\delta_i \in \mathbb{R}$ we have $\overline{p}_i\delta_i^+ - \underline{p}_i(-\delta_i)^+ \geq \pi_i\delta_i$. Therefore, every feasible solution of the original auxiliary problem for the upper bound,

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \delta_j \\ \text{s.t.} \quad & \sum_{i \in \mathcal{I}} \left(\overline{p}_i\delta_i^+ - \underline{p}_i(-\delta_i)^+ \right) \leq U, \\ & -\sum_{i' \in \mathcal{A}(i)} \delta_{i'} \leq R_i, \qquad \forall i \in \mathcal{I}, \end{aligned}$$

is feasible for the linear relaxation

$$\begin{aligned} \max_{\boldsymbol{\delta}} \quad & \delta_j \\ \text{s.t.} \quad & \sum_{i \in \mathcal{I}} \pi_i\delta_i \leq U, \\ & -\sum_{i' \in \mathcal{A}(i)} \delta_{i'} \leq R_i, \qquad \forall i \in \mathcal{I}. \end{aligned}$$

Therefore, any finite upper bound for this relaxation is also a valid upper bound for the original auxiliary problem. The same dual certificate as in the equality case applies to this relaxation, with p_i and q_m replaced by π_i and q_m^{π}. Thus δ_j is bounded above.

The lower bound follows analogously by replacing the objective by $-\delta_j$.

Similarly, for SLC functions we obtained the robust reformulation in (16), which satisfies the convergence assumptions:

- The objective function is proper, closed and convex, while each constraint function is an SLC function in the form of (14).
- The uncertainty sets $\tilde{\mathcal{C}}_{ij}$ are nonempty, compact and convex.
- The feasible region in $(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\lambda})$ is convex and bounded (see reasoning convex functions).

D Extensive derivations

D.1 Dual of RPT relaxation ASP

The full RPT relaxation of the distributionally robust ASP in (19) is given by

$$\begin{aligned}
\min_{\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\lambda}} \quad & \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \Big(\overline{p}_i \, \kappa_i - \underline{p}_i \, \lambda_i \Big) \\
\text{s.t.} \quad & \mathbf{x} \geq 0, \quad \mathbf{1}^\top \mathbf{x} = T, \\
& \boldsymbol{\kappa} \geq 0, \quad \boldsymbol{\lambda} \geq 0, \\
& \mathrm{Tr}\Big(\mathbf{T}^\top \mathbf{V}^i \Big) - {\mathbf{w}^i}^\top \mathbf{T} \mathbf{x} - \sum_{i' \in \mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) - {\mathbf{z}^i}^\top \mathbf{A}^\top \boldsymbol{\beta} \leq 0, \ \forall (\mathbf{w}^i, \mathbf{z}^i, \mathbf{V}^i) \in \hat{\Theta}_i, \ \forall i \in \mathcal{I}.
\end{aligned}$$

Here, a separate copy $(\mathbf{w}^i, \mathbf{z}^i, \mathbf{V}^i)$ is introduced for each robust constraint indexed by $i \in \mathcal{I}$. Using duality theory from Beck and Ben-Tal (2009), the dual problem introduces one dual multiplier ρ^i for each i, yielding the following joint formulation:

$$\begin{aligned}
\max_{\pi, \{\rho^i, \mathbf{w}^i, \mathbf{z}^i, \mathbf{V}^i\}_{i \in \mathcal{I}}} \quad & \sum_{i \in \mathcal{I}} \rho^i \, \mathrm{Tr}\Big(\mathbf{T}^\top \mathbf{V}^i \Big) - T \, \pi \\
\text{s.t.} \quad & \pi \, \mathbf{1} \geq \sum_{i \in \mathcal{I}} \rho^i \mathbf{T}^\top \mathbf{w}^i, \\
& \overline{\mathbf{p}} \geq \sum_{i \in \mathcal{I}} \rho^i \mathbf{s}^{(i)}, \\
& \sum_{i \in \mathcal{I}} \rho^i \mathbf{s}^{(i)} \geq \underline{\mathbf{p}}, \\
& \sum_{i \in \mathcal{I}} \rho^i \mathbf{A} \mathbf{z}^i = \mathbf{b}, \\
& (\mathbf{w}^i, \mathbf{z}^i, \mathbf{V}^i) \in \hat{\Theta}_i, \qquad \forall i \in \mathcal{I}, \\
& \rho^i \geq 0, \qquad \forall i \in \mathcal{I},
\end{aligned}$$

