On the Optimality of Affine Decision Rules in Robust and Distributionally Robust Optimization

Angelos Georghiou[1], Angelos Tsoukalas[2], and Wolfram Wiesemann[3]

[1] *Department of Business and Public Administration, University of Cyprus, Cyprus*

[2] *Rotterdam School of Management, Erasmus University Rotterdam, The Netherlands*

[3] *Imperial College Business School, Imperial College London, United Kingdom*

May 20, 2021

Abstract

We propose tight conditions under which two-stage robust and distributionally robust optimization problems are optimally solved in affine decision rules. Contrary to previous work, our conditions do not impose any structure on the support of the uncertain problem parameters, and they ensure point-wise (as opposed to worst-case) optimality of affine decision rules. The absence of support restrictions allows us to consider rich classes of uncertainty sets as well as transfer non-linearities to the support via liftings, while the point-wise optimality ensures that decision rules remain optimal for broad classes of distributionally robust optimization problems, including data-driven problems over ϕ-divergence or Wasserstein ambiguity sets. We show that our conditions are met by problems in diverse application domains, such as logistics, inventory and supply chain management, flexible production planning and healthcare scheduling. We also show how problems that 'almost' meet our conditions can sometimes be solved by complementing affine decision rules with methods that isolate the complicating problem structure.

Keywords: Affine Decision Rules; (Distributionally) Robust Optimization.

1 Introduction

Robust and distributionally robust optimization problems faithfully model the uncertainty and ambiguity inherent in practical decision problems. Moreover, their two- and multi-stage extensions account for the dynamics of real-life decision making, where some decisions can be postponed and thus taken under a richer information base. Unfortunately, however, the presence of multiple decision stages leads to significant theoretical and computational challenges. In fact, robust linear programs are NP-hard already when they involve two decision stages (Guslitser, 2002), and the solution schemes for two- and multi-stage robust optimization problems, such as (nested) Benders' decomposition (Jiang et al., 2014; Thiele et al., 2010; Zhao et al., 2013), semi-infinite programming (Zeng and Zhao, 2013; Ayoub and Poss, 2016), uncertainty set partitioning (Bertsimas and Dunning, 2016; Postek and den Hertog, 2016; Georghiou et al., 2020), Fourier-Motzkin elimination (Zhen et al., 2018) and robust dual dynamic programming (Georghiou et al., 2019), often exhibit an unfavourable scaling in the size of the problem.

A popular heuristic for generating suboptimal decisions in two- and multi-stage problems approximates the recourse decisions via *affine decision rules*, which impose an affine dependence of these decisions on the revealed uncertainties. Originally proposed by Charnes et al. (1958) for the production scheduling of heating oil, affine decision rules have been largely neglected by the stochastic programming community due to their suboptimality even in well-structured problem classes as well as the difficulty to meaningfully bound the optimality gap (Garstka and Wets, 1974). They resurfaced several decades later in the robust optimization (Ben-Tal et al., 2004), control theory (Skaf and Boyd, 2010) and stochastic programming (Kuhn et al., 2011) domains, where they have subsequently been generalized to segregated affine (Chen et al., 2008; Chen and Zhang, 2009; Goh and Sim, 2010), piecewise affine (Bertsimas and Georghiou, 2015; Georghiou et al., 2015), polynomial and trigonometric (Bampou and Kuhn, 2011; Bertsimas et al., 2011c) decision rules. We refer to the survey of Delage and Iancu (2015) for a detailed review of the decision rule literature.

In this paper, we develop conditions under which affine decision rules are *optimal* in two-stage robust and distributionally robust optimization problems. It comes at no surprise that such conditions must be restrictive. Broadly speaking, our conditions apply to problems where a part of the first-stage decisions are binary and select which second-stage constraints are binding at optimality. By ensuring that this constraint set imposes an affine dependence of the second-stage decisions on the uncertain problem parameters, we can guarantee that optimal affine decision rules exist. Contrary to prior optimality results for decision rules, we do not impose any assumptions on the support of the uncertain problem parameters. This allows us to model rich classes of dependencies in robust optimization problems, and it also enables us to transfer non-linearities to the support via liftings. Moreover, and again in contrast to the existing optimality results, our conditions ensure point-wise (as opposed to worst-case) optimality of affine decision rules, which implies that our results extend to a broad class of distributionally robust optimization problem for which, to our best knowledge, no prior optimality results exist. Our results also allow us to characterize broader classes of problems that are optimally solved in richer classes of decision rules, such as piecewise affine, polynomial and trigonometric decision rules.

Perhaps surprisingly, our conditions are met by various formulations of logistics, inventory and supply chain management, flexible production planning and medical scheduling problems. Also, our conditions may often be met 'approximately', that is, they would be met if it was not for a small set of complicating variables and/or constraints. Isolating such problem structure allows us to employ optimal affine decision rules for the benign part of the problem, while the complicating part can be dealt with separately, for example through a lifting of the uncertainty/ambiguity set or a K-adaptability formulation. This situation is akin to integer programming, where Lagrangian relaxations often allow us to isolate complicating aspects of the problem, and the remainder of the problem can be solved optimally as a linear program due to the presence of a totally unimodular constraint matrix.

We summarize the main contributions of this work as follows.

(i) We develop optimality conditions for affine decision rules in two-stage robust and distributionally robust optimization problems. Our conditions are tight in the sense that there are problems satisfying all but one of the conditions that do not admit optimal affine decision rules. We are not aware of any prior optimality results for affine decision rules in distributionally robust optimization problems.

(ii) We show how complicating problem structure that precludes the optimality of affine decision rules can be isolated by lifting the support or employing a K-adaptability formulation. These techniques allow us to significantly broaden the class of problems for which affine decision rules are optimal.

(iii) We apply our results to problems from diverse application domains, several of which are solved optimally in affine decision rules for the first time.

Several papers characterize the geometry of recourse decisions in stochastic and (distributionally) robust optimization. Garstka and Wets (1974) investigate the optimal structure of decision rules in stochastic programming. They show that two- and multi-stage stochastic linear programs with right-hand side uncertainty are optimized by piecewise affine decision rules, and they conclude that affine decision rules are very restrictive. Most works in the robust and distributionally robust optimization domain take the suboptimality of constant and affine decision rules as given, and they focus on quantifying the optimality gap of these decision rules for specific problem classes. In one of the earliest attempts, Bertsimas and Goyal (2010) show that *constant* decision rules perform well if either the uncertainty set or the probability distribution is symmetric. The results are extended to multi-stage problems and finite adaptability formulations by Bertsimas et al. (2011b) and Housni and Goyal (2018), to non-linear problems by Bertsimas and Goyal (2013), and to problems with uncertain packing constraints by Bertsimas et al. (2015) and Awasthi et al. (2019). The substantially more flexible affine decision rules, while typically suboptimal as well, allow to significantly reduce the optimality gap compared to constant decision rules. Bertsimas and Goyal (2012) relate the optimality gap of affine decision rules in two-stage robust optimization problems with right-hand side uncertainty to the number of constraints and uncertain parameters. In a similar spirit, Bertsimas and Bidkhori (2015) quantify the optimality gap of affine decision rules by studying the distance of the uncertainty set to the smallest enclosing simplex. In a recent paper, Housni and Goyal (2021) study the performance of affine policies in two-stage robust optimization problems with right-hand side uncertainty where the uncertainty sets constitute intersections of budget sets.

As expected, the cases where the affine decision rules are optimal are rare, and they require a benign problem structure to be present. Bertsimas and Goyal (2012) identify that affine decision rules are optimal in two-stage robust linear optimization problems with right-hand side uncertainty if the uncertainty set is a simplex. This is intuitive as the linear problem structure causes the worst-case parameter realizations to be attained at the extreme points of the uncertainty set, and the degrees of freedom in the affine decision rules match the number of extreme points in the simplex. Bertsimas et al. (2010) show that affine decision rules are optimal in a multi-stage robust inventory management problem that considers a single product and that accounts for ordering, inventory holding and backlogging costs. A crucial assumption in this work is that the uncertainty set for the stage-wise customer demands is a hyperrectangle. The result was later extended by Iancu et al. (2013) to problem instances where the corner points of the uncertainty set form a subset of the extreme points of the $[0, 1]$-hypercube and the objective is convex and supermodular. The authors show that such problems find applications in two-echelon supply chains with inventory capacity investments. In a similar line of research, Ardestani-Jaafari and Delage (2016) show that affine decision rules are optimal in a class of inventory management problems where the uncertainty set is the intersection of the 1- and ∞-norm balls. Finally, Simchi-Levi et al. (2019) show that affine decision rules are optimal in a two-stage robust medical supply chain design problem if the uncertain demands are modelled by a budget-type uncertainty set and the supply network has a tree structure.

All of the previously discussed optimality results have in common that they establish the worst-case optimality of affine decision rules in robust optimization problems, and they rely on the interplay of the worst-case scenarios in the uncertainty sets with the structure of the problem. In contrast, Gounaris et al. (2013) show that affine decision rules are optimal in two-stage robust vehicle routing problems, independent of the geometry of the uncertainty set. It turns out that their result, which is established through an intricate *ad hoc* analysis, emerges naturally as a special case of our conditions, which further show that affine decision rules remain optimal under distributional robustness as well as for several other variants of the problem.

The remainder of the paper proceeds as follows. Section 2 develops tight conditions that guarantee the existence of optimal affine decision rules. Section 3 extends our findings to K-adaptability problems that allow for both continuous and discrete second-stage decisions. We study various applications of our results in Section 4, and we offer concluding remarks in Section 5.

2 Optimality of Affine Decision Rules

We consider an ambiguous probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is the sample space of possible outcomes, $\mathcal{F}$ is a σ-algebra on Ω that specifies the measurable events, and $\mathcal{P}$ is an ambiguity set of probability measures. We denote by $\mathcal{L}$ the set of all extended real-valued random variables on $(\Omega, \mathcal{F}, \mathcal{P})$, that is, the set of all measurable functions $X : \Omega \to \overline{\mathbb{R}}$. We fix a law invariant ambiguous risk measure $\rho = \{\rho_{\mathbb{P}}\}_{\mathbb{P} \in \mathcal{P}}$, which is a collection of law invariant risk measures $\rho_{\mathbb{P}} : \mathcal{L} \to \overline{\mathbb{R}}$, $\mathbb{P} \in \mathcal{P}$.

The focus of our study is the two-stage distributionally robust optimization problem

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\right] \\ \text{subject to} & \boldsymbol{x}\in\mathcal{X}, \end{array} \tag{1a}$$

where $\mathcal{X}\subseteq\mathbb{R}^{n_1}$, $\tilde{\boldsymbol{\xi}}$ is a random vector that is governed by some distribution $\mathbb{P}\in\mathcal{P}$ and that is supported on $\Xi\subseteq\mathbb{R}^k$,[1] and the second-stage cost function $\mathcal{Q}$ satisfies

$$\mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \quad = \quad \left[\begin{array}{ll} \text{minimize} & f(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}\geqslant\boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x}+\boldsymbol{D}\boldsymbol{y}\leqslant\boldsymbol{h}(\boldsymbol{\xi}) \\ & \boldsymbol{y}\in\mathbb{R}^{n_2} \end{array}\right], \tag{1b}$$

where $f:\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}\times\Xi\to\mathbb{R}$ is the objective function, the technology matrices $\boldsymbol{A}:\Xi\to\mathbb{R}^{m_1\times n_1}$ and $\boldsymbol{C}:\Xi\to\mathbb{R}^{m_2\times n_1}$ and the right-hand sides $\boldsymbol{g}:\Xi\to\mathbb{R}^{m_1}$ and $\boldsymbol{h}:\Xi\to\mathbb{R}^{m_2}$ can depend on $\boldsymbol{\xi}$, and the recourse matrices $\boldsymbol{B}\in\mathbb{R}^{m_1\times n_2}$ and $\boldsymbol{D}\in\mathbb{R}^{m_2\times n_2}$ are constant. Here and in the following, we adopt the standard convention that the optimal value of a minimization problem is $+\infty$ $(-\infty)$ whenever the problem is infeasible (unbounded).

We make the blanket assumption that the expression $\rho_{\mathbb{P}}[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})]$ that evaluates the risk of the second-stage costs in (1a) is well-defined for all first-stage decisions $\boldsymbol{x}\in\mathcal{X}$ and all probability measures $\mathbb{P}\in\mathcal{P}$. Sufficient conditions to ensure this are discussed in §2.3.1 of Shapiro et al. (2009).

Problem (1) constitutes a very generic two-stage distributionally robust optimization problem with a possibly nonlinear and non-convex first-stage feasible region $\mathcal{X}$ and a polyhedral (possibly unbounded) second-stage feasible region described by (1b). The objective function f of problem (1) can be nonlinear and non-convex in the decision variables and the uncertain problem parameters. The problem assumes a fixed recourse but allows for uncertainty in the technology matrices and right-hand sides. Special cases of problem (1) include stochastic programs, where $\mathcal{P}=\{\mathbb{P}^0\}$, and robust optimization problems, where $\rho_{\mathbb{P}}=\mathbb{P}\text{-}\operatorname{ess\,sup}$ and $\mathcal{P}=\{\delta_{\boldsymbol{z}}\,:\,\boldsymbol{z}\in\mathcal{Z}\}$ with $\delta_{\boldsymbol{z}}$ being the Dirac measure that places unit probability at $\boldsymbol{z}\in\mathbb{R}^k$ and $\mathcal{Z}\subseteq\mathbb{R}^k$ being a (possibly non-convex) uncertainty set. Problem (1) also encompasses distributionally robust optimization problems with moment and data-driven (*e.g.*, ϕ-divergence or Wasserstein) ambiguity sets.

We will study conditions under which the optimal value as well as the first-stage feasible region of problem (1) do not change if we restrict the second-stage decision $\boldsymbol{y}$ to an affine decision rule.

[1]The support of a random vector is the smallest closed set that attains probability 1 under every measure $\mathbb{P}\in\mathcal{P}$.

This restriction results in the single-stage distributionally robust optimization problem

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}(\tilde{\boldsymbol{\xi}}); \tilde{\boldsymbol{\xi}}) \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{y} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}, \end{array} \tag{2a}$$

where $\boldsymbol{y} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$ indicates that $\boldsymbol{y}$ is an affine function of $\boldsymbol{\xi}$, and where $\mathcal{Q}$ is defined as

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) \quad = \quad \begin{cases} f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) & \text{if } \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \quad \text{and} \\ & \quad \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y} \leqslant \boldsymbol{h}(\boldsymbol{\xi}), \\ +\infty & \text{otherwise.} \end{cases} \tag{2b}$$

Problem (2) offers several distinct advantages over problem (1). First and foremost, problem (2) often admits an equivalent reformulation as a tractable optimization problem that is amenable to a solution with standard software, or it can be solved efficiently by iterative solution schemes. In contrast, the discretization schemes commonly employed for the solution of problem (1) typically do not offer an implementable second-stage decision since the realized value of $\tilde{\boldsymbol{\xi}}$ differs from all discretization points with probability 1. Secondly, the optimal recourse policy in problem (2) has a compact representation that can readily be stored and implemented (*e.g.*, on embedded devices without optimization capabilities). Finally, the simple and explicit structure of the recourse policy in problem (2) facilitates interpretability of the optimization problem and may be useful, among others, for comparative statics.

Our optimality result for affine decision rules makes the following assumptions:

(**R**) The risk measure $\rho_{\mathbb{P}}$ is monotonic for every $\mathbb{P} \in \mathcal{P}$.

(**F**) For every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, $f(\boldsymbol{x}, \cdot; \boldsymbol{\xi})$ is monotonically non-decreasing in $\boldsymbol{y}$.

(**A**) The technology matrices $\boldsymbol{A}$, $\boldsymbol{C}$ and the right-hand sides $\boldsymbol{g}$, $\boldsymbol{h}$ are affine functions of $\boldsymbol{\xi}$.

(**D**) The constraint matrix $\boldsymbol{D}$ is non-negative.

