On the Optimality of Affine Decision Rules in 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*

June 6, 2025

Abstract

We propose conditions under which two-stage distributionally robust optimization problems are optimally solved in affine or K-adaptable affine decision rules. Contrary to previous work, our conditions do not impose any structure on the support of the uncertain parameters, and they ensure point-wise (as opposed to worst-case) optimality of (K-adaptable) affine decision rules. The absence of support restrictions allows us to transfer nonlinearities from the problem description to the support via liftings, while the point-wise optimality implies 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 demonstrate how our conditions can be met in two applications.

Keywords: Affine Decision Rules; K-Adaptability; 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 distributionally robust optimization problems. It comes at no surprise that such conditions must be restrictive. Intuitively, 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 the coefficient matrix of these bindings constraints admits a positive inverse, we are guaranteed an affine dependence of the second-stage decisions on the uncertain problem parameters. A sufficient condition for the existence of a positive inverse is that the binding second-stage constraints establish an acyclic dependence structure between the second-stage decisions, that is, the second-stage decisions allow for a reordering (which itself depends on the choice of the first-stage decisions) so that the optimal decisions could in principle be computed sequentially by a backward substitution, if the optimal first-stage decisions were known. 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 distributionally robust optimization problems, and it also enables us to transfer nonlinearities 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. While point-wise optimality tends to be overly restrictive for classical robust optimization problems, where the recourse decisions only need to be optimal for the worst possible parameter realizations, it allows us to apply our results 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.

We demonstrate how our conditions can be met by a supply chain management and a flexible production planning problem. 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 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 distributionally robust optimization problems. Our conditions are minimal 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 some broader classes of distributionally robust optimization problems, which are not optimally solved in affine decision rules, admit optimal solutions in K-adaptable

affine decision rules. This allows us to significantly broaden the class of problems to which our theory applies.

(iii) We apply our results to stylized formulations in two application areas, both 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 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 nonlinear 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 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.

The remainder of the paper proceeds as follows. Section 2 develops minimal conditions for the existence of optimal affine decision rules. Section 3 generalizes our findings to K-adaptability problems. We study two applications of our results in Section 4, and we conclude 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

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

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 empty or 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 set of optimal first-stage decisions in (1) does not change if we restrict the second-stage decision $\boldsymbol{y}$ to an affine decision rule that is selected in the

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

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

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

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

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

Note that even if problem (2) did not impose an affine dependence of $\boldsymbol{y}$ on $\boldsymbol{\xi}$, the interchange of the supremum over $\mathbb{P}\in\mathcal{P}$ and the minimization over the second-stage decisions $\boldsymbol{y}$ would only be admissible if the risk measure $\rho_{\mathbb{P}}$ is monotonic for every $\mathbb{P}\in\mathcal{P}$ (Shapiro, 2017). For the remainder of the paper, we thus make the blanket assumption that $\rho_{\mathbb{P}}$ is monotonic for every $\mathbb{P}\in\mathcal{P}$. This assumption 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).

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 $\tilde{\boldsymbol{\xi}}$ tends to differ 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 assumes that there is an $\boldsymbol{x}\in\mathcal{X}$ optimal in problem (1) such that the following conditions are met:

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

(**A**) The technology matrix $\boldsymbol{A}$ and the right-hand side $\boldsymbol{g}$ are affine functions of $\boldsymbol{\xi}$.

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

(**B**) 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, and

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

Assumption (**F**) is satisfied, for example, in the linear case when $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y}$ and $\boldsymbol{d}(\boldsymbol{\xi}) \geqslant \boldsymbol{0}$ 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 nonlinear components of $\boldsymbol{A}$ and $\boldsymbol{g}$. 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 there is a subset of n_2 constraints that decide whether a second-stage decision $\boldsymbol{y}$ satisfies the first constraint set in (1b) at an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$. 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. *Assume that problem* (1) *attains its optimal value and that (**F**), (**A**), (**D**) and (**B**) are satisfied. Then the optimal values of problems* (1) *and* (2) *coincide.*

Theorem 1 only establishes a relationship between optimal decisions—as opposed to all feasible solutions—to the problems (1) and (2). Indeed, Section 4.1 provides an example where affine decision rules are only optimal when optimal first-stage decisions are taken. Note that problem (1) may be infeasible, in which case its restriction (2) to affine decision rules will also be infeasible. Finally, problem (1) may be feasible, but its optimal value may not be attained. In this case, Theorem 1 cannot make any statements as the condition (**B**) is only required to hold at an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$. Sufficient conditions for problem (1) to attain its optimal value are discussed, among others, by Shapiro et al. (2009).

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 nonlinearities 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 (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 $\boldsymbol{\xi} \in \Xi$. For suitably designed ambiguity sets, this can be 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,*

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

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 (**F**), (**A**), (**D**) and (**B**) are not only sufficient for the optimality of affine decision rules in problem (1), but they are also minimal in the sense that counterexamples can be constructed where all but one of the assumptions are satisfied.

Proposition 1. *The assumptions (**F**), (**A**), (**D**) and (**B**) in Theorem 1 are minimal in the sense that the conclusion of the theorem ceases to hold whenever one of the assumptions is removed.*

We note that despite Proposition 1, the conditions (**F**), (**A**), (**D**) and (**B**) are not necessary for the optimality of affine decision rules. To see this, consider an instance of problem (1) without first-stage decisions, a singleton ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ whose distribution $\mathbb{P}$ places probability 1/2 on each of the two realizations $\xi \in \{-1, 1\}$, $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$ and a second-stage problem that satisfies

$$\mathcal{Q}(\xi) \;=\; \min_{y \in \mathbb{R}} \left\{ |y| \,:\, y \geqslant \xi^2, \;\; y \geqslant 2\xi, \;\; -y \leqslant \xi/2 \right\}.$$

One readily verifies that this instance simultaneously violates all of the conditions (**F**), (**A**), (**D**) and (**B**), yet for every realization ξ the second-stage problem is optimized by $y(\xi) = (\xi + 3)/2$.

While the assumptions (**F**), (**A**) and (**D**) are transparent and easy to verify, assumption (**B**) is less intuitive and appears cumbersome to confirm in practice. We next discuss a sufficient condition 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*

$$y_j \geqslant \boldsymbol{\alpha}_{jk}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{jk}^\top \boldsymbol{y} + \gamma_{jk}(\boldsymbol{\xi}) \qquad \forall j = 1, \ldots, n_2, \ \forall k = 1, \ldots, s_j, \tag{3}$$

where $\boldsymbol{\beta}_{jk} \geqslant \mathbf{0}$ *for all* j *and* k*. Assume further that at an optimal* $\boldsymbol{x} \in \mathcal{X}$*, we have:*

(i) *For every* $j = 1, \ldots, n_2$ *there is a constraint* $k_j \in \{1, \ldots, s_j\}$ *in* (3) *that weakly dominates the other* $s_j - 1$ *constraints for* j *in* (3) *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**) at* $\boldsymbol{x} \in \mathcal{X}$*.*

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 (3) 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 at the optimal $\boldsymbol{x} \in \mathcal{X}$, 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 at the optimal $\boldsymbol{x} \in \mathcal{X}$, 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, there is an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ for which the second-stage decisions can be re-ordered in such a way that y_1 only depends on the realization $\boldsymbol{\xi}$, y_2 may depend on both the realization $\boldsymbol{\xi}$ and the value of y_1, y_3 may depend on $\boldsymbol{\xi}$, y_1 and y_2, and so on. Thus, at the optimal 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.

For the constraint system (3) with $\boldsymbol{\beta}_{jk} \geqslant \mathbf{0}$ for all j and k, condition *(i)* in Proposition 2 is by itself necessary but not sufficient for assumption (**B**), whereas condition *(ii)* is by itself neither necessary nor sufficient. Proposition 2 shows that taken together, both conditions are sufficient to ensure that assumption (**B**) holds. In fact, similar to Theorem 1 and Proposition 1, one can show that the conditions of Proposition 2 are minimal in the sense that the conclusion of the proposition ceases to hold if either condition is removed from the statement. Conversely, one can show that any constraint system (3) with $\boldsymbol{\beta}_{jk} \geqslant \mathbf{0}$ that satisfies assumption (**B**) automatically satisfies condition *(i)* of Proposition 2, whereas it may or may not satisfy condition *(ii)*.

3 Optimality of K-Adaptable Affine Decision Rules

We now relax the two assumptions (**A**) and (**B**) from Section 2 by assuming that there is an $\boldsymbol{x} \in \mathcal{X}$ optimal in problem (1) such that:

(**A'**) There is a covering $\Xi = \Xi_1^{\mathrm{A}} \cup \ldots \cup \Xi_{K_{\mathrm{A}}}^{\mathrm{A}}$ of Ξ such that the technology matrix $\boldsymbol{A}$ and the right-hand side $\boldsymbol{g}$ are affine functions of $\boldsymbol{\xi}$ over each subset Ξ_k^{A}, $k = 1, \ldots, K_{\mathrm{A}}$.

(**B'**) There is a covering $\Xi = \Xi_1^{\mathrm{B}} \cup \ldots \cup \Xi_{K_{\mathrm{B}}}^{\mathrm{B}}$ of Ξ with associated constraint index sets $\mathcal{I}_k \subseteq \{1, \ldots, m_1\}$, $|\mathcal{I}_k| = n_2$, such that each restriction $\boldsymbol{B}_{\mathcal{I}_k}$ of $\boldsymbol{B}$ to those constraints is invertible with a positive inverse, and

$$\Big[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \iff \boldsymbol{A}_{\mathcal{I}_k}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}_k}\boldsymbol{y} \geqslant \boldsymbol{g}_{\mathcal{I}_k}(\boldsymbol{\xi})\Big] \qquad \forall \boldsymbol{\xi} \in \Xi_k^{\mathrm{B}},\ \forall k = 1, \ldots, K_{\mathrm{B}}.$$

Assumption (**A'**) stipulates that the problem parameters $\boldsymbol{A}$ and $\boldsymbol{g}$ are piecewise affine in $\boldsymbol{\xi}$, and it recovers our original assumption (**A**) when $K_{\mathrm{A}} = 1$. Similarly, condition (**B'**) recovers our original condition (**B'**) when $K_{\mathrm{B}} = 1$.

Unfortunately, the assumptions (**F**), (**A'**), (**D**) and (**B'**) are *not* sufficient to guarantee that problem (1) is optimized by an affine decision rule $\boldsymbol{y} : \Xi \xrightarrow{\mathrm{a}} \mathbb{R}^{n_2}$ as the next example shows.

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

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\left[\min\left\{y \,:\, y \geqslant 1/2 - \tilde{\xi},\ \ y \geqslant 0,\ \ y \in \mathbb{R}\right\}\right],$$

*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. This problem satisfies (**F**), (**A'**) and (**D**), and (**B'**) is satisfied for the covering $\Xi = \Xi_1^{\mathrm{B}} \cup \Xi_2^{\mathrm{B}}$ with $(\Xi_1^{\mathrm{B}}, \mathcal{I}_1) = ([0, 1/2], \{1\})$ and $(\Xi_2^{\mathrm{B}}, \mathcal{I}_2) = ([1/2, 1], \{2\})$. Nevertheless, the unique optimal second-stage policy is given by $y^\star(\xi) = \max\{1/2 - \xi,\, 0\}$, which is evidently not affine.*

We next consider the K-adaptability formulation associated with problem (1),

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k(\tilde{\boldsymbol{\xi}})\}_{k=1}^K; \tilde{\boldsymbol{\xi}}) \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{y}_k : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}, \; k = 1, \ldots, K, \end{array} \tag{4}$$

where $\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k\}_{k=1}^K; \boldsymbol{\xi}) = \min\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k; \boldsymbol{\xi}) \; : \; k = 1, \ldots, K\}$ for the second-stage cost function $\mathcal{Q}$ defined in (2b). Problem (4) determines K candidate affine decision rules $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K$ for the second-stage decision $\boldsymbol{y}$ in problem (1) here-and-now and subsequently implements the best of these decision rules once the value of $\tilde{\boldsymbol{\xi}}$ has been observed. Problem (4) thus replaces full flexibility in the second-stage decision with a restricted choice among pre-selected affine decision rules. This renders the solution of the K-adaptability problem (4) easier to analyze and interpret, and it makes the problem admissible to a variety of monolithic formulations and iterative solution schemes that are not available for the generic two-stage distributionally robust optimization problem (1), see Bertsimas and Caramanis (2010), Bertsimas et al. (2011b), Hanasusanto et al. (2015, 2016) and Subramanyam et al. (2020). Note that by construction, problem (4) with $K = 1$ recovers the affine decision rule formulation (2).

We will now show that under (**F**), (**A'**), (**D**) and (**B'**), the optimal value of problem (1) does not change if we instead solve the K-adaptability problem (4) with K sufficiently large relative to K_{A} and K_{B} for which the conditions (**A'**) and (**B'**) hold.

Theorem 2. *Assume that problem* (1) *attains its optimal value and that (**F**), (**A'**), (**D**) and (**B'**) are satisfied. Then the optimal values of problems* (1) *and* (4) *coincide whenever* $K \geqslant K_{\text{A}} \cdot K_{\text{B}}$.

To our best knowledge, the existing optimality results for K-adaptability problems are restricted to problems with a binary recourse, and they consider the pathological cases where none of the second-stage constraints are subject to uncertainty or where K is chosen large enough to cover any possibly feasible second-stage decision (Hanasusanto et al., 2015). In contrast, Theorem 2 studies optimality conditions for problems with a continuous recourse, and our case studies in Section 4 will demonstrate that the conditions of Theorem 2 apply to non-trivial choices of K. In analogy to Proposition 1, we can show that the assumptions (**F**), (**A'**), (**D**) and (**B'**) are minimal 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 application.

Example 2. *Appendix A employs Theorem 2 to study conservative approximations of the two-stage distributionally robust optimization problem*

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

with the second-stage problem

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

*where the new second-stage decisions $\boldsymbol{z}$ have a feasible region $\mathcal{Z}(\boldsymbol{x})$ that may be non-convex (*e.g.*, enforcing integer requirements) and/or depend on the first-stage decisions $\boldsymbol{x}$ in a nonlinear fashion, and where the new recourse matrices $\boldsymbol{E}$ and $\boldsymbol{F}$ for the decisions $\boldsymbol{z}$ may depend on the random problem parameters $\tilde{\boldsymbol{\xi}}$. Theorem 2 implies that in this case, the K-adaptability formulation offers a conservative approximation where the decision rules for $\boldsymbol{y}$ are indeed optimal and the suboptimality is solely caused by an approximation of the new decision $\boldsymbol{z}$.*

Similar to the assumption (**B**) from Section 2, the condition (**B'**) tends to be difficult to verify directly in practice. In the following, we provide a sufficient condition for (**B'**) to hold.

Proposition 3. *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 in* (3)*, where again $\boldsymbol{\beta}_{jk}\geqslant\boldsymbol{0}$ for all j and k. Assume further that at an optimal $\boldsymbol{x}\in\mathcal{X}$, we have:*

(i) *For every $j=1,\ldots,n_2$ there is a subset of constraints $\mathcal{K}_j\subseteq\{1,\ldots,s_j\}$ in* (3) *that weakly dominates the other $s_j-|\mathcal{K}_j|$ constraints for j in* (3) *under every parameter realization $\boldsymbol{\xi}\in\Xi$.*

(ii) *The vectors $\{\boldsymbol{\beta}_{j,k_j}\}_{j=1}^{n_2}$ form a partial order for all $(k_1,\ldots,k_{n_2})\in\mathcal{K}_1\times\ldots\times\mathcal{K}_{n_2}$.*

Then the corresponding instance of problem (1) *satisfies assumption (**B'**) with $K_{\mathrm{B}}=|\mathcal{K}_1|\cdot\ldots\cdot|\mathcal{K}_{n_2}|$.*

Proposition 3 recovers Proposition 2 from the previous section if the constraint subsets $\mathcal{K}_j$ are all singletons; in this case, we obtain $K_{\mathrm{B}}=1$. Proposition 3 will enable us to incorporate pre-shipments in our supply chain management example (*cf.* Section 4.1) as well as supply-side online substitutions in our flexible production planning application (*cf.* Section 4.2).

Assume now that the second-stage decisions $\boldsymbol{y}$ in problem (1) can be decomposed into $\boldsymbol{y}^\top = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_P^\top)$, $\boldsymbol{y}_p \in \mathbb{R}^{n_2^p}$ for $p = 1, \ldots, P$, such that the objective function in (1b) is separable, $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) = \sum_{p=1}^P f_p(\boldsymbol{x}, \boldsymbol{y}_p; \boldsymbol{\xi})$, and the recourse matrix is block diagonal, $\boldsymbol{B} = \operatorname{diag}(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_P)$ with $\boldsymbol{B}_p \in \mathbb{R}^{m_1^p} \times \mathbb{R}^{n_2^p}$ for $p = 1, \ldots, P$. We also partition the technology matrix $\boldsymbol{A}$ and the right-hand side vector $\boldsymbol{g}$ row-wise into $\boldsymbol{A}_p \in \mathbb{R}^{m_1^p \times n_1}$ and $\boldsymbol{g}_p \in \mathbb{R}^{m_1^p}$, $p = 1, \ldots, P$ (without imposing any block diagonality on $\boldsymbol{A}$). We update the invertibility condition (**B'**) as follows.

(**B"**) For every $p = 1, \ldots, P$, there is a covering $\Xi = \Xi_{p,1}^{\mathrm{B}} \cup \ldots \cup \Xi_{p,K_{\mathrm{B}}^p}^{\mathrm{B}}$ with associated constraint index sets $\mathcal{I}_{pk} \subseteq \{1, \ldots, m_1^p\}$, $|\mathcal{I}_{pk}| = n_2^p$, such that each restriction $[\boldsymbol{B}_p]_{\mathcal{I}_{pk}}$ of $\boldsymbol{B}_p$ to those constraints is invertible with a positive inverse, and

$$\Big[\boldsymbol{A}_p(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_p\boldsymbol{y}_p \geqslant \boldsymbol{g}_p(\boldsymbol{\xi}) \iff [\boldsymbol{A}_p]_{\mathcal{I}_{pk}}(\boldsymbol{\xi})\boldsymbol{x} + [\boldsymbol{B}_p]_{\mathcal{I}_{pk}}\boldsymbol{y}_p \geqslant [\boldsymbol{g}_p]_{\mathcal{I}_{pk}}(\boldsymbol{\xi})\Big] \qquad \forall \boldsymbol{\xi} \in \Xi_{pk}^{\mathrm{B}},\ \forall k = 1, \ldots, K_{\mathrm{B}}^p.$$

Consider the following variant of the K-adaptability formulation (4),

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}\Big(\boldsymbol{x}, \bigtimes_{p=1}^{P} \{\boldsymbol{y}_{pk}(\tilde{\boldsymbol{\xi}})\}_{k=1}^{K_p}; \tilde{\boldsymbol{\xi}}\Big) \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \ \boldsymbol{y}_{pk} : \Xi \xrightarrow{\mathrm{a}} \mathbb{R}^{n_2^p},\ k = 1, \ldots, K_p \text{ and } p = 1, \ldots, P, \end{array} \tag{5}$$

where $\mathcal{Q}$ is defined as before. The problem evaluates the second-stage problem for all $\prod_{p=1}^P K_p$ combinations of affine second-stage decision rules $\boldsymbol{y}(\boldsymbol{\xi}) = (y_{1,k_1}, \ldots, y_{P,k_P})(\boldsymbol{\xi})$, $k_p \in \{1, \ldots, K_p\}$ for all $p = 1, \ldots, P$, but it merely optimizes over $\sum_{p=1}^P K_p$ constituent affine decision rules $\boldsymbol{y}_{pk}$.

Proposition 4. *Assume that problem* (1) *attains its optimal value and that (**F**), (**A'**), (**D**) and (**B"**) are satisfied. Then the optimal values of problems* (1) *and* (5) *coincide whenever* $K_p \geqslant K_{\mathrm{A}} \cdot K_{\mathrm{B}}^p$ *for all* $p = 1, \ldots, P$.

