Noname manuscript No.
(will be inserted by the editor)

Primal and dual linear decision rules in stochastic and robust optimization

Daniel Kuhn · Wolfram Wiesemann · Angelos Georghiou

Received: date / Accepted: date

Abstract Linear stochastic programming provides a flexible toolbox for analyzing real-life decision situations, but it can become computationally cumbersome when recourse decisions are involved. The latter are usually modeled as decision rules, i.e., functions of the uncertain problem data. It has recently been argued that stochastic programs can quite generally be made tractable by restricting the space of decision rules to those that exhibit a linear data dependence. In this paper, we propose an efficient method to estimate the approximation error introduced by this rather drastic means of complexity reduction: we apply the linear decision rule restriction not only to the primal but also to a dual version of the stochastic program. By employing techniques that are commonly used in modern robust optimization, we show that both arising approximate problems are equivalent to tractable linear or semidefinite programs of moderate sizes. The gap between their optimal values estimates the loss of optimality incurred by the linear decision rule approximation. Our method remains applicable if the stochastic program has random recourse and multiple decision stages. It also extends to cases involving ambiguous probability distributions.

Keywords Linear decision rules · stochastic optimization · robust optimization · error bounds · semidefinite programming

PACS 90C15

1 Introduction

Stochastic programming researchers have always felt that dynamic decision problems under uncertainty are very difficult. Only recently, this common perception has received a theoretical underpinning. Dyer and Stougie prove that linear two-stage stochastic programming problems are #P-hard [16]. As pointed out by Shapiro and Nemirovski [26], however, reasonably accurate *approximate* solutions can be calculated efficiently via Monte Carlo sampling techniques. In contrast to two-stage models, linear multistage stochastic programs are even hard to solve approximately. Shapiro and Nemirovski argue that 'multistage stochastic programs generically are computationally intractable already when medium-accuracy solutions are sought' [26]. Complexity results of this type indicate that stochastic programming problems need to undergo some (maybe drastic) simplification in

D. Kuhn
Department of Computing
Imperial College of Science, Technology, and Medicine
Tel.: +44-(0)20-7594-8334
Fax: +44-(0)20-7581-8024
E-mail: dkuhn@imperial.ac.uk

order to gain computational tractablility. A suitable approximation is, for instance, to impose a linear structure on the recourse decisions. This *linear decision rule* approximation has attracted considerable interest in recent years since it enables scalability to multistage models.

A survey and critical assessment of early results on decision rules in stochastic programming has been assembled by Garstka and Wets in 1974 [18]. After an extended period of neglect, linear decision rules experienced a recent revival in the context of robust optimization. Ben-Tal et al. use them to derive tractable approximations for multistage decision problems affected by non-stochastic uncertainty [4]. This min-max approach was shown to have distinct advantages over dynamic programming when applied to a two-echelon dynamic supply chain problem [3]. Other promising application areas for linear decision rule-based robust optimization include project scheduling [1] and capacity expansion of networks [21]. To avoid overly risk-averse decisions, Ben-Tal et al. later refined the worst-case approach in [4] to allow for controlled constraint violations [2]. Boosted by their success in robust optimization, linear decision rules have recently gained renewed interest from the stochastic programming community. Their potential for complexity reduction in multistage stochastic programming has first been highlighted by Shapiro and Nemirovski [26], and their suitability for solving chance-constrained stochastic programs has been investigated by Chen et al. [13]. To reduce the inherent approximation error, Chen et al. later coined the concepts of *deflected* and *segregated* linear decision rules, which are more flexible than ordinary linear decision rules but retain their favorable scalability properties [14]. A similar approach based on an intelligent re-parameterization of the uncertain variables was proposed by Chen and Zhang [15]. In a stochastic programming context, Calafiore recently applied linear decision rules to solve multistage portfolio optimization problems with many securities and trading periods [12].

While linear decision rules are very effective at reducing computational complexity, they may incur a considerable loss of optimality.[1] In this paper we attempt to gain a deeper grasp of this tradeoff between tractability and optimality. To this end, we apply the linear decision rule approximation not only to the primal stochastic program (as is done in all of the existing literature) but also to a dual version of it. Recall that the use of *primal* linear decision rules leads to a conservative approximation of the original problem and thus *under*estimates the decision maker's flexibility. In contrast, the use of *dual* linear decision rules results in a progressive approximation which *over*estimates the available flexibility. We show in this paper that both (primal and dual) linear decision rule approximations lead to problems that (in 'many' cases) have reformulations as tractable conic programs. The main benefits of the outlined approach are the following.

- The gap between the optimal values of the conservative and progressive approximate problems estimates the loss of optimality incurred by using linear decision rules. If this gap is small, there is little room for improvement. In contrast, a large optimality gap indicates that one could (and maybe should) improve on the linear decision rules, e.g. by using deflected or segregated linear decision rules as in [14, 15]. Note that the optimality gap can be computed efficiently as it is based on the solution of two tractable conic programs of moderate sizes.
- The primal and dual linear decision rule-based approximations permit scalability to multistage models. Moreover, they only require information about the support and the first two (sometimes the first four) moments of the uncertain problem parameters. This is a desirable feature since full distributional information is scarcely available in reality.
- Our *dual* linear decision rule approximation generates semi-infinite-type problems with finitely many expectation constraints and a continuum of nonnegative decision vari-

[1] Linear decision rules are known to be optimal in control problems with linear dynamics and quadratic costs [8]. Even for linear stochastic programming problems, however, they are generically strictly suboptimal [18].

ables. These problems are of interest in their own right. Based on mild assumptions, our analysis illustrates how they can be reformulated as low-dimensional linear or semidefinite programs. This implies that they are amenable to efficient numerical solution.

The rest of this paper develops as follows. In each of the following three sections we elaborate decision rule-based approximations for a specific class of decision problems under uncertainty. Sections 2 and 3 treat linear one-stage stochastic programs with fixed and random recourse, respectively, while extensions to multistage stochastic programs are investigated in Section 4. Emphasis is put on elaborating novel *progressive* (i.e., dual) approximations. In Section 5 we assess the appropriateness of using linear decision rules for solving a multistage multi-factory inventory problem from the literature [4].

Notation In this paper uncertainty is modeled by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$. The elements of the sample space $\mathbb{R}^k$ are denoted by ξ and will be referred to as outcomes or observations. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ is the set of events that are assigned probabilities by the probability measure $\mathbb{P}$. For notational simplicity, we denote by $\mathcal{L}^2_{k,n} := \mathcal{L}^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P}; \mathbb{R}^n)$ the space of all Borel measurable, square-integrable functions from $\mathbb{R}^k$ to $\mathbb{R}^n$, while Ξ denotes the support of $\mathbb{P}$, that is, the smallest closed subset of $\mathbb{R}^k$ which has probability 1. Finally, $\mathbb{E}(\cdot)$ denotes the expectation operator with respect to $\mathbb{P}$.

By a slight abuse of notation, for $A, B \in \mathbb{R}^{m\times n}$, the relation $A \geq B$ denotes componentwise inequality. Moreover, for $C, D \in \mathbb{R}^{n\times n}$, the relation $C \succeq D$ implies that $C - D$ is positive semidefinite. The converse inequalities $A \leq B$ and $C \preceq D$ are defined in the obvious way. Finally, for every $C \in \mathbb{R}^{n\times n}$ we let $\mathrm{Tr}(C)$ denote the trace of C.

2 One-stage stochastic programs with fixed recourse

This section studies decision problems under uncertainty of the following generic type. A decision maker first observes an element ξ of the sample space $\mathbb{R}^k$. Then, a decision $x(\xi) \in \mathbb{R}^n$ is selected subject to the constraints $A(\xi)x(\xi) \leq b(\xi)$ and at a cost $c(\xi)^\top x(\xi)$. The goal is to choose the function $x \in \mathcal{L}^2_{k,n}$ so as to minimize the expected cost. By convention, x is referred to as a *decision rule*, *strategy*, or *policy*. The decision problem outlined above can be formulated as the following linear one-stage stochastic program.

$$\begin{array}{ll} \underset{x\in\mathcal{L}^2_{k,n}}{\text{minimize}} & \mathbb{E}\left(c(\xi)^\top x(\xi)\right) \\ \text{subject to} & A(\xi)x(\xi) \leq b(\xi) \quad \mathbb{P}\text{-a.s.} \end{array} \tag{$\mathcal{SP}$}$$

In order for $\mathcal{SP}$ to be well-defined, one needs appropriate assumptions about the underlying problem data. We first assume that the objective function coefficients and the right hand side vector depend linearly on the uncertain parameters. Formally speaking, we postulate that $c(\xi) = C\xi$ for some $C \in \mathbb{R}^{n\times k}$ and $b(\xi) = B\xi$ for some $B \in \mathbb{R}^{m\times k}$. This linearity assumption is nonrestrictive since we are free to redefine ξ such that it contains $c(\xi)$ and $b(\xi)$ as subvectors. Moreover, we assume the recourse matrix to be independent of ξ, that is, $A(\xi) \equiv A \in \mathbb{R}^{m\times n}$. The case of random recourse is more intricate and will be investigated in Section 3. Finally, we assume the support of the probability measure $\mathbb{P}$ to be a nonempty compact polyhedron of the form

$$\Xi = \left\{\xi \in \mathbb{R}^k : W\xi \geq h\right\} \tag{2.1a}$$

for some matrix $W \in \mathbb{R}^{l\times k}$ and a vector $h \in \mathbb{R}^l$. Without loss of generality, we will assume that

$$W = (e_1, -e_1, \hat{W}^\top)^\top \quad \text{and} \quad h = (1, -1, \underbrace{0, \ldots, 0}_{l-2})^\top, \tag{2.1b}$$

where $\hat{W} \in \mathbb{R}^{(l-2)\times k}$ is a submatrix of W, and e_1 denotes the basis vector in $\mathbb{R}^k$ whose first element is 1 while all the others are 0. This specification guarantees that the first component of every $\xi \in \Xi$ is equal to 1. It is easy to see that every compact convex polyhedron in the hyperplane $\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1\}$ is representable in the form (2.1). Modeling ξ_1 as a degenerate dummy outcome that is equal to 1 almost surely allows us to represent affine functions of the nondegenerate outcomes $(\xi_2, \dots, \xi_k)$ in a compact manner as linear functions of $\xi = (\xi_1, \dots, \xi_k)$. This trick will have distinct notational advantages below. In addition to being compact, we further require Ξ to span the whole sample space, that is, we assume the linear hull of Ξ to coincide with $\mathbb{R}^k$. By (2.1), Ξ spans $\mathbb{R}^k$ iff it has dimension $k-1$ or, equivalently, the system $\hat{W}\xi \geq 0$ is strictly feasible. This nonrestrictive extra requirement can always be enforced by reducing the dimension of ξ, if necessary.

The stipulated regularity conditions and the restriction to square-integrable decision rules imply that the expectation in the objective of $\mathcal{SP}$ is well-defined. For the further argumentation it proves useful to convert the inequality constraints in $\mathcal{SP}$ to equality constraints by introducing slack variables $s \in \mathcal{L}^2_{k,m}$.

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left(c(\xi)^\top x(\xi)\right) \\ \text{subject to} & x \in \mathcal{L}^2_{k,n}\,,\ s \in \mathcal{L}^2_{k,m} \\ & \left.\begin{array}{l} A(\xi)x(\xi) + s(\xi) = b(\xi) \\ s(\xi) \geq 0 \end{array}\right\} \quad \mathbb{P}\text{-a.s.} \end{array} \tag{2.2}$$

It is known that finding the optimal value of problem $\mathcal{SP}$ or its reformulation (2.2) is #P-hard even if $\mathbb{P}$ is a uniform distribution on the unit cube in $\mathbb{R}^k$, see [16, Theorem 3.2]. Thus, there can be no efficient algorithm to solve $\mathcal{SP}$ exactly unless P=NP. While reasonably accurate approximate solutions for $\mathcal{SP}$ can be obtained efficiently by using the sample average approximation [26], we will now study an alternative approximation that remains tractable in the multistage case.