(**B**) For every $\boldsymbol{x} \in \mathcal{X}$, there is an index set of constraints $\mathcal{I} \subseteq \{1, \ldots, m_1\}$, $|\mathcal{I}| = n_2$, such that the restriction $\boldsymbol{B}_{\mathcal{I}}$ of $\boldsymbol{B}$ to those constraints is invertible with a positive inverse, as well as

$$\left[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \iff \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y} \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) \right] \qquad \forall \boldsymbol{\xi} \in \Xi.$$

Assumption (**R**) is satisfied by many risk measures, including all coherent risk measures (Artzner et al., 1999) and the value-at-risk. It is not satisfied, for instance, by the mean (semi-)moment and the mean deviation risk measures (Shapiro et al., 2009). Assumption (**F**) is satisfied, for example, in the linear case where $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y}$ and $\boldsymbol{d}(\boldsymbol{\xi}) \geqslant \mathbf{0}$ for all $\boldsymbol{\xi} \in \Xi$. Note that if $f(\boldsymbol{x}, \cdot; \boldsymbol{\xi})$ is monotonically non-*increasing* in some or all components of $\boldsymbol{y}$ for every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, then a simple change of the affected variables $y_i \leftarrow -y_i$ satisfies assumption (**F**); care must be taken, however, that the other assumptions remain satisfied by the reformulation. Assumption (**A**) can always be satisfied by lifting the parameter vector $\boldsymbol{\xi}$ so that it contains the non-linear components of $\boldsymbol{A}$, $\boldsymbol{C}$, $\boldsymbol{g}$ and $\boldsymbol{h}$. Thus, assumption (**A**) is non-restrictive for our optimality result; it is nevertheless important as the resulting problem reformulation may involve a non-convex support Ξ, and hence the tractability of the affine decision rule problem (2) may be impacted. Together with assumption (**F**), assumption (**D**) implies that the second constraint set in (1b) imposes upper bounds on the decisions $\boldsymbol{y}$. In some cases, assumption (**D**) can be satisfied by multiplying both sides of a constraint in the second constraint set of (1b) with -1 and thus effectively converting the constraint into a member of the first constraint set (due to the inversion of the inequality). Assumption (**B**), finally, stipulates that for every fixed first-stage decision $\boldsymbol{x} \in \mathcal{X}$, there is a subset of n_2 constraints that decide whether a second-stage decision $\boldsymbol{y}$ satisfies the first constraint set in (1b). The assumption also states that the restriction of the recourse matrix $\boldsymbol{B}$ to those n_2 constraints has a positive inverse, which will be crucial for our optimality proof. Compared to the other assumptions, condition (**B**) is less transparent and appears cumbersome to verify in practice. Later in this section, we will elaborate on more easily verifiable conditions that imply (but are typically not implied by) assumption (**B**). We emphasize that we do not impose a relatively complete recourse in our results.

Theorem 1. *Under the assumptions (**R**), (**F**), (**A**), (**D**) and (**B**), the optimal value and the set of feasible (optimal) first-stage decisions $\boldsymbol{x}$ in problems* (1) *and* (2) *coincide.*

Proof. If problem (1) is infeasible, then its restriction (2) to affine second-stage decisions remains infeasible. In this case, both problems share the same (empty) sets of feasible and optimal solutions, and by our earlier convention both problems attain the same optimal value of $+\infty$. In the following, we thus assume that problem (1) is feasible. We show that for every fixed first-stage decision $\boldsymbol{x} \in \mathcal{X}$, we can construct an affine decision rule $\boldsymbol{y}^\ell : \Xi \xrightarrow{\text{a}} \mathbb{R}^n$ such that

$$\sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] = \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}^\ell(\tilde{\boldsymbol{\xi}}); \tilde{\boldsymbol{\xi}}) \right], \tag{3}$$

where the cost functions on the left-hand side and right-hand side are defined in (1b) and (2b), respectively. Equation (3) immediately implies the statement of the theorem.

To show that equation (3) holds, fix any first-stage decision $\boldsymbol{x} \in \mathcal{X}$, together with an index set $\mathcal{I}$ that satisfies assumption (**B**). Define $\operatorname{dom} \mathcal{Q} = \{\boldsymbol{\xi} \in \Xi : \mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}) < +\infty\}$ as the set of parameter realizations $\boldsymbol{\xi}$ for which $\boldsymbol{x}$ admits a feasible second-stage decision. For any $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q}$, any feasible second-stage decision $\boldsymbol{y}(\boldsymbol{\xi})$ has to satisfy

$$\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{g}(\boldsymbol{\xi}),$$

and assumption (**B**) implies that this is equivalent to

$$\boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}). \tag{4}$$

Since $\boldsymbol{B}_{\mathcal{I}}$ admits a positive inverse, the satisfaction of the constraint set (4) implies that

$$\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{B}_{\mathcal{I}}^{-1}\Big[\boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x}\Big], \tag{5}$$

but not vice versa. Indeed, since $\boldsymbol{B}_{\mathcal{I}}^{-1} \geqslant \boldsymbol{0}$, the constraints in (5) constitute non-negative linear combinations of the constraints in (4), and thus the constraint system (5) is a relaxation of the constraint set (4). Consider now the solution

$$\boldsymbol{y}^{\ell}(\boldsymbol{\xi}) = \boldsymbol{B}_{\mathcal{I}}^{-1}\Big[\boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x}\Big] \qquad \forall \boldsymbol{\xi} \in \Xi, \tag{6}$$

which satisfies the relaxed constraint set (5) as equality for all $\boldsymbol{\xi} \in \Xi$ and which is evidently affine in $\boldsymbol{\xi}$. This solution satisfies the constraint set (4) and thus also the first constraint set in (1b) over $\operatorname{dom} \mathcal{Q}$ (but not over the possibly non-empty set $\Xi \backslash \operatorname{dom} \mathcal{Q}$). To see that $\boldsymbol{y}^{\ell}(\boldsymbol{\xi})$ also satisfies the second constraint set in (1b) over $\operatorname{dom} \mathcal{Q}$, we note that for all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q}$, we have that

$$\boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y}^{\ell}(\boldsymbol{\xi}) \quad \leqslant \quad \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y}(\boldsymbol{\xi})$$

for any feasible second-stage decision $\boldsymbol{y}(\boldsymbol{\xi})$. Here, the inequality follows from assumption (**D**) as well as the fact that $\boldsymbol{y}^{\ell}(\boldsymbol{\xi}) \leqslant \boldsymbol{y}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q}$. Indeed, we have observed that any feasible second-stage solution $\boldsymbol{y}(\boldsymbol{\xi})$ must satisfy the relaxed constraint set (5), and $\boldsymbol{y}^{\ell}(\boldsymbol{\xi})$ is the point-wise smallest decision satisfying (5) according to its definition in (6).

To see that $\boldsymbol{y}^{\ell}(\boldsymbol{\xi})$ is point-wise optimal over Ξ, finally, we note that for all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q}$ and any second-stage decision $\boldsymbol{y}(\boldsymbol{\xi})$ feasible for $\boldsymbol{\xi}$, we have $f(\boldsymbol{x}, \boldsymbol{y}^{\ell}(\boldsymbol{\xi}); \boldsymbol{\xi}) \leqslant f(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{\xi}); \boldsymbol{\xi})$ due to assumption

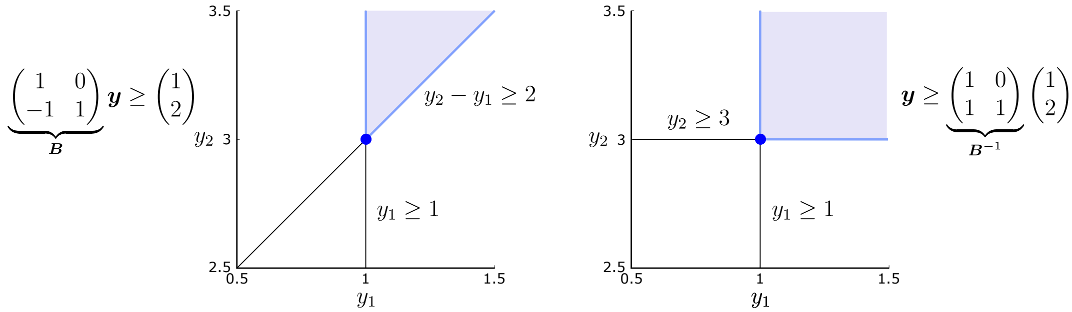

Figure 1. The feasible region imposed by the constraint system (4), left, is not equivalent to that of (5), right, but both share the same coordinate-wise minimal point $(1, 3)$.

(**F**) as well as our earlier finding that $\boldsymbol{y}^{\ell}(\boldsymbol{\xi}) \leqslant \boldsymbol{y}(\boldsymbol{\xi})$. Moreover, $\boldsymbol{y}^{\ell}(\boldsymbol{\xi})$ is only infeasible for the realizations $\boldsymbol{\xi} \in \Xi \backslash \mathrm{dom}\, \mathcal{Q}$ for which any second-stage decision is infeasible. Assumption (**R**) then implies equation (3), which concludes the proof. □

Crucial to our proof of Theorem 1 is the existence of a non-negative inverse $\boldsymbol{B}_{\mathcal{I}}^{-1}$ thanks to assumption (**B**), which ensures that the constraint system (5) is a relaxation of the first set of constraints in (1b) that, if strengthened to equalities as in (6), imposes an affine structure on the second-stage decisions $\boldsymbol{y}(\boldsymbol{\xi})$. Note that the constraint system (5) is *not* equivalent to the first set of constraints in (1b), however. Indeed, the feasible region formed by the constraints $y_1 \geqslant 1$ and $y_2 \geqslant y_1 + 2$ can be interpreted as an instance of the second-stage problem (1b) satisfying condition (**B**), but it does not coincide with the feasible region formed by the constraints $y_1 \geqslant 1$, $y_3 \geqslant 3$ of the associated equation (5), see Figure 1. Crucially, however, both feasible regions share the same component-wise minimal point $(y_1^{\star}, y_2^{\star}) = (1, 3)$, which is what we exploit in the proof.

Remark 1 (More General Classes of Decisions Rules)**.** *We can generalize assumption (**A**) as follows. If the technology matrix* $\boldsymbol{A}$ *and the right-hand side vector* $\boldsymbol{g}$ *belong to a function class* $\mathcal{C}$ *that is closed under linear combinations (* e.g.*, piecewise affine, polynomial or trigonometric functions), then Theorem 1 continues to hold if we replace the class of affine decision rules in problem* (2) *with the broader class of decision rules in* $\mathcal{C}$*. Note that the non-linearities in* $\boldsymbol{A}$ *and* $\boldsymbol{g}$ *can be absorbed in the definition of the support* Ξ*, see Georghiou et al. (2015) and Bertsimas et al. (2019), and thus our optimality result for affine decision rules immediately extends to this broader class of problems.*

Remark 2 (Nonlinear Objective Functions and Epigraph Reformulations)**.** *Even if the objective function* f *in problem* (1) *is piecewise affine, the associated affine decision rule problem* (2) *cannot* *be linearized through an epigraph formulation without affecting our optimality result. To see this, consider the following instance of problem* (1)*:*

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big] \quad \text{with} \quad \mathcal{Q}(\tilde{\xi}) \;=\; \min\big\{ \max\{y,\, 0\} \,:\, y \geqslant \xi,\;\; y \in \mathbb{R}\big\}$$

This instance contains no first-stage decision, its ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ *is a singleton set that contains the distribution* $\mathbb{P}$ *under which* $\tilde{\xi}$ *follows a univariate uniform distribution over the interval* $[-1, 1]$*, and the risk measure* $\rho_{\mathbb{P}}$ *is the expected value. The instance evidently satisfies the assumptions **(R)**, **(F)**, **(A)**, **(D)** and **(B)** of Theorem 1, and its optimal value* $1/4$ *is attained by the affine decision rule* $y^\star(\xi) = \xi$*. The restriction of the epigraphical reformulation*

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big] \quad \text{with} \quad \mathcal{Q}(\tilde{\xi}) \;=\; \min\{\tau \,:\, \tau \geqslant \max\{y,\, 0\}\,,\;\; y \geqslant \xi,\;\; \tau, y \in \mathbb{R}\}$$

to affine decision rules $\tau(\xi)$ *and* $y(\xi)$*, however, attains the strictly larger objective value of* $1/2$*.*

Remark 3 (Equality Constraints)**.** *Consider a variant of problem* (1) *with the second-stage problem*

$$\mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \;=\; \left[\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{E}\boldsymbol{z} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y} + \boldsymbol{F}\boldsymbol{z} \leqslant \boldsymbol{h}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^{n_2},\;\; \boldsymbol{z} \in \mathcal{Z}(\boldsymbol{x};\boldsymbol{\xi}) \end{array}\right]$$

as well as its corresponding single-stage counterpart in affine decision rules. Assume that for each first-stage decision $\boldsymbol{x} \in \mathcal{X}$*, we have* $|\mathcal{Z}(\boldsymbol{x};\boldsymbol{\xi})| \leqslant 1$ *for all* $\boldsymbol{\xi} \in \Xi$*, and that there is an affine mapping* $\boldsymbol{\xi} \mapsto \boldsymbol{\zeta}(\boldsymbol{\xi})$ *satisfying* $\boldsymbol{\zeta}(\boldsymbol{\xi}) \in \mathcal{Z}(\boldsymbol{x};\boldsymbol{\xi})$ *whenever* $\mathcal{Z}(\boldsymbol{x};\boldsymbol{\xi}) \neq \emptyset$, $\boldsymbol{\xi} \in \Xi$*. In practical applications, the set* $\mathcal{Z}(\cdot,\cdot)$ *would typically be characterized by equality constraints that are (de-)activated based on logical conditions involving the first-stage decisions* $\boldsymbol{x}$*. One can verify that Theorem 1 extends to this more general setting if we replace the assumptions **(F)** and **(B)** with*

(F') *For every* $\boldsymbol{x} \in \mathcal{X}$ *and* $\boldsymbol{\xi} \in \Xi$*,* $f(\boldsymbol{x}, \cdot, \boldsymbol{\zeta}(\boldsymbol{\xi}); \boldsymbol{\xi})$ *is monotonically non-decreasing in* $\boldsymbol{y}$*.*

(B') *For every* $\boldsymbol{x} \in \mathcal{X}$*, there is an index set of constraints* $\mathcal{I} \subseteq \{1, \ldots, m_1\}$*,* $|\mathcal{I}| = n_2$*, such that* $\boldsymbol{B}_{\mathcal{I}}$ *is invertible with a positive inverse, as well as*

$$\Big[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} + \boldsymbol{E}\boldsymbol{z} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \;\iff\; \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y} + \boldsymbol{E}_{\mathcal{I}}\boldsymbol{z} \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi})\Big] \qquad \forall \boldsymbol{\xi} \in \Xi.$$

Remark 4 (Solution Methods)**.** *Assume that for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $\boldsymbol{y} : \Xi \xrightarrow{a} \mathbb{R}^{n_2}$*, the employed risk measure satisfies* $\sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}(\tilde{\boldsymbol{\xi}}); \tilde{\boldsymbol{\xi}})\big] = \infty$ *whenever* $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{\xi}); \boldsymbol{\xi}) = \infty$ *for some* $\boldsymbol{\xi} \in \Xi$*. This is satisfied, among others, by the expected value, the conditional value-at-risk and the essential supremum, but it is typically not satisfied by the value-at-risk. An affine decision rule* $\boldsymbol{y}$ *is then feasible in problem* (2) *if and only if it is feasible point-wise over the support, that is,*

$$\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \;\; \textit{and} \;\; \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y}(\boldsymbol{\xi}) \leqslant \boldsymbol{h}(\boldsymbol{\xi}) \qquad \forall \boldsymbol{\xi} \in \Xi,$$

which admits an efficient reformulation via standard robust optimization techniques whenever the support Ξ *of the random vector* $\tilde{\boldsymbol{\xi}}$ *is polyhedral (Ben-Tal et al., 2009; Bertsimas et al., 2011a). The resulting reformulation of problem* (2) *constitutes a single-stage problem for which a range of solution techniques have been developed, including monolithic reformulations via scenario fans (Shapiro et al., 2009; Birge and Louveaux, 2011) or duality theory (Ben-Tal et al., 2009; Bertsimas et al., 2011a; Ben-Tal et al., 2013; Wiesemann et al., 2014; Mohajerin Esfahani and Kuhn, 2018) as well as iterative solution schemes based on Benders decomposition (Shapiro et al., 2009; Birge and Louveaux, 2011) and semi-infinite programming (Blankenship and Falk, 1976; Mutapcic and Boyd, 2009; Gorissen and den Hertog, 2013; Bertsimas et al., 2016).*

We next show that the imposed assumptions (**R**), (**F**), (**A**), (**D**) and (**B**) are not only sufficient but also necessary for the optimality of affine decision rules in problem (1).

Proposition 1. *The assumptions (**R**), (**F**), (**A**), (**D**) and (**B**) in Theorem 1 are tight in the sense that the conclusion of the theorem ceases to hold whenever problem* (1) *violates any one of the assumptions, even if all other assumptions are satisfied.*

Proof. As for assumption (**R**), consider the instance

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big] + \frac{3}{2} \cdot \mathbb{E}_{\mathbb{P}}\big[\big|\mathcal{Q}(\tilde{\xi}) - \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big]\big|\big] \tag{7}$$

of problem (1) that contains no first-stage decision, whose ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is a singleton set that contains the distribution $\mathbb{P}$ under which $\tilde{\xi}$ follows a univariate uniform distribution over the interval $[-1, 1]$, henceforth abbreviated by $\tilde{\xi} \sim \mathcal{U}[-1, 1]$, and whose risk measure $\rho_{\mathbb{P}}$ is the weighted mean-mean absolute deviation. The second-stage problem of this instance satisfies

$$\mathcal{Q}(\xi) \quad = \quad \min \{ y \, : \, y \geqslant \xi, \;\; y \in \mathbb{R} \} \, .$$

Note that although the mean-mean absolute deviation risk measure violates assumption (**R**), the other assumptions (**F**), (**A**), (**D**) and (**B**) of Theorem 1 are all satisfied. One readily verifies that the affine decision rule formulation (2) associated with problem (7) is optimized by $y^\star(\xi) = \xi$, which results in an optimal value of 3/4, whereas the second-stage policy $y(\xi) = \max\{\xi, (\xi+1)/4\}$ is feasible in problem (7) and attains a lower objective value of 2/3.