Proposition 4 shows that if the objective function decomposes and the technology matrix $\boldsymbol{B}$ is block diagonal, then it is sufficient to select a parsimonious set of $K_{\mathrm{A}} \cdot K_{\mathrm{B}}^p$ decision rules for each subvector $\boldsymbol{y}_p \in \mathbb{R}^{n_2^p}$, as opposed to selecting $\prod_p K_{\mathrm{A}} \cdot K_{\mathrm{B}}^p$ different full-dimensional decision rules $\boldsymbol{y} \in \mathbb{R}^{n_2}$, to recover an optimal solution to the two-stage distributionally robust optimization problem (1). This holds true even though the subvectors $\boldsymbol{y}_p$ participate jointly in the upper bound constraints of (1b). Beyond its analytical relevance, which we showcase in our case study in Section 4.1, Proposition 4 may enable the development of tailored solution schemes that avoid the combinatorial complexity that often plagues solution schemes for K-adaptability problems.

4 Applications

We next apply our theory from Sections 2 and 3 to a supply chain management problem (Section 4.1) and a flexible production planning problem (Section 4.2). For ease of exposition, we do not explicitly state the problems in the standard form of formulation (1) in this section. However, second-stage 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 Supply Chain Management

We 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 $g \in \mathcal{G}$. Without loss of generality, we assume that no demand $\tilde{\xi}_{gi}$ is zero w.p. 1 under every probability measure $\mathbb{P} \in \mathcal{P}$. The company wishes to build one specialized production facility for each good as well as up to W warehouses that can carry any combination of goods. 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 construction costs for a transportation link from location i to location j amount to $b_{ij} \in \mathbb{R}_{++}$, while the per-unit flow costs on this link are $t_{gij} \in \mathbb{R}_{++}$ and may vary across the goods $g \in \mathcal{G}$. Note that we do not require these costs to satisfy the triangle inequality, which implies that an optimal solution may ship goods between multiple warehouses before they reach a distribution center. We assume that each node $i \in \mathcal{N}$ has a processing bound C which limits its inbound transshipments across all goods. 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 costs.

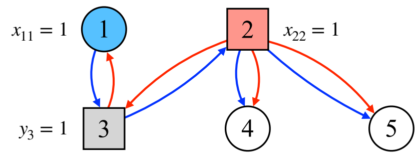

Figure 1. Supply chain design instance with two goods and two warehouses. The blue and red nodes represent the production facilities for goods 1 and 2, respectively, while the blue and red arcs illustrate their corresponding distribution networks $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. The square nodes house warehouses in addition to their distribution centers. A feasible set of potential values is given by $\boldsymbol{p}_1 = (1, 3, 2, 4, 4)^\top$ and $\boldsymbol{p}_2 = (3, 1, 2, 2, 2)^\top$.

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

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{i,j \in \mathcal{N}} b_{ij} c_{ij} + \sup_{\mathbb{P} \in \mathcal{P}} \ \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{p}, \boldsymbol{c}; \tilde{\boldsymbol{\xi}}) \right] & \\
\text{subject to} & \displaystyle\sum_{i \in \mathcal{N}} x_{gi} = 1 & \forall g \in \mathcal{G} \\
& \displaystyle\sum_{i \in \mathcal{N}} y_i \leqslant W & \\
& \displaystyle\sum_{i \in \mathcal{N}} z_{gij} \leqslant 1 & \forall g \in \mathcal{G}, \ \forall j \in \mathcal{N} \\
& p_{gj} \geqslant p_{gi} + 1 - \mathrm{M} \cdot (1 - z_{gij}) & \forall g \in \mathcal{G}, \ \forall i, j \in \mathcal{N} \\
& z_{gij} \leqslant x_{gi} + y_i & \forall g \in \mathcal{G}, \ \forall i, j \in \mathcal{N} \\
& \boldsymbol{x} \in \{0,1\}^{GN}, \ \ \boldsymbol{y} \in \{0,1\}^{N}, \ \ \boldsymbol{z} \in \{0,1\}^{GN^2}, \ \ \boldsymbol{p} \in \mathbb{R}_+^{GN}, \ \ \boldsymbol{c} \in \mathbb{R}_+^{N^2}, &
\end{array}
\tag{6a}
$$

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 c_{ij} record the capacity assigned to the link $(i, j) \in \mathcal{N} \times \mathcal{N}$, z_{gij} determines whether the link (i, j) is part of the distribution network for good $g \in \mathcal{G}$, and the potential variables p_{gi} exclude cycles in the distribution networks. 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 and fourth constraint stipulate that the distribution network $\boldsymbol{z}_g = (z_{gij})_{i,j}$ for each good $g \in \mathcal{G}$ is an arborescence, that is, they preclude networks where a node receives the same product from multiple sources as well as cycles. This is not a business requirement, but it will turn out to play a crucial role for the optimality of affine decision rules. The last constraint ensures that only production facilities and warehouses ship goods. The constant M is chosen to be sufficiently large but finite. Figure 1 illustrates the notation and constraints of our supply chain design problem.

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

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{g \in \mathcal{G}} \sum_{i,j \in \mathcal{N}} t_{gij} f_{gij} & \\ \text{subject to} & \displaystyle \mathrm{M} \cdot x_{gj} + \sum_{i \in \mathcal{N}} f_{gij} \geqslant \sum_{i \in \mathcal{N}} f_{gji} + \xi_{gj} & \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\ & \displaystyle\sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{N}} f_{gij} \leqslant C & \forall j \in \mathcal{N} \\ & \displaystyle\sum_{g \in \mathcal{G}} f_{gij} \leqslant c_{ij} & \forall i, j \in \mathcal{N} \\ & f_{gij} \leqslant \mathrm{M} \cdot z_{gij} & \forall g \in \mathcal{G},\ \forall i, j \in \mathcal{N} \\ & \boldsymbol{f} \in \mathbb{R}_+^{GN^2} & \end{array} \tag{6b}$$

Here, the decision variables f_{gij} record the transshipments of product $g \in \mathcal{G}$ across the locations $i, j \in \mathcal{N}$. The objective function minimizes the overall transportation costs. The constraints, from top to bottom, ensure that product flows are conserved across the nodes, the node and link capacities are obeyed and the transshipments are limited to the distribution networks $(z_{gij})_{i,j}$, $g \in \mathcal{G}$. Note that an instance of problem (6) may be infeasible if the nodal capacity C is insufficient to serve the customer demands.

Observation 1. *Problem* (6) *is optimally solved in affine decision rules.*

We emphasize that the proof of Observation 1 crucially relies on the restriction of the product-wise distribution networks $\boldsymbol{z}_g$ to arborescences, that is, the third and fourth constraint in problem (6a). We now show that in certain cases, this restriction can be lifted without impacting the optimality of affine decision rules. To this end, we define the *unrestricted* version of our supply chain management problem as the variant of problem (6) where the third and fourth constraint in the first stage (6a) as well as the potential variables $\boldsymbol{p} \in \mathbb{R}_+^{GN}$ are removed.

Proposition 5. *Assume that* (i) *all distributions* $\mathbb{P} \in \mathcal{P}$ *share the same rectangular support* Ξ*; that* (ii) *the ambiguous risk measure* ρ *satisfies the condition of Remark 2; and that* (iii) *there is* $\boldsymbol{\tau} \in \mathbb{R}_+^G$ *such that* $t_{gij} = \tau_g \cdot b_{ij}$ *for all* $g \in \mathcal{G}$ *and* $i, j \in \mathcal{N}$*. Then the unrestricted version of problem* (6) *is optimally solved in affine decision rules.*

The existence of $\boldsymbol{\tau} \in \mathbb{R}_+^G$ such that $t_{gij} = \tau_g \cdot b_{ij}$ for all $g \in \mathcal{G}$ and $i, j \in \mathcal{N}$ implies that the second-stage transportation costs are proportional to the first-stage construction costs. One can show that the assumptions in Proposition 5 are minimal in the sense that counterexamples for the optimality of affine decision rules exist that satisfy all but one of these assumptions.

We emphasize that Proposition 5 only ensures the existence of optimal affine decision rules for *optimal* first-stage decisions. Indeed, consider an unrestricted supply chain management problem with a single good ($G = 1$), $N = 3$ nodes with the production facility at node 1, a single warehouse at node 2 and the only non-zero demand arising at node 3. Assume that $\mathcal{P} = \{\mathbb{P}\}$ and that $\tilde{\xi}_3$ is governed by a uniform distribution supported on $[0, 10]$ under $\mathbb{P}$. Consider the (suboptimal) first-stage decision under which node 3 can receive up to 5 units of the good from node 1 and node 2 each. Assuming that $t_{113} < t_{123}$, for each realization ξ_3 the unique minimizer of the second-stage problem satisfies $f_{113}(\xi_3) = \min\{\xi_3, 5\}$ and $f_{123}(\xi_3) = \max\{\xi_3 - 5, 0\}$ and is thus nonlinear.

While affine decision rules tend to be suboptimal in the unrestricted supply chain management problem (except for cases where the conditions of Proposition 5 are satisfied), we can show that under mild conditions, the suboptimality of affine decision rules remains small if there are optimal solutions whose distribution networks are sufficiently close to arborescences. To this end, note that the distribution network of the restricted supply chain management problem (6) comprises $N - 1$ arcs for each of the G arborescences, that is, $G(N - 1)$ arcs in total.

Proposition 6. *Assume that the unrestricted supply chain management problem has an optimal solution whose distribution network has at most $G(N - 1) + \eta$ arcs $z_{gij} = 1$, $(g, i, j) \in \mathcal{G} \times \mathcal{N}^2$. Assume further that the ambiguous risk measure ρ is translation invariant, and that it satisfies the condition of Remark 2. If we regard the per-unit construction and flow costs as constant, then the relative suboptimality of any optimal affine decision rule solution is bounded from above by $\mathcal{O}\big(\eta GC \,/\, \max_{\boldsymbol{\xi} \in \Xi} \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \xi_{gj}\big)$.*

Note that the unrestricted supply chain management problem may have multiple optimal solutions. In particular, if $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c})$ is feasible in the problem, then so is any $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}', \boldsymbol{c})$ with $z'_{gij} \in [z_{gij},\, x_{gi} + y_i]$, $(g, i, j) \in \mathcal{G} \times \mathcal{N}^2$, and $\boldsymbol{z}' \in \{0, 1\}^{GN^2}$. Proposition 6 applies to all such optimal solutions, and its statement is strongest when applied to optimal solutions with minimum numbers of distribution links $z_{gij} = 1$.

We demonstrate the findings of Proposition 6 with a numerical experiment. Consider a family of $10,000$ unrestricted supply chain management instances with a single good ($G = 1$), $N = 15$ nodes and $W = 5$ warehouses. The per-unit construction and flow costs are selected uniformly

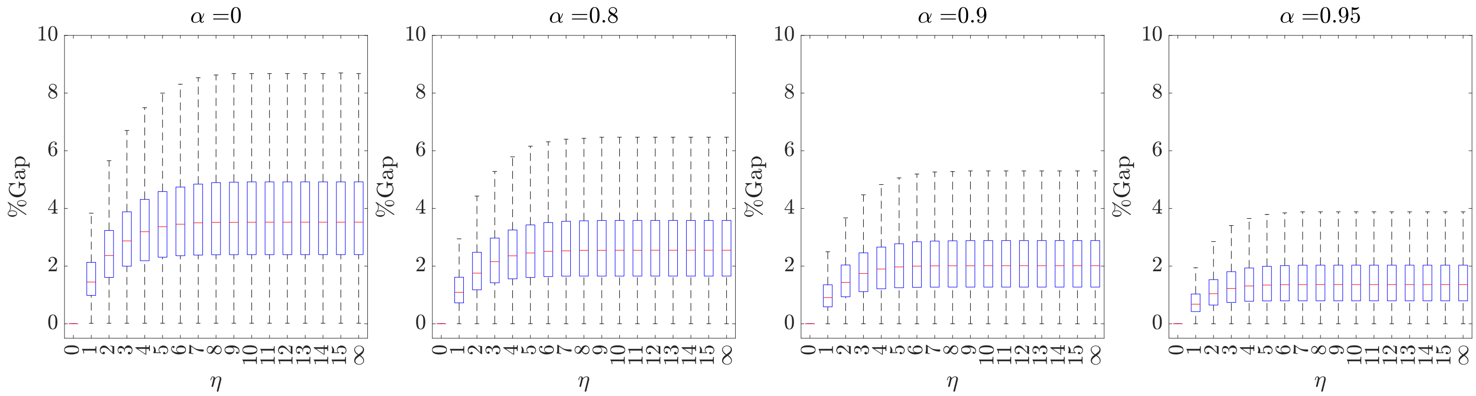

Figure 2. Suboptimality of imposing an arborescence structure in the unrestricted supply chain management problem (6). The suboptimality of arborescences is expressed relative to solutions that are allowed to introduce η additional distribution links. Shown are the results for the CVaR risk measure at risk thresholds $\alpha = 0, 0.8, 0.9, 0.95$.

at random from the interval $[0, 1]$. The uncertain nodal demands are described by a singleton ambiguity set with a (non-ambiguous) discrete probability distribution that places equal weight on 100 scenarios. Each scenario corresponds to a nodal demand vector whose components are selected uniformly at random from the interval $[0, 1]$. We use the conditional value-at-risk (CVaR) as risk measure ρ, and we disregard the nodal capacities C. Figure 2 visualizes the percentage cost savings of allowing for η additional distribution links to be introduced on top of the arborescence structure in the restricted supply chain management problem (6). As expected, the incremental benefits of deviating from an arborescence structure are decreasing, and the overall benefits saturate at around 3.5% on average. Additionally, we observe that the benefit of introducing additional distribution links diminishes as the risk aversion increases, that is, as α increases.

While we know from Observation 1 that the restricted supply chain management problem (6) is optimally solved in affine decision rules, a decision maker may not be required to select a distribution network with an arborescence structure. Instead, she may prefer to solve the *unrestricted* supply chain management problem in affine decision rules. To this end, Figure 3 visualizes the suboptimality of the best affine decision rule solution in the unrestricted supply chain management problem (6), measured relative to an optimal solution in the same (unrestricted) problem. Note that we can determine optimal solutions numerically for this family of instances thanks to the small problem size and the use of non-ambiguous discrete probability distributions; the same would not

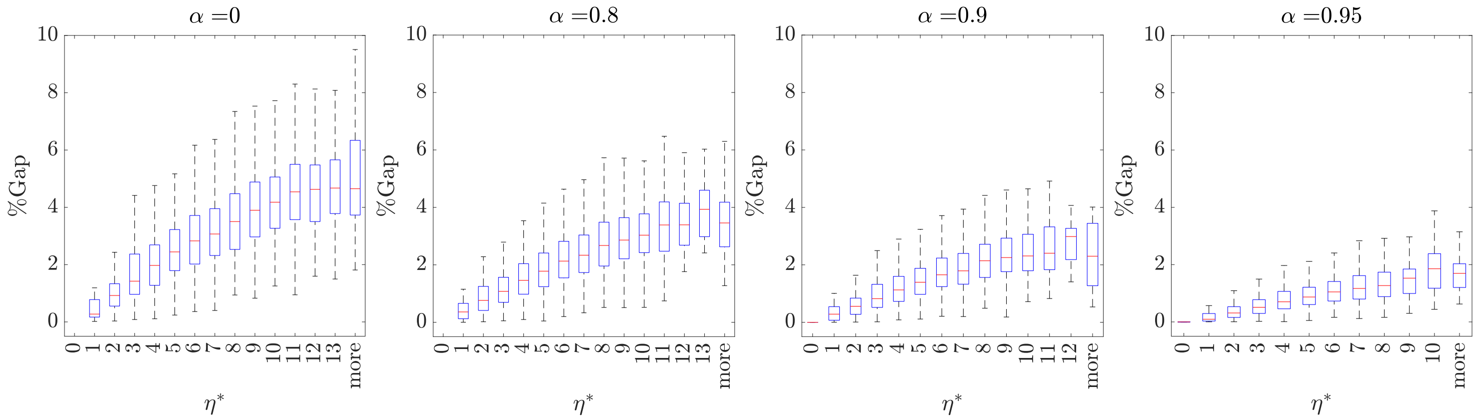

Figure 3. Suboptimality of imposing affine decision rules in the unrestricted supply chain management problem (6). The suboptimality of affine decision rules is computed relative to optimal unrestricted solutions on those instances where the optimal solutions introduce $\eta^\star$ additional links. Shown are the results for the CVaR risk measure at risk thresholds $\alpha = 0, 0.8, 0.9, 0.95$.

hold true for larger and/or distributionally robust problems. Figure 3 shows that—as predicted by Proposition 6—the suboptimality of affine decision rules in the unrestricted problem variant exhibits a clear correlation with the number $\eta^\star$ of additional distribution links introduced by optimal solutions (on top of the arborescence structure), and that this suboptimality remains small across the considered instances. As in the previous experiment, we find that the suboptimality of affine decision rules decreases as the decision maker's risk aversion increases. This is in line with our intuition that affine decision rules tend to perform particularly well in robust problems (that correspond to case of maximally risk-averse decision makers).

Returning to problem (6) with the restriction of the product-wise distribution networks to arborescences, we now assume that the company does not have to meet all demands with probability 1, that is, lost sales are admissible but penalized at a cost of σ per unit. To simplify the analysis, we drop the nodal capacity constraints and assume that the support Ξ is a hyperrectangle. In this case, the first stage of our supply chain design problem (6) remains unchanged, whereas the second-stage costs $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{p}, \boldsymbol{c}; \boldsymbol{\xi})$ now become the optimal value of the following second-stage problem.

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{g \in \mathcal{G}} \sum_{i,j \in \mathcal{N}} t_{gij} f_{gij} + \sigma \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} l_{gj} & \\ \text{subject to} & \mathrm{M} \cdot x_{gj} + \displaystyle\sum_{i \in \mathcal{N}} f_{gij} + l_{gj} \geqslant \sum_{i \in \mathcal{N}} f_{gji} + \xi_{gj} & \forall g \in \mathcal{G}, \ \forall j \in \mathcal{N} \\ & \displaystyle\sum_{g \in \mathcal{G}} f_{gij} \leqslant c_{ij} & \forall i, j \in \mathcal{N} \\ & f_{gij} \leqslant \mathrm{M} \cdot z_{gij} & \forall g \in \mathcal{G}, \ \forall i, j \in \mathcal{N} \\ & \boldsymbol{f} \in \mathbb{R}_+^{GN^2}, \ \ \boldsymbol{l} \in \mathbb{R}_+^{GN} & \end{array} \tag{7}$$

Here, the non-negative decision variables l_{gj} record the lost sales of good $g \in \mathcal{G}$ at the distribution center $j \in \mathcal{N}$. The objective function accounts for the cumulative lost sales across all goods and distribution centers, and the first constraint now allows for the original flow balances to be violated at the expense of non-zero lost sales.

Our revised supply chain design problem with first stage (6a) and second stage (7) does not satisfy the conditions of Theorem 1, and one can indeed readily construct problem instances that are not optimally solved in affine decision rules. We next show that the second-stage problem (7) admits the equivalent reformulation

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{g \in \mathcal{G}} \sum_{i,j \in \mathcal{N}} t_{gij} f_{gij} + \sigma \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \max\{l_{gj}, 0\} & \\ \text{subject to} & \mathrm{M} \cdot x_{gj} + l_{gj} + \displaystyle\sum_{i \in \mathcal{N}} f_{gij} \geqslant \sum_{i \in \mathcal{N}} f_{gji} + \xi_{gj} & \forall g \in \mathcal{G}, \ \forall j \in \mathcal{N} \\ & \displaystyle\sum_{g \in \mathcal{G}} f_{gij} \leqslant c_{ij} & \forall i, j \in \mathcal{N} \\ & f_{gij} \leqslant \mathrm{M} \cdot z_{gij} & \forall g \in \mathcal{G}, \ \forall i, j \in \mathcal{N} \\ & \boldsymbol{f} \in \mathbb{R}_+^{GN^2}, \ \ \boldsymbol{l} \in \mathbb{R}^{GN}. & \end{array} \tag{7'}$$

Problem (7') differs from problem (7) in two respects: the lost sales decisions l_{gj} are no longer required to be non-negative, and the lost sales expression in the objective function of problem (7') is nonlinear.