where $\mathbf{s}^{(i)} \in \{0,1\}^I$ denotes the incidence vector of $\mathcal{A}(i)$, i.e., $s_j^{(i)} = 1$ if $j \in \mathcal{A}(i)$ and $s_j^{(i)} = 0$ otherwise. This problem contains bilinear terms of the form $\rho^i \mathbf{w}^i$, $\rho^i \mathbf{z}^i$, and $\rho^i \mathbf{V}^i$ in both the objective and the constraints. Since the multipliers satisfy $\rho^i \geq 0$, the problem is solved using the hidden convexity framework in Gorissen et al. (2025). In particular, we introduce

$$\tilde{\mathbf{w}}^i = \rho^i \mathbf{w}^i, \qquad \tilde{\mathbf{z}}^i = \rho^i \mathbf{z}^i, \qquad \tilde{\mathbf{V}}^i = \rho^i \mathbf{V}^i.$$

Substituting these variables eliminates the bilinear terms and yields an equivalent formulation that is linear in $(\rho^i, \tilde{\mathbf{w}}^i, \tilde{\mathbf{z}}^i, \tilde{\mathbf{V}}^i)$. The constraints defining $\hat{\Theta}_i$ then become perspective constraints of the original set, which preserve convexity because $\hat{\Theta}_i$ is convex. Consequently, the resulting problem can be solved as a convex optimization problem.

In the special case $I = 1$, we have by assumption $\overline{p}_1 = \underline{p}_1 = 1$ and $\mathbf{s}^{(1)} = 1$. The second and third dual constraint imply $\rho^1 = 1$, and the dual problem simplifies to:

$$\begin{aligned} \max_{\pi, \mathbf{w}, \mathbf{z}, \mathbf{V}} \quad & \mathrm{Tr}\Big(\mathbf{T}^\top \mathbf{V}\Big) - T\,\pi \\ \text{s.t.} \quad & \pi\,\mathbf{1} \geq \mathbf{T}^\top \mathbf{w}, \\ & \mathbf{A}\mathbf{z} = \mathbf{b}, \\ & (\mathbf{w}, \mathbf{z}, \mathbf{V}) \in \hat{\Theta}_1. \end{aligned}$$

D.2 Concave conjugate capital budgeting

We derive the concave conjugate of the cost function

$$v(\mathbf{x}, \mathbf{z}) = \sum_{n=1}^{N} x_n \sum_{t=0}^{T} -\frac{F_{nt}}{(1+z_n)^t},$$

which can be written as

$$v(\mathbf{x}, \mathbf{z}) = \sum_{n=1}^{N} x_n f_n(z_n), \quad \text{where} \quad f_n(z_n) := \sum_{t=0}^{T} f_{nt}(z_n) \quad \text{and} \quad f_{nt}(z_n) := -\frac{F_{nt}}{(1+z_n)^t}.$$

Only the components z_n with $n = 1, \ldots, N$ enter the objective; any additional auxiliary uncertainty components do not affect the present derivation.

The function v is linear in $\mathbf{x}$ and separable in the uncertain parameters z_n. Furthermore, each function f_n is a sum of functions f_{nt}. These properties allow us to use the homogeneity and convolution properties of concave conjugates.

We now compute the individual concave conjugates $(f_{nt})_*(s)$, assuming $z_n \geq 0$.