In view of assumption (**F**), consider the following instance of problem (1):

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big] \tag{8}$$

This instance again contains no first-stage decision, its ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is again a singleton set such that $\tilde{\xi} \sim \mathcal{U}[-1,1]$ under $\mathbb{P}$, and $\rho_{\mathbb{P}}$ is the expected value. The second-stage problem satisfies

$$\mathcal{Q}(\xi) \quad = \quad \min\{-y \,:\, y \geqslant -10, \;\; y \leqslant \xi, \;\; y \leqslant -\xi, \;\; y \in \mathbb{R}\}.$$

Although the objective function fails to be monotonically non-decreasing in y and hence violates assumption (**F**), the other assumptions (**R**), (**A**), (**D**) and (**B**) of Theorem 1 are all satisfied. One readily verifies that the affine decision rule formulation (2) associated with problem (8) is optimized by $y^\star(\xi) = -1$, which results in an optimal value of 1, whereas the second-stage policy $y(\xi) = \min\{\xi, -\xi\}$ is feasible in problem (8) and attains a lower objective value of 1/2.

As for assumption (**A**), consider the instance of problem (1) with objective function (8), that is, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, no first-stage decision, and the ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is such that $\tilde{\xi} \sim \mathcal{U}[-1,1]$ under $\mathbb{P}$. The second-stage problem satisfies

$$\mathcal{Q}(\xi) \quad = \quad \min\big\{y \,:\, y \geqslant \xi^2, \;\; y \in \mathbb{R}\big\}.$$

Although the constraint right-hand side exhibits a nonlinear dependence on ξ and hence violates assumption (**A**), the other assumptions (**R**), (**F**), (**D**) and (**B**) of Theorem 1 are all satisfied. One readily verifies that the associated affine decision rule formulation (2) is optimized by $y^\star(\xi) = 1$, which results in an optimal value of 1, whereas the second-stage policy $y(\xi) = \xi^2$ is feasible in problem (1) and attains a lower objective value of 1/3.

In view of assumption (**D**), consider the instance of problem (1) with objective function (8), that is, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, no first-stage decision, and the ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is such that $\tilde{\xi} \sim \mathcal{U}[-1,1]$ under $\mathbb{P}$. The second-stage problem satisfies

$$\mathcal{Q}(\xi) \quad = \quad \min\{y \,:\, y \geqslant \xi, \;\; -y \leqslant \xi, \;\; y \in \mathbb{R}\}.$$

Although the minus sign on the left-hand side of the second constraint implies that assumption (**D**) is violated, the other assumptions (**R**), (**F**), (**A**) and (**B**) of Theorem 1 are all satisfied. One readily verifies that the associated affine decision rule problem (2) is optimized by $y^\star(\xi) = 1$, which results in an optimal value of 1, whereas the second-stage policy $y(\xi) = \max\{\xi, -\xi\}$ is feasible in problem (1) and attains a lower objective value of 1/2.

In view of assumption (**B**), finally, consider the instance of problem (1) with objective function (8), that is,, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, no first-stage decision, and the ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is such that $\tilde{\xi} \sim \mathcal{U}[-1, 1]$ under $\mathbb{P}$. The second-stage problem satisfies

$$\mathcal{Q}(\xi) \quad = \quad \min\left\{y_1 + 2y_2 \,:\, y_1 + y_2 \geqslant \xi + 1, \;\; y_1, y_2 \geqslant 0, \;\; y_1, y_2 \leqslant 1, \;\; \boldsymbol{y} \in \mathbb{R}^2\right\}.$$

Note that for every $\xi > -1$, the index 1 of the first constraint $y_1 + y_2 \geqslant \xi$ must be contained in the index set $\mathcal{I}$ defined in assumption (**B**). As a result, however, none of the coefficient matrices

$$\boldsymbol{B}_{\mathcal{I}} \in \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\}$$

have a positive inverse, that is, assumption (**B**) is violated. In constrast, the other assumptions (**R**), (**F**), (**A**) and (**D**) of Theorem 1 are all satisfied. One can verify that the associated affine decision rule formulation (2) is optimized by $y_1^\star(\xi) = y_2^\star(\xi) = (\xi + 1)/2$, which results in an optimal value of 3/2, whereas the second-stage policy $y_1(\xi) = \min\{\xi + 1, 1\}$ and $y_2(\xi) = \max\{\xi, 0\}$ is feasible in problem (1) and attains a lower objective value of 5/4. □

While the assumptions (**R**), (**F**), (**A**) and (**D**) of Theorem 1 are transparent and easy to verify, the assumption (**B**) is less intuitive and appears cumbersome to confirm in practice. We next discuss sufficient (but not necessary) conditions for this assumption to be satisfied. To this end, we recall that a matrix $\boldsymbol{Z} \in \mathbb{R}^{n \times n}$ is called a *Z-matrix* if all of its off-diagonal elements are less than or equal to zero, that is, if $Z_{ij} \leqslant 0$ for $i \neq j$. Moreover, a Z-matrix is called an *M-matrix* if all of its eigenvalues have a non-negative real part. The M-matrices form an important subclass of the inverse positive matrices, that is, the matrices that possess a component-wise non-negative inverse. The study of M-matrices has a long history in linear algebra, and M-matrices have found applications, among others, in game theory, Markov chains and economics (Berman and Plemmons, 1994; Bapat and Raghavan, 1997).

We say that a collection of vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n \in \mathbb{R}^n$ form a *partial order* if there is a permutation $\pi(1), \ldots, \pi(n)$ of $1, \ldots, n$ such that $[\boldsymbol{z}_{\pi(j)}]_\ell = 0$ for all $\ell = \pi(j), \ldots, \pi(n)$ and all $j = 1, \ldots, n$. In other words, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n \in \mathbb{R}^n$ form a partial order if there is a permutation matrix $\boldsymbol{\Pi} \in \mathbb{R}^{n \times n}$ such that the matrix $\boldsymbol{\Pi}\big[\boldsymbol{z}_1 \ \ldots \ \boldsymbol{z}_n\big]\boldsymbol{\Pi}^\top$ is upper triangular with zeros on the diagonal. The permutation $\pi(1) = 2$, $\pi(2) = 3$ and $\pi(3) = 1$ certifies that the vectors $\boldsymbol{z}_1 = (0,1,1)^\top$, $\boldsymbol{z}_2 = (0,0,0)^\top$ and $\boldsymbol{z}_3 = (0,1,0)^\top$ form a partial order, for example, since $z_{22} = z_{23} = z_{21} = 0$, $z_{33} = z_{31} = 0$ and $z_{11} = 0$. One readily verifies that the associated permutation matrix $\boldsymbol{\Pi} = \big[\mathbf{e}_{\pi(1)}\, \mathbf{e}_{\pi(2)}\, \mathbf{e}_{\pi(3)}\big]^\top$ satisfies that $\boldsymbol{\Pi}\big[\boldsymbol{z}_1\, \boldsymbol{z}_2\, \boldsymbol{z}_3\big]\boldsymbol{\Pi}^\top$ is upper triangular with zeros on the diagonal.

Proposition 2. *Assume that the first constraint set $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi})$ in* (1b) *can be written as*

$$y_j \geqslant \boldsymbol{\alpha}_{jk}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{jk}^\top \boldsymbol{y} + \gamma_{jk}(\boldsymbol{\xi}) \qquad \forall j = 1, \ldots, n_2,\ \forall k = 1, \ldots, s_j, \tag{9}$$

where $\boldsymbol{\beta}_{jk} \geqslant \mathbf{0}$ for all j and k, such that for every $\boldsymbol{x} \in \mathcal{X}$, we have:

(i) *For every $j = 1, \ldots, n_2$ there is a constraint $k_j \in \{1, \ldots, s_j\}$ in* (9) *that weakly dominates the other $s_j - 1$ constraints for j in* (9) *under every parameter realization $\boldsymbol{\xi} \in \Xi$.*

(ii) *The vectors $\{\boldsymbol{\beta}_{j,k_j}\}_{j=1}^{n_2}$ form a partial order.*

Then the corresponding instance of problem (1) *satisfies assumption **(B)**.*

Note that in *(i)*, the indices $\{k_j\}_{j=1}^{n_2}$ of the dominant constraints may differ for each $\boldsymbol{x} \in \mathcal{X}$.

Proof of Proposition 2. The constraint system (9) can be written as $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi})$ by setting

$$\boldsymbol{A}(\boldsymbol{\xi})^\top = \big[\boldsymbol{A}_1(\boldsymbol{\xi})^\top \ \ldots \ \boldsymbol{A}_{n_2}(\boldsymbol{\xi})^\top\big] \in \mathbb{R}^{n_1 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{A}_j(\boldsymbol{\xi})^\top = \big[-\boldsymbol{\alpha}_{j,1}(\boldsymbol{\xi}) \ \ldots \ -\boldsymbol{\alpha}_{j,s_j}(\boldsymbol{\xi})\big] \in \mathbb{R}^{n_1 \times s_j},$$

$$\boldsymbol{B}^\top = \big[\boldsymbol{B}_1^\top \ \ldots \ \boldsymbol{B}_{n_2}^\top\big] \in \mathbb{R}^{n_2 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{B}_j^\top = \big[\mathbf{e}_j - \boldsymbol{\beta}_{j,1} \ \ldots \ \mathbf{e}_j - \boldsymbol{\beta}_{j,s_j}\big] \in \mathbb{R}^{n_2 \times s_j},$$

$$\boldsymbol{g}(\boldsymbol{\xi})^\top = \big[\boldsymbol{g}_1(\boldsymbol{\xi})^\top \ \ldots \ \boldsymbol{g}_{n_2}(\boldsymbol{\xi})^\top\big] \in \mathbb{R}^{1 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{g}_j(\boldsymbol{\xi})^\top = \big[\gamma_{j,1}(\boldsymbol{\xi}) \ \ldots \ \gamma_{j,s_j}(\boldsymbol{\xi})\big] \in \mathbb{R}^{1 \times s_j}.$$

Fix any $\boldsymbol{x} \in \mathcal{X}$, and choose k_j as stipulated in condition *(i)* of the statement. We set

$$\mathcal{I} = \bigcup_{j=1}^{n_2} \left\{ \left[\sum_{i=1}^{j-1} s_i\right] + k_j \right\}, \quad \text{implying that} \quad \boldsymbol{B}_{\mathcal{I}} = \begin{bmatrix} \mathbf{e}_1^\top - \boldsymbol{\beta}_{1,k_1}^\top \\ \vdots \\ \mathbf{e}_{n_2}^\top - \boldsymbol{\beta}_{n_2,k_{n_2}}^\top \end{bmatrix} \in \mathbb{R}^{n_2 \times n_2}.$$

The claim of the proposition holds if $\boldsymbol{B}_{\mathcal{I}}$ has a positive inverse and if

$$\Big[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \iff \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y} \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi})\Big] \qquad \forall \boldsymbol{\xi} \in \Xi,$$

that is, if

$$
\begin{aligned}
& \left[y_j \geqslant \boldsymbol{\alpha}_{jk}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{jk}^\top \boldsymbol{y} + \gamma_{jk}(\boldsymbol{\xi}) \qquad \forall j = 1, \ldots, n_2,\ \forall k = 1, \ldots, s_j\right] \\
\iff \quad & \left[y_j \geqslant \boldsymbol{\alpha}_{j,k_j}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{j,k_j}^\top \boldsymbol{y} + \gamma_{j,k_j}(\boldsymbol{\xi}) \ \ \forall j = 1, \ldots, n_2\right] \qquad\qquad \forall \boldsymbol{\xi} \in \Xi.
\end{aligned}
$$

Note that the above equivalence immediately follows from condition *(i)* of the statement. We now show that $\boldsymbol{B}_{\mathcal{I}}$ constitutes an M-matrix, which implies that it also has a positive inverse.

We claim that for any permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{n_2 \times n_2}$, $\boldsymbol{B}_{\mathcal{I}}$ is an M-matrix if and only if $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ is an M-matrix. Indeed, the row and column permutations conducted by $\mathbf{\Pi}$ and $\mathbf{\Pi}^\top$, respectively, ensure that the diagonal (off-diagonal) elements of $\boldsymbol{B}_{\mathcal{I}}$ remain diagonal (off-diagonal) elements in $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ and vice versa, which implies that $\boldsymbol{B}_{\mathcal{I}}$ is an Z-matrix if and only if $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ is an Z-matrix. Moreover, $\boldsymbol{B}_{\mathcal{I}}$ and $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ share the same eigenvalues as the matrices are similar, which implies that $\boldsymbol{B}_{\mathcal{I}}$ is an M-matrix if and only if $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ is an M-matrix.

Define now the permutation matrix $\mathbf{\Pi} = \left[\mathbf{e}_{\pi(1)} \ \ldots \ \mathbf{e}_{\pi(n_2)}\right]^\top \in \mathbb{R}^{n_2 \times n_2}$, where π is the permutation that establishes the partial order in condition *(ii)* of the statement. We then have

$$
\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top \ = \ \mathbf{\Pi} \begin{bmatrix} \mathbf{e}_1^\top - \boldsymbol{\beta}_{1,k_1}^\top \\ \vdots \\ \mathbf{e}_{n_2}^\top - \boldsymbol{\beta}_{n_2,k_{n_2}}^\top \end{bmatrix} \mathbf{\Pi}^\top \ = \ \mathbf{\Pi} \begin{bmatrix} \mathbf{e}_1^\top \\ \vdots \\ \mathbf{e}_{n_2}^\top \end{bmatrix} \mathbf{\Pi}^\top - \mathbf{\Pi} \begin{bmatrix} \boldsymbol{\beta}_{1,k_1}^\top \\ \vdots \\ \boldsymbol{\beta}_{n_2,k_{n_2}}^\top \end{bmatrix} \mathbf{\Pi}^\top \ = \ \mathbf{I} - \mathbf{\Pi} \begin{bmatrix} \boldsymbol{\beta}_{1,k_1}^\top \\ \vdots \\ \boldsymbol{\beta}_{n_2,k_{n_2}}^\top \end{bmatrix} \mathbf{\Pi}^\top,
$$

where $\mathbf{e}_i$ is the i-th canonical basis vector in $\mathbb{R}^{n_2}$ and $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{n_2 \times n_2}$, respectively. Note that by definition of the permutation π in condition *(ii)* of the statement, the second matrix on the right-hand side of the last identity is a non-negative upper triangular matrix with zeros on the diagonal. Thus, $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ is a Z-matrix with ones on the diagonal. Since the eigenvalues of a triangular matrix coincide with its diagonal elements, we conclude that all eigenvalues of $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$ are one, and thus $\mathbf{\Pi} \boldsymbol{B}_{\mathcal{I}} \mathbf{\Pi}^\top$—and therefore $\boldsymbol{B}_{\mathcal{I}}$—is an M-matrix, which concludes the proof. □

Since $\boldsymbol{\beta}_{jk} \geqslant \mathbf{0}$ for all $j = 1, \ldots, n_2$ and $k = 1, \ldots, s_j$, the structure of the constraint set (9) in Proposition 2 implies that for each second-stage decision variable y_j there are s_j alternative lower bounds. The first condition of Proposition 2 then guarantees that for each $\boldsymbol{x} \in \mathcal{X}$, only one lower bound matters for each second-stage decision variable y_j, irrespective of the parameter realization $\boldsymbol{\xi} \in \Xi$. In practice, this is achieved by a big-M formulation that de-activates all but one of the constraints for each decision y_j based on the value of the first-stage decision $\boldsymbol{x}$. The second condition of Proposition 2, on the other hand, ensures that for each $\boldsymbol{x} \in \mathcal{X}$, the dependence structure between the second-stage decisions y_j, as expressed by the vectors $\boldsymbol{\beta}_{j,k_j}$ of the weakly dominant constraints, is acyclic. Thus, for a fixed first-stage decision $\boldsymbol{x} \in \mathcal{X}$, the second-stage decisions can be re-ordered in such a way that y_1 only depends on the realization of $\boldsymbol{\xi}$, y_2 may depend on both the realization of $\boldsymbol{\xi}$ and the value of y_1, y_3 may depend on $\boldsymbol{\xi}$, y_1 and y_2, and so on. Thus, for a fixed first-stage decision $\boldsymbol{x} \in \mathcal{X}$, the determination of the point-wise optimal second-stage decision $\boldsymbol{y}^\star$ becomes simple: After the aforementioned re-ordering of the indices, we can set $y_1^\star(\boldsymbol{\xi}) = \boldsymbol{\alpha}_{1,k_1}(\boldsymbol{\xi})^\top \boldsymbol{x} + \gamma_{1,k_1}(\boldsymbol{\xi})$, $y_2^\star(\boldsymbol{\xi}) = \boldsymbol{\alpha}_{2,k_2}(\boldsymbol{\xi})^\top \boldsymbol{x} + \beta_{2,k_2,1} y_1 + \gamma_{2,k_2}(\boldsymbol{\xi})$, and so on. Of course the re-ordering will typically depend on the first-stage decision $\boldsymbol{x} \in \mathcal{X}$, which is why the second-stage decision $\boldsymbol{y}$ cannot easily be substituted out of the problem.