Observation 2. *The revised supply chain design problem with first stage* (6a) *and second stage* (7) *is equivalent to the revised supply chain design problem with first stage* (6a) *and second stage* (7') *in the following sense:*

(i) *Any (possibly nonlinear) feasible solution to* (6a) *and* (7) *gives rise to a (possibly nonlinear) feasible solution to* (6a) *and* (7') *with at most the same objective value, and vice versa.*

(ii) *Any feasible solution to* (6a) *and* (7) *with affine flows* $\boldsymbol{f}$ *gives rise to a fully affine feasible solution to* (6a) *and* (7') *with at most the same objective value.*

(iii) *Any feasible solution to* (6a) *and* (7') *with affine flows* $\boldsymbol{f}$ *gives rise to a feasible solution to* (6a) *and* (7) *with affine flows with at most the same objective value.*

Observation 2 shows that solving the first-stage problem (6a) with the second-stage problem (7') in affine decision rules for both $\boldsymbol{f}$ and $\boldsymbol{l}$ is tantamount to solving the first-stage problem (6a) with the second-stage problem (7) in affine flow decisions $\boldsymbol{f}$ but generic nonlinear lost sales decisions $\boldsymbol{l}$. The next theorem exploits this equivalence to bound the suboptimality of affine decision rules in the first-stage problem (6a) with the second-stage problem (7').

Theorem 3. *The relative suboptimality of affine decision rules in the revised supply chain design problem with first stage* (6a) *and second stage* (7') *is bounded from above by*

$$\left(\bar{t}/\underline{t} + \sigma/\underline{t}\right) \cdot \max_{\boldsymbol{\xi} \in \Xi} \frac{\sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} l^{\star}_{gj}(\boldsymbol{\xi})}{\sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \xi_{gj}},$$

where $\boldsymbol{l}^{\star}$ *is the lost sales policy of any optimal (nonlinear) solution* $(\boldsymbol{f}^{\star}, \boldsymbol{l}^{\star})$ *and* $\bar{t} = \max\{t_{gij} : g \in \mathcal{G},\ i,j \in \mathcal{N}\}$ *and* $\underline{t} = \min\{t_{gij} : g \in \mathcal{G},\ i,j \in \mathcal{N}\}$ *denote the maximum and minimum unit transportation costs in the network, respectively.*

Theorem 3 shows that for fixed per-unit transportation and lost sales costs, the relative suboptimality of affine decision rules is bounded by the worst-case percentage of lost sales in any optimal second-stage policy. In particular, affine decision rules become optimal if the percentage of lost sales in an optimal second-stage policy converges to zero.

We complement the qualitative insights of Theorem 3 with a numerical experiment. Consider a family of 10,000 supply chain instances with a single good (G=1), $N = 15$ nodes and $W = 5$ warehouses. The N nodes are located uniformly at random on the $[0,1]^2$ square. The per-unit transportation costs t_{gij} are set to the Euclidean distances between the corresponding nodes, irrespective of the good $g \in \mathcal{G}$. The per-unit construction costs are set to $b_{ij} = M_b \cdot t_{gij}$, where the scalar M_b is chosen uniformly at random from the interval $[0,5]$. The per-unit lost sales costs are set to $\sigma = (M_b\sqrt{2} + 1) + M_\sigma$, where M_σ is chosen uniformly at random from the interval $[0,5]$ and captures the relative difference between lost sales and investment/flow costs in each instance. The

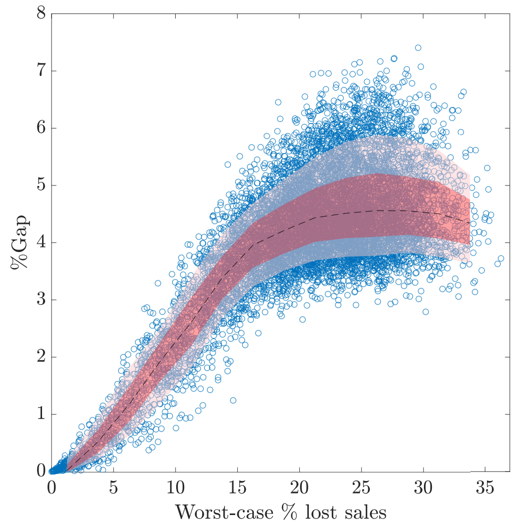

Figure 4. Relative suboptimality of affine decision rules in the supply chain management problem with lost sales. The circles represent the 10,000 instances, the dotted line reports the median gap, the dark red shaded region represents the 25th-75th percentile range, and the light red region visualizes the 10th-90th percentile range of the gap, respectively.

uncertain nodal demands are described by a singleton ambiguity set with a (non-ambiguous) discrete probability distribution that places equal weight on 100 scenarios. Each scenario corresponds to a nodal demand vector whose components are selected uniformly at random from the interval $[0,1]$. We use the expectation operator as risk measure ρ. Figure 4 reports the relative suboptimality of affine decision rules as a function of the worst-case percentage of lost sales in the optimal solution. The figure numerically confirms the qualitative findings of Theorem 3: the suboptimality of affine decision rules increases when the worst-case percentage of lost sales in the optimal solution increases. At the same time, the optimality gap remains moderate for all of the considered problem instances.

Returning once more to problem (6) without lost sales, we now assume that there is the option to pre-ship goods to a subset of at most L nodes, $L \ll N$, at discounted transshipment costs $\lambda \cdot t_{gij}$, $\lambda \in (0,1)$. To this end, we introduce the first-stage transshipment decisions $\boldsymbol{s} \in \mathbb{R}_+^{GN^2}$, we augment the first-stage objective function with the term $\sum_{g \in \mathcal{G}} \sum_{i,j \in \mathcal{N}} \lambda \cdot t_{gij} s_{gij}$ and the first-stage constraints with

$$\begin{aligned}
&\sum_{i \in \mathcal{N}} s_{gij} \geqslant \sum_{i \in \mathcal{N}} s_{gji} && \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
&\sum_{i \in \mathcal{N}} s_{gij} \leqslant \sum_{i \in \mathcal{N}} s_{gji} + \mathrm{M} \cdot w_j && \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
&\sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{N}} s_{gij} \leqslant C && \forall j \in \mathcal{N} \\
&\sum_{g \in \mathcal{G}} s_{gij} \leqslant c_{ij} && \forall i, j \in \mathcal{N} \\
&\sum_{i \in \mathcal{N}} w_i \leqslant L,
\end{aligned} \tag{8a}$$

where the auxiliary decision variables $\boldsymbol{w} \in \{0,1\}^N$ record which of the nodes receive a pre-shipment. We also replace the second-stage flow conservation constraints with

$$\mathrm{M} \cdot x_{gj} + \sum_{i \in \mathcal{N}} (s_{gij} + f_{gij}) \geqslant \xi_{gj} + \sum_{i \in \mathcal{N}} (s_{gji} + f_{gji}) \qquad \forall g \in \mathcal{G},\ \forall j \in \mathcal{N}. \tag{8b}$$

Note that we do *not* require the pre-shipments to form an arborescence, and that all nodes $j \in \mathcal{N}$ can act as transshipment nodes with $s_{gij}, s_{gji} > 0$ (as opposed to just L of them).

Observation 3. *Problem* (6) *with the updated constraints* (8) *is optimally solved in K-adaptable affine decision rules when $K \geqslant 2^{LG}$.*

Intuitively, for each good $g \in \mathcal{G}$ and each node $j \in \mathcal{N}$ that can receive a pre-shipment (*i.e.*, for which $w_j = 1$), either the revised flow conservation constraint (8b) or the non-negativity constraint $f_{gij} \geqslant 0$ is binding for the unique node $i \in \mathcal{N}$ for which $z_{gij} = 1$. The bound of Observation 3 accounts for all combinations that emerge from these LG binary choices.

A closer look at problem (6) with the updated constraints (8) reveals that the second-stage decisions $\boldsymbol{f} \in \mathbb{R}_+^{GN^2}$ actually decompose into product-wise flow graphs $\boldsymbol{f}_g \in \mathbb{R}_+^{N^2}$, $g \in \mathcal{G}$, such that the conditions (**F**), (**A'**), (**D**) and (**B"**) from Section 3 are satisfied. This allows us to invoke Proposition 4 and conclude that 2^L affine decision rules per good $g \in \mathcal{G}$ are sufficient, that is, we can tighten the bound of Observation 3 as follows.

Observation 4. *Problem* (6) *with the updated constraints* (8) *is optimally solved in decomposed K_g-adaptable affine decision rules for $\boldsymbol{f}_g \in \mathbb{R}_+^{N^2}$ when $K_{\mathrm{B}}^g \geqslant 2^L$ for all $g \in \mathcal{G}$.*

We emphasize that problem (6) with the updated constraints (8) does *not* decouple into separate problems for each good $g \in \mathcal{G}$ due to the presence of the warehouse locations $\boldsymbol{y}$ as well as the nodal and link capacity constraints that couple all goods $g \in \mathcal{G}$.

One can use arguments similar to the ones in this section to confirm the optimality of affine decision rules in several variants of the supply chain design problem, such as instances where factories and warehouses incur location-dependent construction costs or where factories can produce multiple products (possibly with a penalty for diversification). 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 1. 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. Likewise, measuring the distance of optimal solutions to those auxiliary structures may provide means to derive suboptimality bounds of affine decision rule solutions in the more general, unconstrained problem variants.

Remark 3 (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.

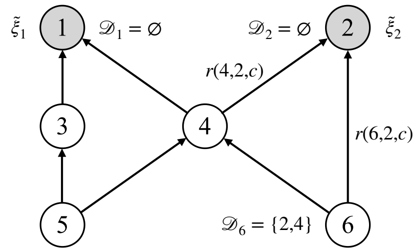

Figure 5. Flexible production planning instance with six entities and a fixed set of configurations. The gray nodes represent final products without descendants, $\mathcal{D}_e = \varnothing$, and with uncertain demands $\tilde{\xi}_e$. In contrast to the supply chain design problem from the previous section, the bill of materials is acyclic but not necessarily arborescent.

4.2 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 descendants as $\mathcal{D}_e = \{d \in \mathcal{E} \,:\, r(e, d, c) > 0 \text{ for some } c \in \mathcal{C}_d\}$. We require that for each choice of configurations $\{c_e\}_{e \in \mathcal{E}}$, the directed production graph with nodes $\mathcal{E}$ and arcs $\{(e, d) \in \mathcal{E} \times \mathcal{E} \,:\, r(e, d, c_d) > 0\}$ is acyclic (but—in contrast to our supply chain management problem from Section 4.1—not necessarily an arborescence).

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:

$$\begin{array}{lll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] & \\ \text{subject to} & \sum_{c \in \mathcal{C}_e} x_{ec} = 1 & \forall e \in \mathcal{E} \\ & x_{ec} \in \{0, 1\} & \forall e \in \mathcal{E},\ \forall c \in \mathcal{C}_e \end{array} \tag{9a}$$

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

$$\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{e\in\mathcal{E}}\sum_{c\in\mathcal{C}_e} p(e,c)\cdot x_{ec}\cdot y_e & \\
\text{subject to} & \displaystyle y_e + \mathrm{M}\cdot\sum_{d\in\mathcal{D}_e}(1-x_{d,c_d}) \geqslant \xi_e + \sum_{d\in\mathcal{D}_e} r(e,d,c_d)\cdot y_d & \forall e\in\mathcal{E},\ \forall \boldsymbol{c}\in \bigtimes_{d\in\mathcal{D}_e}\mathcal{C}_d \\
& y_e \geqslant \xi_e & \forall e\in\mathcal{E}\,:\,\mathcal{D}_e=\emptyset \\
& y_e\in\mathbb{R},\ e\in\mathcal{E}, &
\end{array} \tag{9b}$$

where M is a sufficiently large but finite constant, and the decision y_e determines the quantity of entity $e \in \mathcal{E}$ to produce or procure. The objective function of (9b) minimizes the overall production costs. The first constraint ensures that the quantities y_e of all entities $e \in \mathcal{E}$ with descendants are sufficient to serve both the direct demands as well as the input requirements of all immediate descendants $d \in \mathcal{D}_e$, while the second constraint ensures that the quantities y_e of all entities $e \in \mathcal{E}$ without descendants are sufficient to serve the direct demands. Although the first constraint group comprises $\prod_{d\in\mathcal{D}_e}|\mathcal{C}_d|$ different constraints for each entity $e \in \mathcal{E}$, only the constraint whose configuration vector $\boldsymbol{c}$ satisfies $x_{d,c_d} = 1$ for all $d \in \mathcal{D}_e$ 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* descendants of e. Thus, the size of the formulation remains moderate as long as the number of immediate descendants is small for every entity. 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 $\boldsymbol{r}$. Figure 5 illustrates the notation of our flexible production planning problem.

We now argue that problem (9) is optimally solved in 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, the next observation makes use of Proposition 2.

Observation 5. *Problem* (9) *is optimally solved in affine decision rules.*

Observation 5 ensures that affine decision rules are optimal in problem (9). Even with this insight, however, problem (9) 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.

Note that we can generalize the second-stage objective function in (9b) to

$$\sum_{e \in \mathcal{E}} \sum_{c \in \mathcal{C}_e} x_{ec} \cdot p_{ec}(y_e)$$

without violating assumption (**F**) as long as the cost functions $p_{ec} : \mathcal{E} \to \mathbb{R}_+$ are monotonically non-decreasing. In practice, one can envision both concave cost functions (modeling decreasing marginal production costs due to economies of scale) and convex cost functions (modeling increasing marginal production costs due to reliance on overtime and/or less favorable supply contracts). We emphasize, however, that even in the case of piecewise affine cost functions p_{ec}, we cannot linearize the objective function in (9b) via an epigraph reformulation without sacrificing optimality of affine decision rules. 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 (**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 $1/2$. In other words, Theorem 1 would be less general if we were to restrict our attention to affine objective functions f in assumption (**F**) and rely on epigraphical reformulations instead.

Consider now a variant of the flexible production planning problem (9) that incorporates supply-side online substitutions (Balakrishnan and Geunes, 2000):

$$\begin{array}{lll}
\text{minimize} & \sup\limits_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\{\boldsymbol{x}^k\}_k; \tilde{\boldsymbol{\xi}})\right] & \\
\text{subject to} & \sum\limits_{c \in \mathcal{C}_e} x^k_{ec} = 1 & \forall k = 1, \ldots, K, \; \forall e \in \mathcal{E} \\
& \dfrac{1}{2} \sum\limits_{e \in \mathcal{E}} \sum\limits_{c \in \mathcal{C}_e} |x^k_{ec} - x^l_{ec}| \leqslant B & \forall 1 \leqslant k < l \leqslant K \\
& x^k_{ec} \in \{0, 1\} & \forall k = 1, \ldots, K, \; \forall e \in \mathcal{E}, \; \forall c \in \mathcal{C}_e
\end{array} \tag{10}$$

Here, the second-stage costs satisfy $\mathcal{Q}(\{\boldsymbol{x}^k\}_k; \boldsymbol{\xi}) = \min\{\mathcal{Q}(\boldsymbol{x}^k; \boldsymbol{\xi}) \,:\, k = 1, \ldots K\}$ with $\mathcal{Q}(\boldsymbol{x}^k; \boldsymbol{\xi})$ representing the optimal value of the second-stage problem (9b) from before. In contrast to the original formulation (9), problem (10) now selects K different configuration vectors $\boldsymbol{x}^k$, $k = 1, \ldots, K$, the best of which is chosen reactively once the demands $\boldsymbol{\xi}$ have been observed. The second constraint set in (10) ensures that each pair of configuration vectors differ across at most B entities, where the constant B trades off the competing goals of flexibility and complexity of the production process.

Observation 6. *Problem* (10) *is optimally solved in K-adaptable affine decision rules.*

The vigilant reader may have noted that the product-wise distribution networks of our supply chain management case study needed to be arborescences, whereas the production graphs in this section need to be acyclic but may contain nodes with multiple incoming arcs. The reason for this is subtle: Once a configuration vector $\boldsymbol{c}$ has been selected in the first stage of our production planning problem, manufacturing one unit of entity $e \in \mathcal{E}$ requires precisely $r(a, e, c_e)$ units of *each* ancestor entity $a \in \mathcal{E}$ satisfying $r(a, e, c_e) > 0$ – in particular, it is not possible to substitute a shortage of one ancestor $a \in \mathcal{N}$ through an abundance of another ancestor $a' \in \mathcal{N}$. In contrast, a demand ξ_{gj} for good $g \in \mathcal{G}$ in node $j \in \mathcal{N}$ of our supply chain management problem could be served through any combination of inflows $f_{gij} > 0$, $i \in \mathcal{N}$, that is, the different flows can substitute each other. Modelling such substitutability requires nonlinear decision rules, which is also why our extension of pre-shipments (*cf.* Section 4.1) required the use of K-adaptable affine decision rules. Note that imposing an acyclic graph structure in this section is essentially non-restrictive. Specifically, if a selected choice of configurations in a multi-level bill of materials were to contain a cycle, this would imply that a part A is used to produce a part B, while part B is simultaneously used to produce part A. Such a scenario would only be meaningful if parts A and B are perfect substitutes. In that case, however, either of the two parts can be eliminated from the problem altogether.

The problem formulations studied in this section serve as a template for different production planning problems. One can immediately conceive variants with uncertain production costs $\tilde{p}(e, c)$ or, more generally, uncertain production cost functions $p_{ec} : \mathcal{E} \times \Xi \to \mathbb{R}_+$. As long as the uncertain production cost functions are monotonically non-decreasing in their first argument for every fixed value of the second argument, our results continue to hold. Our theory also extends to problem variants that model dependencies between the admissible configurations for different entities or individual and/or joint upper bounds on the quantities y_e (modelling, *e.g.*, inventory or resource constraints).

5 Conclusion

Problems known to be solved optimally by affine decision rules are rare and were, to our best knowledge, limited 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 can in fact be optimal in two-stage distributionally robust optimization problems *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 or provide analytical insight. We believe that the optimality conditions put forward in this paper may serve as a useful method in a modeller's toolbox to those ends. Our supply chain management problem from Section 4.1, for example, is solved optimally in affine decision rules if we impose the additional assumption of product-wise acyclic distribution networks. The resulting solution may be implemented as is, or it can serve as a basis to study the suboptimality of affine decision rules in more generic problem formulations that violate our optimality conditions (*cf.* Proposition 6). 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.

Acknowledgements

The authors would like to sincerely thank the department editor, the associate editor as well as the three referees for their constructive and insightful feedback that has led to substantial improvements

of the manuscript.

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.

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, 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.

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.

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.

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. 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.

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*, 46(2):674–711, 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.

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.

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.

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. 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. Interchangeability principle and dynamic equations in risk averse stochastic programming. *Operations Research Letters*, 45(4):377–381, 2017.

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.

Appendix A: Conservative K-Adaptability Approximations

Consider the following generalization of the two-stage distributionally robust optimization problem (1),

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{11a}$$

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

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

Here, the new second-stage decisions $\boldsymbol{z}$ have a feasible region $\mathcal{Z}(\boldsymbol{x})$ that may be non-convex (*e.g.*, due to integer requirements) 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 assumption (**D**) from Section 2 as well as, for an $\boldsymbol{x} \in \mathcal{X}$ optimal in (11):

($\mathbf{F}^K$) For every $(\boldsymbol{z}, \boldsymbol{\xi}) \in \mathcal{Z}(\boldsymbol{x}) \times \Xi$, $f(\boldsymbol{x}, \cdot, \boldsymbol{z}; \boldsymbol{\xi})$ is monotonically non-decreasing in $\boldsymbol{y}$.