2.1 Primal Approximation

A tractable approximation for problem (2.2) is obtained by restricting the functional form of the decision rules to be linear, that is, by reducing the space of admissible decision rules to those which are representable as $x(\xi) = X\xi$ and $s(\xi) = S\xi$ for some $X \in \mathbb{R}^{n\times k}$ and $S \in \mathbb{R}^{m\times k}$, respectively.[2] This radical but effective approach to complexity reduction has been proposed by Ben-Tal et al. in a robust optimization context [4] and was later extended to the realm of stochastic programming by Shapiro and Nemirovski [26] and Chen et al. [14]. The resulting approximate problem is of semi-infinite type as it involves only a finite number of decision variables (the matrices X and S) while still exhibiting an infinite number of constraints (m equality and nonnegativity constraints for $\mathbb{P}$-almost every ξ).

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k} \\ & \left.\begin{array}{l} AX\xi + S\xi = B\xi \\ S\xi \geq 0 \end{array}\right\} \quad \mathbb{P}\text{-a.s.} \end{array} \tag{$\mathcal{SP}^u$}$$

Notice that $\mathcal{SP}^u$ provides an upper bound (i.e., a conservative approximation) for $\mathcal{SP}$ since it was obtained by reducing the underlying feasible set. The objective function of $\mathcal{SP}^u$ is expressed in terms of the symmetric second-order moment matrix associated with $\mathbb{P}$ which is defined as

$$M := \mathbb{E}\left(\xi\,\xi^\top\right).$$

[2] Recall that $\xi_1 = 1$ almost surely, and therefore $X\xi$ and $S\xi$ represent affine functions of the nondegenerate outcomes $(\xi_2, \dots, \xi_k)$ on the support of $\mathbb{P}$.

2.2 Tractable reformulation of primal approximation

Although $\mathcal{SP}^u$ has only finitely many decision variables, it still seems to be intractable as it involves infinitely many constraints. By using techniques that are commonly employed in modern robust optimization, however, the semi-infinite constraint system can be reexpressed in terms of a finite number of linear constraints. We first observe that due to their continuity in ξ, the almost sure constraints in $\mathcal{SP}^u$ hold in fact for all ξ in the support of $\mathbb{P}$, that is, for all $\xi \in \Xi$. The equality constraints in $\mathcal{SP}^u$ thus imply that the linear hull of Ξ belongs to the null space of the linear operator $AX + S - B$. Recall next that the uncertainty set Ξ spans the whole of $\mathbb{R}^k$, and therefore we may equivalently require $AX+S = B$ in $\mathcal{SP}^u$. Simplification of the more intricate semi-infinite inequality constraints relies on the following proposition which is at the heart of the robust optimization paradigm due to Ben-Tal and Nemirovski [5, 6].

Proposition 1 *For any $z \in \mathbb{R}^k$ the following statements are equivalent:*

(i) $z^\top \xi \geq 0$ *for all* $\xi \in \Xi$*;*
(ii) $\exists \lambda \in \mathbb{R}^l$ *with* $\lambda \geq 0$, $W^\top \lambda = z$, *and* $h^\top \lambda \geq 0$

Remark 1 Proposition 1 can be viewed as a special case of [6, Theorem 3.1] or [4, Theorem 3.2]. The proof is repeated here to keep this paper self-contained.

Proof (Proof of Proposition 1) By using a standard duality argument, we find

$$\begin{aligned} & z^\top \xi \geq 0 \text{ for all } \xi \text{ subject to } W\xi \geq h \\ \Longleftrightarrow\ & 0 \leq \min_{\xi \in \mathbb{R}^k} \left\{ z^\top \xi : W\xi \geq h \right\} \\ \Longleftrightarrow\ & 0 \leq \max_{\lambda \in \mathbb{R}^l} \left\{ h^\top \lambda : W^\top \lambda = z,\, \lambda \geq 0 \right\} \\ \Longleftrightarrow\ & \exists \lambda \in \mathbb{R}^l \text{ with } W^\top \lambda = z,\, h^\top \lambda \geq 0,\, \lambda \geq 0 \end{aligned}$$

The equivalence in the third line of the above expression follows from strong linear programming duality, which holds since the primal minimization problem has a nonempty feasible set. Thus, the claim follows. □

If $z_\mu^\top$ denotes the μth row of the matrix S in problem $\mathcal{SP}$, then we can use Proposition 1 to replace the semi-infinite inequality constraint $z_\mu^\top \xi \geq 0\ \forall \xi \in \Xi$ by a finite set of linear constraints in (z_μ, λ_μ) for some new decision vector $\lambda_\mu \in \mathbb{R}^l$, $\mu = 1, \ldots, m$. Interpreting $\lambda_\mu^\top$ as the μth row of a matrix $\Lambda \in \mathbb{R}^{m \times l}$, problem $\mathcal{SP}^u$ thus simplifies to the following linear program.

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, \Lambda \in \mathbb{R}^{m \times l} \\ & AX + \Lambda W = B \\ & \Lambda h \geq 0 \\ & \Lambda \geq 0 \end{array} \tag{2.3}$$

One can directly verify that the slack variables $s(\xi) = S\xi = \Lambda W \xi$ corresponding to any (X, Λ) feasible in (2.3) are indeed nonnegative on Ξ, that is,

$$S\xi = \Lambda W \xi = \Lambda(W\xi - h) + \Lambda h \geq 0 \quad \text{whenever} \quad W\xi \geq h\,.$$

The striking advantage of using linear decision rules is that the resulting approximate problem $\mathcal{SP}^u$ can be solved very efficiently as it is equivalent to a linear program whose size is polynomial in k, l, m, and n, that is, the size of the description of the original problem $\mathcal{SP}$ and the underlying uncertainty set Ξ.

2.3 Dual approximation

Similar techniques that were used to derive the linear program (2.3) can also be used to find a computationally tractable lower bound on $\mathcal{SP}$. To this end, we first restrict certain *dual* decision rules to be linear functions of ξ. Subsequently, we apply a *dual* version of Proposition 1 to convert the resulting problem to a tractable linear program. Before we can embark on this pathway, we have to introduce a min-max reformulation of problem (2.2) in which the equality constraints are dualized.

$$\begin{array}{ll} \underset{x\in\mathcal{L}^2_{k,n},\, s\in\mathcal{L}^2_{k,m}}{\text{minimize}} & \underset{y\in\mathcal{L}^2_{k,m}}{\sup} \ \mathbb{E}\left(c(\xi)^\top x(\xi) + y(\xi)^\top\left[Ax(\xi)+s(\xi)-b(\xi)\right]\right) \\ \text{subject to} & s(\xi)\geq 0 \quad \mathbb{P}\text{-a.s.} \end{array} \tag{2.4}$$

Note that the maximization over the dual decisions $y \in \mathcal{L}^2_{k,m}$ imposes an infinite penalty on every primal decision $(x,s) \in \mathcal{L}^2_{k,n} \times \mathcal{L}^2_{k,m}$ which violates the equality constraints $Ax(\xi) + s(\xi) = b(\xi)$ on a set of strictly positive probability. A formal proof establishing the equivalence of (2.2) and (2.4) is provided in [28, § 4]. In the following derivation, we use the shorthand notation '$\inf_{x,s}$' to denote the infimum operator over all $x \in \mathcal{L}^2_{k,n}$ and over all $s \in \mathcal{L}^2_{k,m}$ that are almost surely nonnegative. Similarly, '$\sup_y$' and '$\sup_Y$' stand for the supremum operators over all $y \in \mathcal{L}^2_{k,m}$ and $Y \in \mathbb{R}^{m\times k}$, respectively. Using the equivalence of (2.2) and (2.4) we obtain

$$\begin{aligned} \inf \mathcal{SP} &= \inf_{x,s}\ \sup_{y}\ \mathbb{E}\left(c(\xi)^\top x(\xi) + y(\xi)^\top\left[Ax(\xi)+s(\xi)-b(\xi)\right]\right) \\ &\geq \inf_{x,s}\ \sup_{Y}\ \mathbb{E}\left(c(\xi)^\top x(\xi) + \xi^\top Y^\top\left[Ax(\xi)+s(\xi)-b(\xi)\right]\right) \\ &= \inf_{x,s}\ \sup_{Y}\ \mathbb{E}\left(c(\xi)^\top x(\xi)\right) + \operatorname{Tr}\left[Y^\top \mathbb{E}\left(\left[Ax(\xi)+s(\xi)-b(\xi)\right]\xi^\top\right)\right]. \end{aligned}$$

In the second line of the above expression we require the dual decisions to be representable as $y(\xi) = Y\xi$ for some $Y \in \mathbb{R}^{m\times k}$. Thus, we effectively restrict the dual feasible set to contain only *linear* decision rules. The maximization in the third line can be carried out explicitly, which implies that the optimal value of $\mathcal{SP}$ is bounded below by that of the following approximate problem.

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left(c(\xi)^\top x(\xi)\right) \\ \text{subject to} & x\in\mathcal{L}^2_{k,n}\,,\ s\in\mathcal{L}^2_{k,m} \\ & \left.\begin{array}{r} \mathbb{E}\left(\left[Ax(\xi)+s(\xi)-b(\xi)\right]\xi^\top\right)=0 \\ s(\xi)\geq 0 \end{array}\right\} \quad \mathbb{P}\text{-a.s.} \end{array} \tag{$\mathcal{SP}^l$}$$

It is easy to verify directly that $\mathcal{SP}^l$ represents indeed a relaxation of $\mathcal{SP}$. Any (x,s) feasible in (2.2) satisfies $Ax(\xi) + s(\xi) - b(\xi) = 0$ for $\mathbb{P}$-almost all ξ and will therefore also satisfy the less restrictive expectation constraint in $\mathcal{SP}^l$. This confirms that the optimal value of problem $\mathcal{SP}^l$ provides a lower bound on the optimal value of $\mathcal{SP}$. Notice that $\mathcal{SP}^l$ involves only finitely many equality constraints. Thus, it is certainly not harder to solve than $\mathcal{SP}$. However, $\mathcal{SP}^l$ still appears to be intractable as it involves a continuum of decision variables and inequality constraints.

2.4 Tractable reformulation of dual approximation

We now show that $\mathcal{SP}^l$ has a reformulation as a simple linear program whenever $\mathcal{SP}$ is strictly feasible. The techniques used in this section are developed ad hoc. An important ingredient for our reformulation is a dual version of Proposition 1, see Proposition 3 below. For the further argumentation, we need the following technical result about the second-order moment matrix M.

Proposition 2 *The matrix M is positive definite and invertible.*

Proof It is clear that M is positive semidefinite. Assume now that M is not invertible. Thus, there is a $v \in \mathbb{R}^k$, $v \neq 0$, such that $v^\top M v = 0$. By definition of M, we have $v^\top \xi = 0$ with probability one. Thus, v is orthogonal to each $\xi \in \Xi$, which implies that Ξ cannot span $\mathbb{R}^k$. This is a contradiction. □

A first step towards simplifying problem $\mathcal{SP}^l$ consists in introducing new decision variables $X \in \mathbb{R}^{n\times k}$ and $S \in \mathbb{R}^{m\times k}$ which are completely determined by the original decisions $x \in \mathcal{L}^2_{k,n}$ and $s \in \mathcal{L}^2_{k,m}$ through the relations

$$XM = \mathbb{E}\left(x(\xi)\,\xi^\top\right) \quad \text{and} \quad SM = \mathbb{E}\left(s(\xi)\,\xi^\top\right), \tag{2.5}$$

respectively. Note that if x is a linear decision rule, $x(\xi) = \hat{X}\xi$ for some coefficient matrix $\hat{X} \in \mathbb{R}^{n\times k}$, then we have $X = \hat{X}$. A similar statement holds for the slack variables. By using (2.5), the objective function in $\mathcal{SP}^l$ can be reformulated as $\mathrm{Tr}\left(MC^\top X\right)$, while the equality constraint in $\mathcal{SP}^l$ reduces to $AXM + SM - BM = 0$. Since M is invertible, the latter restriction is equivalent to $AX + S = B$. Consequently, $\mathcal{SP}^l$ is equivalent to the following optimization problem.