3 The K-Adaptability Problem

We now consider a generalization of the two-stage distributionally robust optimization problem (1) where some of the second-stage decisions may be integer, subjected to a random recourse and/or violate the assumptions (**F**), (**D**) and (**B**). We thus consider the problem

$$\begin{array}{ll} \text{minimize} & \sup\limits_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{10a}$$

where the second-stage problem $\mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi})$ is now defined as

$$\mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}) \quad = \quad \left[\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{\xi})\boldsymbol{z} + \boldsymbol{D}\boldsymbol{y} \leqslant \boldsymbol{h}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^{n_2}, \;\; \boldsymbol{z} \in \mathcal{Z}(\boldsymbol{x}) \end{array} \right]. \tag{10b}$$

Note that the new second-stage decisions $\boldsymbol{z}$ have a feasible region $\mathcal{Z}(\boldsymbol{x})$ that may be non-convex and/or depend on the first-stage decisions $\boldsymbol{x}$ in a nonlinear fashion. Also, contrary to the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$, the recourse matrices $\boldsymbol{E}$ and $\boldsymbol{F}$ for the decisions $\boldsymbol{z}$ may depend on the random problem parameters $\tilde{\boldsymbol{\xi}}$. We impose the assumptions (**R**) and (**D**) from the previous section as well as

(**F'**) For every $\boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{z} \in \mathcal{Z}(\boldsymbol{x})$ and $\boldsymbol{\xi} \in \Xi$, $f(\boldsymbol{x}, \cdot, \boldsymbol{z}; \boldsymbol{\xi})$ is monotonically non-decreasing in $\boldsymbol{y}$.

(**A'**) The technology matrices $\boldsymbol{A}$, $\boldsymbol{C}$, $\boldsymbol{E}$, $\boldsymbol{F}$ and the right-hand sides $\boldsymbol{g}$, $\boldsymbol{h}$ are affine functions of $\boldsymbol{\xi}$.

(**B'**) For every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{Z}(\boldsymbol{x})$, there is an index set of constraints $\mathcal{I} \subseteq \{1, \ldots, m_1\}$, $|\mathcal{I}| = n_2$, such that $\boldsymbol{B}_{\mathcal{I}}$ is invertible with a positive inverse, as well as

$$\Big[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \iff \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{z} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y} \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi})\Big] \qquad \forall \boldsymbol{\xi} \in \Xi.$$

We emphasize that under the new set of assumptions, the second-stage decisions $\boldsymbol{z}$ only have to satisfy the weaker conditions that have previously been imposed on the first-stage decisions $\boldsymbol{x}$. In particular, the objective function f may fail to be monotone in $\boldsymbol{z}$, the recourse matrices $\boldsymbol{E}$ and $\boldsymbol{F}$ associated with $\boldsymbol{z}$ may be random and contain arbitrary coefficients, and the existence of a positive inverse is restricted to the coefficient matrix $\boldsymbol{B}$ of the second-stage decisions $\boldsymbol{y}$.

Unfortunately, the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**) are *not* sufficient to guarantee that problem (10) is optimized by an affine decision rule $\boldsymbol{y} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$ as the next example shows.

Example 1. *Consider the following instance of problem* (10)*,*

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\left[\min\left\{y + z \;:\; y \geqslant z, \;\; z \geqslant 1/2 - \tilde{\xi}, \;\; y \in \mathbb{R}, \;\; z \in \mathbb{R}_+\right\}\right],$$

*which does not involve any first-stage decisions $\boldsymbol{x}$, whose ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is a singleton set that contains the uniform distribution supported on $[0,1]$, and that employs the expected value as a risk measure. One readily verifies that this problem satisfies (**R**), (**F'**), (**A'**), (**D**) and (**B'**), but the unique optimal second-stage policy is given by $y^\star(\xi) = z^\star(\xi) = \max\{1/2 - \xi, 0\}$.*

We next consider the K-adaptability formulation associated with problem (10):

$$\begin{array}{ll} \text{minimize} & \sup\limits_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{z}_k\}_k; \tilde{\boldsymbol{\xi}})\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \;\; \boldsymbol{z}_k \in \mathcal{Z}(\boldsymbol{x}), \; k = 1, \ldots, K, \end{array} \tag{11a}$$

where $\{\boldsymbol{z}_k\}_k = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}$ and $\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}) = \min\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi}) \;:\; k = 1, \ldots, K\}$ with

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi}) \;\; = \;\; \left[\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_k; \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{D}\boldsymbol{y} \leqslant \boldsymbol{h}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^{n_2} \end{array}\right]. \tag{11b}$$

Problem (11) determines K candidate first-stage decisions $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ for the second-stage decision $\boldsymbol{z}$ in problem (10) here-and-now and subsequently implements the best of these decisions once the value of $\tilde{\boldsymbol{\xi}}$ has been observed. By construction, the K-adaptability problem (11) constitutes a conservative approximation of the two-stage distributionally robust optimization problem (10). For a detailed analysis of problem (11), we refer to Bertsimas and Caramanis (2010), Bertsimas et al. (2011b), Hanasusanto et al. (2015, 2016) and Subramanyam et al. (2020).

We will now show that under the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**), the optimal value as well as the first-stage feasible region of problem (11) do not change if we restrict the second-stage decision $\boldsymbol{y}$ to *a collection* of affine decision rules, that is, if we instead solve the single-stage problem

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k(\tilde{\boldsymbol{\xi}})\}_k, \{\boldsymbol{z}_k\}_k; \tilde{\boldsymbol{\xi}})\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \ \ \boldsymbol{y}_k : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2} \text{ and } \boldsymbol{z}_k \in \mathcal{Z}(\boldsymbol{x}),\ k = 1, \ldots, K, \end{array} \tag{12a}$$

where $\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k\}_k, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}) = \min\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{z}_k; \boldsymbol{\xi}) \,:\, k = 1, \ldots, K\}$ with

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{z}_k; \boldsymbol{\xi}) \;=\; \begin{cases} f(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{z}_k; \boldsymbol{\xi}) & \text{if } \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{B}\boldsymbol{y}_k \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \quad \text{and} \\ & \quad \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{D}\boldsymbol{y}_k \leqslant \boldsymbol{h}(\boldsymbol{\xi}), \\ +\infty & \text{otherwise.} \end{cases} \tag{12b}$$

Problem (12) can be solved through iterative solution schemes, see Bertsimas and Caramanis (2010) and Subramanyam et al. (2020).

Theorem 2. *Under the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**), the optimal value and the set of feasible (optimal) first-stage decisions $\boldsymbol{x}$ in problems* (11) *and* (12) *coincide.*

Proof. Since problem (12) constitutes a restriction of problem (11), it is infeasible whenever problem (11) is. In the remainder, we may thus assume that problem (11) is feasible. We show that for every fixed first-stage decisions $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z}_k \in \mathcal{Z}(\boldsymbol{x})$, $k = 1, \ldots, K$, we can construct a collection of K affine decision rules $\boldsymbol{y}_k^\ell : \Xi \xrightarrow{\text{a}} \mathbb{R}^n$, $k = 1, \ldots, K$, such that

$$\sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{z}_k\}_k; \tilde{\boldsymbol{\xi}})\right] \;=\; \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell(\tilde{\boldsymbol{\xi}})\}_k, \{\boldsymbol{z}_k\}_k; \tilde{\boldsymbol{\xi}})\right], \tag{13}$$

where the cost functions on the left-hand side and right-hand side are defined in (11b) and (12b), respectively. Equation (13) immediately implies the statement of the theorem.

To prove that (13) holds, we show that

$$\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell\}_k, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}) \;\leqslant\; \mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}) \qquad \forall \boldsymbol{\xi} \in \Xi \tag{14}$$

for the aforementioned set of affine decision rules $\{\boldsymbol{y}_k^\ell\}_k$. Note that the left-hand side of this inequality refers to the objective of the second-stage problem of (12) whose recourse actions are restricted to $\{\boldsymbol{y}_k^\ell\}_k$, whereas the right-hand side refers to the objective of the second-stage problem of (11) under the optimal recourse actions. Thus, the inequality reverse to the one in (14) holds by construction, and together with assumption (**R**), equation (14) then implies equation (13).

To see that equation (14) holds, note first that the assumptions (**F'**), (**A'**), (**D**) and (**B'**) allow us to conclude that for each $k = 1, \ldots, K$, we have

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k^\ell(\boldsymbol{\xi}), \boldsymbol{z}_k; \boldsymbol{\xi}) \;\;\leqslant\;\; \mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi}) \qquad \forall \boldsymbol{\xi} \in \Xi \tag{15}$$

for the affine decision rule $\boldsymbol{y}_k^\ell : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$ defined via $\boldsymbol{y}_k^\ell(\boldsymbol{\xi}) = \boldsymbol{B}_{\mathcal{I}}^{-1}\left[\boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} - \boldsymbol{E}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{z}_k\right]$. The justification of equation (15) is the same as in the proof of Theorem 1 and is thus omitted. Note that both sides of the inequality in (15) may evaluate to ∞.

Fix any $\boldsymbol{\xi} \in \Xi$ and assume that the right-hand side of equation (14) is attained by the constituent function $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi})$, where $k \in \{1, \ldots, K\}$. In that case, equation (15) implies that

$$\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell\}_k, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}) \;\;\leqslant\;\; \mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k^\ell(\boldsymbol{\xi}), \boldsymbol{z}_k; \boldsymbol{\xi}) \;\;\leqslant\;\; \mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi}) \;\;=\;\; \mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{z}_k\}_k; \boldsymbol{\xi}).$$

Since the parameter selection $\boldsymbol{\xi} \in \Xi$ was arbitrary, equation (14) and, *a fortiori*, the statement of the theorem, follow. This concludes the proof. □

In analogy to Proposition 1, we can show that the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**) are tight in the sense that if any of these assumptions is violated, then the statement of Theorem 2 ceases to hold in general even if all other assumptions are satisfied. Since the proof does not require any new ideas over those in the proof of Proposition 1, we omit the details.

We illustrate the statement of Theorem 2 with an example.

Example 2. *Consider the problem of minimizing* $\mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\boldsymbol{\xi}})\big]$, *which has no first-stage decisions* $\boldsymbol{x}$, *whose ambiguity set* $\mathcal{P} = \{\mathbb{P}\}$ *is a singleton set that contains the uniform distribution supported on* $[0,1]^2$ *and whose risk measure is the expected value. The second-stage problem* $\mathcal{Q}(\boldsymbol{\xi})$ *is given as*

$$\mathcal{Q}(\boldsymbol{\xi}) \;\;=\;\; \left[\begin{array}{ll} \text{minimize} & y_1 + y_2 + y_3 + 5z \\ \text{subject to} & y_1 + z \geqslant \xi_1 + \xi_2 \\ & y_2 + z \geqslant \xi_1 - \xi_2 \\ & y_3 + z \geqslant \xi_2 - \xi_1 \\ & y_1 \leqslant 1, \;\; y_2, y_3 \leqslant 0, \;\; z \in [0,1] \end{array}\right],$$

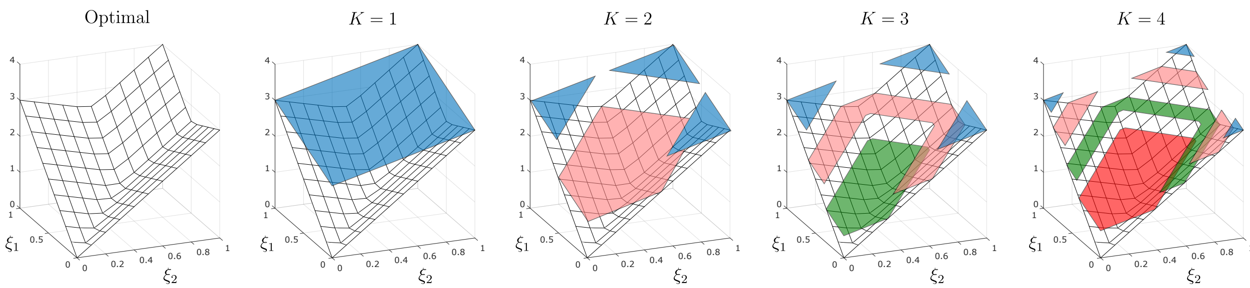

Figure 2. Objective values corresponding to the optimal second-stage policy and the optimal affine decision rules for $K = 1, \ldots, 4$ pre-selected candidate decisions $\{z_k\}_{k=1}^K$. Different colors correspond to realizations of $\boldsymbol{\xi}$ where a different affine policy is optimal.

*and thus the instance satisfies the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**). Figure 2 illustrates the optimal value of the second-stage problem $\mathcal{Q}(\boldsymbol{\xi})$ as well as the optimal value of the K-adaptability problem $\mathcal{Q}(\{\boldsymbol{y}_k\}_k, \{z_k\}_k; \boldsymbol{\xi})$ for $K = 1, \ldots, 4$, where $\{\boldsymbol{y}_k\}_k$ and $\{z_k\}_k$ are chosen optimally.*

We close this section with two immediate consequences of Theorem 2.

Remark 5 (Optimality of the K-Adaptability Problem)**.** *Assume that $|\bigcup_{\boldsymbol{x} \in \mathcal{X}} \mathcal{Z}(\boldsymbol{x})| < \infty$, which holds, for example, if both $\boldsymbol{x}$ and $\boldsymbol{z}$ are discrete decision vectors that are restricted to bounded sets. In that case, Theorem 2 implies that the K-adaptability problem* (12) *recovers an optimal solution to the original two-stage distributionally robust optimization problem* (10) *for sufficiently large K, given that the assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**) are satisfied. To our best knowledge, this is the first optimality result for instances of the K-adaptability problem where the second-stage variables are not exclusively discrete.*

Remark 6 (Suboptimality of Affine Decision Rules in Problem (1))**.** *Consider an instance of the two-stage distributionally robust optimization problem* (1) *from Section 2 where the second-stage decisions can be decomposed into vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ such that the weaker set of assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**) is satisfied. We can then interpret the affine decision rule problem* (2) *as a* 1*-adaptability approximation to problem* (1) *where the candidate decision $\boldsymbol{z}_1$ is an affine decision rule whose dependence on $\boldsymbol{\xi}$ is absorbed in the recourse matrices $\boldsymbol{E}$ and $\boldsymbol{F}$. Theorem 2 then implies that for the fixed first-stage decision $\boldsymbol{x}$ and the fixed affine decision rule $\boldsymbol{z}_1$, the affine decision rule $\boldsymbol{y}$ is optimal. In other words, if an instance of the two-stage distributionally robust optimization problem* (1) *from Section 2 satisfies the weaker assumptions (**R**), (**F'**), (**A'**), (**D**) and (**B'**) under* *which affine decision rules are not optimal, then the suboptimality is solely caused by the affine decision rule $\boldsymbol{z}_1$, whereas the affine decision rule $\boldsymbol{y}$ is optimal. This opens room for tailored solution approaches that directly address the suboptimality of $\boldsymbol{z}_1$.*

4 Applications

We next apply our theory from Sections 2 and 3 to different domains. The dual purpose of this section is to demonstrate the breadth of applications that are amenable to our optimality results as well as highlight different technical aspects of our earlier findings.

Section 4.1 demonstrates how our optimality results encompass and extend previously developed results about the optimality of affine decision rules in vehicle routing problems. Section 4.2 shows how a supply chain management problem that is by itself not amenable to our optimality results can be solved optimally in affine decision rules by imposing an additional assumption on the distribution network. Section 4.3 illustrates the versatility of the objective functions that are supported by our optimality result in the context of a healthcare scheduling problem. Section 4.4 applies Remark 3 on second-stage equality constraints to inventory management problems. Section 4.5 discusses a production planning problem where affine decision rules are optimal despite the potential presence of multiple cycles in the production graph. Section 4.6, finally, exploits the structure inherent in two-stage robust optimization problems so that our optimality results impose conditions on the geometry of the uncertainty set, as opposed to the structure of the second-stage constraints.

For ease of exposition, we do not always convert the problem formulations into the standard form of problem (1) in this section. However, constraints belonging to the first set in (1b), $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi})$, will always be written as greater or equal constraints, while constraints of the second set in (1b), $\boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{D}\boldsymbol{y} \leqslant \boldsymbol{h}(\boldsymbol{\xi})$, will always be formulated as less or equal constraints.