($\mathbf{A}^K$) The technology matrices $\boldsymbol{A}$ and $\boldsymbol{E}$ and the right-hand side $\boldsymbol{g}$ are affine functions of $\boldsymbol{\xi}$.

($\mathbf{B}^K$) For every $\boldsymbol{z} \in \mathcal{Z}(\boldsymbol{x})$, 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

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

We emphasize that under the new set of assumptions, the second-stage decisions $\boldsymbol{z}$ only have to satisfy weaker conditions akin to those 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 $\boldsymbol{y}$.

Similar to Example 1 from Section 3, one readily verifies that the assumptions ($\mathbf{F}^K$), ($\mathbf{A}^K$), (**D**) and ($\mathbf{B}^K$) are not sufficient to guarantee that problem (11) is optimized by a single affine decision rule $\boldsymbol{y} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$. We next consider the K-adaptability formulation associated with problem (11):

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

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

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k; \boldsymbol{\xi}) \;\; = \;\; \left[\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_k; \boldsymbol{\xi}) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \\ & \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{D}\boldsymbol{y} \leqslant \boldsymbol{h}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^{n_2} \end{array}\right]. \tag{12b}$$

Problem (12) determines K candidate first-stage decisions $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ for the second-stage decision $\boldsymbol{z}$ in problem (11) here-and-now and subsequently implements the best of these decisions once the value of $\tilde{\boldsymbol{\xi}}$ has been observed. Since problem (12) restricts the flexibility of the recourse decision in problem (11), it constitutes a conservative approximation and will typically not attain the same optimal value.

We will now show that under the assumptions ($\mathbf{F}^K$), ($\mathbf{A}^K$), ($\mathbf{D}$) and ($\mathbf{B}^K$), Theorem 2 implies that the optimal value of problem (12) does not change if we restrict $\boldsymbol{y}$ to a collection of affine decision rules, that is, if we instead solve the single-stage problem

$$\begin{array}{ll} \text{minimize} & \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k(\tilde{\boldsymbol{\xi}})\}_k, \{\boldsymbol{z}_k\}_k; \tilde{\boldsymbol{\xi}})\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \;\; \boldsymbol{y}_k : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2} \text{ and } \boldsymbol{z}_k \in \mathcal{Z}(\boldsymbol{x}),\, k = 1, \ldots, K, \end{array} \tag{13a}$$

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

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{z}_k; \boldsymbol{\xi}) \;\; = \;\; \begin{cases} f(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{z}_k; \boldsymbol{\xi}) & \begin{array}{l} \text{if } \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{E}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{B}\boldsymbol{y}_k \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \quad \text{and} \\ \quad \boldsymbol{C}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{\xi})\boldsymbol{z}_k + \boldsymbol{D}\boldsymbol{y}_k \leqslant \boldsymbol{h}(\boldsymbol{\xi}), \end{array} \\ +\infty & \text{otherwise.} \end{cases} \tag{13b}$$

Corollary 1. *Assume that problem* (12) *attains its optimal value and that (*$\boldsymbol{F}^K$*), (*$\boldsymbol{A}^K$*), (*$\boldsymbol{D}$*) and (*$\boldsymbol{B}^K$*) are satisfied. Then the optimal values of problems* (12) *and* (13) *coincide.*

The proof of Corollary 1 transforms the problems (12) and (13) into instances of the problems (1) and (4) and then applies Theorem 2 from Section 3. In analogy to Proposition 1, we can show that the assumptions ($\mathbf{F}^K$), ($\mathbf{A}^K$), ($\mathbf{D}$) and ($\mathbf{B}^K$) are minimal in the sense that if any of these

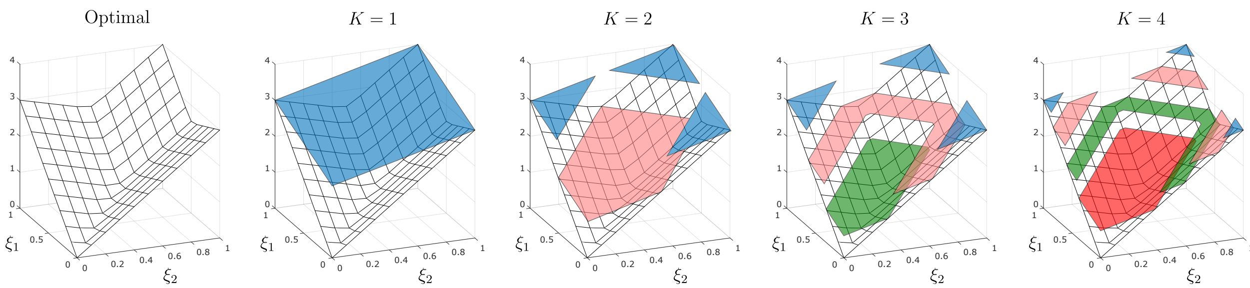

Figure 6. Objective values corresponding to the optimal second-stage policy and the optimal affine decision rules for $K = 1, \ldots, 4$ pre-selected candidate decisions $\{z_k\}_{k=1}^K$. Different colors correspond to realizations of $\boldsymbol{\xi}$ where a different affine policy is optimal.

assumptions is violated, then the statement of Corollary 1 ceases to hold in general even if all other assumptions are satisfied.

We illustrate the statement of Corollary 1 with an example.

Example 3. *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*

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

and thus the instance satisfies the assumptions ($\boldsymbol{F}^K$), ($\boldsymbol{A}^K$), ($\boldsymbol{D}$) and ($\boldsymbol{B}^K$). Figure 6 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 appendix with two immediate consequences of Corollary 1.

Remark 4 (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, Corollary 1 implies that the K-adaptability problem* (13) *recovers an optimal solution to the original two-stage distributionally robust optimization problem* (11) *for sufficiently large K, given that the assumptions ($\boldsymbol{F}^K$), ($\boldsymbol{A}^K$), ($\boldsymbol{D}$) and ($\boldsymbol{B}^K$) are satisfied.*

Remark 5 (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 (*$\boldsymbol{F}^K$*), (*$\boldsymbol{A}^K$*), (*$\boldsymbol{D}$*) and (*$\boldsymbol{B}^K$*) 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* $\boldsymbol{F}$*. Corollary 1 then implies that for an optimal 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 (*$\boldsymbol{F}^K$*), (*$\boldsymbol{A}^K$*), (*$\boldsymbol{D}$*) and (*$\boldsymbol{B}^K$*) 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. More generally, the suboptimality of the conservative approximation* (13) *to problem* (11) *can be fully ascribed to the finite adaptability of the second-stage decisions* $\boldsymbol{z}$*. This opens room for tailored solution approaches that directly address the suboptimality of these decisions.*

Appendix B: Proofs

Proof of Theorem 1. We show that for an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A**), (**D**) and (**B**) in problem (1), we can construct an affine decision rule $\boldsymbol{y}^{\ell} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$ such that

$$\sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] \;=\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}^{\ell}(\tilde{\boldsymbol{\xi}}); \tilde{\boldsymbol{\xi}}) \right], \tag{14}$$

where the cost functions on the left-hand side and right-hand side are defined in (1b) and (2b), respectively. The resulting solution $(\boldsymbol{x}, \boldsymbol{y}^{\ell})$ is feasible in problem (2), which implies that the optimal value of (2) is less than or equal to the optimal value of (1). On the other hand, the optimal value of problem (2) also weakly exceeds the optimal value of problem (1) since the former problem constitutes a restriction of the latter one. Thus, equation (14) implies that the optimal values of (1) and (2) coincide, as desired.

To show that equation (14) holds, fix an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A**), (**D**) and (**B**), 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

$$\boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}}\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}). \tag{15}$$

Since $\boldsymbol{B}_{\mathcal{I}}$ admits a positive inverse, the satisfaction of the constraint set (15) implies that

$$\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{B}_{\mathcal{I}}^{-1} \Big[\boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} \Big], \tag{16}$$

but not vice versa. Indeed, since $\boldsymbol{B}_{\mathcal{I}}^{-1} \geqslant \boldsymbol{0}$, the constraints in (16) constitute non-negative linear combinations of the constraints in (15), and thus the constraint system (16) is a relaxation of the constraint set (15). Consider now the solution

$$\boldsymbol{y}^{\ell}(\boldsymbol{\xi}) = \boldsymbol{B}_{\mathcal{I}}^{-1} \Big[\boldsymbol{g}_{\mathcal{I}}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}}(\boldsymbol{\xi})\boldsymbol{x} \Big] \qquad \forall \boldsymbol{\xi} \in \Xi, \tag{17}$$

which satisfies the relaxed constraint set (16) as equality for all $\boldsymbol{\xi} \in \Xi$ and which is evidently affine in $\boldsymbol{\xi}$. This solution satisfies the constraint set (15) and thus also the first constraint set in (1b)

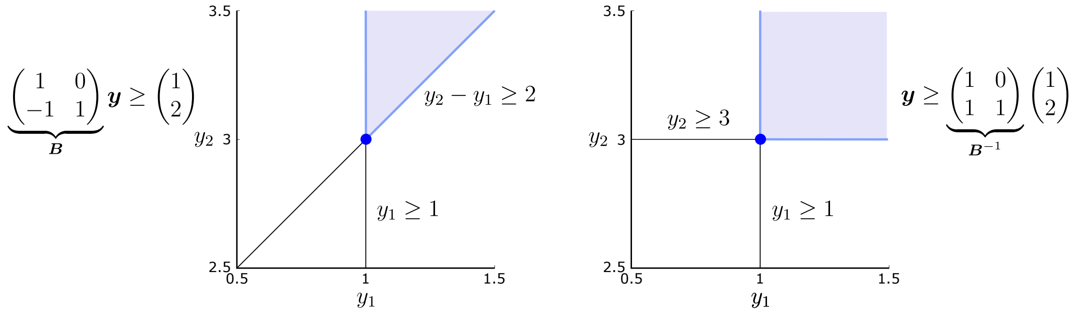

Figure 7. The feasible region imposed by the constraint system (15), left, is not equivalent to that of (16), right, but both share the same coordinate-wise minimal point $(1, 3)$.

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 (16), and $\boldsymbol{y}^{\ell}(\boldsymbol{\xi})$ is the point-wise smallest decision satisfying (16) according to its definition in (17).

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 (**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 $\boldsymbol{\xi} \in \Xi \backslash \operatorname{dom} \mathcal{Q}$ for which any second-stage decision is infeasible. Equation (14) thus follows, 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 (16) is a relaxation of the first constraint set in (1b) that, if strengthened to equalities as in (17), imposes an affine structure on the second-stage decisions $\boldsymbol{y}(\boldsymbol{\xi})$. Note that the constraint system (16) 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 $y_1 \geqslant 1$, $y_2 \geqslant 3$ of the associated equation (16), see Figure 7. 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.

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

$$\text{minimize} \quad \mathbb{E}_{\mathbb{P}}\big[\mathcal{Q}(\tilde{\xi})\big] \tag{18}$$

This instance contains no first-stage decision, its ambiguity set $\mathcal{P} = \{\mathbb{P}\}$ is a singleton set such that $\tilde{\xi} \sim \mathcal{U}[-1, 1]$ follows a univariate uniform distribution over the interval $[-1, 1]$ 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 (**A**), (**D**) and (**B**) of Theorem 1 are all satisfied. One readily verifies that the affine decision rule formulation (2) associated with problem (18) is optimized by $y^\star(\xi) = -1$, which results in an objective value of 1, whereas for every realization ξ, the second-stage problem $\mathcal{Q}(\xi)$ in (18) is optimized by $y(\xi) = \min\{\xi, \, -\xi\}$, resulting in a lower objective value of 1/2.

As for assumption (**A**), consider the instance of problem (1) with objective function (18), that is, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, there is 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 (**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 objective value of 1, whereas for every realization ξ, the second-stage problem $\mathcal{Q}(\xi)$ in (1) is optimized by $y(\xi) = \xi^2$, resulting in a lower objective value of 1/3.

In view of assumption (**D**), consider the instance of problem (1) with objective function (18), that is, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, there is 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 (**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 objective value of 1, whereas for every realization ξ, the second-stage problem $\mathcal{Q}(\xi)$ in (1) is optimized by $y(\xi) = \max\{\xi, -\xi\}$, resulting in a lower objective value of 1/2.

In view of assumption (**B**), finally, consider the instance of problem (1) with objective function (18), that is, the risk measure satisfies $\rho_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}$, there is 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 $y_1 + y_2 \geqslant \xi + 1$ 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 (**F**), (**A**) and (**D**) of Theorem 1 are all satisfied. One can verify that the associated affine decision rule formulation (2) is optimized by $y_1^\star(\xi) = y_2^\star(\xi) = (\xi + 1)/2$, resulting in an objective value of 3/2, whereas for every realization ξ, the second-stage problem $\mathcal{Q}(\xi)$ in (1) is optimized by $y_1(\xi) = \min\{\xi + 1, 1\}$ and $y_2(\xi) = \max\{\xi, 0\}$, resulting in a lower objective value of 5/4. □

Proof of Proposition 2. The constraint system (3) can be written as $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi})$ by setting

$$\begin{gathered} \boldsymbol{A}(\boldsymbol{\xi})^\top = \left[\boldsymbol{A}_1(\boldsymbol{\xi})^\top \, \ldots \, \boldsymbol{A}_{n_2}(\boldsymbol{\xi})^\top\right] \in \mathbb{R}^{n_1 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{A}_j(\boldsymbol{\xi})^\top = \left[-\boldsymbol{\alpha}_{j,1}(\boldsymbol{\xi}) \, \ldots \, -\boldsymbol{\alpha}_{j,s_j}(\boldsymbol{\xi})\right] \in \mathbb{R}^{n_1 \times s_j}, \\ \boldsymbol{B}^\top = \left[\boldsymbol{B}_1^\top \, \ldots \, \boldsymbol{B}_{n_2}^\top\right] \in \mathbb{R}^{n_2 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{B}_j^\top = \left[\mathbf{e}_j - \boldsymbol{\beta}_{j,1} \, \ldots \, \mathbf{e}_j - \boldsymbol{\beta}_{j,s_j}\right] \in \mathbb{R}^{n_2 \times s_j}, \\ \boldsymbol{g}(\boldsymbol{\xi})^\top = \left[\boldsymbol{g}_1(\boldsymbol{\xi})^\top \, \ldots \, \boldsymbol{g}_{n_2}(\boldsymbol{\xi})^\top\right] \in \mathbb{R}^{1 \times \sum_{j=1}^{n_2} s_j} \text{ with } \boldsymbol{g}_j(\boldsymbol{\xi})^\top = \left[\gamma_{j,1}(\boldsymbol{\xi}) \, \ldots \, \gamma_{j,s_j}(\boldsymbol{\xi})\right] \in \mathbb{R}^{1 \times s_j}, \end{gathered} \tag{19}$$

where $\mathbf{e}_j$ is the j-th canonical basis vector in $\mathbb{R}^{n_2}$. Fix the optimal $\boldsymbol{x} \in \mathcal{X}$ from the statement of the proposition, and choose k_j, $j = 1, \ldots, n_2$, 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

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

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 $\mathbf{\Pi} = \big[\mathbf{e}_{\pi(1)} \ \ldots \ \mathbf{e}_{\pi(n_2)}\big]^\top \in \mathbb{R}^{n_2 \times n_2}$, 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 $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{n_2 \times n_2}$. 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. □

Proof of Theorem 2. Following a similar argument as in the proof of Theorem 1, we show that for an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A'**), (**D**) and (**B'**) in problem (1), we can construct a collection of $K = K_{\mathrm{A}} \cdot K_{\mathrm{B}}$ affine decision rules $\boldsymbol{y}_k^\ell : \Xi \xrightarrow{\mathrm{a}} \mathbb{R}^{n_2}$, $k = 1, \ldots, K$, such that

$$\sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] \;=\; \sup_{\mathbb{P} \in \mathcal{P}} \rho_{\mathbb{P}} \left[\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell(\tilde{\boldsymbol{\xi}})\}_{k=1}^K; \tilde{\boldsymbol{\xi}}) \right], \tag{20}$$

where the left cost function is defined in (1b), whereas the right cost function satisfies $\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell(\boldsymbol{\xi})\}_{k=1}^K; \boldsymbol{\xi}) = \min\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k^\ell; \boldsymbol{\xi}) \, : \, k = 1, \ldots, K\}$ with $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k^\ell; \boldsymbol{\xi})$ defined in (2b). The resulting solution $(\boldsymbol{x}, \{\boldsymbol{y}_k^\ell\}_{k=1}^K)$ is thus feasible in problem (4), and a similar argument as in the proof of Theorem 1 then implies that the optimal values of both problems coincide.

To show that (20) holds, fix an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A'**), (**D**) and (**B'**), together with a collection of index sets $\mathcal{I}_k^{\mathrm{B}}$, $k = 1, \ldots, K_{\mathrm{B}}$, that satisfies assumption (**B'**). Using assumption (**A'**), construct the common refinement $\Xi = \bigcup_{k=1}^K \Xi_k$ with $\Xi_{(i-1)\cdot K_{\mathrm{B}}+j} = \Xi_i^{\mathrm{A}} \cap \Xi_j^{\mathrm{B}}$ as well as $\mathcal{I}_{(i-1)\cdot K_{\mathrm{B}}+j} = \mathcal{I}_j^{\mathrm{B}}$, $i = 1, \ldots, K_{\mathrm{A}}$ and $j = 1, \ldots, K_{\mathrm{B}}$.