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k} \\ & AX + S = B \\ & \exists\, x \in \mathcal{L}^2_{k,n} : XM = \mathbb{E}\left(x(\xi)\,\xi^\top\right) \\ & \exists\, s \in \mathcal{L}^2_{k,m} : SM = \mathbb{E}\left(s(\xi)\,\xi^\top\right),\ s(\xi) \geq 0\ \mathbb{P}\text{-a.s.} \end{array} \tag{2.6}$$

The penultimate constraint in (2.6) is redundant and may be deleted without affecting the problem's feasible set. Indeed, for any $X \in \mathbb{R}^{n\times k}$ the linear decision rule $x(\xi) := X\xi$ satisfies the postulated conditions. However, the last constraint looks difficult as it involves the solution of m moment problems: for a given $S \in \mathbb{R}^{m\times k}$ we have to verify the existence of m nonnegative Borel measures whose vectors of zero- and first-order moments coincide with the rows of SM and which have square-integrable densities with respect to $\mathbb{P}$. This moment feasibility constraint can be viewed as the dual counterpart of the semi-infinite inequality constraint in $\mathcal{SP}^u$. We will now show that it can generically be replaced by a finite number of linear constraints. To this end, we need a *dual* version of Proposition 1.

Proposition 3 *Consider the following two convex cones in $\mathbb{R}^k$:*

$$\begin{aligned} \mathcal{K} &:= \left\{z \in \mathbb{R}^k : \left(W - h\,e_1^\top\right) z \geq 0\right\}, \\ \mathcal{K}_{\mathbb{P}} &:= \left\{z \in \mathbb{R}^k : \exists\, s \in \mathcal{L}^2_{k,1} \text{ with } \mathbb{E}(s(\xi)\,\xi) = z \text{ and } s(\xi) \geq 0\ \mathbb{P}\text{-a.s.}\right\}. \end{aligned}$$

Then, $\emptyset \neq \operatorname{int} \mathcal{K} \subset \mathcal{K}_{\mathbb{P}} \subset \mathcal{K}$.

Remark 2 Note that the last constraint in (2.6) is equivalent to the requirement that every row of the matrix $SM \in \mathbb{R}^{m\times k}$ is an element of $\mathcal{K}_{\mathbb{P}}$.

Proof (Proof of Proposition 3) Fix an arbitrary $z \in \mathcal{K}_{\mathbb{P}}$, and let s be a corresponding scalar function that satisfies the conditions in the definition of $\mathcal{K}_{\mathbb{P}}$. Then, we have

$$\left(W - h\,e_1^\top\right) z = \mathbb{E}\left[\left(W - h\,e_1^\top\right) \xi\, s(\xi)\right] \geq 0\,.$$

The inequality follows from the fact that $e_1^\top \xi = 1$ while both $W\xi - h$ and $s(\xi)$ are nonnegative on the support of $\mathbb{P}$. As the choice of $z \in \mathbb{R}^k$ was arbitrary, the postulated inclusion $\mathcal{K}_{\mathbb{P}} \subset \mathcal{K}$ follows.

Denote now by $\mathcal{M}^+$ the set of all nonnegative finite measures on $(\Xi, \mathcal{B}(\Xi))$ with finite second moments, and let $\mathcal{M}^+_{\mathbb{P}}$ be the subset of measures in $\mathcal{M}^+$ that have a square-integrable density with respect to $\mathbb{P}$. Note that $\mathcal{M}^+_{\mathbb{P}}$ is dense in $\mathcal{M}^+$ with respect to the weak topology since Ξ constitutes the support of $\mathbb{P}$ (see e.g. [10] for a discussion weak topologies in spaces of measures). Next, define

$$\tilde{\mathcal{K}} := \left\{ \int_{\Xi} \xi \, \mu(\mathrm{d}\xi) : \mu \in \mathcal{M}^+ \right\} \quad \text{and} \quad \tilde{\mathcal{K}}_{\mathbb{P}} := \left\{ \int_{\Xi} \xi \, \mu(\mathrm{d}\xi) : \mu \in \mathcal{M}^+_{\mathbb{P}} \right\},$$

which represent convex cones in $\mathbb{R}^k$. Since the density of any $\mu \in \mathcal{M}^+_{\mathbb{P}}$ with respect to $\mathbb{P}$ is a function in $\mathcal{L}^2_{k,1}$, it is clear that $\tilde{\mathcal{K}}_{\mathbb{P}} = \mathcal{K}_{\mathbb{P}} \subset \tilde{\mathcal{K}}$. Below, we will argue that $\emptyset \neq \operatorname{int} \tilde{\mathcal{K}} \subset \tilde{\mathcal{K}}_{\mathbb{P}}$, and $\tilde{\mathcal{K}} = \mathcal{K}$. This will establish $\emptyset \neq \operatorname{int} \mathcal{K} \subset \mathcal{K}_{\mathbb{P}}$.

For each $\xi \in \Xi$, the cone $\mathcal{M}^+$ contains the Dirac measure δ_ξ which concentrates unit mass at the point ξ. Therefore, $\tilde{\mathcal{K}}$ is easily seen to coincide with the convex closed cone generated by Ξ. This cone has nonempty interior since Ξ is assumed to span $\mathbb{R}^k$. As $\mathcal{M}^+_{\mathbb{P}}$ is weakly dense in $\mathcal{M}^+$ and the identity mapping $\xi \mapsto \xi$ is continuous, $\tilde{\mathcal{K}}_{\mathbb{P}}$ is dense in $\tilde{\mathcal{K}}$. Keeping in mind that $\tilde{\mathcal{K}}_{\mathbb{P}}$ is also convex, the above findings imply

$$\emptyset \neq \operatorname{int} \tilde{\mathcal{K}} \subset \tilde{\mathcal{K}}_{\mathbb{P}} \subset \tilde{\mathcal{K}}. \tag{2.7}$$

The geometry of Ξ further implies that each $z \in \tilde{\mathcal{K}}$, $z \neq 0$, satisfies $e_1^\top z > 1$. By [23, Theorem 6.20] the convex closed cone $\tilde{\mathcal{K}}$ coincides with the intersection of all half-spaces that contain Ξ and have the origin as a boundary point. Thus, $z \in \tilde{\mathcal{K}}$ iff

$$0 \leq \inf_{\pi \in \mathbb{R}^k} \left\{ \pi^\top z : \pi^\top \xi \geq 0 \; \forall \xi \in \Xi \right\}.$$

Proposition 1 allows us to reexpress the semi-infinite constraint in the above minimization problem as a simple linear constraint. Hence, $z \in \tilde{\mathcal{K}}$ iff

$$\begin{aligned} 0 &\leq \inf_{\pi \in \mathbb{R}^k, \lambda \in \mathbb{R}^l} \left\{ \pi^\top z : \pi = W^\top \lambda, \, h^\top \lambda \geq 0, \, \lambda \geq 0 \right\} \\ &= \inf_{\lambda \in \mathbb{R}^l} \left\{ \lambda^\top W z : h^\top \lambda \geq 0, \, \lambda \geq 0 \right\} \\ &= \sup_{\psi \in \mathbb{R}} \left\{ 0 : Wz - h\psi \geq 0, \, \psi \geq 0 \right\}. \end{aligned}$$

Here, the equality in the last line follows from strong linear programming duality, which holds since $\lambda = 0$ is feasible in the primal problem. By the special structure of W and h imposed in (2.1b), the only feasible dual solution is $\psi = e_1^\top z$. Thus, $z \in \tilde{\mathcal{K}}$ iff

$$(W - h e_1^\top) z \geq 0 \quad \text{and} \quad e_1^\top z \geq 0.$$

Note that the constraint $e_1^\top z \geq 1$ is redundant since $e_1^\top z > 1$ for all $z \in \tilde{\mathcal{K}}$ with $z \neq 0$. Therefore, we conclude that $z \in \tilde{\mathcal{K}}$ iff $z \in \mathcal{K}$. This implies $\tilde{\mathcal{K}} = \mathcal{K}$.

The inclusion (2.7) thus translates to $\emptyset \neq \operatorname{int} \mathcal{K} \subset \mathcal{K}_{\mathbb{P}} \subset \mathcal{K}$. □

In the following, we interpret the Cartesian products $\mathcal{K}^m$ and $\mathcal{K}^m_{\mathbb{P}}$ as cones in the space of $m \times k$-matrices. Recall that the last constraint in (2.6) is equivalent to $SM \in \mathcal{K}^m_{\mathbb{P}}$, see also Remark 2. By removing the (redundant) penultimate constraint in (2.6), reexpressing the last constraint as $SM \in \mathcal{K}^m$, and using the definition of $\mathcal{K}$, we obtain the following finite-dimensional linear program whose size is polynomial in k, l, m, and n.

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(M C^\top X \right) \\ \text{subject to} & X \in \mathbb{R}^{n \times k}, S \in \mathbb{R}^{m \times k} \\ & AX + S = B \\ & \left(W - h e_1^\top \right) M S^\top \geq 0 \end{array} \tag{2.8}$$

By Proposition 3 we have $\mathcal{K}^m_{\mathbb{P}} \subset \mathcal{K}^m$, which implies that (2.8) constitutes a relaxation of (2.6). Thus, the optimal value of (2.8) provides a lower bound on the optimal value of (2.6) (as well as on the optimal value of $\mathcal{SP}^l$). Proposition 3 further asserts that $\operatorname{int}\mathcal{K}^m \subset \mathcal{K}^m_{\mathbb{P}}$. Thus, sharpening all inequalities in (2.8) to be strict leads to a problem that is more restrictive than (2.6). The optimal values of (2.8) and its modification with strict inequalities therefore bracket the optimal value of (2.6). However, all three optimal values will coincide whenever the original problem $\mathcal{SP}$ satisfies a strict feasibility condition.

Proposition 4 *Suppose that problem $\mathcal{SP}$ is strictly feasible, that is, there exists a tolerance $\varepsilon > 0$ and a pair of decision rules $\bar{x} \in \mathcal{L}^2_{k,n}$ and $\bar{s} \in \mathcal{L}^2_{k,m}$ with*

$$A\bar{x}(\xi) + \bar{s}(\xi) = b(\xi) \quad \textit{and} \quad \bar{s}(\xi) \geq \varepsilon\, e \quad \mathbb{P}\text{-a.s.}, \tag{2.9}$$

where e is the vector in $\mathbb{R}^m$ all of whose components are equal to 1. Then, the optimal values of (2.6) and (2.8) coincide.

Proof We first show that the linear program (2.8) inherits strict feasibility from the original problem $\mathcal{SP}$. To this end, we define $\bar{X} \in \mathbb{R}^{n\times k}$ and $\bar{S} \in \mathbb{R}^{m\times k}$ via

$$\bar{X}M = \mathbb{E}\left(\bar{x}(\xi)\xi^\top\right) \quad \text{and} \quad \bar{S}M = \mathbb{E}\left(\bar{s}(\xi)\xi^\top\right).$$

By the postulated properties of $\bar{x}$ and $\bar{s}$, it is easy to see that $A\bar{X} + \bar{S} = B$. As the support of $\mathbb{P}$ coincides with the convex set Ξ, we must have $\mathbb{E}(\xi) \in \operatorname{int}\Xi$, which implies

$$(W - he_1^\top)M\bar{S}^\top = \mathbb{E}\left((W - he_1^\top)\xi\,\bar{s}(\xi)^\top\right) \geq \varepsilon\,(W\mathbb{E}(\xi) - h)\,e^\top > 0\,.$$

Thus, $(\bar{X}, \bar{S})$ is strictly feasible in (2.8).