4.1 Logistics

Consider a complete, directed graph $G = (V, A)$ whose nodes $V = \{0, 1, \ldots, n\}$ represent a unique depot $i = 0$ and geographically dispersed customers $i \in V_C = \{1, \ldots, n\}$ with uncertain demands $\tilde{\xi}_i$ for a single good that are governed by some probability distribution $\mathbb{P} \in \mathcal{P}$. We assume that $\Xi \subseteq \mathbb{R}^n_+$, that is, the customer demands are non-negative, and $\mathbb{P}\big[\tilde{\xi}_j > 0\big] > 0$ for all $j \in V_C$ and $\mathbb{P} \in \mathcal{P}$, that is, no customer demand vanishes $\mathbb{P}$-almost surely under any $\mathbb{P} \in \mathcal{P}$. A company has at its disposal m

homogeneous vehicles of capacity Q that can traverse the arcs $(i,j) \in A = \{(i,j) \in V \times V \,:\, i \neq j\}$ at transportation costs $c_{ij} \in \mathbb{R}_+$. The company wishes to determine a route for each vehicle so that all customer demands are satisfied without split deliveries (*i.e.*, each customer is served by exactly one vehicle) at minimum transportation cost. The deterministic version of this problem is known as the capacitated vehicle routing problem (CVRP), and it has been studied intensively ever since its inception in the 1950s (Dantzig and Ramser, 1959).

We can formulate the problem as the two-stage distributionally robust optimization problem

$$\begin{array}{lll} \text{minimize} & \sup\limits_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] & \\ \text{subject to} & \sum\limits_{i \in V} x_{ij} = \sum\limits_{i \in V} x_{ji} = \begin{cases} m & \text{if } j = 0, \\ 1 & \text{otherwise} \end{cases} & \forall j \in V \\ & x_{ij} \in \{0,1\},\ (i,j) \in A, & \end{array} \tag{16a}$$

where the decision variable x_{ij} indicates whether one of the vehicles traverses the arc $(i,j) \in A$, and where the constraint ensures that each vehicle leaves and enters the depot node, whereas each customer node is visited (and subsequently left) by exactly one vehicle. For each first-stage decision $\boldsymbol{x}$ and for each realization $\boldsymbol{\xi}$ of the customer demands, the second-stage problem is defined as

$$\mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}) = \left[\begin{array}{lll} \text{minimize} & \sum\limits_{(i,j) \in A} c_{ij} x_{ij} & \\ \text{subject to} & y_j + \mathrm{M}(1 - x_{ij}) \geqslant y_i + \xi_j & \forall j \in V_C,\ \forall i \in V,\ i \neq j \\ & y_j \leqslant Q & \forall j \in V_C \\ & y_j \in \mathbb{R}_+,\ j \in V & \end{array} \right], \tag{16b}$$

where M is a sufficiently large positive constant so that the first constraint is redundant whenever $x_{ij} = 0$ (our subsequent discussion will reveal that $\mathrm{M} = Q$ is sufficient). The objective function minimizes the overall transportation costs. The constraints ensure that for each customer $j \in V_C$, the decision variable y_j is at least as large as the cumulative customer demands served by the vehicle that visits customer j, immediately after it has left customer j. Since $\mathbb{P}\big[\tilde{\xi}_j > 0\big] > 0$ for all $j \in V_C$ and $\mathbb{P} \in \mathcal{P}$, the second-stage constraints ensure that no vehicle can return to a previously visited customer, that is, they eliminate subtours that do not involve the depot node. Since each y_j, $j \in V_C$, is also bounded from above by Q, the constraints further ensure that the vehicles' capacities are obeyed. Problem (16) follows the idea of the Miller-Tucker-Zemlin formulation that has first been proposed for the deterministic CVRP by Kulkarni and Bhave (1985).

We now argue that problem (16) is optimized by affine decision rules. Indeed, one readily verifies that the second-stage problem (16b) satisfies the assumptions (**F**), (**A**) and (**D**) of Section 2. To see that assumption (**B**) is satisfied as well, we show that the conditions of Proposition 2 are met.

Observation 1. *The second-stage problem* (16b) *satisfies the conditions of Proposition 2.*

Proof. The first condition of Proposition 2 is trivially satisfied for y_0, which is only restricted by the lower bound $y_0 \geqslant 0$. Consider now any other decision y_j, $j \in V_C$, and let $p \in V$ be the unique node that satisfies $x_{pj} = 1$. In that case, the constraint $y_j \geqslant y_p + \xi_j - \mathrm{M}(1 - x_{pj})$ weakly dominates the other lower bounds on y_j since $x_{ij} = 0$ for all $i \in V \backslash \{p, j\}$ while $x_{pj} = 1$, and the non-negativity of $\boldsymbol{y}$ and $\boldsymbol{\xi}$ implies that the right-hand side $y_p + \xi_j - \mathrm{M}(1 - x_{pj})$ of this inequality is non-negative.

In view of the second condition of Proposition 2, we note that the arcs $(p, j) \in A$ satisfying $x_{pj} = 1$ induce a partition of the customer set V_C into m routes $\boldsymbol{R}_k = (R_{k,0}, R_{k,1}, \ldots, R_{k,n_k}, R_{k,n_k+1})$, $k = 1, \ldots, m$, satisfying $R_{k,0} = R_{k,n_k+1} = 0$ and $R_{k,l} \in V_C$, $l = 1, \ldots, n_k$, such that $x_{pj} = 1$ if and only if $p = R_{k,l}$ and $j = R_{k,l+1}$ for some $k \in \{1, \ldots, m\}$ and $l \in \{0, \ldots, n_k\}$, and this partition is unique up to a reordering of the routes. Thus, each weakly dominant constraint $y_j \geqslant y_p + \xi_j - \mathrm{M}(1 - x_{pj})$ from the previous paragraph links two nodes p and j satisfying $p = R_{k,l}$ and $j = R_{k,l+1}$ for some $k \in \{1, \ldots, m\}$ and $l \in \{0, \ldots, n_k - 1\}$. In this case, the permutation $\pi(0) = 0$ and

$$\pi(R_{k,l}) \;=\; \left[\sum_{\kappa=1}^{k-1} n_\kappa\right] + l \qquad \forall k = 1, \ldots, m, \; \forall l = 1, \ldots, n_k,$$

which enumerates the customers $j \in V_C$ in order of ascending vehicle indices k and in order of their position within the k-th route, certifies that the right-hand side coefficient vectors $\boldsymbol{\beta}_{jp} = \mathbf{e}_p$ of the weakly dominant lower bounds $y_j \geqslant y_p + \xi_j - \mathrm{M}(1 - x_{pj})$ form a partial order. Here, $\mathbf{e}_p$ denotes the p-th canonical basis vector in $\mathbb{R}^n$. □

Note that the objective function $f(\boldsymbol{x}; \boldsymbol{\xi}) = \sum_{(i,j) \in A} c_{ij} x_{ij}$ in (16b) involves neither the second-stage decisions $\boldsymbol{y}$ nor the uncertain parameters $\boldsymbol{\xi}$. As a result, the feasible region and optimal value of problem (16) do not depend on the risk measure $\rho_{\mathbb{P}}$ as long as $\sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}})\big] = \infty$ whenever $\mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}) = \infty$ for some $\boldsymbol{\xi} \in \Xi$, which in turn is typically required to derive a tractable reformulation for the problem (*cf.* Remark 4). The choice of the risk measure becomes important, however, once the objective function in (16b) balances the weight across the vehicles by accounting for the maximum vehicle load $\max\{y_j \,:\, j \in V_C\}$. Since assumption (**F**) continues to be satisfied, affine decision rules remain optimal as long as the employed risk measure satisfies (**R**).

The optimality of affine decision rules is not restricted to the two-stage Miller-Tucker-Zemlin formulation (16). In fact, arguments similar to those in the proof of Observation 1 show that affine decision rules are also optimal in the two-stage 1-commodity (Gouveia, 1995) and 2-commodity (Baldacci et al., 2004) flow formulations. Moreover, affine decision rules are optimal in two-stage capacitated arc routing problems, where the distribution or collection of goods appear on the arcs of the network (Golden and Wong, 1981), and they remain optimal in two-stage capacitated location routing problems (Contardo et al., 2013), which simultaneously optimize over depot locations and delivery routes. In all of these problems, our results from Section 3 allow us to postpone some or all of the routing decisions to the second stage (after the demand has been observed). While the postponed routing decisions will be suboptimal (unless K is chosen large enough), the affine decision rules for $\boldsymbol{y}$ will remain optimal in the emerging K-adaptability formulations.

Remark 7 (Bibliographical Notes)**.** *The optimality of affine decision rules in the Miller-Tucker Zemlin and the commodity flow formulations has first been shown by Gounaris et al. (2013) through intricate* ad hoc *arguments in the context of the two-stage robust CVRP. To our best knowledge, the optimality in the problem variant that balances the vehicles' loads, the optimality in two-stage capacitated arc routing and location routing problems as well as the optimality in K-adaptability formulations has not been discussed in the literature.*

4.2 Supply Chain Management

We next study a multi-echelon supply chain design problem faced by a company that sells multiple goods $g \in \mathcal{G} = \{1, \ldots, G\}$. Let $\mathcal{N} = \{1, \ldots, N\}$ be a set of nodes, where each node $i \in \mathcal{N}$ corresponds to a location with a distribution center that faces an uncertain demand $\tilde{\xi}_{gi}$ for every good $g \in \mathcal{G}$. The company wishes to build one specialized production facility for each good as well as up to W warehouses. Each warehouse can carry a combination of goods as long as their combined storage requirements, computed from the per-unit sizes s_g for each good g, do not exceed the warehouse capacity S. Each good is transported either directly from its production facility to a distribution center, or it is transported indirectly via one or multiple warehouses. The per-unit transportation costs from location i to location j for good g amount to c_{gij} for transshipments between factories and warehouses and to $d_{gij} > c_{gij}$ for transshipments to distribution centers, thus reflecting different modes of transportation. Note that we do not require the transportation costs to satisfy the triangle inequality, which implies that an optimal solution may ship goods between multiple warehouses before they reach a distribution center. The company wishes to determine the distribution network upfront, that is, before the demands are known, whereas the actual transshipment quantities can be selected once the demands have been observed. The objective is to serve all demands at minimum overall transportation costs.

We can formulate the problem as the two-stage distributionally robust optimization problem

$$
\begin{array}{lll}
\text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}; \tilde{\boldsymbol{\xi}}) \right] & \\
\text{subject to} & \sum_{i \in \mathcal{N}} x_{gi} = 1 & \forall g \in \mathcal{G} \\
& \sum_{i \in \mathcal{N}} y_i \leqslant W & \\
& \sum_{i \in \mathcal{N}} z_{gij} \leqslant 1 & \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& \sum_{i \in \mathcal{N}} w_{gij} = 1 & \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& z_{gii} = 0 & \forall g \in \mathcal{G},\ \forall i \in \mathcal{N} \\
& \boldsymbol{x} \in \{0,1\}^{GN},\ \ \boldsymbol{y} \in \{0,1\}^{N},\ \ \boldsymbol{z} \in \{0,1\}^{GN^2}\ \ \boldsymbol{w} \in \{0,1\}^{GN^2} & ,
\end{array}
\tag{17a}
$$

where the decisions x_{gi} and y_i determine whether a production facility for good $g \in \mathcal{G}$ or a warehouse should be erected at node $i \in \mathcal{N}$, respectively, the decisions z_{gij} record whether the link $(i, j) \in \mathcal{N} \times \mathcal{N}$ is part of the distribution network for good $g \in \mathcal{G}$, and w_{gij} decides whether the distribution center $j \in \mathcal{N}$ receives its stock of good $g \in \mathcal{G}$ from a production facility or warehouse located at node $i \in \mathcal{N}$. The first constraint ensures that exactly one production facility is built for each product, and the second constraint allows for up to W warehouses to be erected. The third constraint stipulates that the distribution network $\boldsymbol{z}_g = (z_{gij})_{i,j}$ for each good $g \in \mathcal{G}$ is a tree, that is, it precludes networks where a warehouse receives the same product from multiple sources. This is not a business requirement, but it will turn out to be crucial for the optimality of affine decision rules. The fourth constraint matches each distribution center to a single production facility or warehouse.

The second-stage costs $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}; \boldsymbol{\xi})$ of our supply chain management problem can be cast as the optimal value of the following second-stage problem.

$$\begin{aligned}
\text{minimize} \quad & \sum_{g \in \mathcal{G}} \sum_{i,j \in \mathcal{N}} [c_{gij} f_{gij} + d_{gij} \phi_{gij}] \\
\text{subject to} \quad & \mathrm{M} \cdot x_{gj} + \sum_{i \in \mathcal{N}} f_{gij} \geqslant \sum_{i \in \mathcal{N}} f_{gji} + \sum_{i \in \mathcal{N}} \phi_{gji} && \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{N}} s_g f_{gij} \leqslant S \cdot y_j && \forall j \in \mathcal{N} \\
& \sum_{i \in \mathcal{N}} \phi_{gij} \geqslant \xi_{gj} && \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& f_{gij} \leqslant \mathrm{M} \cdot z_{gij}, \quad \phi_{gij} \leqslant \mathrm{M} \cdot w_{gij} && \forall g \in \mathcal{G},\ \forall i, j \in \mathcal{N} \\
& f_{gij}, \phi_{gij} \in \mathbb{R}_+,\ g \in \mathcal{G} \text{ and } i, j \in \mathcal{N}.
\end{aligned} \tag{17b}$$

In this problem, the decision variables f_{gij} and ϕ_{gij} record the internal (*i.e.*, between production facilities and warehouses) and external (*i.e.*, from warehouses and production facilities to distribution centers) transshipments of product $g \in \mathcal{G}$ across the locations $i, j \in \mathcal{N}$, respectively. The objective function minimizes the overall transportation costs. The constraints, from top to bottom, ensure that product flows are conserved across the nodes, the warehouse capacities are obeyed, the customer demands are satisfied and the transshipments are limited to the distribution network $\boldsymbol{z}$ and the matching $\boldsymbol{w}$ selected in the first stage.

Observation 2. *Problem* (17) *is optimally solved in affine decision rules if (**R**) is satisfied.*

Proof. We proceed in three steps. The first two steps develop equivalent reformulations for the first and the third constraint in (17b), which by themselves do *not* satisfy assumption (**B**) from Section 2. The third step employs Proposition 2 to prove the statement of the observation.

In view of the first step, we claim that for each $g \in \mathcal{G}$ and $j \in \mathcal{N}$, the first constraint in (17b) is equivalent to the following set of constraints, one for each nodal subset $\mathcal{T} \subseteq \mathcal{N}$:

$$\begin{cases}
x_{gj} = 0 \wedge z_{gji} = 1\ \forall i \in \mathcal{T} \wedge z_{glj} = 1 & \Longrightarrow \quad f_{glj} \geqslant \sum_{i \in \mathcal{T}} f_{gji} + \sum_{i \in \mathcal{N}} \phi_{gji} \quad \forall l \in \mathcal{N} \\
x_{gj} = 0 \wedge z_{gji} = 1\ \forall i \in \mathcal{T} \wedge z_{gij} = 0\ \forall i \in \mathcal{N} & \Longrightarrow \quad 0 \geqslant \sum_{i \in \mathcal{T}} f_{gji} + \sum_{i \in \mathcal{N}} \phi_{gji}
\end{cases} \tag{18}$$

Indeed, fix any first-stage decision $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})$ and any uncertainty realization $\boldsymbol{\xi} \in \Xi$. We claim that any second-stage decision $(\boldsymbol{f}, \boldsymbol{\phi})$ satisfying the fourth and fifth constraint of (17b) satisfies the first constraint of (17b) if and only if it satisfies (18) for all $\mathcal{T} \subseteq \mathcal{N}$. To this end, assume first that $\sum_{i \in \mathcal{N}} z_{gij} = 1$. In that case, the first constraint in (17b) simplifies to

$$\mathrm{M} \cdot x_{gj} + f_{glj} \geqslant \sum_{i \in \mathcal{T}} f_{gji} + \sum_{i \in \mathcal{N}} \phi_{gji}$$

for the unique $l \in \mathcal{N}$ satisfying $z_{glj} = 1$ and $\mathcal{T} = \{i \in \mathcal{N} : z_{gji} = 1\}$, which one readily recognizes to be equivalent to the weakly dominant constraint in (18). Assume next that $\sum_{i \in \mathcal{N}} z_{gij} = 0$; according to the third constraint in (17a), this is the only alternative case to consider. In that case, the first constraint in (17b) simplifies to

$$\mathrm{M} \cdot x_{gj} \geqslant \sum_{i \in \mathcal{T}} f_{gji} + \sum_{i \in \mathcal{N}} \phi_{gji}$$

for $\mathcal{T} = \{i \in \mathcal{N} : z_{gji} = 1\}$, which again coincides with the weakly dominant constraint in (18). Note that (18) can be readily linearized by adding the expressions $\mathrm{M} \cdot (x_{gj} + |\mathcal{T}| - \sum_{i \in \mathcal{T}} z_{gji} + 1 - z_{glj})$ and $\mathrm{M} \cdot (x_{gj} + |\mathcal{T}| - \sum_{i \in \mathcal{T}} z_{gji} + \sum_{i \in \mathcal{N}} z_{gij})$ to the constraint left-hand sides, respectively.