Similar arguments as in the proof of Theorem 1 show that for each $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_k$, $k = 1, \ldots, K$, any feasible second-stage decision $\boldsymbol{y}(\boldsymbol{\xi})$ has to satisfy

$$\boldsymbol{y}(\boldsymbol{\xi}) \geqslant \boldsymbol{B}_{\mathcal{I}_k}^{-1} \Big[\boldsymbol{g}_{\mathcal{I}_k}(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}_k}(\boldsymbol{\xi}) \boldsymbol{x} \Big]. \tag{21}$$

By construction, $\boldsymbol{A}_{\mathcal{I}_k}$ and $\boldsymbol{g}_{\mathcal{I}_k}$ are affine over Ξ_k (but not necessarily over Ξ). Fix any matrix- and vector-valued functions $\boldsymbol{A}_{\mathcal{I}_k}^\ell$ and $\boldsymbol{g}_{\mathcal{I}_k}^\ell$ that are affine over the entire support Ξ and that coincide with $\boldsymbol{A}_{\mathcal{I}_k}$ and $\boldsymbol{g}_{\mathcal{I}_k}$ over Ξ_k, and consider the collection of affine decision rules

$$\boldsymbol{y}_k^\ell(\boldsymbol{\xi}) = \boldsymbol{B}_{\mathcal{I}_k}^{-1} \Big[\boldsymbol{g}_{\mathcal{I}_k}^\ell(\boldsymbol{\xi}) - \boldsymbol{A}_{\mathcal{I}_k}^\ell(\boldsymbol{\xi}) \boldsymbol{x} \Big], \qquad k = 1, \ldots, K, \tag{22}$$

each of which is defined over the entire support Ξ. Similar arguments as in the proof of Theorem 1 show that for each $k = 1, \ldots, K$, $\boldsymbol{y}_k^\ell(\boldsymbol{\xi})$ is the point-wise smallest feasible decision in the second-stage problem (1b) for all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_k$. Assumption (**F**) then implies that $\boldsymbol{y}_k^\ell(\boldsymbol{\xi})$ is point-wise optimal in the second-stage problem (1b) over $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_k$, $k = 1, \ldots, K$. Since $\Xi = \bigcup_{k=1}^K \Xi_k$, we thus have for all $k = 1, \ldots, K$ and all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_k$ that

$$\mathcal{Q}(\boldsymbol{x}, \{\boldsymbol{y}_{k'}^\ell(\boldsymbol{\xi})\}_{k'=1}^K; \boldsymbol{\xi}) \;=\; \min\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_{k'}^\ell(\boldsymbol{\xi}); \boldsymbol{\xi}) \, : \, k' = 1, \ldots, K\} \;\leqslant\; \mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}_k^\ell(\boldsymbol{\xi}); \boldsymbol{\xi}) \;\leqslant\; \mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}),$$

and equation (20) follows from the monotonicity of $\rho_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$. $\square$

Proof of Proposition 3. Using the same notation as in (19), the constraint system (3) can be written as $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi})$. For the remainder of the proof, fix the optimal $\boldsymbol{x} \in \mathcal{X}$ from the statement of the proposition. For $j = 1, \ldots, n_2$, choose $\mathcal{K}_j$ as stipulated in condition *(i)* of the statement and let Ξ_{jk}, $k \in \mathcal{K}_j$, be the set of parameter realizations $\boldsymbol{\xi} \in \Xi$ for which the constraint

$$y_j \geqslant \boldsymbol{\alpha}_{jk}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{jk}^\top \boldsymbol{y} + \gamma_{jk}(\boldsymbol{\xi})$$

weakly dominates all other constraints

$$y_j \geqslant \boldsymbol{\alpha}_{jk'}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{\beta}_{jk'}^\top \boldsymbol{y} + \gamma_{jk'}(\boldsymbol{\xi}),$$

$k' \neq k$, with ties between indices $k, k' \in \mathcal{K}_j$ broken arbitrarily. Condition *(i)* in the statement of the proposition implies that $\Xi = \bigcup_{k \in \mathcal{K}_j} \Xi_{jk}$ for all $j = 1, \ldots, n_2$, while our construction guarantees that $\Xi_{jk} \cap \Xi_{jk'} = \emptyset$ for all $k, k' \in \mathcal{K}_j$, $k \neq k'$. We claim that assumption (**B'**) is satisfied for the covering $\Xi = \bigcup_{\boldsymbol{k} \in \boldsymbol{\mathcal{K}}} \Xi_{\boldsymbol{k}}^{\mathrm{B}}$ with $\Xi_{\boldsymbol{k}}^{\mathrm{B}} = \bigcap_{j=1}^{n_2} \Xi_{jk_j}$ for $\boldsymbol{k} = (k_1, \ldots, k_{n_2}) \in \boldsymbol{\mathcal{K}} = \mathcal{K}_1 \times \ldots \times \mathcal{K}_{n_2}$ as well as the associated constraint index sets

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

Note that $|\boldsymbol{\mathcal{K}}| = K_{\mathrm{B}}$ and that $\bigcup_{\boldsymbol{k} \in \boldsymbol{\mathcal{K}}} \Xi_{\boldsymbol{k}}^{\mathrm{B}}$ is indeed a covering of Ξ since

$$\bigcup_{\boldsymbol{k} \in \boldsymbol{\mathcal{K}}} \Xi_{\boldsymbol{k}}^{\mathrm{B}} \;=\; \bigcup_{\boldsymbol{k} \in \boldsymbol{\mathcal{K}}} \bigcap_{j=1}^{n_2} \Xi_{jk_j} \;=\; \bigcap_{j=1}^{n_2} \bigcup_{k_j \in \mathcal{K}_j} \Xi_{jk_j} \;=\; \bigcap_{j=1}^{n_2} \Xi \;=\; \Xi.$$

Also, each constraint index set $\mathcal{I}_{\boldsymbol{k}}$ satisfies $|\mathcal{I}_{\boldsymbol{k}}| = n_2$ by construction for all $\boldsymbol{k} \in \boldsymbol{\mathcal{K}}$. Moreover, condition *(i)* in the statement of the proposition implies that for each $\boldsymbol{k} \in \boldsymbol{\mathcal{K}}$, we have

$$\Big[\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geqslant \boldsymbol{g}(\boldsymbol{\xi}) \iff \boldsymbol{A}_{\mathcal{I}_{\boldsymbol{k}}}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}_{\mathcal{I}_{\boldsymbol{k}}}\boldsymbol{y} \geqslant \boldsymbol{g}_{\mathcal{I}_{\boldsymbol{k}}}(\boldsymbol{\xi}) \Big] \qquad \forall \boldsymbol{\xi} \in \Xi_{\boldsymbol{k}}^{\mathrm{B}}.$$

Finally, similar arguments as in the proof of Proposition 2, combined with the fact that condition *(ii)* of the statement holds for every $\boldsymbol{k} \in \boldsymbol{\mathcal{K}}$, show that $\boldsymbol{B}_{\mathcal{I}_{\boldsymbol{k}}}$ has a positive inverse for each $\boldsymbol{k} \in \boldsymbol{\mathcal{K}}$. $\square$

Proof of Proposition 4. Following a similar argument as in the proof of Theorem 2, we show that for an optimal first-stage decision $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A'**), (**D**) and (**B"**) in problem (1), we can construct for every $p = 1, \ldots, P$ a collection of $K_p = K_\text{A} \cdot K_\text{B}^p$ affine decision rules $\boldsymbol{y}_{pk}^\ell : \Xi \xrightarrow{\text{a}} \mathbb{R}^{n_2}$, $k \in \mathcal{K}_p = \{1, \ldots, K_p\}$, such that

$$\sup_{\mathbb{P} \in \mathcal{P}} \rho_\mathbb{P} \left[\mathcal{Q}(\boldsymbol{x}; \tilde{\boldsymbol{\xi}}) \right] = \sup_{\mathbb{P} \in \mathcal{P}} \rho_\mathbb{P} \left[\mathcal{Q}\Big(\boldsymbol{x}, \bigtimes_{p=1}^P \{\boldsymbol{y}_{pk}^\ell(\tilde{\boldsymbol{\xi}})\}_{k=1}^{K_p}; \tilde{\boldsymbol{\xi}} \Big) \right], \tag{23}$$

where the left cost function is defined in (1b) and the right cost function satisfies

$$\mathcal{Q}\Big(\boldsymbol{x}, \bigtimes_{p=1}^P \{\boldsymbol{y}_{pk}^\ell(\boldsymbol{\xi})\}_{k=1}^{K_p}; \boldsymbol{\xi} \Big) = \min \Big\{ \mathcal{Q}(\boldsymbol{x}, (\boldsymbol{y}_{1,k_1}^\ell(\boldsymbol{\xi}), \ldots, \boldsymbol{y}_{P,k_P}^\ell(\boldsymbol{\xi})); \boldsymbol{\xi}) \, : \, \boldsymbol{k} \in \boldsymbol{\mathcal{K}} \Big\}$$

with $\mathcal{Q}$ defined in (2b), where $\boldsymbol{k} = (k_1, \ldots, k_P)$ and $\boldsymbol{\mathcal{K}} = \mathcal{K}_1 \times \ldots \times \mathcal{K}_P$. The resulting solution $(\boldsymbol{x}, \times_{p=1}^P \{\boldsymbol{y}_{pk}^\ell\}_{k=1}^{K_p})$ is thus feasible in problem (5), and a similar argument as in the proof of Theorem 2 then implies that the optimal values of both problems coincide.

To show that (23) holds, fix an optimal $\boldsymbol{x} \in \mathcal{X}$ satisfying our conditions (**F**), (**A'**), (**D**) and (**B''**), together with a collection of index sets $\mathcal{I}_{pk}^\text{B}$, $p = 1, \ldots, P$ and $k = 1, \ldots, K_\text{B}^p$, that satisfies assumption (**B''**). For every $p = 1, \ldots, P$, construct the common refinement $\Xi = \bigcup_{k=1}^{K_p} \Xi_{pk}$ with $\Xi_{p,(i-1)\cdot K_\text{B}^p + j} = \Xi_i^\text{A} \cap \Xi_{pj}^\text{B}$ as well as $\mathcal{I}_{p,(i-1)\cdot K_\text{B}^p + j} = \mathcal{I}_{pj}^\text{B}$, $i = 1, \ldots, K_\text{A}$ and $j = 1, \ldots, K_\text{B}^p$, where the sets Ξ_i^A are from assumption (**A'**). For $\boldsymbol{k} \in \boldsymbol{\mathcal{K}}$, we also define $\Xi_{\boldsymbol{k}} = \bigcap_{p=1}^P \Xi_{p,k_p}$; similar arguments as in the proof of Proposition 3 show that the sets $\Xi_{\boldsymbol{k}}$ cover Ξ, that is, $\Xi = \bigcup_{\boldsymbol{k} \in \boldsymbol{\mathcal{K}}} \Xi_{\boldsymbol{k}}$.

Similar arguments as in the proofs of Theorems 1 and 2 show that for each $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_{pk}$, $p = 1, \ldots, P$ and $k \in \mathcal{K}_p$, any feasible second-stage decision $\boldsymbol{y}_p(\boldsymbol{\xi})$ has to satisfy

$$\boldsymbol{y}_p(\boldsymbol{\xi}) \geqslant [\boldsymbol{B}_p]_{\mathcal{I}_{pk}}^{-1} \Big[[\boldsymbol{g}_p]_{\mathcal{I}_{pk}}(\boldsymbol{\xi}) - [\boldsymbol{A}_p]_{\mathcal{I}_{pk}}(\boldsymbol{\xi}) \boldsymbol{x} \Big].$$

By construction, $[\boldsymbol{A}_p]_{\mathcal{I}_{pk}}$ and $[\boldsymbol{g}_p]_{\mathcal{I}_{pk}}$ are affine over Ξ_{pk} (but not necessarily over Ξ). Fix any matrix- and vector-valued functions $[\boldsymbol{A}_p^\ell]_{\mathcal{I}_{pk}}$ and $[\boldsymbol{g}_p^\ell]_{\mathcal{I}_{pk}}$ that are affine over the entire support Ξ and that coincide with $[\boldsymbol{A}_p]_{\mathcal{I}_{pk}}$ and $[\boldsymbol{g}_p]_{\mathcal{I}_{pk}}$ over Ξ_{pk}, and consider the collection of affine decision rules

$$\boldsymbol{y}_{pk}^\ell(\boldsymbol{\xi}) = [\boldsymbol{B}_p]_{\mathcal{I}_{pk}}^{-1} \Big[[\boldsymbol{g}_p^\ell]_{\mathcal{I}_{pk}}(\boldsymbol{\xi}) - [\boldsymbol{A}_p^\ell]_{\mathcal{I}_{pk}}(\boldsymbol{\xi}) \boldsymbol{x} \Big], \qquad p = 1, \ldots, P \text{ and } k \in \mathcal{K}_p,$$

each of which is defined over the entire support Ξ. Similar arguments as in the proofs of Theorems 1 and 2 show that for each $p = 1, \ldots, P$ and $k \in \mathcal{K}_p$, $\boldsymbol{y}_{pk}^\ell(\boldsymbol{\xi})$ is the point-wise smallest feasible decision in the second-stage problem (1b) for all $\boldsymbol{\xi} \in \operatorname{dom} \mathcal{Q} \cap \Xi_{pk}$. Assumption (**F**) and the fact that $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) = \sum_{p=1}^P f_p(\boldsymbol{x}, \boldsymbol{y}_p; \boldsymbol{\xi})$ then imply that $\boldsymbol{y}_{pk}^\ell(\boldsymbol{\xi})$ is point-wise optimal in the second-stage problem (1b) over $\boldsymbol{\xi} \in \text{dom}\, \mathcal{Q} \cap \Xi_{pk}$, $p = 1, \ldots, P$ and $k \in \mathcal{K}_p$. Since $\Xi = \bigcup_{\boldsymbol{k} \in \mathcal{K}} \Xi_{\boldsymbol{k}}$, we thus have for all $\boldsymbol{k} \in \mathcal{K}$ and all $\boldsymbol{\xi} \in \text{dom}\, \mathcal{Q} \cap \Xi_{\boldsymbol{k}}$ that

$$
\begin{aligned}
\mathcal{Q}\Big(\boldsymbol{x}, \bigtimes_{p=1}^{P} \{\boldsymbol{y}^{\ell}_{pk'}(\boldsymbol{\xi})\}_{k'=1}^{K_p}; \boldsymbol{\xi}\Big) &= \min \Big\{ \mathcal{Q}(\boldsymbol{x}, (\boldsymbol{y}^{\ell}_{1,k'_1}(\boldsymbol{\xi}), \ldots, \boldsymbol{y}^{\ell}_{P,k'_P}(\boldsymbol{\xi})); \boldsymbol{\xi}) \,:\, \boldsymbol{k}' \in \mathcal{K} \Big\} \\
&\leqslant \mathcal{Q}(\boldsymbol{x}, (\boldsymbol{y}^{\ell}_{1,k_1}(\boldsymbol{\xi}), \ldots, \boldsymbol{y}^{\ell}_{P,k_P}(\boldsymbol{\xi})); \boldsymbol{\xi}) \\
&\leqslant \mathcal{Q}(\boldsymbol{x}; \boldsymbol{\xi}).
\end{aligned}
$$

Equation (23) now follows from the monotonicity of $\rho_{\mathbb{P}}$, $\mathbb{P} \in \mathcal{P}$. □

Proof of Observation 1. We employ Theorem 1 to show that problem (6) is optimally solved in affine decision rules. Observe first that affine decision rules are vacuously optimal in problem (6) if the problem is infeasible (*i.e.*, if the nodal capacity C is too small). Assume therefore that problem (6) is feasible, fix any optimal first-stage decision $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{p}^\star, \boldsymbol{c}^\star)$, let $\mathcal{T}_g = \{(i,j) \in \mathcal{N} \times \mathcal{N} \,:\, z^\star_{gij} = 1\}$ denote the distribution network for product $g \in \mathcal{G}$, and consider the following reformulation of (6b),

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{g \in \mathcal{G}} \sum_{(i,j) \in \mathcal{T}_g} t_{gij} f_{gij} & \\
\text{subject to} & \mathrm{M} \cdot x^\star_{gj} + f_{g,i(g,j),j} \geqslant \displaystyle\sum_{(j,i) \in \mathcal{T}_g} f_{gji} + \xi_{gj} & \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& \displaystyle\sum_{g \in \mathcal{G}} f_{g,i(g,j),j} \leqslant C & \forall j \in \mathcal{N} \\
& \displaystyle\sum_{\substack{g \in \mathcal{G}: \\ (i,j) \in \mathcal{T}_g}} f_{gij} \leqslant c^\star_{ij} & \forall (i,j) \in \displaystyle\bigcup_{g \in \mathcal{G}} \mathcal{T}_g \\
& f_{gij} \in \mathbb{R}_+,\ g \in \mathcal{G} \text{ and } (i,j) \in \mathcal{T}_g, &
\end{array}
\qquad \text{(6b')}
$$

where for $g \in \mathcal{G}$ and $j \in \mathcal{N}$, $i(g,j) \in \mathcal{N}$ is the index satisfying $(i(g,j),j) \in \mathcal{T}_g$. Note that $i(g,j)$ is uniquely defined for every $g \in \mathcal{G}$ and $j \in \mathcal{N}$ due to the third constraint in (6a). Problem (6b') emerges from problem (6b) if we remove the flow decisions f_{gij} that are forced to zero by the fourth constraint in (6b). Problem (6b') is equivalent to (6b) in the sense that their objective functions as well as their feasible regions coincide when projected onto the remaining flow decisions $\{f_{gij} \,:\, g \in \mathcal{G} \text{ and } (i,j) \in \mathcal{T}_g\}$.

Problem (6b') clearly satisfies the conditions (**F**), (**A**) and (**D**). To see that condition (**B**) is satisfied as well, we invoke Proposition 2. To this end, fix any $g \in \mathcal{G}$ and $(i,j) \in \mathcal{T}_g$, and note that $i = i(g,j)$. As for condition *(i)* of Proposition 2, note that the decision f_{gij} participates in exactly two lower bound constraints:

$$f_{gij} \geqslant \sum_{(j,l)\in\mathcal{T}_g} f_{gjl} + \xi_{gj} - \mathrm{M} \cdot x^{\star}_{gj} \qquad \text{and} \qquad f_{gij} \in \mathbb{R}_+$$

If $x^{\star}_{gj} = 1$, then the non-negativity constraint is weakly dominant for all realizations of $\boldsymbol{\xi}$; otherwise, if $x^{\star}_{gj} = 0$, then the flow constraint is weakly dominant for all realizations of $\boldsymbol{\xi}$ since the flow decisions f_{gjl} on the right-hand side as well as the demand ξ_{gj} are guaranteed to be non-negative. In view of condition *(ii)* of Proposition 2, observe that the first constraint set of problem (6b') has an arborescence structure thanks to $\mathcal{T}_g$, which implies that their associated coefficient vectors admit a partial order. □

The proof of Proposition 5 utilizes the following technical result, which we state and prove first.

Lemma 1. *Problem* (1) *satisfies the following weak duality realationship:*

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\big] \;\; \geqslant \;\; \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\min_{\boldsymbol{x}\in\mathcal{X}} \mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\right]. \tag{24}$$

Proof of Lemma 1. Note that for all $\boldsymbol{x} \in \mathcal{X}$ and all $\mathbb{P} \in \mathcal{P}$, we have

$$\rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\big] \;\; \geqslant \;\; \min_{\boldsymbol{x}'\in\mathcal{X}} \left\{\rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x}';\tilde{\boldsymbol{\xi}})\big]\right\} \;\; \geqslant \;\; \rho_{\mathbb{P}}\left[\min_{\boldsymbol{x}'\in\mathcal{X}} \mathcal{Q}(\boldsymbol{x}';\tilde{\boldsymbol{\xi}})\right],$$

where the first inequality is immediate and the second one holds since $\mathcal{Q}(\boldsymbol{x}';\boldsymbol{\xi}) \geqslant \min_{\boldsymbol{x}''\in\mathcal{X}} \mathcal{Q}(\boldsymbol{x}'';\boldsymbol{\xi})$ for all realizations $\boldsymbol{\xi}$ and the assumed monotonicity of $\rho_{\mathbb{P}}$. For all $\boldsymbol{x} \in \mathcal{X}$, we thus have

$$\sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\big[\mathcal{Q}(\boldsymbol{x};\tilde{\boldsymbol{\xi}})\big] \;\; \geqslant \;\; \sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}}\left[\min_{\boldsymbol{x}'\in\mathcal{X}} \mathcal{Q}(\boldsymbol{x}';\tilde{\boldsymbol{\xi}})\right]$$

and therefore, as desired, (24). □

Proof of Proposition 5. Affine decision rules are vacuously optimal if the unrestricted version of problem (6) is infeasible (*i.e.*, if the nodal capacity C is too small). Assume therefore that the problem is feasible, and fix the facility location components $(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star})$ of any optimal first-stage decision $(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star}, \boldsymbol{z}^{\star}, \boldsymbol{c}^{\star})$ in the unrestricted version of problem (6). We show that $(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star})$ can be complemented with first-stage decisions $(\boldsymbol{z}', \boldsymbol{c}')$ such that there is an optimal policy $\boldsymbol{f}^{\star} : \Xi \rightarrow \mathbb{R}_+^{GN^2}$ for the second-stage decision $\boldsymbol{f}$ in (6b) that exhibits an affine dependence on $\boldsymbol{\xi}$. To simplify the exposition, we define

$$\mathcal{F}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{c};\boldsymbol{\xi}) \;=\; \left\{ \boldsymbol{f} \in \mathbb{R}_+^{GN^2} \;:\; \left[\begin{array}{ll} \mathrm{M}\cdot x_{gj} + \sum\limits_{i\in\mathcal{N}} f_{gij} \geqslant \sum\limits_{i\in\mathcal{N}} f_{gji} + \xi_{gj} & \forall g\in\mathcal{G},\ \forall j\in\mathcal{N} \\ \sum\limits_{g\in\mathcal{G}}\sum\limits_{i\in\mathcal{N}} f_{gij} \leqslant C & \forall j\in\mathcal{N} \\ \sum\limits_{g\in\mathcal{G}} f_{gij} \leqslant c_{ij} & \forall i,j\in\mathcal{N} \\ f_{gij} \leqslant \mathrm{M}\cdot z_{gij} & \forall g\in\mathcal{G},\ \forall i,j\in\mathcal{N} \end{array}\right]\right\}$$

as the set of all second-stage flow decisions $\boldsymbol{f}$ that are feasible in (6b) under the parameter realization $\boldsymbol{\xi}$ for the fixed first-stage decisions $\boldsymbol{x}$, $\boldsymbol{z}$ and $\boldsymbol{c}$.