Strict feasibility guarantees that for any (X, S) feasible in (2.8) there exist strictly feasible matrices (X_ν, S_ν), $\nu \in \mathbb{N}$, which converge to (X, S). In fact, (X_ν, S_ν) can be obtained by taking suitable convex combinations of $(\bar{X}, \bar{S})$ and (X, S). By Proposition 3 we have $S_\nu \in \operatorname{int}\mathcal{K}^m \subset \mathcal{K}^m_{\mathbb{P}}$ for each $\nu \in \mathbb{N}$. Thus, all the (X_ν, S_ν) are feasible in (2.6). As (X, S) was chosen freely within the feasible set of (2.8), and since (2.6) and (2.8) share the same continuous objective function, we conclude that the optimal value of (2.6) is no larger than that of (2.8). The converse inequality follows immediately from Proposition 3. □

Remark 3 One can directly verify that (2.8) is less restrictive than the upper bounding problem (2.3). For any (X, Λ) feasible in (2.3) we set $S := \Lambda W$. Then, the objective value of (X, S) in (2.8) coincides with the objective value of (X, Λ) in (2.3). Moreover, the decision (X, S) thus constructed is feasible in (2.8). Indeed, the equality $AX + S = B$ follows from the definition of S and feasibility of (X, Λ) in (2.3). Furthermore, we have

$$\left(W - h\,e_1^\top\right)MS^\top = \mathbb{E}\left[\left(W - h\,e_1^\top\right)\xi\,\xi^\top W^\top \Lambda^\top\right] \geq 0\,,$$

where the inequality holds since $\Lambda \geq 0$ and $\Lambda h \geq 0$, while $W\xi \geq h$ and $e_1^\top \xi = 1$ for all ξ in the support of $\mathbb{P}$.

Remark 4 Our lower bound approximation is based on dualizing the equality constraints in problem (2.2). Alternatively, we could have directly dualized the inequality constraints in the original problem $\mathcal{SP}$ (without slack variables). It can be shown that this alternative approach results in the same lower bound. However, this alternative derivation of the linear program (2.8) permits no clear separation between the application of the decision rule approximation, which transforms $\mathcal{SP}$ to a semi-infinite problem of the type $\mathcal{SP}^l$, and the elimination of the semi-infinite constraints.

Remark 5 Different lower bounds can be obtained from alternative min-max reformulations of problem (2.2). Instead of (2.4), our derivation could start from the min-max problem

$$\begin{array}{ll} \underset{x\in\mathcal{L}^2_{k,n},\, s\in\mathcal{L}^2_{k,m}}{\text{minimize}} & \sup_{y\in\mathcal{L}^2_{k,m}} \mathbb{E}\left(c(\xi)^\top x(\xi)\right) + \mathbb{E}^{\mathbb{Q}}\left(y(\xi)^\top \left[Ax(\xi) + s(\xi) - b(\xi)\right]\right) \\ \text{subject to} & s(\xi) \geq 0 \quad \mathbb{P}\text{-a.s.,} \end{array} \tag{2.10}$$

where $\mathbb{Q}$ is a probability measure equivalent to $\mathbb{P}$ with bounded density function $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}(\xi)$, while $\mathbb{E}^{\mathbb{Q}}(\cdot)$ denotes expectation with respect to $\mathbb{Q}$. However, restricting the dual decisions in (2.10) to be linear in ξ yields a relaxation of problem $\mathcal{SP}$ whose optimal value is always smaller than that of $\mathcal{SP}^l$. The best lower bound is in fact obtained by setting $\mathbb{Q} = \mathbb{P}$. Details are omitted for brevity of exposition.

For later reference we pool the central insights of this section in the following main theorem.

Theorem 1 *If $\mathbb{P}$ has a polyhedral support of the type* (2.1) *while $\mathcal{SP}$ has fixed recourse and is strictly feasible, then $\mathcal{SP}^u$ and $\mathcal{SP}^l$ are equivalent to the linear programs* (2.3) *and* (2.8)*, respectively. The sizes of these linear programs are polynomial in k, l, m, and n, implying that they are efficiently solvable.*

3 One-stage stochastic programs with random recourse

Consider again the one-stage stochastic program $\mathcal{SP}$ with linearly parameterized cost coefficients $c(\xi) = C\xi$ and right hand side vector $b(\xi) = B\xi$. In contrast to Section 2, we assume here that the recourse matrix $A(\xi)$ depends also linearly on the uncertain parameters. The μth row of $A(\xi)$ is thus representable as $\xi^\top A_\mu$ for some matrix $A_\mu \in \mathbb{R}^{k\times n}$, where μ ranges from 1 to m. Moreover, as a notational convention, the μth row of the matrix B is denoted by $b_\mu^\top$.

In the following, we equip the space $\mathbb{R}^{k\times k}$ with the trace scalar product defined through $\langle R, S\rangle := \mathrm{Tr}(R^\top S)$, $R, S \in \mathbb{R}^{k\times k}$. Furthermore, we denote by $\mathbb{S}$ the subspace of symmetric matrices in $\mathbb{R}^{k\times k}$. For technical reasons, the support of the probability measure $\mathbb{P}$ is now assumed to be representable as

$$\varXi = \left\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1,\, \xi^\top W_\ell\, \xi \geq 0,\, \ell = 1, \ldots, l\right\} \tag{3.11}$$

for some matrices $W_\ell \in \mathbb{S}$. By construction, the first component of every $\xi \in \varXi$ is again equal to 1. As in Section 2 we assume $\varXi$ to be nonempty and bounded, while the linear hull of $\varXi$ is assumed to coincide with $\mathbb{R}^k$. Note that $\varXi$ spans $\mathbb{R}^k$ iff it has dimension $k-1$ or, equivalently, the system $\xi^\top W_\ell\, \xi \geq 0$, $\ell = 1, \ldots, l$, is strictly feasible. The next proposition shows that the quadratically constrained uncertainty sets of the type (3.11) are indeed more general than the polyhedral ones considered in Section 2.

Proposition 5 *The uncertainty sets of the form* (3.11) *cover all compact convex polytopes in the hyperplane* $\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1\}$.

Remark 6 A similar result is reported in [6, Remark 3.1] in the context of robust optimization.

Proof (Proof of Proposition 5) We denote by ξ_1 the first component and by ξ_{-1} the subvector of the $k-1$ last components of $\xi \in \mathbb{R}^k$, respectively. Any two real numbers $a < b$ and any vector $v_{-1} \in \mathbb{R}^{k-1}$ determine a stripe-shaped subset of $\mathbb{R}^{k-1}$ of the form

$$C := \left\{\xi_{-1} \in \mathbb{R}^{k-1} : a \leq (v_{-1})^\top \xi_{-1} \leq b\right\}. \tag{3.12}$$

In order to reexpress C in terms of quadratic constraints, we define

$$v := \begin{pmatrix} -\frac{a+b}{2} \\ v_{-1} \end{pmatrix} \quad \text{and} \quad W := \frac{(b-a)^2}{4}\, e_1 e_1^\top - v\, v^\top .$$

By construction, W is an element of $\mathbb{S}$, and the submatrix obtained by removing the first row and the first column of W is negative semidefinite. The following equivalences hold for all $\xi \in \mathbb{R}^k$.

$$\begin{aligned} \xi^\top W \xi \geq 0 \iff & \left(v^\top \xi\right)^2 \leq \frac{(b-a)^2}{4} \left(e_1^\top \xi\right)^2 \\ \iff & -\frac{b-a}{2}\,|\xi_1| \leq v^\top \xi \leq \frac{b-a}{2}\,|\xi_1| \\ \iff & -\frac{b-a}{2}\,|\xi_1| + \frac{a+b}{2}\,\xi_1 \leq (v_{-1})^\top \xi_{-1} \leq \frac{b-a}{2}\,|\xi_1| + \frac{a+b}{2}\,\xi_1 \end{aligned}$$

For $\xi_1 = 1$, the last expression is equivalent to the double constraint in (3.12). Therefore, we may identify C with the projection of

$$\left\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1,\ \xi^\top W \xi \geq 0\right\}$$

to the $k-1$ last components of ξ. Since every compact convex polytope in $\mathbb{R}^{k-1}$ is representable as a finite intersection of stripe-shaped sets of the form (3.12), the claim follows. □

Using the terminology of Ben-Tal and Nemirovski [5], we remark that the sets of the type (3.11) cover not only all polyhedral but also all ∩-ellipsoidal uncertainty sets restricted to the hyperplane $\{\xi \in \mathbb{R}^k : e_1^\top \xi = 1\}$. The proof of this statement is omitted for brevity of exposition.

3.1 Primal approximation

It seems clear that by allowing the recourse matrix to depend on ξ and by considering quadratically constrained uncertainty sets, the computational tractability of problem $\mathcal{SP}$ and its equivalent reformulation (2.2) is further reduced. In order to make (2.2) tractable, we again restrict attention to the subspace of linear decision rules. Thus, we require that $x(\xi) = X\xi$ for some matrix $X \in \mathbb{R}^{n\times k}$. With this simplification, however, the product term $A(\xi)x(\xi)$ becomes a *quadratic* function of the uncertain parameters. Since the right hand side vector $b(\xi)$ is *linear* in ξ, the equality constraints of the underlying stochastic program are only satisfiable if the slack variables exhibit a *quadratic* dependence on ξ. Hence, we must require that $s_\mu(\xi) = \xi^\top S_\mu \xi$ for some (without loss of generality symmetric) matrices $S_\mu \in \mathbb{S}$, where μ ranges from 1 to m. With these conventions, problem (2.2) reduces to

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k},\ S = (S_1,\ldots,S_m) \in \mathbb{S}^m \\ & \left.\begin{array}{l} \xi^\top A_\mu X \xi + \xi^\top S_\mu \xi = b_\mu^\top \xi \\ \xi^\top S_\mu \xi \geq 0 \end{array}\right\} \mathbb{P}\text{-a.s., } \mu = 1,\ldots,m . \end{array} \qquad (\mathcal{SP}^u)$$

3.2 Tractable reformulation of primal approximation

Although the number of decision variables is now finite, the quadratic semi-infinite constraints in $\mathcal{SP}^u$ look severely intractable. Fortunately, robust optimization technology can again provide remedy. As all constraint functions are continuous in ξ, the μth equality constraint in $\mathcal{SP}^u$ is equivalent to

$$\xi^\top H_\mu \xi = 0 \quad \forall \xi \in \Xi\,, \tag{3.13}$$

where $H_\mu \in \mathbb{S}$ is defined as

$$H_\mu := \frac{1}{2}\left(A_\mu X + X^\top A_\mu^\top - e_1 b_\mu^\top - b_\mu e_1^\top\right) + S_\mu\,.$$

Equation (3.13) not only holds on Ξ, but also on cone(Ξ), that is, the cone generated by Ξ. Hence, the Hessian of the mapping $\xi \mapsto \xi^\top H_\mu \xi$ vanishes on the interior of cone(Ξ). Since the Hessian of the above mapping is given by $2H_\mu$ and the interior of cone(Ξ) is nonempty (recall that Ξ is assumed to span $\mathbb{R}^k$), we conclude that $H_\mu = 0$. Therefore, the semi-infinite equality constraints in $\mathcal{SP}^u$ are equivalent to the requirement that $H_\mu = 0$ for all $\mu = 1, \ldots, m$.