As for the second step, a similar argument as in the previous paragraph shows that for each $g \in \mathcal{G}$ and $j \in \mathcal{N}$, the third constraint in (17b) is equivalent to the constraint set

$$\mathrm{M} \cdot (1 - w_{gij}) + \phi_{gij} \geqslant \xi_{gj} \qquad \forall i \in \mathcal{N}; \tag{19}$$

we omit the details of this step for the sake of brevity.

So far, we have shown that the objective function, the decision variables as well as the feasible region of problem (17b) remain unchanged for all feasible first-stage decisions as well as all uncertainty realizations if we replace the first and third constraint in (17b) by (18) and (19), respectively. Thus, affine decision rules are optimal in the original problem (17b) if and only if they are optimal in the reformulated problem with constraints (18) and (19). Moreover, the reformulated problem clearly satisfies the assumptions (**F**), (**A**) and (**D**). To conclude Step 3 of the proof, we make use of Proposition 2 to show that assumption (**B**) is satisfied as well.

In view of condition *(i)* of Proposition 2, consider first the second-stage decision f_{glj}, $g \in \mathcal{G}$ and $l, j \in \mathcal{N}$ with $l \neq j$. If $z_{glj} = 0$ or $x_{gj} = 1$, the non-negativity constraint $f_{glj} \geqslant 0$ is weakly dominant. If $z_{glj} = 1$ and $x_{gj} = 0$, on the other hand, then the first constraint in (18) with $\mathcal{T} = \{i \in \mathcal{N} : z_{gji} = 1\}$ weakly dominates all other constraints of (18) as well as the non-negativity constraint. Similar arguments apply to f_{gjj}, $g \in \mathcal{G}$ and $j \in \mathcal{N}$, as well as ϕ_{gij}, $g \in \mathcal{G}$ and $i, j \in \mathcal{N}$.

In view of condition *(ii)* of Proposition 2, recall that (17a) imposes a tree structure on the distribution network $\boldsymbol{z}_g = (z_{gij})_{i,j}$ for each good $g \in \mathcal{G}$, and that the matching $\boldsymbol{w}_g = (w_{gij})_{i,j}$ assigns each distribution center $j \in \mathcal{N}$ to a single node of $\boldsymbol{z}_g$. Since every tree imposes a partial order, we can determine a permutation of the variables f_{gij} and ϕ_{gij}, $i, j \in \mathcal{N}$, such that each weakly dominant constraint from the previous paragraph has a left-hand side variable whose order is higher than the order of all right-hand side variables. This implies that the right-hand side coefficient vectors of these constraints form a partial order, which completes the proof. □

One can readily adapt the proof of Observation 2 to confirm the optimality of affine decision rules in several variants of the problem, such as instances where factories and warehouses incur location-dependent construction costs, where factories can produce multiple products (possibly with a penalty for diversification) or where the warehouse capacities are decision variables (which incur costs in the objective function).

We emphasize that the product-wise tree structure of the distribution network, which we imposed in addition to the original business requirements, is crucial to ensure the optimality of affine decision rules in Observation 2. A similar approach of imposing additional structure onto a problem (such as separate distribution channels for different products or acyclicity) may prove useful to obtain optimality guarantees for affine decision rules in other application domains as well.

Remark 8 (Bibliographical Notes)**.** *A rich body of literature is devoted to two-stage robust network design problems where the decision maker selects arc capacities in the first stage, then observes the uncertain supplies and demands and finally responds with flows that balance the network. The problem is typically solved by projecting the feasible region onto the first-stage variables through an iterative cut separation, which obviates the need to explicitly model the second-stage decisions. Minoux (2010) proves the NP-hardness of this problem as well as its separation problem. Cacchiani et al. (2016) solve the problem exactly via a branch-and-cut algorithm. Atamtürk and Zhang (2007), Ordóñez and Zhao (2007) and Minoux (2010) prove polynomial solvability of specific instances, such as those whose graphs contain a single supply-demand pair, admit a total order or form an arborescence, and those whose uncertainty sets have polynomially many extreme points or constitute hyperrectangles. Babonneau et al. (2013), Poss and Raack (2013), Poss (2014) and Mattia and Poss (2018) solve the problem suboptimally in affine decision rules and alternative policy classes.*

Suboptimal affine decision rules have also been applied to an emergency response and evacuation traffic flow problem by Ben-Tal et al. (2011) and to a lot sizing problem as well as a facility location problem on a bipartite graph by Bertsimas and de Ruiter (2016), respectively.

4.3 Accident & Emergency Scheduling

Medical appointment scheduling has a rich history in operations research (Cayirli and Veral, 2003; Gupta and Denton, 2008). To account for the uncertainty inherent in the appointment durations, a number of distributionally robust optimization formulations have been developed recently. Under the assumption that only the means and the covariances of the appointment durations are known, Kong et al. (2013) derive a co-positive program that minimizes the worst-case expected waiting and overtime costs, and they show that this problem admits a tractable relaxation as a semi-definite program. If only some marginal moments of the appointment durations are known, Mak et al. (2015) solve the same problem exactly by means of a conic program. Bertsimas et al. (2019) employ lifted affine decision rules to conservatively minimize the worst-case expected waiting and overtime costs over all duration distributions with known marginal first and second moments. Accounting for all duration distributions with known partial cross-moments, Zhen et al. (2018) develop an alternative conservative approximation that combines a partially executed Fourier-Motzkin elimination with affine decision rules. Extensions to Wasserstein ambiguity sets and to the no-show behavior of patients have been reported by Jiang et al. (2019) and Kong et al. (2020).

Here, we consider a patient scheduling problem faced by an accident & emergency (A&E) department. At the beginning of a shift, n patients in need of acute care arrive without prior appointments. Although an initial screening is performed immediately upon arrival, the broad range of illnesses and injuries implies that the estimated treatment times are subject to high uncertainty. Moreover, some patients might arrive in life threatening conditions and must thus be prioritized. The goal is to determine an order in which the patients are seen so that the overall health outcome is maximized. Note that in practice, not all patients arrive at the same time. In this case, the problem described here can be resolved (with updated waiting times for those patients that are already present) every time a new patient arrives, or after a batch of new patients has arrived.

We formulate the problem as the two-stage distributionally robust optimization problem

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\right] \\ \text{subject to} & \sum_{i\neq j} x_{ij} \leqslant 1, \;\; \sum_{i\neq j} x_{ji} \leqslant 1 \qquad \forall j = 1,\ldots,n \\ & \sum_{i=1}^{n}\sum_{j\neq i} x_{ij} = n-1 \\ & x_{ij} \in \{0,1\},\; i,j = 1,\ldots,n \text{ and } i \neq j, \end{array} \tag{20a}$$

where the decision variable x_{ij} indicates whether patient j is seen immediately after patient i. The first constraint set ensures that each patient is preceded and succeeded by at most one other patient, and the second constraint ensures that all patients form a single ordered sequence. For each first-stage scheduling decision $\boldsymbol{x}$ and for each realization $\boldsymbol{\xi}$ of the patients' treatment times, the second-stage problem can be formulated as follows.

$$\mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \;=\; \left[\begin{array}{lll} \text{minimize} & f(\boldsymbol{g}(\boldsymbol{y})) & \\ \text{subject to} & y_j + \mathrm{M}(1-x_{ij}) \geqslant y_i + \xi_i & \forall i,j = 1,\ldots,n \text{ with } i \neq j \\ & \boldsymbol{y} \in \mathbb{R}^n_+ & \end{array}\right] \tag{20b}$$

Here, M is a positive constant that is sufficiently large so as to make the first constraint redundant whenever $x_{ij} = 0$ (the essential supremum of the sum of all treatment times is sufficient). The constraints ensure that each y_j weakly exceeds the earliest possible start time of patient j. The constraints of our A&E scheduling problem are structurally similar to those of the logistics problem (16) studied in Section 4.1, and thus the next result is stated without formal proof.

Observation 3. *Problem* (20b) *is solved in affine decision rules if (**F**) and (**R**) hold.*

The second-stage objective function is composed of two functions. The inner function $\boldsymbol{g} : \mathbb{R}^n_+ \to \mathbb{R}^n$ maps the patients' waiting times to their health outcomes. We assume that $\boldsymbol{g}$ is component-wise non-decreasing, that is, longer waiting times correspond to worse health outcomes. In particular, a patient $j \in \{1,\ldots,n\}$ in need of immediate attention is represented by a quickly increasing component function g_j, whereas the component functions of patients with mild conditions will increase less rapidly. The outer function $f : \mathbb{R}^n \to \mathbb{R}$ aggregates the n individual health outcomes $g_1(\boldsymbol{y}),\ldots,g_n(\boldsymbol{y})$ to an overall health outcome. We list some possible choices for this function.

(i) **Sum of all health outcomes:** $f_1(\boldsymbol{q}) = \mathbf{e}^\top \boldsymbol{q}$. Since f_1 and $\boldsymbol{g}$ are monotonically non-decreasing and monotonicity is preserved under compositions, the objective function $f_1(\boldsymbol{g}(\boldsymbol{y}))$ satisfies condition (**F**). For the special case where $\boldsymbol{g}$ is affine, the ambiguity set $\mathcal{P}$ is a singleton and the risk measure is the expectation, the 'weighted shortest expected processing time first' rule optimizes problem (20), see Theorem 10.1.1 of Pinedo (2012).

(ii) **Sum of the K worst health outcomes:** $f_2(\boldsymbol{q}) = \sum_{k=1}^{K} q_{[k]}$, where $q_{[k]}$ is the k-th largest component of $\boldsymbol{q}$ (with ties broken arbitrarily). A similar argument as in *(i)* shows that

$f_2(\boldsymbol{g}(\boldsymbol{y}))$ satisfies condition (**F**). Note that f_2 has the piecewise affine representation

$$f_2(\boldsymbol{q}) \;=\; \max\left\{\sum_{k=1}^{K} q_{i_k} \;:\; 1 \leqslant i_1 < i_2 < \ldots < i_K \leqslant n\right\}$$

that consists of $\binom{n}{K}$ affine pieces. Despite this large number, the affine decision rule formulation of problem (20) with objective function $f_2(\boldsymbol{g}(\boldsymbol{y}))$ can often be solved efficiently via iterative solution schemes that only consider the relevant pieces of f_2, see Remark 4. Note also that the function f_1 from *(i)* is a special case of the function f_2 where $K = n$.

(iii) **Rank-weighted sum of all health outcomes:** $f_3(\boldsymbol{q}) = \sum_{k=1}^{n} w_k \cdot q_{[k]}$, where the weights satisfy $w_1 \geqslant w_2 \geqslant \ldots \geqslant w_n \geqslant 0$ and the notation $q_{[k]}$ coincides with that of *(ii)*. One readily verifies that the function $f_3(\boldsymbol{g}(\boldsymbol{y}))$ satisfies condition (**F**). Moreover, f_3 can be represented as

$$f_3(\boldsymbol{q}) \;=\; \max\left\{\sum_{k=1}^{n} w_k \cdot q_{i_k} \;:\; 1 \leqslant i_1 \neq i_2 \neq \ldots \neq i_n \leqslant n\right\},$$

which consists of $n!$ affine pieces. As before, the corresponding affine decision rule formulation of problem (20) can be solved iteratively. Note also that the function f_2 from *(ii)* is a special case of the function f_3 where $w_1 = \ldots = w_K = 1$ and $w_{K+1} = \ldots = w_n = 0$.

We can envision several variants of problem (20) where affine decision rules remain optimal. For example, one could consider different treatment options that can be selected subject to resource constraints and that lead to different treatment times and/or health outcomes. The treatment options could be selected either here-and-now or, by utilizing a K-adaptability formulation, wait-and-see once the patient is seen by a doctor. Likewise, we can generalize problem (20) to multiple patient queues that are seen by different doctors (possibly restricted by additional matching constraints), where the assignment to a doctor can again impact the treatment time and/or health outcome, and where the schedule is additionally restricted by shift and break constraints.

4.4 Inventory Management

Inventory management has been among the first modern applications of affine decision rules. Conservative approximations of multi-stage stochastic and robust inventory management problems with affine, quadratic and cubic decision rules have been reported, among others, by Ben-Tal et al. (2004, 2005), Bertsimas et al. (2011c) and Kuhn et al. (2011). An exact Benders' decomposition scheme for the multi-stage robust inventory management problem, which can be interpreted as modelling certain adjustable decisions via piecewise affine decision rules, has been developed by Bienstock and Özbay (2008). To our best knowledge, Bertsimas et al. (2010) were the first to prove that affine decision rules are in fact optimal in a multi-stage single product inventory management problem with ordering, inventory holding and backlogging costs when robust decisions are sought in view of period-wise uncertain customer demands that are only known to be contained in stage-wise rectangular interval uncertainty sets. The result has later been extended to two-echelon supply chains with inventory capacity investments and demand uncertainty sets whose corner points form subsets of the extreme points of the $[0,1]$-hypercube by Iancu et al. (2013). In the following, we complement those findings by showing that affine decisions rules are also optimal in a two-stage distributionally robust inventory management problem with multiple products whose uncertain demands are supported on arbitrary polyhedra and where a here-and-now base stock policy is sought.

The first stage of our inventory management problem reads as follows.

$$\begin{array}{lll} \text{minimize} & \sup\limits_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x},\boldsymbol{p};\tilde{\boldsymbol{\xi}})\right] & \\ \text{subject to} & \sum\limits_{t=1}^{T} x_t \leqslant B & \\ & p_{ti} \leqslant \mathrm{M} x_t & \forall t=1,\ldots,T,\ \forall i=1,\ldots,I \\ & \boldsymbol{x}\in\{0,1\}^T,\ \ \boldsymbol{p}\in\mathbb{R}_+^{T\times N} & \end{array} \tag{21a}$$

Here, M is a sufficiently large number (such as the essential supremum of the sum of all demands). In (21a), the binary decision variables x_t select up to B ordering periods. In each period $t=1,\ldots,T$ with $x_t=1$, where $T\in\mathbb{N}$ denotes the finite time horizon of our problem, the company orders up to p_{ti} units of each product $i=1,\ldots,N$. Upon observing the demands ξ_{ti} for all products across all periods, the company solves the second-stage problem

$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{p};\boldsymbol{\xi}) = \left[\begin{array}{lll} \text{minimize} & \sum\limits_{t=1}^{T}\sum\limits_{i=1}^{N} h_{ti} I_{ti} & \\ \text{subject to} & I_{t+1,i} = \begin{cases} I_{ti}-\xi_{ti} & \text{if } x_t=0 \\ p_{ti}-\xi_{ti} & \text{if } x_t=1 \end{cases} & \begin{array}{l}\forall t=1,\ldots,T,\\ \forall i=1,\ldots,N\end{array} \\ & I_{1i}=I_i^0 & \forall i=1,\ldots,N \\ & \boldsymbol{I}\in\mathbb{R}_+^{(T+1)\times N} & \end{array}\right], \tag{21b}$$

where the decision variables I_{ti} record the inventory levels for each product i across all time periods

t. The objective function minimizes the inventory holding costs across all products. From top to bottom, the constraints enforce the inventory evolution across subsequent time periods, the initial inventory levels I_i^0 as well as the non-negativity of the inventories. The inventory evolution constraints replenish the product inventories up to the levels p_{ti} in the ordering periods t satisfying $x_t = 1$. While we have formulated this constraint as a logical constraint to ease exposition, it is straightforward to re-express it by a set of linear equality constraints using a big-M formulation. Note that we do not exclude the case where $p_{ti} < I_{ti}$ for some realizations of ξ_{ti}, that is, where the inventory is actually depleted rather than replenished. The non-negativity of the inventory levels $I_{t+1,i}$ in the artificial time period $T+1$ ensures that the demands ξ_{Ti} of period T are fully served.

Due to the presence of the first-stage timing decisions $\boldsymbol{x}$ in the inventory evolution constraints, it appears difficult to reformulate problem (21) as a single-stage distributionally robust optimization problem of compact size. Nevertheless, we now show that (21) is solved by affine decision rules.