Define $\overline{\boldsymbol{\xi}} \in \mathbb{R}^{GN}$ component-wise via $\overline{\xi}_{gj} = \max\{\xi_{gj} \,:\, \boldsymbol{\xi}\in\Xi\}$, $g\in\mathcal{G}$ and $j\in\mathcal{N}$, and observe that $\overline{\boldsymbol{\xi}}\in\Xi$ due to assumption *(i)* of the proposition. The feasible region of the unrestricted version of problem (6) does not change if we include in its first stage (6a) the constraint that there needs to be $\overline{\boldsymbol{f}} \in \mathcal{F}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{c};\overline{\boldsymbol{\xi}})$; indeed, assumption *(ii)* ensures that this constraint is already implied by the second stage (6b).

Next, we observe that due to Lemma 1, the following problem bounds the optimal value of the unrestricted version of problem (6) with fixed facility locations $(\boldsymbol{x}^\star,\boldsymbol{y}^\star)$ from below:

$$\sup_{\mathbb{P}\in\mathcal{P}} \rho_{\mathbb{P}} \left[\begin{array}{lll} \text{minimize} & \sum\limits_{i,j\in\mathcal{N}} b_{ij}c_{ij} + \sum\limits_{g\in\mathcal{G}}\sum\limits_{i,j\in\mathcal{N}} t_{gij}f_{gij} & \\ \text{subject to} & z_{gij} \leqslant x^\star_{gi} + y^\star_i & \forall g\in\mathcal{G},\ \forall i,j\in\mathcal{N} \\ & \boldsymbol{f}\in\mathcal{F}(\boldsymbol{x}^\star,\boldsymbol{z},\boldsymbol{c};\tilde{\boldsymbol{\xi}}),\ \ \overline{\boldsymbol{f}}\in\mathcal{F}(\boldsymbol{x}^\star,\boldsymbol{z},\boldsymbol{c};\overline{\boldsymbol{\xi}}) & \\ & \boldsymbol{z}\in\{0,1\}^{GN^2},\ \ \boldsymbol{c}\in\mathbb{R}_+^{N^2},\ \ \boldsymbol{f},\overline{\boldsymbol{f}}\in\mathbb{R}_+^{GN^2} & \end{array}\right] \tag{25}$$

We show that there is $(\boldsymbol{z}',\boldsymbol{c}',\overline{\boldsymbol{f}}')$ that optimizes the minimization problem in (25) simultaneously for every $\boldsymbol{\xi}\in\Xi$ if supplemented with a suitable $\boldsymbol{f}'$ that may itself depend on $\boldsymbol{\xi}\in\Xi$. It then follows that $(\boldsymbol{x}^\star,\boldsymbol{y}^\star,\boldsymbol{z}',\boldsymbol{c}')$ is also optimal in the unrestricted version of problem (6). Moreover, our solution will satisfy $p'_{gj} \geqslant p'_{gi} + 1 - \mathrm{M}\cdot(1-z'_{gij})$ and $\sum_{i\in\mathcal{N}} z'_{gij} \leqslant 1$ for all $g\in\mathcal{G}$ and $i,j\in\mathcal{N}$ if supplemented with a suitable $\boldsymbol{p}'\in\mathbb{R}_+^{GN}$, which implies that these constraints can be added to the unrestricted version of problem (6) without impacting its optimal value. Using the same arguments as in the proof of Observation 1, we can then conclude that affine decision rules are optimal.

By the superadditivity of the minimum operator, the optimal value of the minimization problem in (25) is bounded from below for each realization $\boldsymbol{\xi} \in \Xi$ by the optimal value of the problem

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{i,j \in \mathcal{N}} b_{ij} c_{ij} & \\
\text{subject to} & z_{gij} \leqslant x^\star_{gi} + y^\star_i & \forall g \in \mathcal{G},\ \forall i,j \in \mathcal{N} \\
& \boldsymbol{f} \in \mathcal{F}(\boldsymbol{x}^\star, \boldsymbol{z}, \boldsymbol{c}; \boldsymbol{\xi}),\ \ \overline{\boldsymbol{f}} \in \mathcal{F}(\boldsymbol{x}^\star, \boldsymbol{z}, \boldsymbol{c}; \overline{\boldsymbol{\xi}}) & \\
& \boldsymbol{z} \in \{0,1\}^{GN^2},\ \ \boldsymbol{c} \in \mathbb{R}^{N^2}_+,\ \ \boldsymbol{f}, \overline{\boldsymbol{f}} \in \mathbb{R}^{GN^2}_+ &
\end{array}
\tag{26}
$$

plus the optimal values of the G problems, one for each good $g \in \mathcal{G}$,

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{i,j \in \mathcal{N}} t_{gij} f_{gij} & \\
\text{subject to} & z_{gij} \leqslant x^\star_{gi} + y^\star_i & \forall g \in \mathcal{G},\ \forall i,j \in \mathcal{N} \\
& \boldsymbol{f} \in \mathcal{F}(\boldsymbol{x}^\star, \boldsymbol{z}, \boldsymbol{c}; \boldsymbol{\xi}),\ \ \overline{\boldsymbol{f}} \in \mathcal{F}(\boldsymbol{x}^\star, \boldsymbol{z}, \boldsymbol{c}; \overline{\boldsymbol{\xi}}) & \\
& \boldsymbol{z} \in \{0,1\}^{GN^2},\ \ \boldsymbol{c} \in \mathbb{R}^{N^2}_+,\ \ \boldsymbol{f}, \overline{\boldsymbol{f}} \in \mathbb{R}^{GN^2}_+. &
\end{array}
\tag{27}
$$

We can replace each network decision z_{gij} in problem (26) with its maximum feasible value $\min\{x^\star_{gi} + y^\star_i, 1\}$ and subsequently remove it by restricting all flows f_{gij} and $\overline{f}_{gij}$ to the arcs in $\mathcal{E}_g = \{(i,j) \in \mathcal{N} \times \mathcal{N} \,:\, x^\star_{gi} = 1 \text{ or } y^\star_i = 1\}$, $g \in \mathcal{G}$. Likewise, we can identify each flow f_{gij} in (26) with $\overline{f}_{gij}$, and we can replace each capacity c_{ij} with its lower bound $\sum_{g \in \mathcal{G}} \overline{f}_{gij}$. With those changes, problem (26) simplifies to

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{g \in \mathcal{G}} \sum_{(i,j) \in \mathcal{E}_g} b_{ij} \overline{f}_{gij} & \\
\text{subject to} & \mathrm{M} \cdot x^\star_{gj} + \displaystyle\sum_{(i,j) \in \mathcal{E}_g} \overline{f}_{gij} \geqslant \sum_{(j,i) \in \mathcal{E}_g} \overline{f}_{gji} + \overline{\xi}_{gj} & \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& \displaystyle\sum_{g \in \mathcal{G}} \sum_{(i,j) \in \mathcal{E}_g} f_{gij} \leqslant C & \forall j \in \mathcal{N} \\
& \overline{f}_{gij} \in \mathbb{R}_+,\ g \in \mathcal{G} \text{ and } (i,j) \in \mathcal{E}_g. &
\end{array}
$$

For each $g \in \mathcal{G}$, let $\mathcal{T}_g$ be any minimum cost arborescence on the directed graph $(\mathcal{N}, \mathcal{E}_g)$ with root $i \in \mathcal{I}$ satisfying $x^\star_i = 1$ and arc weights $\boldsymbol{b}$. Then the simplified problem is optimally solved by $\overline{\boldsymbol{f}}'$ where each $\overline{f}'_{gij}$ coincides with the sum of those demands $\overline{\xi}_{gl}$ for which $\mathcal{T}_g$ contains a directed path from i to l via j.

Similarly, in each instance $g \in \mathcal{G}$ of problem (27) we can remove the network decisions z_{gij} as in the previous paragraph, identify each flow $\overline{f}_{gij}$ with $\overline{f}'_{gij}$ from the previous paragraph and set

$c_{ij} = \sum_{g \in \mathcal{G}} \overline{f}'_{gij}$ as before. With those changes, problem (27) simplifies to

$$\begin{array}{lll} \text{minimize} & \tau_g \cdot \displaystyle\sum_{(i,j) \in \mathcal{E}_g} b_{ij} f_{gij} & \\ \text{subject to} & \mathrm{M} \cdot x^\star_{gj} + \displaystyle\sum_{(i,j) \in \mathcal{E}_g} f_{gij} \geqslant \sum_{(j,i) \in \mathcal{E}_g} f_{gji} + \xi_{gj} & \forall j \in \mathcal{N} \\ & \displaystyle\sum_{g \in \mathcal{G}} \sum_{(i,j) \in \mathcal{E}_g} f_{gij} \leqslant C & \forall j \in \mathcal{N} \\ & f_{gij} \in \mathbb{R}_+, \; (i,j) \in \mathcal{E}_g. & \end{array}$$

The simplified problem is optimally solved by $\boldsymbol{f}'$ where each f'_{gij} coincides with the sum of those demands ξ_{gl} for which the arborescence $\mathcal{T}_g$ from the previous paragraph contains a directed path from i to l via j. Indeed, this is the case due to assumption *(iii)*, which ensures that the minimum cost arborescence problems underlying (26) and (27) share the same arc weights up to a constant scaling.

Consider now the solution $(\boldsymbol{z}', \boldsymbol{c}', \boldsymbol{f}', \overline{\boldsymbol{f}}')$ to the lower bound problem (25) where $z'_{gij} = 1$ if $(i,j) \in \mathcal{E}_g$ and $\overline{f}'_{gij} > 0$; $= 0$ otherwise, $c'_{ij} = \sum_{g \in \mathcal{G}} \overline{f}_{gij}$ and $\boldsymbol{f}'$ and $\overline{\boldsymbol{f}}'$ are defined as in the preceding two paragraphs. This solution is optimal in both (26) and (27), and it must therefore also optimize the minimization problem in (25) with the same optimal value. Furthermore, this solution satisfies the tree constraints $\sum_{i \in \mathcal{N}} z'_{gij} \leqslant 1$ by construction, and it can be readily complemented with potential values $\boldsymbol{p}' \in \mathbb{R}^{GN}_+$ that satisfy $p'_{gj} \geqslant p'_{gi} + 1 - \mathrm{M} \cdot (1 - z'_{gij})$ for all $g \in \mathcal{G}$ and all $i, j \in \mathcal{N}$. The statement now follows from the fact that neither $\boldsymbol{c}'$ nor $\boldsymbol{z}'$, $\boldsymbol{p}'$ or $\overline{\boldsymbol{f}}'$ depend on the actual parameter realization $\boldsymbol{\xi} \in \Xi$. □

Proof of Proposition 6. Denote by OPT the optimal value of the unrestricted supply chain management problem across all feasible second-stage policies $\boldsymbol{f} : \Xi \rightarrow \mathbb{R}^{GN^2}_+$, and let ADR denote the optimal value of the unrestricted supply chain management problem where the second-stage policies $\boldsymbol{f} : \Xi \xrightarrow{\text{a}} \mathbb{R}^{GN^2}_+$ are restricted to be affine in the uncertain demands $\boldsymbol{\xi} \in \Xi$. Note that

$$\text{OPT} \;\geqslant\; \left[\min_{i,j \in \mathcal{N}} \; b_{ij} \right] \cdot \max_{\boldsymbol{\xi} \in \Xi} \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \xi_{gj}$$

since any feasible solution to the unrestricted supply chain management problem must construct sufficient transportation capacity to serve the maximum demand, thanks to our assumption that the employed risk measure satisfies the condition of Remark 2, namely that $\sup_{\mathbb{P} \in \mathcal{P}} \; \rho_{\mathbb{P}} \big[\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{f}(\tilde{\boldsymbol{\xi}}); \tilde{\boldsymbol{\xi}}) \big] =$

∞ whenever $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{f}(\boldsymbol{\xi}); \boldsymbol{\xi}) = \infty$ for some $\boldsymbol{\xi} \in \Xi$. In the remainder of the proof, we will show that

$$\text{ADR} - \text{OPT} \;\leqslant\; \eta \cdot \left[\max_{g \in \mathcal{G}} \max_{i,j \in \mathcal{N}} \; G \cdot b_{ij} + 2t_{gij} \right] \cdot C, \tag{28}$$

where η bounds from above the number of arcs in the distribution network that exceed the number of arcs $G(N-1)$ required to form G arborescences. The statement of the proposition then follows since

$$\frac{\text{ADR} - \text{OPT}}{\text{OPT}} \;\leqslant\; \frac{\eta \cdot \left[\max_{g \in \mathcal{G}} \max_{i,j \in \mathcal{N}} \; G \cdot b_{ij} + 2t_{gij} \right] \cdot C}{\left[\min_{i,j \in \mathcal{N}} \; b_{ij} \right] \cdot \max_{\boldsymbol{\xi} \in \Xi} \; \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \xi_{gj}} \;\propto\; \frac{\eta \cdot G \cdot C}{\max_{\boldsymbol{\xi} \in \Xi} \; \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{N}} \xi_{gj}},$$

where we exploited that b_{ij} and t_{gij} are regarded as constants and that $b_{ij} > 0$ for all $i, j \in \mathcal{N}$.

To see that equation (28) holds, fix any first-stage decision $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{c}^\star)$ feasible in the unrestricted variant of problem (6a), together with any associated second-stage policy $\boldsymbol{f}^\star : \Xi \to \mathbb{R}_+^{GN^2}$ feasible in (6b), that attain the optimal value OPT in the unrestricted variant of (6) and whose distribution network $\boldsymbol{z}^\star$ contains at most $G(N-1) + \eta$ arcs $z^\star_{gij} = 1$, $(g, i, j) \in \mathcal{G} \times \mathcal{N}^2$. In the following, we will construct a new first-stage decision $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', \boldsymbol{p}', \boldsymbol{c}')$ feasible in the *restricted* problem (6a), together with an associated second-stage policy $\boldsymbol{f}' : \Xi \to \mathbb{R}_+^{GN^2}$ feasible in (6b), that attain an objective value in (6) that does not exceed OPT by more than the right-hand side of equation (28). Since we know from Observation 1 that the restricted problem (6) is optimally solved in affine decision rules, there must be an affine decision rule solution to the restricted problem (6) whose objective value does not exceed OPT by more than the right-hand side of equation (28) either. Since the unrestricted variant of problem (6) constitutes a relaxation of the restricted problem (6), we conclude that any optimal affine decision rule solution to the *unrestricted* variant of problem (6) must attain an objective value not exceeding OPT by more than the right-hand side of equation (28) either. This will then verify the validity of equation (28) and thus conclude the proof.

To construct the desired solution $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', \boldsymbol{p}', \boldsymbol{c}'; \boldsymbol{f}')$, denote for each good $g \in \mathcal{G}$ by $\mathcal{N}_g = \{j \in \mathcal{N} : \sum_{i \in \mathcal{N}} z^\star_{gij} > 1\}$ the set of nodes $j \in \mathcal{N}$ that possess multiple incoming links in the distribution network of the optimal solution. Note that $\sum_{g \in \mathcal{G}} |\mathcal{N}_g| \leqslant \eta$ by construction. Our goal is to construct the revised solution $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', \boldsymbol{p}', \boldsymbol{c}'; \boldsymbol{f}')$ such that for each good $g \in \mathcal{G}$, each node in $\mathcal{N}_g$ has a single incoming link in the distribution network. To this end, we set $\boldsymbol{x}' = \boldsymbol{x}^\star$ and $\boldsymbol{y}' = \boldsymbol{y}^\star$ as well as

$$z'_{gij} = \begin{cases} x_{gi} & \text{if } j \in \mathcal{N}_g, \\ z^\star_{gij} & \text{otherwise,} \end{cases} \qquad c'_{ij} = \begin{cases} \sum_{k \in \mathcal{N}} c^\star_{kj} & \text{if } j \in \mathcal{N}_g \text{ for some } g \in \mathcal{G} \text{ with } x_{gi} = 1, \\ c^\star_{ij} & \text{otherwise} \end{cases}$$

and

$$f'_{gij}(\boldsymbol{\xi}) = \begin{cases} x_{gi} \cdot \sum_{k \in \mathcal{N}} f^\star_{gkj}(\boldsymbol{\xi}) & \text{if } j \in \mathcal{N}_g, \\ f^\star_{gij}(\boldsymbol{\xi}) & \text{otherwise} \end{cases} \qquad \forall \boldsymbol{\xi} \in \Xi.$$

Thus, the revised solution retains the original placement of the production facilities and warehouses, but it removes those links $z^\star_{gij} = 1$, $(g, i, j) \in \mathcal{G} \times \mathcal{N}^2$, from the distribution network where the successor node j has multiple predecessors for the same good g (that is, $j \in \mathcal{N}_g$) and where the particular predecessor i is not a production facility for good g (that is, $x_{gi} = 0$). The capacities of and the flows along the remaining links are adjusted so that any demands are served from the new distribution network. We defer the construction of node potentials $\boldsymbol{p}'$ that satisfy the fourth constraint in (6a) to the next paragraph.