Next, we show how to approximate the semi-infinite inequality constraints by a system of linear matrix inequalities (LMI). This approach was first developed in a robust optimization context [4, 7] and relies partly on the following important result from matrix analysis, see e.g. [11].

Lemma 1 ($\mathcal{S}$-lemma) *Consider two matrices $W, S \in \mathbb{S}$ and assume that the inequality $\xi^\top W \xi \geq 0$ is strictly feasible, that is, $\bar{\xi}^\top W \bar{\xi} > 0$ for some $\bar{\xi} \in \mathbb{R}^k$. Then, the following equivalence holds:*

$$\left[\xi^\top W \xi \geq 0 \;\Rightarrow\; \xi^\top S \xi \geq 0\right] \quad \Longleftrightarrow \quad \exists \lambda \geq 0 :\; S \succeq \lambda W\,.$$

The $\mathcal{S}$-lemma is a central ingredient for the following proposition, which is inspired by Theorem 4.1 in [4].

Proposition 6 *Consider the following two statements for some fixed $S \in \mathbb{S}$:*

(i) $\exists \lambda \in \mathbb{R}^l$ *with* $\lambda \geq 0$ *and* $S - \sum_{\ell=1}^{l} \lambda_\ell W_\ell \succeq 0$*;*
(ii) $\xi^\top S \xi \geq 0$ $\mathbb{P}$*-a.s.*

For any $l \in \mathbb{N}$, (i) implies (ii). The converse implication holds if $l = 1$.

Proof Select any $\xi \in \Xi$. Under the assumptions of statement (i) we have

$$0 \leq \xi^\top \left[S - \sum_{\ell=1}^{l} \lambda_\ell W_\ell\right] \xi \;= \xi^\top S \xi - \sum_{\ell=1}^{l} \lambda_\ell \xi^\top W_\ell \xi \leq \xi^\top S \xi\,,$$

where the first inequality follows from positive semidefiniteness of the matrix in square brackets, and the second inequality holds since $\lambda_\ell \geq 0$ while $\xi^\top W_\ell \xi \geq 0$ for all $\xi \in \Xi$. Since the choice of $\xi \in \Xi$ was arbitrary, statement (ii) follows.

We show now that (ii) implies (i) if $l = 1$. Since quadratic functions are continuous, statement (ii) effectively asserts that $\xi^\top S \xi \geq 0$ for all $\xi \in \Xi$. In fact, this inequality readily extends to the double cone generated by Ξ. As Ξ is assumed to be bounded and nonempty, there exists no $\xi \neq 0$ with $\xi^\top W_1 \xi \geq 0$ and $e_1^\top \xi = 0$ (note that a vector with these properties would constitute a recession direction along which Ξ would be unbounded). Hence, the double cone generated by Ξ coincides with the feasible set of the inequality $\xi^\top W_1 \xi \geq 0$. So far we have thus shown that $\xi^\top W_1 \xi \geq 0$ implies $\xi^\top S \xi \geq 0$.

As Ξ has nonempty relative interior, the inequality $\xi^\top W_1 \xi \geq 0$ is strictly feasible. The $\mathcal{S}$-Lemma then implies that there exists $\lambda_1 \geq 0$ such that $S - \lambda_1 W_1$ is positive semidefinite. □

Remark 7 It seems unsatisfactory that we failed to reexpress the semi-infinite inequality constraints in problem $\mathcal{SP}^u$ in terms of a tractable constraint system in the case $l > 1$. If this was possible, however, we could devise an efficient algorithm for the NP-hard problem of checking whether a given square matrix is copositive, see [4, Example 3.1].

We can use Proposition 6 to replace the semi-infinite inequality constraint $\xi^\top S_\mu \xi \geq 0$ $\mathbb{P}$-a.s. by a set of semidefinite constraints in (S_μ, λ_μ) for some new decision vector $\lambda_\mu \in \mathbb{R}^l$, $\mu = 1, \ldots, m$. Interpreting $\lambda_\mu^\top$ as the μth row of a matrix $\Lambda \in \mathbb{R}^{m \times l}$, problem $\mathcal{SP}^u$ simplifies to the following semidefinite program (SDP).

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n \times k},\ S = (S_1, \ldots, S_m) \in \mathbb{S}^m,\ \Lambda \in \mathbb{R}^{m \times l} \\ & \frac{1}{2}\left(A_\mu X + X^\top A_\mu^\top\right) + S_\mu = \frac{1}{2}\left(e_1 b_\mu^\top + b_\mu e_1^\top\right) \quad \forall\, \mu = 1, \ldots, m \\ & S_\mu - \sum_{\ell=1}^{l} \Lambda_{\mu\ell} W_\ell \succeq 0 \quad \forall\, \mu = 1, \ldots, m \\ & \Lambda \geq 0 \end{array} \tag{3.14}$$

Proposition 6 implies that (3.14) constitutes a conservative approximation for $\mathcal{SP}^u$ whenever $l > 1$. Equivalence of (3.14) and $\mathcal{SP}^u$ only holds in the special case $l = 1$. The benefit of using linear decision rules lies again in the fact that the size of the SDP (3.14) grows only polynomially with k, l, m, and n, while SDPs can be solved efficiently by means of modern interior-point algorithms [29].

3.3 Dual approximation

Similar techniques that were used to derive the SDP (3.14) can also be used to find a computationally tractable lower bound on $\mathcal{SP}$. To this end, we follow the same general strategy as in Section 2, considering first the min-max version (2.4) of problem (2.2), applying a suitable restriction to the set of *dual* decision rules, and simplifying the arising problem by means of a *dual* version of Proposition 6. In the following derivation, we let f_μ be a shorthand notation for the μth constraint function, which is defined through

$$f_\mu : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}\,, \quad f_\mu(x, s_\mu, \xi) := \xi^\top A_\mu x + s_\mu - b_\mu^\top \xi\,.$$

Moreover, we let '$\inf_{x,s}$' denote the infimum operator over all $x \in \mathcal{L}^2_{k,n}$ and over all $s = (s_1, \ldots, s_m) \in \mathcal{L}^2_{k,m}$ that are almost surely nonnegative. Similarly, '$\sup_y$' and '$\sup_Y$' stand for the supremum operators over all $y = (y_1, \ldots, y_m) \in \mathcal{L}^2_{k,m}$ and $Y = (Y_1, \ldots, Y_m) \in \mathbb{S}^m$, respectively. The equivalence of $\mathcal{SP}$ and (2.4) implies

$$\begin{aligned} \inf \mathcal{SP} &= \inf_{x,s} \sup_{y} \mathbb{E}\left(c(\xi)^\top x(\xi) + \sum_{\mu=1}^{m} y_\mu(\xi)\ f_\mu\left[x(\xi), s_\mu(\xi), \xi\right]\right) \\ &\geq \inf_{x,s} \sup_{Y} \mathbb{E}\left(c(\xi)^\top x(\xi) + \sum_{\mu=1}^{m} \xi^\top Y_\mu \xi\ f_\mu\left[x(\xi), s_\mu(\xi), \xi\right]\right) \\ &= \inf_{x,s} \sup_{Y} \mathbb{E}\left(c(\xi)^\top x(\xi)\right) + \sum_{\mu=1}^{m} \mathrm{Tr}\left[Y_\mu \mathbb{E}\left(\xi\ f_\mu\left[x(\xi), s_\mu(\xi), \xi\right]\ \xi^\top\right)\right]. \end{aligned}$$

Here, in the second line we require the dual decisions to be representable as $y_\mu(\xi) = \xi^\top Y_\mu \xi$ for some $Y_\mu \in \mathbb{S}$, where μ ranges from 1 to m. Thus, we restrict the dual feasible set to contain only *quadratic* decision rules. The maximization in the last line can be carried out explicitly, which implies that the optimal value of $\mathcal{SP}$ is bounded below by that of the following approximate problem.

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left(c(\xi)^\top x(\xi)\right) \\ \text{subject to} & x \in \mathcal{L}^2_{k,n}\,,\; s = (s_1, \ldots, s_m) \in \mathcal{L}^2_{k,m} \\ & \left.\begin{array}{r} \mathbb{E}\left[\xi\, f_\mu\left(x(\xi), s_\mu(\xi), \xi\right)\, \xi^\top\right] = 0 \\ s_\mu(\xi) \geq 0 \end{array}\right\} \;\mathbb{P}\text{-a.s., } \mu = 1 \ldots, m\,. \end{array} \tag{$\mathcal{SP}^l$}$$

Observe that any (x, s) feasible in (2.2) satisfies $f_\mu(x(\xi), s(\xi), \xi) = 0$ $\mathbb{P}$-a.s. and will therefore also satisfy the less restrictive expectation constraint in $\mathcal{SP}^l$. This confirms a posteriori that the optimal value of $\mathcal{SP}^l$ provides a lower bound on the optimal value of $\mathcal{SP}$.

3.4 Tractable reformulation of dual approximation

The relaxed problem $\mathcal{SP}^l$ still involves functional decision variables and a continuum of inequality constraints. In the remainder of this section we will argue that $\mathcal{SP}^l$ can be progressively approximated by a tractable SDP. This approximation relies on techniques that we develop ad hoc; in particular, we invoke Proposition 8, which constitutes a dual version of Proposition 6. Our derivation will involve the linear mapping $Q : \mathbb{R}^{k\times k} \to \mathbb{R}^{k\times k}$ defined through

$$e_\alpha^\top Q(e_\beta e_\gamma^\top) e_\delta = \mathbb{E}\left(\xi_\alpha\, \xi_\beta\, \xi_\gamma\, \xi_\delta\right) \quad \text{for } 1 \leq \alpha, \beta, \gamma, \delta \leq k\,,$$

where $\{e_\alpha\}_{\alpha=1}^k$ denotes the standard basis of $\mathbb{R}^k$. Notice that Q can be interpreted as the totally symmetric tensor of all moments of $\mathbb{P}$ up to the fourth order. Since $e_1^\top \xi = 1$ for all ξ in the support of $\mathbb{P}$, we have $Q(e_1 e_1^\top) = M$, that is, Q maps $e_1 e_1^\top$ to the matrix of second order moments of $\mathbb{P}$. Moreover, Q is self-adjoint with respect to the trace scalar product,

$$\mathrm{Tr}\left[A^\top Q(B)\right] = \mathrm{Tr}\left[Q(A)^\top B\right] \quad \forall\, A, B \in \mathbb{R}^{k\times k},$$

and invariant under transposition of its argument, $Q(A) = Q(A^\top)$ $\forall\, A \in \mathbb{R}^{k\times k}$. The following result will be crucial for our subsequent analysis.