Observation 4. *The second-stage problem* (21b) *satisfies the conditions (**A**) and (**D**) of Theorem 1 as well as the conditions (**F**') and (**B**') of Remark 3.*

Proof. Define $\tau^-(t) = \max\{\tau = 1, \ldots, t-1 \,:\, x_\tau = 1\}$ as the last ordering period before period t, and set $\tau^-(t) = -\infty$ if there is no ordering period before t. The only candidate decision $\boldsymbol{I}$ satisfying the first constraint of (21b) is then defined via

$$I_{ti}(\boldsymbol{\xi}) \;=\; \begin{cases} p_{\tau^-(t),i} - \displaystyle\sum_{\tau=\tau^-(t)}^{t-1} \xi_{\tau i} & \text{if } \tau^-(t) \neq -\infty, \\ I_i^0 - \displaystyle\sum_{\tau=1}^{t-1} \xi_{\tau i} & \text{if } \tau^-(t) = -\infty, \end{cases} \qquad \forall t = 1, \ldots, T+1,\ \forall i = 1, \ldots, N,$$

which is evidently affine in $\boldsymbol{\xi}$. Note that this decision is infeasible if it violates the non-negativity constraints. The assumptions (**A**), (**D**), (**F**') and (**B**') are then trivially satisfied since problem (21b) contains no second-stage decisions $\boldsymbol{y}$ other than the inventory decisions $\boldsymbol{I}$. □

Since the second stage of our inventory problem has a feasible region with at most one solution for every first-stage decision $(\boldsymbol{x}, \boldsymbol{p})$ and each demand realization $\boldsymbol{\xi} \in \Xi$, the choice of the risk measure $\rho_{\mathbb{P}}$ has no impact on the optimality of affine decision rules. Condition (**R**) must be satisfied, however, once we add second-stage decisions that are not uniquely determined by the constraints.

The inventory management problem discussed in this section can be extended in various directions. For example, the first-stage problem (21a) can be modified to account for minimum time gaps between successive ordering periods, time-dependent lower and upper bounds on the order up-to levels p_{ti} or restrictions on the number of different products that can be ordered in each ordering period. The second-stage problem (21b), on the other hand, can be modified to account for time-dependent lower and upper bounds on the inventory levels of individual items or subsets of items, they can accommodate ramping constraints for the inventory evolution that disallow large variations across subsequent time periods, or they can relate the inventory levels to the demands (*e.g.*, the demand in each period must not exceed 50% of the available inventory). A less immediate extension allows for backlogging by removing the non-negativity constraints on I_{ti} and amending the second-stage objective function (21b) with the expression

$$\sum_{t=1}^{T} \sum_{i=1}^{N} (b_{ti} + h_{ti}) \cdot \max\{-I_{ti}, 0\},$$

where b_{ti} accounts for the per-unit backlogging costs of product i in time period t. Similar arguments as in the proof of Observation 4 show that the emerging problem variant continues to be solved by affine decision rules. Likewise, the inventory problem of this section can be combined with aspects of the supply chain management problem of Section 4.2 so that the second stage contains both inventory decisions that are uniquely defined by equality constraints as well as transshipment decisions that are restricted (but not uniquely determined) by inequality constraints involving the inventory decisions. Finally, we can develop K-adaptability formulations which model the base stock decisions p_{ti}, $t = 2, \ldots, T$, in (21a) as second-stage decisions that are taken after the demands have been observed. Note that in this case, the principle of non-anticipativity is violated since the decisions p_{ti} at time $t = 2, \ldots, T-1$ are taken under the knowledge of the demands $\xi_{\tau i}$ of future time periods $\tau > t$. In practice, this issue can be alleviated by a rolling horizon implementation that resolves the problem with updated information after the (non-anticipative, and hence implementable) first-stage decisions have been taken.

4.5 Flexible Production Planning

We study the production planning problem faced by a manufacturer who produces multiple products that are related through a configurable multi-level bill of materials (Balakrishnan and Geunes, 2000; Lamothe et al., 2006). To this end, we denote by $\mathcal{E}$ the set of considered entities (such as raw materials, intermediate parts or end products). For each entity $e \in \mathcal{E}$, the manufacturer can choose exactly one configuration $c \in \mathcal{C}_e$, which is characterized by a resource function $r(\cdot, e, c) : \mathcal{E} \to \mathbb{R}_+$ describing the quantity $r(e', e, c)$ of each entity $e' \in \mathcal{E}$ that is required to produce one unit of entity e, as well as a price $p(e, c) \in \mathbb{R}_+$ that describes the per-unit cost of implementing the configuration. In particular, a configuration $c \in \mathcal{C}_e$ satisfying $r(e', e, c) = 0$ for all $e' \in \mathcal{E}$ and $p(e, c) > 0$ represents the external purchase of an entity and thus models a make-or-buy decision. For each entity $e \in \mathcal{E}$, we define the set of immediate ancestors as $\mathcal{A}_e = \{a \in \mathcal{E} : r(e, a, c) > 0 \text{ for some } c \in \mathcal{C}_a\}$. We require that for each configuration $\{c_e\}_{e \in \mathcal{E}}$, the directed production graph with nodes $\mathcal{E}$ and arcs $\{(e, a) \in \mathcal{E} \times \mathcal{E} : r(e, a, c_a) > 0\}$ is acyclic (but—in contrast to our supply chain management problem from Section 4.2—not necessarily a tree).

The manufacturer wishes to serve the uncertain demands $\tilde{\xi}_e$ for the entities $e \in \mathcal{E}$, which are assumed to be non-negative, at lowest overall costs, and she thus solves the following problem:

$$\begin{array}{lll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] & \\ \text{subject to} & \sum_{c \in \mathcal{C}_e} x_{ec} = 1 & \forall e \in \mathcal{E} \\ & x_{ec} \in \{0, 1\} & \forall e \in \mathcal{E}, \ \forall c \in \mathcal{C}_e \end{array} \tag{22a}$$

Here, the decision x_{ec} determines whether or not to choose configuration $c \in \mathcal{C}_e$ for entity $e \in \mathcal{E}$, and the second-stage costs $\mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi})$ coincide with the optimal value of the second-stage problem

$$\begin{array}{lll} \text{minimize} & \sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}_e} p(e, c) \cdot x_{ec} \cdot y_e & \\ \text{subject to} & y_e + \mathrm{M} \sum_{a \in \mathcal{A}_e} (1 - x_{a, c_a}) \geqslant \xi_e + \sum_{a \in \mathcal{A}_e} r(e, a, c_a) \cdot y_a & \forall e \in \mathcal{E}, \ \forall \boldsymbol{c} \in \bigtimes_{a \in \mathcal{A}_e} \mathcal{C}_a \\ & y_e \geqslant \xi_e & \forall e \in \mathcal{E} : \mathcal{A}_e = \emptyset \\ & y_e \in \mathbb{R}, \ e \in \mathcal{E}, & \end{array} \tag{22b}$$

where the decision y_e determines the quantity of entity $e \in \mathcal{E}$ to produce or procure. The objective function of (22b) minimizes the overall production costs. The first constraint ensures that the quantities y_e of all non-root entities $e \in \mathcal{E}$ are sufficient to serve both the direct demands as well as the input requirements of all immediate ancestors $a \in \mathcal{A}_e$, while the second constraint ensures that the quantities y_e of all root entities are sufficient to serve the direct demands. Although the first constraint group comprises $\prod_{a \in \mathcal{A}_e} |\mathcal{C}_a|$ different constraints for each entity $e \in \mathcal{E}$, only the constraint whose configuration vector $\boldsymbol{c}$ satisfies $x_{a, c_a} = 1$ for all $a \in \mathcal{A}_e$ will be active. Note that for each entity $e \in \mathcal{E}$, the number of constraints is combinatorial in the number of *immediate* but not in the number of *transitive* ancestors of each entity. Thus, the size of the formulation remains moderate as long as the number of immediate ancestors is small for every entity (in a tree, for example, each entity has at most one immediate ancestor). Note also that the non-negativity of $\boldsymbol{y}$ is enforced implicitly through the non-negativity of the demands $\tilde{\xi}_e$ and the production requirements $\boldsymbol{r}$.

We now argue that problem (22) is solved by affine decision rules. Indeed, one readily verifies that the assumptions (**F**), (**A**) and (**D**) of Section 2 are satisfied. To see that assumption (**B**) is satisfied as well, we make use of Proposition 2.

Observation 5. *The second-stage problem* (22b) *satisfies the conditions of Proposition 2.*

Proof. We first show that condition *(i)* of Proposition 2 is satisfied. To this end, fix any entity $e \in \mathcal{E}$. Condition *(i)* is trivially satisfied if $\mathcal{A}_e = \emptyset$. Assume next that $\mathcal{A}_e \neq \emptyset$, and fix the configuration vector $\boldsymbol{c} \in \times_{a \in \mathcal{A}_e} \mathcal{C}_a$ satisfying $x_{a,c_a} = 1$ for all $a \in \mathcal{A}_e$. This vector $\boldsymbol{c}$ is guaranteed to exist by the constraint of the first-stage problem (22a). The constraint

$$y_e + \mathrm{M} \sum_{a \in \mathcal{A}_e} (1 - x_{a,c_a}) \geqslant \xi_e + \sum_{a \in \mathcal{A}_e} r(e, a, c_a) \cdot y_a$$

then weakly dominates all other lower bounds on y_e imposed by configuration vectors $\boldsymbol{c'} \in \times_{a \in \mathcal{A}_e} \mathcal{C}_a$ since for each of them, at least one of the binary variables x_{a,c'_a}, $a \in \mathcal{A}_e$, must evaluate to zero, again due to the constraint of the first-stage problem (22a).

In view of the second condition of Proposition 2, we recall that for a fixed configuration $\{c_e\}_{e \in \mathcal{E}}$, the graph with nodes $\mathcal{E}$ and arcs $\{(e, a) \in \mathcal{E} \times \mathcal{E} \, : \, r(e, a, c_a) > 0\}$ is acyclic. Since a directed acyclic graph admits a topological ordering, there is a permutation $\pi : \mathcal{E} \to \mathcal{E}$ such that $\pi(a) < \pi(e)$ for all $e, a \in \mathcal{E}$ satisfying $r(e, a, c_a) > 0$. Hence, each weakly dominant constraint

$$y_e \geqslant \xi_e + \sum_{a \in \mathcal{A}_e} r(e, a, c_a) \cdot y_a - \mathrm{M} \sum_{a \in \mathcal{A}_e} (1 - x_{a,c_a})$$

corresponding to an entity $e \in \mathcal{E}$ with $\mathcal{A}_e \neq \emptyset$ links e only with its immediate ancestors $a \in \mathcal{A}_e$ satisfying $\pi(a) < \pi(e)$. We therefore conclude that the right-hand side coefficient vectors $\boldsymbol{\beta_{ec}}$, $e \in \mathcal{E}$ and $\boldsymbol{c} \in \times_{a \in \mathcal{A}_e} \mathcal{C}_a$ with $x_{a,c_a} = 1$ for all $a \in \mathcal{A}_e$, with elements $r(e, a, c_a)$ form a partial order for all entities $e \in \mathcal{E}$ with $\mathcal{A}_e \neq \emptyset$. The statement then follows since the right-hand side coefficient vectors of the root entities $e \in \mathcal{E}$ with $\mathcal{A}_e = \emptyset$ are zero, which allows us to include them anywhere in the partial order. □

Observation 5 ensures that for any risk measure satisfying assumption (**R**), affine decision rules are optimal in problem (22). Even with this insight, however, problem (22) appears to be computationally challenging due to its non-convex objective function that involves products of the decision variables x_{ec} and y_e. Fortunately, however, the configuration decisions x_{ec} are binary, which allows us to linearize the objective function exactly with standard techniques. Since this linearization is applied *after* our restriction to affine decision rules, it does not impact the optimality of affine decision rules. We emphasize that the generality of assumption (**F**), which allows the objective function to be non-convex as long as it is monotone in $\boldsymbol{y}$ for each value of $\boldsymbol{x}$ and each realization of $\tilde{\boldsymbol{\xi}}$, is crucial for this reformulation.

Problem (22) serves as a template for different production planning problems. One can immediately conceive variants with uncertain production costs $\tilde{p}(e, c)$ or dependencies between the admissible configurations for different entities. Also, the possibility to switch (some of) the configurations $c \in \mathcal{C}_e$ after observing the demands can be modelled through a K-adaptability formulation.

4.6 Robust Optimization

Our results from Section 2 immediately apply to robust optimization problems, which constitute a subclass of problem (1) where the ambiguity set $\mathcal{P}$ contains all Dirac distributions supported on some uncertainty set and where the risk measure $\rho_{\mathbb{P}}$ is the essential supremum. In this case, the conditions (**F**), (**A**), (**D**) and (**B**) restrict the choice of admissible objective functions and constraints in order to guarantee optimality of affine decision rules. Alternatively, the specific structure of robust optimization problems allow us to employ convex duality to 'swap' the characterizations of the uncertainty set and the second-stage feasible region. If we do so, we obtain an alternative two-stage robust optimization problem whose optimal objective value and first-stage feasible region coincide with those of the original problem, but where our optimality results now impose restrictions on the structure of the uncertainty set (as opposed to the second-stage objective and constraints).

To illustrate this idea, consider the two-stage robust optimization problem

$$\begin{array}{ll} \text{minimize} & \max_{\boldsymbol{\xi} \in \Xi} \; \mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{23a}$$

where $\mathcal{X} \subseteq \mathbb{R}^{n_1}$, $\boldsymbol{\xi}$ is an uncertain parameter vector supported on the uncertainty set $\Xi = [-1, 1]^k$, and where the second-stage cost function $\mathcal{Q}$ satisfies

$$\mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \quad = \quad \begin{bmatrix} \text{minimize} & \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y} \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^{n_2} \end{bmatrix} \tag{23b}$$

with $\boldsymbol{c}(\boldsymbol{\xi}) = \boldsymbol{C}\boldsymbol{\xi} + \boldsymbol{c}$, $\boldsymbol{A}(\boldsymbol{\xi}) = \boldsymbol{A} + \boldsymbol{A}_1\xi_1 + \ldots + \boldsymbol{A}_k\xi_k$ and $\boldsymbol{g}(\boldsymbol{\xi}) = \boldsymbol{G}\boldsymbol{\xi} + \boldsymbol{g}$ affine in $\boldsymbol{\xi}$ and $\boldsymbol{d}$ and $\boldsymbol{B}$ deterministic. In contrast to our results from Section 2, we thus impose conditions on the shape of the uncertainty set Ξ, but we do not impose any further restrictions on the objective function or the constraints of the second-stage problem (23b). In particular, no sign restrictions apply to the objective coefficients, the constraint matrices and the right-hand sides, and the constraints in (23b) can thus impose arbitrary lower and upper bounds on the first- and second-stage decisions.

Problem (23) describes a two-stage robust optimization problem with a mixed-integer first stage and a linear second stage, respectively, where the uncertainty is described by a factor model. Indeed, we can think of the components ξ_k as independent factors, and the uncertain problem parameters $\boldsymbol{c}(\boldsymbol{\xi})$, $\boldsymbol{A}(\boldsymbol{\xi})$ and $\boldsymbol{g}(\boldsymbol{\xi})$ emerge from affine combinations of those factors. We now show that problem (23) has an equivalent single-stage representation.

Observation 6. *The two-stage robust optimization problem* (23) *has the same optimal value as well as the same first-stage feasible region as the single-stage robust optimization problem*

$$\begin{array}{ll} \text{minimize} & \max\limits_{\boldsymbol{\lambda}\in\Lambda} \left\{ (\boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{\lambda})^\top \boldsymbol{x} + \boldsymbol{g}^\top\boldsymbol{\lambda} + \left\| \boldsymbol{G}^\top\boldsymbol{\lambda} + \boldsymbol{C}^\top\boldsymbol{x} - \boldsymbol{A}(\boldsymbol{\lambda})^\top\boldsymbol{x} \right\|_1 \right\} \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \end{array} \tag{23'}$$

over the uncertainty set $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^{m_1} \,:\, \boldsymbol{B}^\top\boldsymbol{\lambda} = \boldsymbol{d}\}$, *where* $\boldsymbol{A}(\boldsymbol{\lambda})^\top = (\boldsymbol{A}_1^\top\boldsymbol{\lambda}, \ldots, \boldsymbol{A}_k^\top\boldsymbol{\lambda})^\top$.