We now show that the revised solution $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', \boldsymbol{p}', \boldsymbol{c}'; \boldsymbol{f}')$ is feasible in the restricted problem (6). To this end, note that the first two constraints in (6a) only involve $\boldsymbol{x}'$ and $\boldsymbol{y}'$, and that the values of these decisions have not changed. The third constraint in (6a) holds thanks to our construction of $\boldsymbol{z}'$: For pairs $(g, j) \in \mathcal{G} \times \mathcal{N}$ with $j \in \mathcal{N}_g$, we have $\sum_{i \in \mathcal{N}} z'_{gij} = \sum_{i \in \mathcal{N}} x_{gi} = 1$ thanks to the first constraint in (6a), whereas for pairs $(g, j) \in \mathcal{G} \times \mathcal{N}$ with $j \notin \mathcal{N}_g$, the definition of $\mathcal{N}_g$ implies that $\sum_{i \in \mathcal{N}} z'_{gij} = \sum_{i \in \mathcal{N}} z^\star_{gij} \leqslant 1$. In view of the fourth constraint in (6a), we note that satisfaction of the third constraint implies that the distribution network $(z'_{gij})_{i,j}$ for each good $g \in \mathcal{G}$ forms a directed acyclic graph. We can thus identify $\boldsymbol{p}'$ with any partial order on $\mathbb{N}_0$ that is compatible with this directed acyclic graph. As for the last constraint set in (6a), finally, note that $z'_{gij} \leqslant z^\star_{gij}$ unless $j \in \mathcal{N}_g$ and $x_{gi} = 1$, and $z^\star_{gij}$ satisfies the constraint by assumption. If $x_{gi} = 1$, on the other hand, the constraint is vacuously satisfied. To see that the second-stage constraints hold for all $\boldsymbol{\xi} \in \Xi$, note that the sum on the left-hand side of the first constraint in (6b) can only change for pairs $(g, j) \in \mathcal{G} \times \mathcal{N}$ satisfying $j \in \mathcal{N}_g$, and for such pairs we have

$$\sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) \;=\; \sum_{i \in \mathcal{N}} x_{gi} \cdot \sum_{k \in \mathcal{N}} f^\star_{gkj}(\boldsymbol{\xi}) \;=\; \sum_{k \in \mathcal{N}} f^\star_{gkj}(\boldsymbol{\xi}).$$

On the other hand, the sum on the right-hand side of the first constraint in (6b) evaluates to

$$\sum_{i\in\mathcal{N}} f'_{gji}(\boldsymbol{\xi}) \;=\; \sum_{i\in\mathcal{N}:i\in\mathcal{N}_g} f'_{gji}(\boldsymbol{\xi}) \;+\; \sum_{i\in\mathcal{N}:i\notin\mathcal{N}_g} f'_{gji}(\boldsymbol{\xi}) \;=\; \sum_{i\in\mathcal{N}:i\in\mathcal{N}_g} x_{gj}\cdot\sum_{k\in\mathcal{N}} f^{\star}_{gki}(\boldsymbol{\xi}) \;+\; \sum_{i\in\mathcal{N}:i\notin\mathcal{N}_g} f^{\star}_{gji}(\boldsymbol{\xi})$$

$$\leqslant\; \begin{cases} \sum_{i\in\mathcal{N}} f^{\star}_{gji}(\boldsymbol{\xi}) & \text{if } x_{gj}=0,\\ \mathrm{M} & \text{if } x_{gj}=1. \end{cases}$$

Thus, if $x_{gj}=0$, then the right-hand side of the first constraint in (6b) has either decreased or remained unchanged, whereas the left-hand side of the same constraint evaluates to at least M if $x_{gj}=1$. Our earlier finding that $\sum_{i\in\mathcal{N}} f'_{gij}(\boldsymbol{\xi}) = \sum_{k\in\mathcal{N}} f^{\star}_{gkj}(\boldsymbol{\xi})$ for all pairs $(g,j)\in\mathcal{G}\times\mathcal{N}$ immediately implies that the second constraint in (6b) is satisfied as well. As for the third constraint in (6b), we note that the left-hand side can only increase for node pairs $(i,j)\in\mathcal{N}^2$ satisfying $j\in\mathcal{N}_g$ for some $g\in\mathcal{G}$ with $x_{gi}=1$, in which case our construction of c'_{ij} ensures that the right-hand side has increased sufficiently to cater for the additional flow. In view of the last constraint in (6b), finally, we observe that the left-hand side can only increase for triples $(g,i,j)\in\mathcal{G}\times\mathcal{N}^2$ satisfying $j\in\mathcal{N}_g$ with $x_{gi}=1$, in which case we have $z'_{gij}=1$.

To conclude the proof, we show that the objective value of the revised solution $(\boldsymbol{x}',\boldsymbol{y}',\boldsymbol{z}',\boldsymbol{p}',\boldsymbol{c}';\boldsymbol{f}')$ does not exceed OPT by more than the right-hand side of equation (28). To this end, observe first that

$$\begin{aligned}&\left(\sum_{i,j\in\mathcal{N}} b_{ij}c'_{ij} + \sup_{\mathbb{P}\in\mathcal{P}}\ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f'_{gij}(\tilde{\boldsymbol{\xi}})\right]\right) \;-\; \left(\sum_{i,j\in\mathcal{N}} b_{ij}c^{\star}_{ij} + \sup_{\mathbb{P}\in\mathcal{P}}\ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right]\right)\\ =\;&\left(\sum_{i,j\in\mathcal{N}} b_{ij}\left[c'_{ij}-c^{\star}_{ij}\right]\right) \;+\; \left(\sup_{\mathbb{P}\in\mathcal{P}}\ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f'_{gij}(\tilde{\boldsymbol{\xi}})\right] \;-\; \sup_{\mathbb{P}\in\mathcal{P}}\ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right]\right).\end{aligned} \tag{29}$$

The first expression in (29) can be bounded from above via

$$\sum_{i,j\in\mathcal{N}} b_{ij}\left[c'_{ij}-c^{\star}_{ij}\right] \;\leqslant\; \left[\max_{i,j\in\mathcal{N}} b_{ij}\right]\cdot\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}_g}\sum_{k\in\mathcal{N}} c^{\star}_{kj} \;\leqslant\; \left[\max_{i,j\in\mathcal{N}} b_{ij}\right]\cdot\eta\cdot G\cdot C, \tag{30}$$

where the first inequality follows from the construction of $\boldsymbol{c}'$ and the fact that there is a unique production facility for each good. The second inequality is due to the fact that $\sum_{g\in\mathcal{G}}|\mathcal{N}_g|\leqslant\eta$ and $\sum_{k\in\mathcal{N}} c^{\star}_{kj}\leqslant G\cdot C$ for all $j\in\mathcal{N}$. Indeed, (6b) ensures that $\sum_{k\in\mathcal{N}} f^{\star}_{gkj}(\boldsymbol{\xi})\leqslant C$ for all $g\in\mathcal{G}$, $j\in\mathcal{N}$ and $\boldsymbol{\xi}\in\Xi$, and thus no optimal solution will invest in a cumulative incoming link capacity $\sum_{k\in\mathcal{N}} c^{\star}_{kj}$ exceeding $G\cdot C$ since all capacity costs $\{b_{kj}\}_k$ are strictly positive.

To upper bound the second expression in (29), we first observe that

$$\begin{aligned}&\sup_{\mathbb{P}\in\mathcal{P}} \ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f'_{gij}(\tilde{\boldsymbol{\xi}})\right] \ - \ \sup_{\mathbb{P}\in\mathcal{P}} \ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] \\ \leqslant \ &\sup_{\mathbb{P}\in\mathcal{P}}\left(\rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f'_{gij}(\tilde{\boldsymbol{\xi}})\right] \ - \ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right]\right).\end{aligned} \tag{31}$$

Indeed, fix any $\mathbb{P}'$ and $\mathbb{P}^{\star}$ that are ϵ-optimal in the first and second supremum on the left-hand side of the inequality, respectively, and note that

$$\rho_{\mathbb{P}'}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] - \rho_{\mathbb{P}^{\star}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] \ \leqslant \ \rho_{\mathbb{P}'}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] - \rho_{\mathbb{P}'}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] + \epsilon.$$

Since ϵ was chosen arbitrarily, (31) follows. Next observe that the second-stage transshipment costs satisfy

$$\left|\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f'_{gij}(\boldsymbol{\xi}) \ - \ \sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\boldsymbol{\xi})\right| \ \leqslant \ \eta\cdot\left[\max_{g\in\mathcal{G}}\max_{i,j\in\mathcal{N}} \ t_{gij}\right]\cdot C \qquad \forall\boldsymbol{\xi}\in\Xi$$

since

$$\left|\left\{(g,i,j)\in\mathcal{G}\times\mathcal{N}^2 \ : \ f'_{gij}(\boldsymbol{\xi}) \neq f^{\star}_{gij}(\boldsymbol{\xi}) \text{ for some } \boldsymbol{\xi}\in\Xi\right\}\right| \ \leqslant \ \sum_{g\in\mathcal{G}}|\mathcal{N}_g| \ \leqslant \ \eta$$

and $|f'_{gij}(\boldsymbol{\xi}) - f^{\star}_{gij}(\boldsymbol{\xi})| \leqslant C$ for all $g\in\mathcal{G}$, $i,j\in\mathcal{N}$ and $\boldsymbol{\xi}\in\Xi$ thanks to the construction of $\boldsymbol{f}'$ as well as the second constraint in (6b). Next, we note that any risk measure $\rho_{\mathbb{P}}$, $\mathbb{P}\in\mathcal{P}$, satisfies $|\rho_{\mathbb{P}}(\tilde{X}) - \rho_{\mathbb{P}}(\tilde{Y})| \leqslant 2\zeta$ for any random variables $\tilde{X},\tilde{Y}$ satisfying $|\tilde{X}-\tilde{Y}| \leqslant \zeta$ $\mathbb{P}$-a.s. since

$$\rho_{\mathbb{P}}(\tilde{Y}) \ \leqslant \ \rho_{\mathbb{P}}(\tilde{X}+\zeta) \ = \ \rho_{\mathbb{P}}(\tilde{X})+\zeta \quad \text{and} \quad \rho_{\mathbb{P}}(\tilde{X}) \ \leqslant \ \rho_{\mathbb{P}}(\tilde{Y}+\zeta) \ = \ \rho_{\mathbb{P}}(\tilde{Y})+\zeta,$$

where the inequalities and equalities follow from the assumed monotonicity (from Section 2) and translation invariance (from the statement of the proposition) of ρ, respectively. In conclusion, we observe that

$$\sup_{\mathbb{P}\in\mathcal{P}}\left(\rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right] \ - \ \rho_{\mathbb{P}}\left[\sum_{g\in\mathcal{G}}\sum_{i,j\in\mathcal{N}} t_{gij} f^{\star}_{gij}(\tilde{\boldsymbol{\xi}})\right]\right) \ \leqslant \ 2\eta\cdot\left[\max_{g\in\mathcal{G}}\max_{i,j\in\mathcal{N}} \ t_{gij}\right]\cdot C.$$

Combining this finding with (29) and (30) and comparing the resulting expression with the right-hand side of equation (28), the statement follows. □

Proof of Observation 2. We make the following observations:

(i) Any feasible solution $(\boldsymbol{f}, \boldsymbol{l})$ to the first stage problem (6a) with second stage (7) is also feasible in the first stage problem (6a) with second stage (7') with the same objective value. Indeed, the constraints in (7) imply those of (7') since they additionally require non-negativity of $\boldsymbol{l}$, and for all those solutions the objective functions of (7) and (7') coincide.

(ii) If the feasible solution $(\boldsymbol{f}, \boldsymbol{l})$ to the first stage problem (6a) with second stage (7) in point *(i)* has affine flows $\boldsymbol{f}$, then we can replace $\boldsymbol{l}$ in (7') with $l'_{gj}(\boldsymbol{\xi}) = \mathrm{M} \cdot x_{gj} + \sum_{i \in \mathcal{N}} f_{gij} - \sum_{i \in \mathcal{N}} f_{gji} + \xi_{gj}$, which is affine. The revised decision $\boldsymbol{l}'$ is weakly smaller than $\boldsymbol{l}$ point-wise, which implies that the objective value in (7') does not increase if we replace $\boldsymbol{l}$ with $\boldsymbol{l}'$.

(iii) Any feasible solution $(\boldsymbol{f}, \boldsymbol{l})$ in the first stage problem (6a) with second stage (7') gives rise to a feasible solution in the first stage problem (6a) with second stage (7) with the same objective value if we replace $\boldsymbol{l}$ with $l'_{gj}(\boldsymbol{\xi}) = [l_{gj}(\boldsymbol{\xi})]_+$. Indeed, the revised decision $\boldsymbol{l}'$ is weakly larger than $\boldsymbol{l}$ point-wise, which implies that the flow constraints in (7) remain satisfied. The revised decision $\boldsymbol{l}'$ is also point-wise non-negative, which implies that the last constraint in (7) is satisfied. Finally, the objective value does not change since the objective function of (7') evaluates $[l_{gj}(\boldsymbol{\xi})]_+$ by construction.

Our first claim from the statement of the observation now follows from observations *(i)* and *(iii)*, the second claim follows from observation *(ii)*, and the third claim follows from observation *(iii)*. □

Proof of Theorem 3. Observation 2 allows us to equivalently study the suboptimality of affine flow decisions in the revised supply chain design problem with first stage (6a) and second stage (7), where the affine flow decisions can be complemented with any (possibly nonlinear) lost sales decisions.

Fix an optimal solution $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{p}^\star, \boldsymbol{c}^\star)$ to the revised supply chain design problem with first stage (6a) and second stage (7), and fix an associated optimal policy $(\boldsymbol{f}^\star, \boldsymbol{l}^\star)$, $\boldsymbol{f}^\star : \Xi \to \mathbb{R}_+^{GN^2}$ and $\boldsymbol{l}^\star : \Xi \to \mathbb{R}_+^{GN}$, to the second-stage problem (7). Define $\mathcal{D}(g, j)$ as the set of all immediate and transitive successors of node $j \in \mathcal{G}$ in the arborescence $\boldsymbol{z}_g^\star$ associated with product $g \in \mathcal{G}$, and let $\mathcal{S}_{gj}(\boldsymbol{\xi}) = \xi_{gj} + \sum_{k \in \mathcal{D}(g,j)} \xi_{gk}$ denote the downstream demands for product g from node j onwards. We suppress the dependence of $\mathcal{D}$ and $\mathcal{S}$ on $\boldsymbol{z}^\star$ to simplify the notation. Consider the second-stage policy $(\boldsymbol{f}', \boldsymbol{l}')$ with affine flows $\boldsymbol{f}' : \Xi \xrightarrow{\text{a}} \mathbb{R}_+^{GN^2}$ and nonlinear lost sales $\boldsymbol{l}' : \Xi \to \mathbb{R}_+^{GN}$ defined via

$$f'_{gij}(\boldsymbol{\xi}) \;=\; \begin{cases} \dfrac{\mathcal{S}_{gj}(\boldsymbol{\xi})}{\max\limits_{\boldsymbol{\xi} \in \Xi} \sum\limits_{g \in \mathcal{G}} \mathcal{S}_{gj}(\boldsymbol{\xi})} \cdot c^\star_{ij} & \text{if } z^\star_{gij} = 1, \\ 0 & \text{otherwise} \end{cases}$$

as well as

$$l'_{gj}(\boldsymbol{\xi}) \;=\; \begin{cases} \left[\xi_{gj} + \sum\limits_{i \in \mathcal{N}} f'_{gji}(\boldsymbol{\xi}) - \sum\limits_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) \right]_+ & \text{if } x^\star_{gj} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

We first claim that the first-stage decision $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{p}^\star, \boldsymbol{c}^\star)$ and the second-stage policy $(\boldsymbol{f}', \boldsymbol{l}')$ are feasible in the revised supply chain design problem with first stage (6a) and second stage (7). To this end, we verify satisfaction of the second-stage constraints (7) one by one. Constraints in the first constraint set are vacuously satisfied whenever $x^\star_{gj} = 1$. When $x^\star_{gj} = 0$, the constraint left-hand side evaluates to

$$\begin{aligned} \sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) + l'_{gj}(\boldsymbol{\xi}) \;&=\; \sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) + \left[\xi_{gj} + \sum_{i \in \mathcal{N}} f'_{gji}(\boldsymbol{\xi}) - \sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) \right]_+ \\ &\geqslant\; \sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) + \xi_{gj} + \sum_{i \in \mathcal{N}} f'_{gji}(\boldsymbol{\xi}) - \sum_{i \in \mathcal{N}} f'_{gij}(\boldsymbol{\xi}) \\ &=\; \xi_{gj} + \sum_{i \in \mathcal{N}} f'_{gji}(\boldsymbol{\xi}), \end{aligned}$$

that is, the constraint is satisfied. The second constraint set is satisfied since

$$\sum_{g \in \mathcal{G}} f'_{gij}(\boldsymbol{\xi}) \;=\; \sum_{\substack{g \in \mathcal{G}: \\ z^\star_{gij} = 1}} \frac{\mathcal{S}_{gj}(\boldsymbol{\xi})}{\max\limits_{\boldsymbol{\xi} \in \Xi} \sum\limits_{g \in \mathcal{G}} \mathcal{S}_{gj}(\boldsymbol{\xi})} \cdot c^\star_{ij} \;\leqslant\; c^\star_{ij}$$

for all $\boldsymbol{\xi} \in \Xi$. Satisfaction of the third and fourth constraint sets, finally, follow immediately from the construction of $\boldsymbol{f}'$ and $\boldsymbol{l}'$.

Consider next the relative suboptimality of $(\boldsymbol{f}', \boldsymbol{l}')$ in the second-stage problem,

$$\begin{aligned} &\frac{\left[\sum\limits_{g \in \mathcal{G}} \sum\limits_{i,j \in \mathcal{N}} t_{gij} f'_{gij}(\boldsymbol{\xi}) + \sigma \sum\limits_{g \in \mathcal{G}} \sum\limits_{j \in \mathcal{N}} l'_{gj}(\boldsymbol{\xi}) \right] - \left[\sum\limits_{g \in \mathcal{G}} \sum\limits_{i,j \in \mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum\limits_{g \in \mathcal{G}} \sum\limits_{j \in \mathcal{N}} l^\star_{gj}(\boldsymbol{\xi}) \right]}{\sum\limits_{g \in \mathcal{G}} \sum\limits_{i,j \in \mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum\limits_{g \in \mathcal{G}} \sum\limits_{j \in \mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})} \\ = \;& \frac{\left[\sum\limits_{g \in \mathcal{G}} \sum\limits_{i,j \in \mathcal{N}} t_{gij} \left[f'_{gij}(\boldsymbol{\xi}) - f^\star_{gij}(\boldsymbol{\xi}) \right] \right] + \left[\sigma \sum\limits_{g \in \mathcal{G}} \sum\limits_{j \in \mathcal{N}} \left[l'_{gj}(\boldsymbol{\xi}) - l^\star_{gj}(\boldsymbol{\xi}) \right] \right]}{\sum\limits_{g \in \mathcal{G}} \sum\limits_{i,j \in \mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum\limits_{g \in \mathcal{G}} \sum\limits_{j \in \mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}. \end{aligned} \tag{32}$$

The relative second-stage suboptimality of $(\boldsymbol{f}', \boldsymbol{l}')$ overestimates the relative suboptimality of $(\boldsymbol{f}', \boldsymbol{l}')$ across both stages since the first-stage costs are non-negative and do not change if we replace $(\boldsymbol{f}^\star, \boldsymbol{l}^\star)$ with $(\boldsymbol{f}', \boldsymbol{l}')$.

In view of the first term in the numerator of (32), we notice that

$$\begin{aligned}
\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} \left[f'_{gij}(\boldsymbol{\xi}) - f^\star_{gij}(\boldsymbol{\xi})\right] &= \sum_{g\in\mathcal{G}} \sum_{\substack{i,j\in\mathcal{N}:\\ z^\star_{gij}=1}} t_{gij} \left[f'_{gij}(\boldsymbol{\xi}) - f^\star_{gij}(\boldsymbol{\xi})\right] \\
\leqslant \sum_{g\in\mathcal{G}} \sum_{\substack{i,j\in\mathcal{N}:\\ z^\star_{gij}=1}} t_{gij} \left[\mathcal{S}_{gj}(\boldsymbol{\xi}) - f^\star_{gij}(\boldsymbol{\xi})\right] &\leqslant \sum_{g\in\mathcal{G}} \sum_{\substack{i,j\in\mathcal{N}:\\ z^\star_{gij}=1}} t_{gij} \sum_{k\in\mathcal{D}(g,i)} l^\star_{gk}(\boldsymbol{\xi}) \\
= \sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} t_{gij} \cdot l^\star_{gk}(\boldsymbol{\xi}),
\end{aligned}$$

where the first identity disregards product-arc combinations without flows, the first inequality overestimates the affine flow with the entire downstream demand, the second inequality overestimates the unserved downstream demand under the optimal policy by the sum of downstream lost sales, and the last identity denotes the production facility node of good $g \in \mathcal{G}$ by $i_g \in \mathcal{N}$ as well as the set of arcs on the path from i_g to k in $\boldsymbol{z}^\star_g$ by $\mathcal{A}(g,k)$, respectively. Plugging this upper bound back into (32), we obtain

$$\begin{aligned}
\frac{\displaystyle\sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} t_{gij} \cdot l^\star_{gk}(\boldsymbol{\xi})}{\displaystyle\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})} &\leqslant \frac{\displaystyle\sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} t_{gij} \cdot l^\star_{gk}(\boldsymbol{\xi})}{\displaystyle\sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} t_{gij} \cdot \xi_{gj}} \\
&\leqslant \frac{\displaystyle\sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} \bar{t} \cdot l^\star_{gk}(\boldsymbol{\xi})}{\displaystyle\sum_{g\in\mathcal{G}} \sum_{k\in\mathcal{N}\setminus\{i_g\}} \sum_{(i,j)\in\mathcal{A}(g,k)} \underline{t} \cdot \xi_{gj}} = \frac{\bar{t}}{\underline{t}} \cdot \frac{\displaystyle\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\displaystyle\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}.
\end{aligned} \tag{33}$$

where the first inequality underestimates the second-stage costs of the optimal policy $(\boldsymbol{f}^\star, \boldsymbol{l}^\star)$ by an infeasible policy that disregards arc capacities and that can thus serve all demands, the second inequality overestimates and underestimates the per-unit transportation costs in the numerator and denominator, respectively, and the identity cancels terms.