Proposition 7 *The restriction of Q to $\mathbb{S}$ is positive definite and invertible.*

Proof Q is positive semidefinite because

$$\langle S, Q(S)\rangle = \mathrm{Tr}(SQ(S)) = \mathbb{E}\left(\left[\xi^\top S\, \xi\right]^2\right) \geq 0 \quad \forall S \in \mathbb{S}\,.$$

It remains to be shown that Q is invertible. Assume now that Q is not invertible. Thus, there is a matrix $S \in \mathbb{S}$, $S \neq 0$, such that $\langle S, Q(S)\rangle = 0$. This implies that $S\xi = 0$ with probability one, that is, Ξ is a subset of the null space of S. Since $S \neq 0$, this null space has not full dimension, and thus Ξ cannot span $\mathbb{R}^k$. This contradicts our assumption about Ξ, and hence Q must be invertible. □

To simplify problem $\mathcal{SP}^l$, we introduce new decision variables $X \in \mathbb{R}^{n\times k}$ and $S_\mu \in \mathbb{S}$, $\mu = 1, \ldots, m$, which are related to the existing functional decision variables through the equations

$$\mathbb{E}\left(\xi\, \xi^\top R\, x(\xi) \xi^\top\right) = Q(RX) \quad \forall\, R \in \mathbb{R}^{k\times n}, \tag{3.15a}$$

$$\mathbb{E}\left(s_\mu(\xi)\, \xi\xi^\top\right) = Q(S_\mu)\,. \tag{3.15b}$$

Note that it is sufficient to enforce the equality in (3.15a) only for a (finite) set of basis matrices. By linearity of Q, the equality will then hold for all $R \in \mathbb{R}^{n\times k}$. The relation (3.15a) allows us now to reexpress the objective function of $\mathcal{SP}^l$ in terms of X. In fact, we have

$$\begin{aligned}
\mathbb{E}\left(c(\xi)^\top x(\xi)\right) &= e_1^\top \mathbb{E}\left(\xi\,\xi^\top C^\top x(\xi)\xi^\top\right) e_1 \\
&= \mathrm{Tr}\left(e_1 e_1^\top Q(C^\top X)\right) \\
&= \mathrm{Tr}\left(Q(e_1 e_1^\top) C^\top X\right) \\
&= \mathrm{Tr}\left(M C^\top X\right),
\end{aligned}$$

where the first equality holds since $e_1^\top \xi = 1$ on the support of $\mathbb{P}$. The other equalities follow from (3.15a), self-adjointness of Q, and the fact that $Q(e_1 e_1^\top)$ corresponds to the second-order moment matrix of $\mathbb{P}$. Next, by virtue of (3.15), the μth equality constraint in $\mathcal{SP}^l$ reduces to

$$\begin{aligned}
0 &= \mathbb{E}\left(\xi\,\xi^\top A_\mu\, x(\xi)\xi^\top\right) + \mathbb{E}\left(s_\mu(\xi)\,\xi\xi^\top\right) - \mathbb{E}\left(\xi\xi^\top\,\xi\xi^\top\, e_1 b_\mu^\top\right) \\
&= Q(A_\mu X) + Q(S_\mu) - Q(e_1 b_\mu^\top) \\
&= Q\left(\tfrac{1}{2}[A_\mu X + X^\top A_\mu^\top] + S_\mu - \tfrac{1}{2}[e_1 b_\mu^\top + b_\mu e_1^\top]\right).
\end{aligned}$$

Here, we appended again an irrelevant factor $\xi^\top e_1$ to the last term in the first line. The third equality of the above expression is based on linearity and invariance of Q under transposition of its argument. Since the restriction of Q to $\mathbb{S}$ is invertible, the μth equality constraint in $\mathcal{SP}^l$ can therefore be represented as

$$\frac{1}{2}\left(A_\mu X + X^\top A_\mu^\top\right) + S_\mu = \frac{1}{2}\left(e_1 b_\mu^\top + b_\mu e_1^\top\right).$$

Hence, problem $\mathcal{SP}^l$ has the following equivalent reformulation.

$$\begin{array}{ll}
\text{minimize} & \mathrm{Tr}\left(M C^\top X\right) \\
\text{subject to} & X \in \mathbb{R}^{n\times k}, S = (S_1, \ldots, S_m) \in \mathbb{S}^m \\
& \tfrac{1}{2}\left(A_\mu X + X^\top A_\mu^\top\right) + S_\mu = \tfrac{1}{2}\left(e_1 b_\mu^\top + b_\mu e_1^\top\right)\ \forall\, \mu = 1, \ldots, m \\
& \exists\, x \in \mathcal{L}^2_{k,n}\ \ : x \text{ and } X \text{ satisfy (3.15a)} \\
& \exists\, s_\mu \in \mathcal{L}^2_{k,1} : s_\mu \text{ and } S_\mu \text{ satisfy (3.15b) and} \\
& \qquad\qquad\quad s_\mu(\xi) \geq 0\ \mathbb{P}\text{-a.s. } \forall\, \mu = 1, \ldots, m
\end{array} \tag{3.16}$$

The first existence constraint in (3.16) is redundant and can be deleted. Indeed, for any $X \in \mathbb{R}^{n\times k}$ the linear decision rule $x(\xi) := X\xi$ satisfies condition (3.15a). However, the second existence constraint is restrictive as it involves m inherent moment problems: $S_\mu \in \mathbb{S}$ is feasible in (3.16) only if there exists a nonnegative Borel measure whose second-order moment matrix is given by $Q(S_\mu)$ and which has a square-integrable density s_μ with respect to $\mathbb{P}$, $\mu = 1, \ldots, m$. We will now elaborate a progressive approximation of this constraint in terms of finitely many LMIs. To this end, we need the following *dual* version of Proposition 6.

Proposition 8 *Consider the following two convex cones in* $\mathbb{S}$*:*

$$\begin{aligned}
\mathcal{C} &:= \{S \in \mathbb{S} : S \succeq 0,\ \mathrm{Tr}\,(W_\ell S) \geq 0\ \forall\, \ell = 1, \ldots, l\}, \\
\mathcal{C}_\mathbb{P} &:= \left\{S \in \mathbb{S} : \exists\, s \in \mathcal{L}^2_{k,1} \text{ with } \mathbb{E}\left(s(\xi)\,\xi\,\xi^\top\right) = S \text{ and } s(\xi) \geq 0\ \mathbb{P}\text{-a.s.}\right\}.
\end{aligned}$$

Then, $\mathcal{C}_\mathbb{P} \subset \mathcal{C}$ *for each* $l \in \mathbb{N}$*, and* $\mathcal{C}_\mathbb{P} \supset \mathrm{int}\,\mathcal{C} \neq \emptyset$ *for* $l = 1$*.*

Remark 8 Note that the last constraint in problem (3.16) is equivalent to the requirement that $Q(S_\mu) \in \mathcal{C}_{\mathbb{P}}$ for each $\mu = 1, \dots, m$.

Proof (Proof of Proposition 8) Fix an arbitrary $S \in \mathcal{C}_{\mathbb{P}}$, and let s be a corresponding scalar function that satisfies the conditions in the definition of $\mathcal{C}_{\mathbb{P}}$. For any $v \in \mathbb{R}^k$ we then have

$$v^\top S\, v = \mathbb{E}\left(s(\xi) \left[v^\top \xi \right]^2 \right) \geq 0\,,$$

where the inequality holds since $s(\xi)$ is nonnegative on the support of $\mathbb{P}$. Thus, S is positive semidefinite. Moreover, we find

$$\mathrm{Tr}\,(W_\ell S) = \mathbb{E}\left(s(\xi)\, \xi^\top W_\ell\, \xi \right) \geq 0 \quad \forall\, \ell = 1, \dots, l$$

since both $s(\xi) \geq 0$ and $\xi^\top W_\ell\, \xi \geq 0$ on the support of $\mathbb{P}$. Hence, $S \in \mathcal{C}$. As the choice of $S \in \mathcal{C}_{\mathbb{P}}$ was arbitrary, the postulated inclusion $\mathcal{C}_{\mathbb{P}} \subset \mathcal{C}$ follows.

Henceforth we assume that $l = 1$. As in the proof of Proposition 3, we denote by $\mathcal{M}^+$ the set of all nonnegative finite measures on $(\varXi, \mathcal{B}(\varXi))$ with finite second moments and let $\mathcal{M}^+_{\mathbb{P}}$ be the subset of measures in $\mathcal{M}^+$ that have a square-integrable density with respect to $\mathbb{P}$. Recall also that $\mathcal{M}^+_{\mathbb{P}}$ is dense in $\mathcal{M}^+$ with respect to the weak topology. Next, define

$$\tilde{\mathcal{C}} := \left\{ \int_{\varXi} \xi\xi^\top \mu(\mathrm{d}\xi) : \mu \in \mathcal{M}^+ \right\} \quad \text{and} \quad \tilde{\mathcal{C}}_{\mathbb{P}} := \left\{ \int_{\varXi} \xi\xi^\top \mu(\mathrm{d}\xi) : \mu \in \mathcal{M}^+_{\mathbb{P}} \right\}.$$

It is immediately clear that $\tilde{\mathcal{C}}_{\mathbb{P}} = \mathcal{C}_{\mathbb{P}} \subset \tilde{\mathcal{C}}$. Below, we will argue that $\tilde{\mathcal{C}}_{\mathbb{P}} \supset \mathrm{int}\, \tilde{\mathcal{C}} \neq \emptyset$, and $\tilde{\mathcal{C}} = \mathcal{C}$. This will establish the relation $\mathcal{C}_{\mathbb{P}} \supset \mathrm{int}\, \mathcal{C} \neq \emptyset$ for $l = 1$.

Observe that $\tilde{\mathcal{C}}$ is certainly a convex cone in $\mathbb{S}$. In the proof of Proposition 7 we have seen that for every $S \in \mathbb{S}$, $S \neq 0$, there exists a $\xi \in \varXi$ with $\mathrm{Tr}(S\, \xi\xi^\top) \neq 0$. Hence, the orthogonal complement of $\mathcal{C}_\delta := \{\xi\xi^\top : \xi \in \varXi\}$ in $\mathbb{S}$ is empty, which implies that $\mathcal{C}_\delta$ spans $\mathbb{S}$. Since $\mathcal{M}^+$ contains all Dirac measures δ_ξ for $\xi \in \varXi$, we conclude that $\mathcal{C}_\delta \subset \tilde{\mathcal{C}}$. Consequently, $\tilde{\mathcal{C}}$ spans $\mathbb{S}$, as well. Convexity of $\tilde{\mathcal{C}}$ then implies that the interior of $\tilde{\mathcal{C}}$ (with respect to the Euclidean topology on $\mathbb{S}$) is nonempty. As $\mathcal{M}^+_{\mathbb{P}}$ is weakly dense in $\mathcal{M}^+$ and the mapping $\xi \mapsto \xi\xi^\top$ is continuous, $\tilde{\mathcal{C}}_{\mathbb{P}}$ is dense in $\tilde{\mathcal{C}}$. Keeping in mind that $\tilde{\mathcal{C}}_{\mathbb{P}}$ is also convex, the above findings imply

$$\emptyset \neq \mathrm{int}\, \tilde{\mathcal{C}} \subset \tilde{\mathcal{C}}_{\mathbb{P}} \subset \tilde{\mathcal{C}}\,. \tag{3.17}$$

It can be shown that reducing the set of measures in the definition of $\tilde{\mathcal{C}}$ to those which are supported on at most $k + 1$ points has no influence on $\tilde{\mathcal{C}}$, see e.g. [24, Lemma 3.1]. Consequently, $\tilde{\mathcal{C}}$ coincides with the convex cone generated by $\mathcal{C}_\delta := \left\{ \xi\xi^\top : \xi \in \varXi \right\}$. Note that $\mathcal{C}_\delta$ is compact as the continuous image of a compact set, and therefore $\tilde{\mathcal{C}}$ is closed. By [23, Theorem 6.20] the closed convex cone $\tilde{\mathcal{C}}$ generated by $\mathcal{C}_\delta$ coincides with the intersection of all half-spaces that contain $\mathcal{C}_\delta$ and have the origin as a boundary point. Thus, $S \in \tilde{\mathcal{C}}$ iff

$$0 \leq \inf_{\varLambda \in \mathbb{S}} \left\{ \mathrm{Tr}(S\varLambda) : \mathrm{Tr}(\xi\xi^\top \varLambda) \geq 0\ \forall\, \xi \in \varXi \right\}.$$

For $l = 1$, the $\mathcal{S}$-lemma allows us to reexpress the semi-infinite constraint in the above minimization problem as a semidefinite constraint. Hence, $S \in \tilde{\mathcal{C}}$ iff

$$\begin{aligned} 0 &\leq \inf_{\varLambda \in \mathbb{S}, \lambda_1 \geq 0} \left\{ \mathrm{Tr}(S\varLambda) : \varLambda - \lambda_1 W_1 \succeq 0 \right\} \\ &= \sup_{\varGamma \in \mathbb{S}} \left\{ 0 : \varGamma = S,\ \mathrm{Tr}(W_1 \varGamma) \geq 0,\ \varGamma \succeq 0 \right\}. \end{aligned}$$

Here, the equality in the second line follows from strong semidefinite programming duality, which holds since the primal minimization problem is strictly feasible [27]. By inspecting the (degenerate) feasible set of the dual maximization problem, we conclude that $S \in \tilde{\mathcal{C}}$ if and only if $S \in \mathcal{C}$. This implies $\tilde{\mathcal{C}} = \mathcal{C}$.