- For $t = 0$, we have $f_{n0}(z_n) = -F_{n0}$, so its concave conjugate is

$$(f_{n0})_*(s) = \begin{cases} F_{n0} & \text{if } s \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

- For $t = 1, \ldots, T$, by definition:

$$(f_{nt})_*(s) = \inf_{z \geq 0} \{sz - f_{nt}(z)\} = \inf_{z \geq 0} \left\{ sz + \frac{F_{nt}}{(1+z)^t} \right\}.$$

For $s > 0$, the first-order condition gives

$$s - t \cdot \frac{F_{nt}}{(1+z)^{t+1}} = 0 \quad \iff \quad z^* = \left(t \cdot \frac{F_{nt}}{s} \right)^{\frac{1}{t+1}} - 1.$$

Substituting this back into the objective yields

$$sz^* + \frac{F_{nt}}{(1+z^*)^t} = (t+1)t^{-\frac{t}{t+1}} F_{nt}^{\frac{1}{t+1}} s^{\frac{t}{t+1}} - s.$$

By continuity, the same expression gives $(f_{nt})_*(0) = 0$. Hence,

$$(f_{nt})_*(s) = \begin{cases} \alpha_{nt} s^{\frac{t}{t+1}} - s & \text{if } s \geq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

where we define $\alpha_{nt} = (t+1)t^{-\frac{t}{t+1}} F_{nt}^{\frac{1}{t+1}}$.

Using the convolution rule, we obtain

$$(f_n)_*(y_n) = \sup_{\{s_{nt}\}_t} \left\{ F_{n0} + \sum_{t=1}^{T} \left(\alpha_{nt} s_{nt}^{\frac{t}{t+1}} - s_{nt} \right) \;\middle|\; \sum_{t=0}^{T} s_{nt} = y_n,\ s_{nt} \geq 0\ \forall t \right\}.$$

By homogeneity, we have $v_*(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{N} x_n (f_n)_*(\frac{y_n}{x_n})$ for indices n such that $x_n > 0$. To obtain a formulation that is well defined when $x_n = 0$, we introduce the variables $r_{nt} = x_n s_{nt}$ for $t = 1, \ldots, T$. Then $\sum_{t=0}^{T} s_{nt} = \frac{y_n}{x_n} \iff x_n s_{n0} + \sum_{t=1}^{T} r_{nt} = y_n$. Since $s_{n0} \geq 0$ and s_{n0} does not appear in the objective, this is equivalent to the condition $\sum_{t=1}^{T} r_{nt} \leq y_n$.

Therefore, the concave conjugate of v can be represented as

$$v_*(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{N} x_n F_{n0} + \sup_{\{r_{nt}\}_{n,t}} \left\{ \sum_{n=1}^{N} \sum_{t=1}^{T} \left(\alpha_{nt} x_n^{\frac{1}{t+1}} r_{nt}^{\frac{t}{t+1}} - r_{nt} \right) \mid \sum_{t=1}^{T} r_{nt} \leq y_n\ \forall n,\ r_{nt} \geq 0\ \forall n, t \right\}.$$

E Instances

E.1 Appointment scheduling

Parameter	**Instance 3**	**Instance 4**
Number of customers N	8	8
Mean service times μ	[53.60, 37.51, 51.32, 58.40, 30.58, 42.15, 37.54, 30.68]	[46.35, 40.86, 59.01, 44.54, 48.37, 52.12, 33.41, 57.13]
Uncertainty level ϵ	0.05	0.085
Standard deviation σ	$\sigma_i = \epsilon \cdot \mu_i$	$\sigma_i = \epsilon \cdot \mu_i$
Planning horizon T_d	344.886	387.613
Overtime penalty γ	2	2
Support bounds $[a_i, b_i]$	[0,100]	[0,100]
Precision tolerance ρ	0.01	0.01
Budget parameter ϕ	$\sum_{i=1}^{N}(\mu_i + \sigma_i)$	$\sum_{i=1}^{N}(\mu_i + \sigma_i)$
Probability $\mathcal{C}_1$ $[\underline{p}_1, \overline{p}_1]$	$[0.9, 1]$	$[0.9, 1]$

Table 8: Parameter settings for instances 3 and 4 for appointment scheduling.