Proof. Using Theorem 1 of Bertsimas and de Ruiter (2016), the two-stage robust optimization problem (23) shares the same optimal objective value and the same first-stage feasible region with the two-stage robust optimization problem

$$\begin{array}{ll} \text{minimize} & \max\limits_{\boldsymbol{\lambda}\in\Lambda} \; \mathcal{Q}'(\boldsymbol{x};\boldsymbol{\lambda}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array}$$

where

$$\mathcal{Q}'(\boldsymbol{x};\boldsymbol{\lambda}) \quad = \quad \begin{bmatrix} \text{minimize} & (\boldsymbol{c} - \boldsymbol{A}^\top\boldsymbol{\lambda})^\top\boldsymbol{x} + \boldsymbol{g}^\top\boldsymbol{\lambda} + \left(\overline{\boldsymbol{\mu}} + \underline{\boldsymbol{\mu}}\right)^\top \mathbf{e} \\ \text{subject to} & \overline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}} = \boldsymbol{G}^\top\boldsymbol{\lambda} + \boldsymbol{C}^\top\boldsymbol{x} - \boldsymbol{A}(\boldsymbol{\lambda})^\top\boldsymbol{x} \\ & \overline{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}} \in \mathbb{R}^k_+ \end{bmatrix},$$

which is derived by dualizing the second-stage problem in (23), interchanging the order of the maximization over the uncertain problem parameters $\boldsymbol{\xi}$ and the newly introduced dual variables $\boldsymbol{\lambda}$, and finally dualizing the new inner maximization over $\boldsymbol{\xi} \in \Xi$. The substitution $\overline{\boldsymbol{\mu}} \leftarrow \underline{\boldsymbol{\mu}} + \boldsymbol{G}^\top \boldsymbol{\lambda} + \boldsymbol{C}^\top \boldsymbol{x} - \boldsymbol{A}(\boldsymbol{\lambda})^\top \boldsymbol{x}$ simplifies the second-stage problem to

$$\mathcal{Q}'(\boldsymbol{x};\boldsymbol{\lambda}) \quad = \quad \begin{bmatrix} \text{minimize} & (\boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{\lambda})^\top \boldsymbol{x} + \boldsymbol{g}^\top \boldsymbol{\lambda} + \left(2\underline{\boldsymbol{\mu}} + \boldsymbol{G}^\top \boldsymbol{\lambda} + \boldsymbol{C}^\top \boldsymbol{x} - \boldsymbol{A}(\boldsymbol{\lambda})^\top \boldsymbol{x}\right)^\top \mathbf{e} \\ \text{subject to} & \underline{\boldsymbol{\mu}} \geqslant \boldsymbol{A}(\boldsymbol{\lambda})^\top \boldsymbol{x} - \boldsymbol{G}^\top \boldsymbol{\lambda} - \boldsymbol{C}^\top \boldsymbol{x} \\ & \underline{\boldsymbol{\mu}} \in \mathbb{R}^k_+ \end{bmatrix}.$$

Contrary to our proof of Theorem 1, the second-stage decisions $\underline{\boldsymbol{\mu}}$ already appear conveniently isolated on the constraint left-hand sides. At the same time, however, they participate in *two* lower bound constraints, neither of which can be identified as dominated. Nevertheless, since the second-stage decisions have non-negative objective coefficients, we readily conclude that the piecewise affine policy $\underline{\boldsymbol{\mu}}^\star = \max\left\{\boldsymbol{A}(\boldsymbol{\lambda})^\top \boldsymbol{x} - \boldsymbol{G}^\top \boldsymbol{\lambda} - \boldsymbol{C}^\top \boldsymbol{x},\, \mathbf{0}\right\}$ is optimal in $\mathcal{Q}'(\boldsymbol{x};\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \Lambda$. Substituting this expression in the objective function of $\mathcal{Q}'(\boldsymbol{x};\boldsymbol{\lambda})$ and re-arranging terms then yields the result. □

Problem (23'), despite being a static robust optimization problem, remains challenging as its objective evaluates the worst case of a convex function. The specific structure present in (23'), however, allows us to employ the iterative solution scheme proposed by Gorissen and den Hertog (2013). This algorithm solves a sequence of increasingly tight relaxations of problem (23') whose objective functions optimize over finite subsets of the 'parameter realizations' $\boldsymbol{\lambda} \in \Lambda$ as well as a limited number of pieces of the piecewise affine 1-norm expression in (23'). Alternatively, we can solve the original two-stage robust optimization problem (23) with the column-and-constraint generation scheme developed by Zeng and Zhao (2013). This algorithm solves a sequence of increasingly tight relaxations of problem (23) that are formed from scenario fans that consider finitely many parameter realizations $\boldsymbol{\xi} \in \Xi$. Both the scheme of Gorissen and den Hertog (2013) and the method proposed by Zeng and Zhao (2013) solve non-convex subproblems in each iteration, but the subproblems of the column-and-constraint generation scheme are significantly more involved as they replace the second-stage problem of (23) by its Karush-Kuhn-Tucker conditions, which tend to be difficult to optimize over. We next examine the potential speed-up in a numerical example.

We consider the two-stage robust inventory management problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\boldsymbol{\xi} \in \Xi} \; \mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n_+, \end{array}$$

	problem (23')			problem (23)		
n	1 min	10 min	60 min	1 min	10 min	60 min
25	21%	0%	0%	0%	0%	0%
30	55%	0%	0%	137%	137%	0%
35	80%	0%	0%	144%	144%	0%
40	97%	1%	0%	152%	152%	150%
45	104%	4%	0%	-	-	153%
50	106%	6%	0%	-	-	154%
75	160%	33%	0%	-	-	-
100	164%	74%	3%	-	-	-

Table 1. Average optimality gaps of the algorithms of Gorissen and den Hertog (2013), left, and Zeng and Zhao (2013), right. Absent entries correspond to experiments where some instances have not completed the first iteration within the given time limit.

borrowed from Section 5.2 of Georghiou et al. (2020), where $\Xi = [-1, 1]^k$ and

$$\mathcal{Q}(\boldsymbol{x};\boldsymbol{\xi}) \quad = \quad \left[\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{n} c_y \cdot y_i + c_h \cdot h_i + c_b \cdot b_i \\ \text{subject to} & \boldsymbol{h} \geqslant \boldsymbol{x} + \boldsymbol{y} - \boldsymbol{D}(\boldsymbol{\xi}) \\ & \boldsymbol{b} \geqslant \boldsymbol{D}(\boldsymbol{\xi}) - \boldsymbol{x} - \boldsymbol{y} \\ & \displaystyle\sum_{i=1}^{n} y_i \leqslant B \\ & \boldsymbol{y}, \boldsymbol{h}, \boldsymbol{b} \in \mathbb{R}^n_+ \end{array} \right].$$

In this problem, the uncertain customer demands $D_1(\boldsymbol{\xi}), \ldots, D_n(\boldsymbol{\xi})$ for the n products are modelled as affine combinations $D_i(\boldsymbol{\xi}) = \boldsymbol{\phi}_i^\top \boldsymbol{\xi} + 1$ of the uncertain factors $\boldsymbol{\xi}$, where the factor weights $\boldsymbol{\phi}_i$ are selected uniformly at random from the set $\Phi = \{\boldsymbol{\phi} \in \mathbb{R}^k : \|\boldsymbol{\phi}\|_1 = 1\}$. The order adjustment costs c_y, the inventory holding costs c_h as well as the backlogging costs c_b are selected uniformly at random from the intervals $[0, 2]$, $[3, 5]$ and $[0, 1/2]$, respectively, and we choose B so that half of the demand can be served via adjustment orders $\boldsymbol{y}$. Table 1 reports the average optimality gaps of CPLEX 20.9 run on an i9-10900 CPU with 2.80GHz clock speed and 64GB of RAM after 1, 10 and 60 minutes over 25 randomly generated instances of varying size. The table shows that solving the static robust optimization problem (23') with the algorithm of Gorissen and den Hertog (2013) indeed appears to be more efficient than solving the generic two-stage robust optimization problem (23) via column-and-constraint generation.

5 Conclusion

Problems known to be solved optimally by affine decision rules are rare and were, to our best knowledge, restricted to a few two-stage robust optimization problems that impose restrictive assumptions on both the geometry of the uncertainty set and the structure of the constraints. We showed in this paper that affine decision rules are in fact optimal in a number of application domains *if* the problem formulations are carefully chosen. As such, our work also sheds light on how seemingly inconsequential differences in modelling assumptions can lead to radically different conclusions about the problem's solvability in affine decision rules. A modeller does not just capture the world as she sees it – she typically has the liberty to disregard certain aspects to ensure tractability. We believe that the optimality conditions put forward in this paper may serve as a useful method in a modeller's toolbox to determine an attractive trade-off between accuracy and tractability. Our supply chain management problem from Section 4.2, for example, is solved optimally in affine decision rules if we impose the additional assumption of a product-wise acyclic distribution network. The resulting solution may be implemented as is, or it can serve as a basis for a heuristic policy which suboptimally solves a more generic problem formulation that violates our optimality conditions. Knowing that the resulting policies are optimal for some well-defined subclasses of the problem instills confidence that the heuristic policies perform satisfactorily also in broader instance classes where our optimality conditions may not be satisfied.

Our work lends itself to several extensions and generalizations. It would be instructive to study how the optimality of affine decision rules can be extended to multi-stage problems. We also see value in exploring alternative optimality criteria, such as conditions under which affine decision rules become asymptotically optimal as the problem size grows, or conditions under which affine decision rules are optimal with high probability, based on a sampling of the problem data.

References

A. Ardestani-Jaafari and E. Delage. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494, 2016.

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

A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.

P. Awasthi, V. Goyal, and B. Y. Lu. On the adaptivity gap in two-stage robust linear optimization under uncertain constraints. *Mathematical Programming*, 173(1–2):313–352, 2019.

J. Ayoub and M. Poss. Decomposition for adjustable robust linear optimization subject to uncertainty polytope. *Computational Management Science*, 13(2):219–239, 2016.

F. Babonneau, J.-P. Vial, O. Klopfenstein, and A. Ouorou. Robust capacity assignment solutions for telecommunications networks with uncertain demands. *Networks*, 62(4):255–272, 2013.

A. Balakrishnan and J. Geunes. Requirements planning with substitutions: Exploiting bill-of-materials flexibility in production planning. *Manufacturing & Service Operations Management*, 2(2):166–185, 2000.

R. Baldacci, E. Hadjiconstantinou, and A. Mingozzi. An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Operations Research*, 52(5):723–738, 2004.

D. Bampou and D. Kuhn. Scenario-free stochastic programming with polynomial decision rules. *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 7806–7812, 2011.

R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

A. Ben-Tal, B. Golany, A. Nemirovski, and J. P. Vial. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

A. Ben-Tal, B. D. Chung, S. R. Mandala, and T. Yao. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation Research Part B*, 45(8):1177–1189, 2011.

A. Ben-Tal, D. den Hertog, A. de Waegenaere, B. Melenbergand, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1994.

D. Bertsimas and H. Bidkhori. On the performance of affine policies for two-stage adaptive optimization: A geometric perspective. *Mathematical Programming*, 153(2):577–594, 2015.

D. Bertsimas and C. Caramanis. Finite adaptibility in multistage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.

D. Bertsimas and F. J. C. T. de Ruiter. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500–511, 2016.

D. Bertsimas and I. Dunning. Multistage robust mixed integer optimization with adaptive partitions. *Operations Research*, 64(4):980–998, 2016.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3):610–627, 2015.

D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, 35(2):284–305, 2010.

D. Bertsimas and V. Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 134(2):491–531, 2012.

D. Bertsimas and V. Goyal. On the approximability of adjustable robust convex optimization under uncertainty. *Mathematical Methods of Operations Research*, 77(3):323–343, 2013.

D. Bertsimas, D. Iancu, and P. Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011a.

D. Bertsimas, V. Goyal, and X. A. Sun. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, 36(1):24–54, 2011b.

D. Bertsimas, D. Iancu, and P. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12):2809–2824, 2011c.

D. Bertsimas, V. Goyal, and B. Y. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, 150(2):281–319, 2015.

D. Bertsimas, I. Dunning, and M. Lubin. Reformulation versus cutting-planes for robust optimization. *Computational Management Science*, 13(2):195–217, 2016.

D. Bertsimas, M. Sim, and M. Zhang. Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618, 2019.

D. Bienstock and N. Özbay. Computing robust basestock levels. *Discrete Optimization*, 5(2):389–414, 2008.

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.

J. W. Blankenship and J. E. Falk. Infinitely constrained optimization problems. *Journal of Optimization Theory and Applications*, 19(2):261–281, 1976.

V. Cacchiani, M. Jünger, F. Liers, A. Lodi, and D. R. Schmidt. Single-commodity robust network design with finite and Hose demand sets. *Mathematical Programming*, 157(1):297–342, 2016.

T. Cayirli and E. Veral. Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12(4):519–549, 2003.

A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.

X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.

X. Chen, M. Sim, P. Sun, and J. Zhang. A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357, 2008.

C. Contardo, J. F. Cordeau, and B. Gendron. A computational comparison of flow formulations for the capacitated location-routing problem. *Discrete Optimization*, 10(4):263–295, 2013.

G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

E. Delage and D. Iancu. Robust multistage decision making. *INFORMS Tutorials in Operations Research*, pages 20–46, 2015.

S. J. Garstka and R. J.-B. Wets. On decision rules in stochastic programming. *Mathematical Programming*, 7(1):117–143, 1974.

A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1):301–338, 2015.

A. Georghiou, A. Tsoukalas, and W. Wiesemann. Robust dual dynamic programming. *Operations Research*, 67(3):813–830, 2019.

A. Georghiou, A. Tsoukalas, and W. Wiesemann. A primal-dual lifting scheme for two-stage robust optimization. *Operations Research*, 68(2):572–590, 2020.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917, 2010.

B. L. Golden and R. T. Wong. Capacitated arc routing problems. *Networks*, 11(3):305–315, 1981.

B. L. Gorissen and D. den Hertog. Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43, 2013.

C. E. Gounaris, W. Wiesemann, and C. A. Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693, 2013.

L. Gouveia. A result on projection for the vehicle routing problem. *Journal of Operational Research Society*, 85(3):610–624, 1995.

D. Gupta and B. Denton. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9):800–819, 2008.

E. Guslitser. Uncertainty-immunized solutions in linear programming. Master's thesis, Technion, 2002.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 63(4):877–891, 2015.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage distributionally robust binary programming. *Operations Research Letters*, 44(1):6–11, 2016.

O. El Housni and V. Goyal. Piecewise static policies for two-stage adjustable robust linear optimization. *Mathematical Programming*, 169(2):649–665, 2018.

O. El Housni and V. Goyal. On the optimality of affine policies for budgeted uncertainty sets. *Mathematics of Operations Research*, Ahead of Print, 2021.

D. Iancu, M. Sharma, and M. Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.

R. Jiang, M. Zhang, G. Li, and Y. Guan. Two-stage network constrained robust unit commitment problem. *European Journal of Operational Research*, 234(3):751–762, 2014.

R. Jiang, M. Ryu, and G. Xu. Data-driven distributionally robust appointment scheduling over Wasserstein balls. 2019. Available at arXiv:1907.03219.

Q. Kong, C.-Y. Lee, C.-P. Teo, and Z. Zheng. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3):711–726, 2013.

Q. Kong, S. Li, N. Liu, C.-P. Teo, and Z. Yan. Appointment scheduling under time-dependent patient no-show behavior. *Management Science*, 66(8):3480–3500, 2020.

D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.

R. V. Kulkarni and P. R. Bhave. Integer programming formulations of vehicle routing problems. *European Journal of Operational Research*, 20(1):58–67, 1985.

J. Lamothe, K. Hadj-Hamou, and M. Aldanondo. An optimization model for selecting a product family and designing its supply chain. *European Journal of Operational Research*, 169(3):1030–1047, 2006.

H.-Y. Mak, Y. Rong, and J. Zhang. Appointment scheduling with limited distributional information. *Management Science*, 61(2):316–334, 2015.

S. Mattia and M. Poss. A comparison of different routing schemes for the robust network loading problem: Polyhedral results and computation. *Computational Optimization and Applications*, 69(3):753–800, 2018.

M. Minoux. Robust network optimization under polyhedral demand uncertainty is NP-hard. *Discrete Applied Mathematics*, 158(5):597–603, 2010.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):1–52, 2018.

A. Mutapcic and S. Boyd. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods and Software*, 24(3):381–406, 2009.

F. Ordóñez and J. Zhao. Robust capacity expansion of network flows. *Networks*, 50(2):136–145, 2007.

M. L. Pinedo. *Scheduling: Theory, Algorithms, and Systems*. Springer, 4th edition, 2012.

M. Poss. A comparison of routing sets for robust network design. *Optimization Letters*, 8(5):1619–1635, 2014.

M. Poss and C. Raack. Affine recourse for the robust network design problem: Between static and dynamic routing. *Networks*, 61(2):180–198, 2013.

K. S. Postek and D. den Hertog. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 28(3):553–574, 2016.

A. Shapiro, D. Dentcheva, and A. Ruszczynski, editors. *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM, 2nd edition, 2009.

D. Simchi-Levi, N. Trichakis, and P. Y. Zhang. Designing response supply chain against bioattacks. *Operations Research*, 67(5):1246–1268, 2019.

J. Skaf and S. P. Boyd. Design of affine controllers via convex optimization. *IEEE Transactions on Automatic Control*, 55(11):2476–2487, 2010.

A. Subramanyam, C. E. Gounaris, and W. Wiesemann. K-adaptability in two-stage mixed-integer robust optimization. *Mathematical Programming Computation*, 12(2):193–224, 2020.

A. Thiele, T. Terry, and M. Epelman. Robust linear optimization with recourse. Technical report, Lehigh University and University of Michigan, 2010.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–561, 2013.

C. Zhao, J. Wang, J.-P. Watson, and Y. Guan. Multi-stage robust unit commitment considering wind and demand response uncertainties. *IEEE Transactions on Power Systems*, 28(3):2708–2717, 2013.

J. Zhen, D. den Hertog, and M. Sim. Adjustable robust optimization via Fourier-Motzkin elimination. *Operations Research*, 66(4):1086–1100, 2018.