For $l = 1$ the inclusion (3.17) thus translates to $\emptyset \neq \mathrm{int}\, \mathcal{C} \subset \mathcal{C}_{\mathbb{P}} \subset \mathcal{C}$. □

By removing the (redundant) penultimate constraint in (3.16) and by replacing the constraints $Q(S_\mu) \in \mathcal{C}_{\mathbb{P}}$ with their relaxed versions $Q(S_\mu) \in \mathcal{C}$ for $\mu = 1, \ldots, m$, we obtain the following SDP.

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, S = (S_1, \ldots, S_m) \in \mathbb{S}^m \\ & \frac{1}{2}\left(A_\mu X + X^\top A_\mu^\top\right) + S_\mu = \frac{1}{2}\left(e_1 b_\mu^\top + b_\mu e_1^\top\right) \quad \forall\, \mu = 1, \ldots, m \\ & \operatorname{Tr}\left(W_\ell\, Q(S_\mu)\right) \geq 0 \quad \forall\, \ell = 1, \ldots, l,\ m = 1, \ldots, m \\ & Q(S_\mu) \succeq 0 \quad \forall\, \mu = 1, \ldots, m \end{array} \tag{3.18}$$

The inclusion $\mathcal{C}_{\mathbb{P}} \subset \mathcal{C}$ implies that the optimal value of (3.18) underestimates the optimal value of (3.16) (as well as that of $\mathcal{SP}^l$). In the special case $l = 1$, however, one can show that the optimal values of (3.16) and (3.18) are generically equal.

Proposition 9 *Suppose that $l = 1$ and that problem $\mathcal{SP}$ is strictly feasible. Then, the optimal values of* (3.16) *and* (3.18) *coincide.*

Proof This result relies on the fact that $\operatorname{int}\mathcal{C} \subset \mathcal{C}_{\mathbb{P}}$ for $l = 1$. Its proof is a simple generalization of the proof of Proposition 4 and can thus be omitted. □

Remark 9 One can directly verify that upper bounding problem (3.14) is more restrictive than the lower bounding problem (3.18). To this end, we choose some (X, S, Λ) feasible in (3.14). As (3.14) and (3.18) have identical objective functions and equality constraints, the claim follows if we can show that (X, S) satisfies the inequality constraints in (3.18). First, we observe that

$$\operatorname{Tr}\left(W_\ell\, Q(S_\mu)\right) = \mathbb{E}\left(\xi^\top W_\ell\, \xi\, \xi^\top S_\mu \xi\right) \geq \sum_{\alpha=1}^{l} \Lambda_{\mu\alpha}\, \mathbb{E}\left(\xi^\top W_\ell\, \xi\, \xi^\top W_\alpha \xi\right) \geq 0\,,$$

where the inequalities hold since $\xi^\top W_\ell\, \xi \geq 0$ on the support of $\mathbb{P}$, while $S_\mu \succeq \sum_{\ell=1}^{l} \Lambda_{\mu\ell}\, W_\ell$. In a similar manner, one can show that

$$v^\top Q(S_\mu) v = \mathbb{E}\left([v^\top \xi]^2\, \xi^\top S_\mu \xi\right) \geq \sum_{\ell=1}^{l} \Lambda_{\mu\ell}\, \mathbb{E}\left([v^\top \xi]^2\, \xi^\top W_\ell\, \xi\right) \geq 0$$

for all $v \in \mathbb{R}^k$. Hence, $Q(S_\mu)$ is positive semidefinite.

A summary of the key results elaborated in this section is provided in the following theorem.

Theorem 2 *If $\mathbb{P}$ has a quadratically constrained support of the type* (3.11)*, then the SDP* (3.14) *conservatively approximates $\mathcal{SP}^u$, while the SDP* (3.18) *progressively approximates $\mathcal{SP}^l$. The sizes of these SDPs are polynomial in k, l, m, and n, implying that they are efficiently solvable. If $l = 1$ and $\mathcal{SP}$ is strictly feasible, then $\mathcal{SP}^u$ and $\mathcal{SP}^l$ are equivalent to* (3.14) *and* (3.18)*, respectively.*

4 Multistage stochastic programs

The dynamic decision problems to be considered in this section are cast in a similar framework as the one-stage stochastic programs of Section 2. In particular, we continue to work with the probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P})$ and assume that $\mathbb{P}$ has a polyhedral support Ξ of the type (2.1) that is nonempty, bounded, and spans $\mathbb{R}^k$. Here, however, we impose a temporal structure: the elements of the sample space are assumed to be representable as $\xi = (\xi_1, \ldots, \xi_T)$ where the subvectors $\xi_t \in \mathbb{R}^{k_t}$ are observed sequentially at time points indexed by $t \in \mathbb{T} := \{1, \dots, T\}$. Without loss of generality, we assume that $k_1 = 1$, which implies that $\xi_1 = 1$ for all $\xi \in \Xi$. The history of observations up to time t is denoted by $\xi^t := (\xi_1, \dots, \xi_t) \in \mathbb{R}^{k^t}$ where $k^t := \sum_{s=1}^t k_s$. Consistency then requires that $\xi^T = \xi$ and $k^T = k$. For notational convenience, we let $\mathbb{E}_t(\cdot)$ denote (a version of) the conditional expectation with respect to $\mathbb{P}$ given the random variable ξ^t. We further introduce truncation operators P_t, $t \in \mathbb{T}$, which are defined through

$$P_t : \mathbb{R}^k \to \mathbb{R}^{k^t}, \quad \xi \mapsto \xi^t.$$

Consider now a sequential decision process in which the decision $x_t(\xi^t) \in \mathbb{R}^{n_t}$ is selected at time t after the outcome history ξ^t has been observed but before the future outcomes $\{\xi_s\}_{s>t}$ have been revealed. The objective is to find a sequence of decision rules $x_t \in \mathcal{L}^2_{k^t,n_t}$, $t \in \mathbb{T}$, which map the available observations to decisions and minimize a linear expected cost function subject to certain linear constraints. The requirement that x_t depends solely on ξ^t reflects the *non-anticipative* nature of the dynamic decision problem at hand and essentially ensures its causality. Decision problems of this type can often be formulated as linear multistage stochastic programs of the following form.

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left(\sum_{t=1}^T c_t(\xi^t)^\top x_t(\xi^t)\right) \\ \text{subject to} & x_t \in \mathcal{L}^2_{k^t,n_t} \ \forall t \in \mathbb{T} \\ & \mathbb{E}_t\left(\sum_{s=1}^T A_{ts} x_s(\xi^s)\right) \leq b_t(\xi^t) \quad \mathbb{P}\text{-a.s. } \forall t \in \mathbb{T} \end{array} \tag{$\mathcal{MSP}$}$$

In order for $\mathcal{MSP}$ to be well-defined, we assume that the objective function coefficients and the right-hand side vectors are non-anticipative linear functions of the random parameters. Thus, we postulate that $c_t(\xi^t) = C_t P_t \xi$ for some $C_t \in \mathbb{R}^{n_t \times k^t}$ while $b_t(\xi^t) = B_t P_t \xi$ for some $B_t \in \mathbb{R}^{m_t \times k^t}$. The constraint matrices $A_{ts} \in \mathbb{R}^{m_t \times n_s}$ are assumed to be independent of ξ. Note that the constraints in $\mathcal{MSP}$ are of the expected value type. If $A_{ts} = 0$ for all $t < s$, then the conditional expectations in the constraints of $\mathcal{MSP}$ become redundant. In this case, the expectation constraints reduce to conventional pointwise constraints. However, our generalized formulation offers more flexibility in modelling the decision maker's risk attitude. For example, if $L \in \mathcal{L}^2_{k,1}$ represents a loss variable, a risk-averse decision maker may want to constrain the expected excess loss over a threshold $\underline{L}$ to be smaller than some tolerance $\overline{L}$. This requirement can be viewed as an integrated chance constraint in the sense of [19], and it can be expressed in terms of linear expectation constraints as

$$\mathbb{E}(\max\{L(\xi) - \underline{L}, 0\}) \leq \overline{L} \iff \exists s \in \mathcal{L}^2_{k,1} : \begin{array}{l} s(\xi) \geq L(\xi) - \underline{L},\ s(\xi) \geq 0 \quad \mathbb{P}\text{-a.s.}, \\ \mathbb{E}(s(\xi)) \leq \overline{L}\,. \end{array}$$

Alternatively, the decision maker may want to constrain the conditional value-at-risk (CVaR) of the loss. CVaR is a popular risk measure in decision theory which quantifies the expected losses exceeding a given percentile of the loss distribution [22]. The requirement that the CVaR at level $\beta \in (0, 1)$ of L should not exceed a given tolerance $\overline{L}$ can also be expressed in terms of linear expectation constraints.

$$\beta\text{-CVaR}(L) \leq \overline{L} \iff \exists s \in \mathcal{L}^2_{k,1},\ \alpha \in \mathbb{R} : \begin{array}{l} s(\xi) \geq L(\xi) - \alpha,\ s(\xi) \geq 0 \quad \mathbb{P}\text{-a.s.}, \\ \alpha + \dfrac{1}{1-\beta}\mathbb{E}(s(\xi)) \leq \overline{L} \end{array}$$

Furthermore, conditional expectation constraints can be used to explicitly enforce non-anticipativity of decisions variables. They can also arise as artefacts of a superordinate approximation based on time discretization [20].

By introducing a sequence of non-anticipative slack variables $s_t \in \mathcal{L}^2_{k^t,m_t}$, $t \in \mathbb{T}$, problem $\mathcal{MSP}$ can be brought to the following standard form.

$$\begin{aligned}
\text{minimize} \quad & \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\
\text{subject to} \quad & x_t \in \mathcal{L}^2_{k^t,n_t},\ s_t \in \mathcal{L}^2_{k^t,m_t}\ \forall t \in \mathbb{T} \\
& \left.\begin{aligned} \mathbb{E}_t\left(\sum_{s=1}^{T} A_{ts} x_s(\xi^s)\right) + s_t(\xi^t) = b_t(\xi^t) \\ s_t(\xi^t) \geq 0 \end{aligned}\right\} \ \mathbb{P}\text{-a.s. } \forall t \in \mathbb{T}
\end{aligned} \tag{4.1}$$

Shapiro and Nemirovski argue that $\mathcal{MSP}$ is generically computationally intractable even if 'medium-accuracy' solutions are sought and even if there are no expectation constraints while the probability measure $\mathbb{P}$ is 'easy to describe' [26]. As a drastic but effective means for complexity reduction they propose the use of linear decision rules. In our notation, this amounts to setting $x_t(\xi^t) = X_t P_t\, \xi$ for some $X_t \in \mathbb{R}^{n_t \times k^t}$ and $s_t(\xi^t) = S_t P_t\, \xi$ for some $S_t \in \mathbb{R}^{m_t \times k^t}$, $t \in \mathbb{T}$. To ensure that this approximation will convert (4.1) to a tractable problem, we require $\mathbb{E}_t(\xi)$ to be almost surely linear in ξ^t. This is the case, for instance, if the random variables $\{\xi_t\}_{t\in\mathbb{T}}$ are mutually independent.[3] Formally speaking, we postulate that for each $t \in \mathbb{T}$ there exists a matrix $M_t \in \mathbb{R}^{k \times k^t}$ such that almost surely $\mathbb{E}_t(\xi) = M_t P_t\, \xi$. Recall also that $M := \mathbb{E}(\xi\, \xi^\top)$. With these conventions, (4.1) reduces to

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \operatorname{Tr}\left(P_t M P_t^\top C_t^\top X_t\right) \\
\text{subject to} \quad & X_t \in \mathbb{R}^{n_t \times k^t},\ S_t \in \mathbb{R}^{m_t \times k^t}\ \forall t \in \mathbb{T} \\
& \left.\begin{aligned} \sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t \xi + S_t P_t \xi = B_t P_t \xi \\ S_t P_t \xi \geq 0 \end{aligned}\right\} \ \mathbb{P}\text{-a.s. } \forall t \in \mathbb{T}\,.
\end{aligned} \tag{$\mathcal{MSP}^u$}$$

This semi-infinite program constitutes a conservative approximation for $\mathcal{MSP}$ and has the same general structure as problem $\mathcal{SP}^u$ in Section 2. By using robust optimization techniques, it can therefore be reformulated as an equivalent linear program. We only show the final result of this transformation without repeating the involved manipulations.