In view of the second term in the numerator of (32), we observe that

$$\begin{aligned}
&\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}\left[l'_{gj}(\boldsymbol{\xi})-l^\star_{gj}(\boldsymbol{\xi})\right]\\
\leqslant\ &\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\left[f^\star_{gi_gj}(\boldsymbol{\xi})-f'_{gi_gj}(\boldsymbol{\xi})\right]_+ \ \leqslant\ \sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\left[\mathcal{S}_{gj}(\boldsymbol{\xi})-\frac{\mathcal{S}_{gj}(\boldsymbol{\xi})}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})}\cdot c^\star_{i_gj}\right]_+\\
=\ &\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\left[\frac{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})-c^\star_{i_gj}}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})}\right]_+\cdot\mathcal{S}_{gj}(\boldsymbol{\xi})\ \leqslant\ \sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\left[\frac{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\left(l^\star_{gj}(\boldsymbol{\xi})+\sum_{k\in\mathcal{D}(g,j)}l^\star_{gk}(\boldsymbol{\xi})\right)}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})}\right]\cdot\mathcal{S}_{gj}(\boldsymbol{\xi}),
\end{aligned}$$

where the first inequality upper bounds the difference in lost sales downstream from the initial node by the difference of flows that is sent from each production facility. Here, we continue to denote by $i_g\in\mathcal{N}$ the production facility node of good $g\in\mathcal{G}$, and $\mathcal{J}_g$ denotes the immediate descendants of i_g in $\boldsymbol{z}^\star_g$. The second inequality overestimates the optimal flow decisions by the downstream demands and replaces $\boldsymbol{f}'$ with its definition. The first identity reorders terms, and the last inequality upper bounds the difference between the downstream demands and the capacity of the arc (i_g,j) by the sum of downstream lost sales. We next observe that

$$\begin{aligned}
&\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\left[\frac{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\left(l^\star_{gj}(\boldsymbol{\xi})+\sum_{k\in\mathcal{D}(g,j)}l^\star_{gk}(\boldsymbol{\xi})\right)}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})}\right]\cdot\mathcal{S}_{gj}(\boldsymbol{\xi})\\
\leqslant\ &\left[\frac{\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\left(l^\star_{gj}(\boldsymbol{\xi})+\sum_{k\in\mathcal{D}(g,j)}l^\star_{gk}(\boldsymbol{\xi})\right)}{\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\mathcal{S}_{gj}(\boldsymbol{\xi})}\right]\cdot\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\mathcal{S}_{gj}(\boldsymbol{\xi})\\
=\ &\left[\frac{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}l^\star_{gj}(\boldsymbol{\xi})}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}\xi_{gj}}\right]\cdot\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{J}_g}\mathcal{S}_{gj}(\boldsymbol{\xi})\ \leqslant\ \left[\frac{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}l^\star_{gj}(\boldsymbol{\xi})}{\max_{\boldsymbol{\xi}\in\Xi}\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}\xi_{gj}}\right]\cdot\sum_{g\in\mathcal{G}}\sum_{j\in\mathcal{N}}\xi_{gj},
\end{aligned}$$

where the first inequality is a consequence of the relationship between the harmonic mean and the arithmetic mean, the identity uses the fact that the sets $\mathcal{D}(g,j)$, $j\in\mathcal{J}_g$, are pairwise disjoint and that Ξ is rectangular, the second inequality upper bounds the sum of downstream demands with the sum of demands (which include the demand at the production facility nodes). Plugging this upper bound back into (32), we obtain

$$\frac{\sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \left[l'_{gj}(\boldsymbol{\xi}) - l^\star_{gj}(\boldsymbol{\xi})\right]}{\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})} \leqslant \frac{\sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \left[l'_{gj}(\boldsymbol{\xi}) - l^\star_{gj}(\boldsymbol{\xi})\right]}{\underline{t} \cdot \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}$$

$$\leqslant \frac{\sigma}{\underline{t}} \cdot \frac{\left[\dfrac{\max_{\boldsymbol{\xi}\in\Xi} \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\max_{\boldsymbol{\xi}\in\Xi} \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}\right] \cdot \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}} = \frac{\sigma}{\underline{t}} \cdot \left[\frac{\max_{\boldsymbol{\xi}\in\Xi} \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\max_{\boldsymbol{\xi}\in\Xi} \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}\right] \leqslant \frac{\sigma}{\underline{t}} \cdot \max_{\boldsymbol{\xi}\in\Xi} \frac{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}}, \tag{34}$$

where the first inequality underestimates the second-stage costs of the optimal policy $(\boldsymbol{f}^\star, \boldsymbol{l}^\star)$ by an infeasible policy that disregards arc capacities and that can thus serve all demands, the second inequality uses our previously derived upper bound in the numerator, the first identity cancels terms, and the last inequality overestimates the ratio of two maxima by a single maximum.

Combining the equations (32), (33) and (34), we obtain

$$\frac{\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} \left[f'_{gij}(\boldsymbol{\xi}) - f^\star_{gij}(\boldsymbol{\xi})\right]}{\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})} + \frac{\sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \left[l'_{gj}(\boldsymbol{\xi}) - l^\star_{gj}(\boldsymbol{\xi})\right]}{\sum_{g\in\mathcal{G}} \sum_{i,j\in\mathcal{N}} t_{gij} f^\star_{gij}(\boldsymbol{\xi}) + \sigma \sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}$$

$$\leqslant \frac{\overline{t}}{\underline{t}} \cdot \frac{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}} + \frac{\sigma}{\underline{t}} \cdot \max_{\boldsymbol{\xi}\in\Xi} \frac{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} l^\star_{gj}(\boldsymbol{\xi})}{\sum_{g\in\mathcal{G}} \sum_{j\in\mathcal{N}} \xi_{gj}},$$

and maximizing the expression on the right-hand side of the inequality over all $\boldsymbol{\xi} \in \Xi$ results in the bound that is reported in the statement of the theorem. □

Proof of Observation 3. We employ Theorem 2 to show that problem (6) with the updated constraints (8) is optimally solved in K-adaptable affine decision rules when $K = 2^{LG}$. In analogy to the proof of Observation 1, we can assume that the problem is feasible. Fix any optimal first-stage decision $(\boldsymbol{x}^\star, \boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{p}^\star, \boldsymbol{c}^\star, \boldsymbol{s}^\star, \boldsymbol{w}^\star)$, let $\mathcal{T}_g = \{(i,j) \in \mathcal{N} \times \mathcal{N} \ : \ z^\star_{gij} = 1\}$ denote the distribution network for product $g \in \mathcal{G}$, and consider the following reformulation of the second-stage problem (6b) with the constraint (8b),

$$\begin{aligned}
&\text{minimize} && \sum_{g \in \mathcal{G}} \sum_{(i,j) \in \mathcal{T}_g} t_{gij} f_{gij} \\
&\text{subject to} && \mathrm{M} \cdot x^\star_{gj} + \sum_{i \in \mathcal{N}} s^\star_{gij} + f_{g,i(g,j),j} \geqslant \sum_{(j,i) \in \mathcal{T}_g} f_{gji} + \xi_{gj} + \sum_{i \in \mathcal{N}} s^\star_{gji} && \forall g \in \mathcal{G},\ \forall j \in \mathcal{N} \\
& && \sum_{g \in \mathcal{G}} f_{g,i(g,j),j} \leqslant C && \forall j \in \mathcal{N} \qquad\qquad \text{(6b")} \\
& && \sum_{\substack{g \in \mathcal{G}: \\ (i,j) \in \mathcal{T}_g}} f_{gij} \leqslant c^\star_{ij} && \forall (i,j) \in \bigcup_{g \in \mathcal{G}} \mathcal{T}_g \\
& && f_{gij} \in \mathbb{R}_+,\ g \in \mathcal{G} \text{ and } (i,j) \in \mathcal{T}_g,
\end{aligned}$$

where, as before, $i(g,j) \in \mathcal{N}$ is the unique index satisfying $(i(g,j),j) \in \mathcal{T}_g$, $g \in \mathcal{G}$ and $j \in \mathcal{N}$. Problem (6b") is equivalent to (6b) with the updated constraint (8b) in the sense that their objective functions as well as their feasible regions coincide when projected onto the remaining flow decisions $\{f_{gij} : g \in \mathcal{G} \text{ and } (i,j) \in \mathcal{T}_g\}$.

Problem (6b") clearly satisfies the conditions (**F**), (**A'**) with $K_{\mathrm{A}} = 1$ and $\Xi^{\mathrm{A}}_1 = \Xi$ as well as (**D**). To see that condition (**B'**) is satisfied as well, we invoke Proposition 3. To this end, fix any $g \in \mathcal{G}$ and $(i,j) \in \mathcal{T}_g$, and note that $i = i(g,j)$. If $w^\star_j = 0$, then $\sum_{i \in \mathcal{N}} s^\star_{gij} = \sum_{i \in \mathcal{N}} s^\star_{gji}$ by the first two constraints in (8a). In this case, as well as the case where $x^\star_{gj} = 1$, the first constraint of (6b") reduces to the first constraint of (6b') in the proof of Observation 1, and the same arguments show that a single constraint is weakly dominant for f_{gij} for all realizations $\boldsymbol{\xi} \in \Xi$, that is, the corresponding subset of constraints $\mathcal{K}_{gij}$ in condition *(i)* of Proposition 3 has cardinality 1. If $x^\star_{gj} = 0$ and $w^\star_j = 1$, on the other hand, then the first constraint of (6b"),

$$\sum_{i \in \mathcal{N}} s^\star_{gij} + f_{g,i(g,j),j} \geqslant \sum_{(j,i) \in \mathcal{T}_g} f_{gji} + \xi_{gj} + \sum_{i \in \mathcal{N}} s^\star_{gji},$$

will be weakly dominant for the realizations $\boldsymbol{\xi} \in \Xi$ under which $\sum_{(j,i) \in \mathcal{T}_g} f_{gji} + \xi_{gj} + \sum_{i \in \mathcal{N}} [s^\star_{gji} - s^\star_{gij}] \geqslant 0$, and the non-negativitiy constraint $f_{gij} \in \mathbb{R}_+$ will be weakly dominant otherwise. In this case, which arises for at most LG flow decisions, the corresponding subset of constraints $\mathcal{K}_{gij}$ in condition *(i)* of Proposition 3 has cardinality 2. In summary, we thus have $|\mathcal{K}_{\mathrm{B}}| = \prod_{g \in \mathcal{G}} \prod_{(i,j) \in \mathcal{T}_g} |\mathcal{K}_{gij}| \leqslant 2^{LG}$. Condition *(ii)* of Proposition 3, finally, follows from the same arguments as in the proof of Observation 1 since the presence of the first-stage decisions $s^\star_{gij}$ in (6b") does not impact the coefficient vectors of the second-stage decisions. □

Proof of Observation 4. As before, we can assume that the problem is feasible. We show that the assumptions of Proposition 4 are satisfied. To this end, note that the second-stage decisions $\boldsymbol{f}$ in problem (6) with the updated constraints (8) decompose across different goods $g \in \mathcal{G}$, that is, $\boldsymbol{f} = (\boldsymbol{f}_g)_{g \in \mathcal{G}}$ with $\boldsymbol{f}_g \in \mathbb{R}_+^{N^2}$. Likewise, the updated objective function separates across the goods $g \in \mathcal{G}$ in a natural way. The recourse matrix $\boldsymbol{B}$ is block diagonal since the lower bound constraints

$$\mathrm{M} \cdot x_{gj} + \sum_{i \in \mathcal{N}} (s_{gij} + f_{gij}) \geqslant \xi_{gj} + \sum_{i \in \mathcal{N}} (s_{gji} + f_{gji}) \qquad \forall g \in \mathcal{G},\ \forall j \in \mathcal{N}$$

decompose across the goods $g \in \mathcal{G}$. Also, as discussed in the proof of Observation 3, the second-stage problem clearly satisfies the conditions (**F**), (**A'**) with $K_{\mathrm{A}} = 1$ and $\Xi_1^{\mathrm{A}} = \Xi$ as well as (**D**).

To see that assumption (**B"**) is satisfied for $K_{\mathrm{B}}^g = 2^L$, $g \in \mathcal{G}$, recall that the proof of Observation 3 shows that the second-stage problem satisfies assumption (**B'**). We can actually apply the argument in that proof separately to the lower bound constraints of each good $g \in \mathcal{G}$ to obtain product-wise partitions $\Xi = \Xi_{g,1}^{\mathrm{B}} \cup \ldots \cup \Xi_{g,K_{\mathrm{B}}^g}^{\mathrm{B}}$, $g \in \mathcal{G}$, that each satisfy assumption (**B'**). This, however, is exactly what is required by assumption (**B"**), and Proposition 4 is therefore applicable. □

Proof of Observation 5. 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 $\mathcal{D}_e = \emptyset$. Assume next that $\mathcal{D}_e \neq \emptyset$, and fix the configuration vector $\boldsymbol{c} \in \times_{d \in \mathcal{D}_e} \mathcal{C}_d$ satisfying $x_{d,c_d} = 1$ for all $d \in \mathcal{D}_e$. This vector $\boldsymbol{c}$ is guaranteed to exist by the constraint of the first-stage problem (9a). The constraint

$$y_e + \mathrm{M} \sum_{d \in \mathcal{D}_e} (1 - x_{d,c_d}) \geqslant \xi_e + \sum_{d \in \mathcal{D}_e} r(e,d,c_d) \cdot y_d$$

then weakly dominates all other lower bounds on y_e imposed by configuration vectors $\boldsymbol{c}' \in \times_{d \in \mathcal{D}_e} \mathcal{C}_d$ since for each of them, at least one of the binary variables x_{d,c'_d}, $d \in \mathcal{D}_e$, must evaluate to zero, again due to the constraint of the first-stage problem (9a).

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

$$y_e \geqslant \xi_e + \sum_{d \in \mathcal{D}_e} r(e,d,c_d) \cdot y_d - \mathrm{M} \sum_{d \in \mathcal{D}_e} (1 - x_{d,c_d})$$

corresponding to an entity $e \in \mathcal{E}$ with $\mathcal{D}_e \neq \emptyset$ links e only with its immediate descendants $d \in \mathcal{D}_e$ satisfying $\pi(d) < \pi(e)$. We therefore conclude that the right-hand side coefficient vectors $\boldsymbol{\beta}_{e\boldsymbol{c}}$, $e \in \mathcal{E}$ and $\boldsymbol{c} \in \times_{d \in \mathcal{D}_e} \mathcal{C}_d$ with $x_{d,c_d} = 1$ for all $d \in \mathcal{D}_e$, with elements $r(e, d, c_d)$ form a partial order for all entities $e \in \mathcal{E}$ with $\mathcal{D}_e \neq \emptyset$. The statement then follows since the right-hand side coefficient vectors of the entities $e \in \mathcal{E}$ with no descendants, that is, $\mathcal{D}_e = \emptyset$, are zero, which allows us to include them anywhere in the partial order. □

Proof of Observation 6. Problem (10) can be interpreted as an instance of problem (12) with $\boldsymbol{z}_k = \boldsymbol{x}^k$, $k = 1, \ldots, K$. One readily verifies that the second-stage problem (9b) satisfies the assumptions ($\mathbf{F}^K$), ($\mathbf{A}^K$) and (**D**). Likewise, the reasoning in the proof of Observation 5 can be applied to every $\boldsymbol{z}_k$, which implies that assumption ($\mathbf{B}^K$) is satisfied as well. We can therefore invoke Corollary 1 from Appendix A to conclude problem (10) is indeed optimally solved in K-adaptable affine decision rules. □

Proof of Corollary 1. We show that under the assumptions of the corollary, we can express problem (12) as an instance of problem (1) satisfying the relaxed assumptions (**F**), (**A'**), (**D**) and (**B'**), and that the associated instance (4) of the K-adaptability problem is equivalent to formulation (13). The statement of the corollary then follows from Theorem 2.

Fix an instance of problem (12) as well as an optimal first-stage decision $(\boldsymbol{x}^\star, \{\boldsymbol{z}_k^\star\}_{k=1}^K)$ satisfying our conditions ($\mathbf{F}^K$), ($\mathbf{A}^K$), (**D**) and ($\mathbf{B}^K$) in (12). Consider the following instance of problem (1),

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

where the second-stage cost function $\mathcal{Q}'$ satisfies

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

with $\Xi = \bigcup_{k=1}^K \Xi_k$ being any covering of Ξ such that for all $k = 1, \ldots, K$, $\mathcal{Q}(\boldsymbol{x}^\star, \boldsymbol{z}_k^\star; \boldsymbol{\xi})$ is minimal among all $\mathcal{Q}(\boldsymbol{x}^\star, \boldsymbol{z}_l^\star; \boldsymbol{\xi})$, $l = 1, \ldots, K$, for all $\boldsymbol{\xi} \in \Xi_k$ (ties can be broken arbitrarily), as well as

$$
\begin{aligned}
f'(\boldsymbol{x}', \boldsymbol{y}; \boldsymbol{\xi}) &= \sum_{k=1}^{K} \mathbf{1}[\boldsymbol{\xi} \in \Xi_k] \cdot f(\boldsymbol{x}, \boldsymbol{z}_k, \boldsymbol{y}; \boldsymbol{\xi}), \\
\boldsymbol{A}'(\boldsymbol{\xi}) &= [\boldsymbol{A}(\boldsymbol{\xi}),\ \mathbf{1}[\boldsymbol{\xi} \in \Xi_1] \cdot \boldsymbol{E}(\boldsymbol{\xi}),\ \ldots,\ \mathbf{1}[\boldsymbol{\xi} \in \Xi_K] \cdot \boldsymbol{E}(\boldsymbol{\xi})], \\
\boldsymbol{C}'(\boldsymbol{\xi}) &= [\boldsymbol{C}(\boldsymbol{\xi}),\ \mathbf{1}[\boldsymbol{\xi} \in \Xi_1] \cdot \boldsymbol{F}(\boldsymbol{\xi}),\ \ldots,\ \mathbf{1}[\boldsymbol{\xi} \in \Xi_K] \cdot \boldsymbol{F}(\boldsymbol{\xi})].
\end{aligned}
$$

Here, $\mathbf{1}[\cdot]$ attains the value 1 (0) if the condition $\cdot$ is (not) met. Note that the assumptions of Theorem 2 are met since (**F**) follows immediately from ($\mathbf{F}^K$), (**D**) holds by construction, and (**A'**) and (**B'**) follow from ($\mathbf{A}^K$) and ($\mathbf{B}^K$) if we choose $\Xi_k^{\mathrm{A}} = \Xi_k$, $k = 1, \ldots, K_{\mathrm{A}} = K$ and $\Xi_k^{\mathrm{B}} = \Xi_k$, $k = 1, \ldots, K_{\mathrm{B}} = K$, respectively. Theorem 2 thus implies that the optimal first-stage decision $(\boldsymbol{x}^\star, \{\boldsymbol{z}_k^\star\}_{k=1}^K)$ in (12) is also optimal in

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

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

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

By replacing f', $\boldsymbol{A}'$ and $\boldsymbol{C}'$ with their definitions, one readily confirms that this instance of problem (4) is indeed equivalent to problem (13). Note that the adaptability of $K_{\mathrm{A}} \cdot K_{\mathrm{B}} = K^2$ required by Theorem 2 can be reduced to an adaptability of K since the coverings $\{\Xi_k^{\mathrm{A}}\}_{k=1}^K$ and $\{\Xi_k^{\mathrm{B}}\}_{k=1}^K$ coincide. □