N	**Mean service times** μ	**Planning horizon** T_d
8	[53.60, 37.51, 51.32, 58.40, 30.58, 42.15, 37.54, 30.68]	344.886
10	[35.23, 36.60, 49.37, 34.24, 57.94, 53.26, 33.05, 50.73, 49.88, 59.35]	470.778
12	[54.90, 37.38, 56.44, 30.99, 49.64, 36.55, 32.38, 41.71, 47.94, 57.91, 43.49, 39.45]	540.468
14	[46.32, 50.82, 54.70, 55.77, 35.88, 30.90, 49.11, 46.73, 51.58, 33.85, 36.93, 38.20, 35.00, 51.31]	629.690
16	[54.03, 47.30, 48.57, 51.10, 35.82, 41.93, 34.06, 52.74, 59.11, 45.06, 52.65, 45.67, 48.74, 39.19, 35.56, 45.49]	750.990
18	[36.09, 35.19, 30.40, 46.86, 33.01, 37.81, 35.70, 49.17, 47.03, 39.74, 40.45, 36.45, 33.72, 55.42, 36.40, 42.20, 55.73, 58.22]	763.098
20	[32.92, 45.41, 32.75, 43.34, 47.25, 59.88, 55.76, 43.09, 52.00, 39.65, 53.69, 44.68, 48.98, 59.97, 31.33, 58.13, 36.42, 55.79, 41.95, 51.45]	950.393
Shared parameters:		
Uncertainty level ϵ		0.15
Standard deviation σ		$\sigma_i = \epsilon \cdot \mu_i$
Overtime penalty γ		2
Support bounds $[a_i, b_i]$		[0,100]
Precision level ρ		0.01

Table 9: Parameter settings for appointment scheduling instances with varying N.

E.2 Capital budgeting

Parameter	Instance
Number of projects N	12
Number of periods T	5
Cash flow matrix $\mathbf{F}$	$\begin{bmatrix} 0.2978 & 0.1588 & 0.3727 & 0.3815 & 0.4812 & 0.1584 \\ 0.1912 & 0.4930 & 0.1213 & 0.3678 & 0.2613 & 0.1472 \\ 0.2022 & 0.4673 & 0.2334 & 0.3467 & 0.2227 & 0.3007 \\ 0.2585 & 0.3262 & 0.1230 & 0.1915 & 0.3566 & 0.3269 \\ 0.2509 & 0.2267 & 0.1734 & 0.3811 & 0.2946 & 0.5024 \\ 0.4986 & 0.4588 & 0.4778 & 0.4618 & 0.2062 & 0.3452 \\ 0.2633 & 0.3725 & 0.4827 & 0.4282 & 0.4490 & 0.3212 \\ 0.4088 & 0.2580 & 0.3913 & 0.2780 & 0.1743 & 0.4874 \\ 0.4042 & 0.3049 & 0.3989 & 0.4782 & 0.1945 & 0.4861 \\ 0.2240 & 0.3117 & 0.2884 & 0.4221 & 0.3745 & 0.2058 \\ 0.2386 & 0.4749 & 0.4800 & 0.3990 & 0.2993 & 0.3867 \\ 0.2407 & 0.3308 & 0.2658 & 0.1514 & 0.1306 & 0.5050 \end{bmatrix}$
Project costs $\mathbf{q}$	[1.4539, 2.3962, 2.5157, 2.7330, 2.5367, 1.7377, 1.0092, 1.0575, 1.0222, 1.9128, 1.5745, 1.1665]
Budget θ	8
Mean discount rates $\boldsymbol{\mu}$	[0.01125, 0.02625, 0.04125, 0.05625, 0.07125, 0.08625, 0.10125, 0.11625, 0.13125, 0.14625, 0.16125, 0.17625]
Support lower bounds $\mathbf{a}$	[0.000, 0.005, 0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055]
Support upper bounds $\mathbf{b}$	[0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.175, 0.200, 0.225, 0.250, 0.275, 0.300]
Thresholds τ_i	$\tau_i = (\mu_i + a_i)/2$
Probability $\mathcal{C}_1$ $[\underline{p}_1, \overline{p}_1]$	[0.65, 0.75]

Table 10: Parameter settings for the capital budgeting instance.