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \operatorname{Tr}\left(P_t M P_t^\top C_t^\top X_t\right) \\
\text{subject to} \quad & \left.\begin{aligned} X_t \in \mathbb{R}^{n_t \times k^t},\ \Lambda_t \in \mathbb{R}^{m_t \times l} \\ \sum_{s=1}^{T} A_{ts} X_s P_s M_t P_t + \Lambda_t W = B_t P_t \\ \Lambda_t h \geq 0 \\ \Lambda_t \geq 0 \end{aligned}\right\} \ \forall t \in \mathbb{T}
\end{aligned} \tag{4.2}$$

A major benefit of using linear decision rules is that the size of the approximating linear program (4.2) is polynomial in k, l, $m := \sum_{t=1}^{T} m_t$, and $n := \sum_{t=1}^{T} n_t$. Note that these numbers are typically linear in T, and hence the size of (4.2) grows typically only polynomially with the number of decision stages. This massive reduction of computational complexity necessarily comes at the cost of reduced accuracy. In order to estimate this loss of accuracy, we next attempt to find an efficiently computable lower bound on $\mathcal{MSP}$.

[3] Stagewise independence is a widely used standard assumption in stochastic programming. It implies that $\mathbb{E}_t(\xi_s)$ is constant and thus almost surely equal to a constant multiple of ξ_1 for all $t < s$.

As a first step towards a lower bound, we use a duality scheme by Wright [28, § 4] to reexpress the standardized stochastic program (4.1) as a min-max problem in which the dual variable corresponding to the tth equality constraint is given by a non-anticipative square-integrable decision rule $y_t \in \mathcal{L}^2_{k^t,m_t}$. In an attempt to reduce the problem complexity we then require these dual decision rules to be representable as $y_t(\xi^t) = Y_t\xi^t = Y_tP_t\,\xi$ for some $Y_t \in \mathbb{R}^{m_t\times k^t}$. Carrying out the maximization over the matrices $\{Y_t\}_{t\in\mathbb{T}}$ yields

$$\begin{array}{ll}\text{minimize} & \mathbb{E}\left(\sum_{t=1}^{T} c_t(\xi^t)^\top x_t(\xi^t)\right) \\ \text{subject to} & x_t \in \mathcal{L}^2_{k^t,n_t},\; s_t \in \mathcal{L}^2_{k^t,m_t}\;\forall\, t\in\mathbb{T} \\ & \left.\begin{array}{r}\mathbb{E}\left(\sum_{s=1}^{T}\left[A_{ts}x_s(\xi^s) + s_t(\xi^t) - b_t(\xi^t)\right][P_t\xi]^\top\right) = 0 \\ s_t(\xi^t) \geq 0\end{array}\right\}\ \mathbb{P}\text{-a.s. } \forall t\in\mathbb{T}\,.\end{array} \qquad (\mathcal{MSP}^l)$$

The derivation of problem $\mathcal{MSP}^l$ was intentionally abbreviated since similar arguments were already employed in Sections 2 and 3. Moreover, by using the law of iterated conditional expectations, one can directly verify that any (x, s) feasible in (4.1) is also feasible in $\mathcal{MSP}^l$ with the same objective value. This confirms that $\mathcal{MSP}^l$ constitutes a progressive approximation for $\mathcal{MSP}$.

Next, we introduce new decision variables $X_t \in \mathbb{R}^{n_t\times k^t}$ and $S_t \in \mathbb{R}^{m_t\times k^t}$ that are related to the original variables $x_t \in \mathcal{L}^2_{k^t,n_t}$ and $s_t \in \mathcal{L}^2_{k^t,m_t}$ through the equations

$$X_tP_tM = \mathbb{E}\left(x_t(\xi^t)\,\xi^\top\right) \quad\text{and}\quad S_tP_tM = \mathbb{E}\left(s_t(\xi^t)\,\xi^\top\right). \tag{4.3}$$

Due to the truncation operators, which are absent in the one-stage case, it is not obvious that the matrices X_t and S_t exist and are uniquely determined by (4.3).

Lemma 2 *For any given $x_t \in \mathcal{L}^2_{k^t,n_t}$ and $s_t \in \mathcal{L}^2_{k^t,m_t}$ there exist unique matrices $X_t \in \mathbb{R}^{n_t\times k^t}$ and $S_t \in \mathbb{R}^{m_t\times k^t}$ satisfying* (4.3)*.*

Proof Define $X_t \in \mathbb{R}^{n_t\times k^t}$ through

$$X_tP_tMP_t^\top = \mathbb{E}\left(x_t(\xi^t)\,(P_t\xi)^\top\right). \tag{4.4}$$

Notice that $P_tMP_t^\top$ is a principal submatrix of M and constitutes an automorphism on $\mathbb{R}^{k^t}$. Thus, X_t is in fact uniquely determined by the above relation. Using the assumption that $\mathbb{E}_t(\xi) = M_tP_t\xi$ $\mathbb{P}$-a.s., we find

$$\begin{aligned}\mathbb{E}\left(x_t(\xi^t)\,\xi^\top\right) &= \mathbb{E}\left(x_t(\xi^t)\,\mathbb{E}_t(\xi)^\top\right) \\ &= \mathbb{E}\left(x_t(\xi^t)\,(P_t\xi)^\top\right)M_t^\top = X_tP_tMP_t^\top M_t^\top = X_tP_tM\,,\end{aligned}$$

where the last equality follows from

$$M_tP_tMP_t^\top = \mathbb{E}(M_tP_t\xi\xi^\top P_t^\top) = \mathbb{E}\left(\mathbb{E}_t(\xi)\,(P_t\xi)^\top\right) = \mathbb{E}\left(\xi\,(P_t\xi)^\top\right) = MP_t^\top.$$

Thus, X_t constructed in (4.4) satisfies the first relation in (4.3). It is easy to show that there is no other matrix in $\mathbb{R}^{n_t\times k^t}$ with this property. The existence and uniqueness of $S_t \in \mathbb{R}^{m_t\times k^t}$ satisfying the second relation in (4.3) is proved in a similar manner. □

As in Section 2, we can now eliminate the infinite-dimensional decision variables x_t and s_t and reformulate problem ($\mathcal{MSP}^l$) in terms of the finite-dimensional decision variables X_t and S_t. Among other things, this gives rise to the constraints

$$\exists x_t \in \mathcal{L}^2_{k^t,m_t} : \ \mathbb{E}(x_t(\xi^t)\,\xi^\top) = X_t P_t M, \tag{4.5a}$$

$$\exists s_t \in \mathcal{L}^2_{k^t,m_t} : \ \mathbb{E}(s_t(\xi^t)\,\xi^\top) = S_t P_t M,\ s_t(\xi^t) \geq 0 \ \mathbb{P}\text{-a.s.} \tag{4.5b}$$

Constraint (4.5a) is redundant. Indeed, for any $X_t \in \mathbb{R}^{n_t \times k^t}$ the linear decision rule $x_t(\xi^t) = X_t \xi^t = X_t P_t \xi$ satisfies the postulated condition. Moreover, we can use Proposition 3 to reformulate (4.5b) in terms of finitely many linear inequalities. Applicability of Proposition 3 is guaranteed by the following result.

Lemma 3 *For any given $S_t \in \mathbb{R}^{m_t \times k^t}$ constraint* (4.5b) *is equivalent to*

$$\exists \tilde{s}_t \in \mathcal{L}^2_{k,m_t} : \ \mathbb{E}(\tilde{s}_t(\xi)\,\xi^\top) = S_t P_t M,\ \tilde{s}_t(\xi^t) \geq 0 \ \mathbb{P}\text{-}a.s. \tag{4.5c}$$

Proof It is clear that (4.5b) implies (4.5c). To establish the converse implication, assume that (4.5c) holds, and define $s_t \in \mathcal{L}^2_{k^t,m_t}$ through $s_t(\xi^t) = \mathbb{E}_t(\tilde{s}_t(\xi))$. Then,

$$\begin{aligned}\mathbb{E}\left(s_t(\xi^t)\,\xi^\top\right) &= \mathbb{E}\left(\tilde{s}_t(\xi)\,\mathbb{E}_t(\xi)^\top\right)\\ &= \mathbb{E}\left(\tilde{s}_t(\xi)\,\xi^\top\right) P_t^\top M_t^\top = S_t P_t M P_t^\top M_t^\top = S_t P_t M\,,\end{aligned}$$

where the last equality follows from the proof of Lemma 2. □

If $\mathcal{MSP}$ is strictly feasible, we can now prove that $\mathcal{MSP}^l$ is equivalent to the linear program

$$\begin{array}{ll}\text{minimize} & \displaystyle\sum_{t=1}^{T} \operatorname{Tr}\left(P_t M P_t^\top C_t^\top X_t\right)\\ \text{subject to} & \left.\begin{array}{l} X_t \in \mathbb{R}^{n_t \times k^t},\ S_t \in \mathbb{R}^{m_t \times k^t}\\ \displaystyle\sum_{s=1}^{T} A_{ts} X_s P_s N_t P_t + S_t P_t = B_t P_t\\ \left(W - h\,e_1^\top\right) M P_t^\top S_t^\top \geq 0 \end{array}\right\}\ \forall t \in \mathbb{T}\end{array} \tag{4.6}$$

where $N_t := M P_t^\top (P_t M P_t^\top)^{-1}$. The derivation of (4.6) largely parallels that of (2.8). Only the equality constraints in (4.6) require a more detailed description. Substituting (4.3) into the tth equality constraint in $\mathcal{MSP}^l$ yields

$$\sum_{s=1}^{T} A_{ts} X_s P_s M P_t^\top + S_t P_t M P_t^\top = B_t P_t M P_t^\top. \tag{4.7}$$

As $P_t M P_t^\top$ inherits invertibility from M, we can multiply (4.7) from the right by $(P_t M P_t^\top)^{-1} P_t$ to obtain the postulated equality constraint

$$\sum_{s=1}^{T} A_{ts} X_s P_s N_t P_t + S_t P_t = B_t P_t\,.$$

In contrast to the one-stage case, the probability measure $\mathbb{P}$ affects the constraints of the approximating linear programs (4.2) and (4.6) not only through its support but also through its moments. This is not surprising since $\mathcal{MSP}$ accommodates expectation constraints. In the absence of expectation constraints, however, the moment dependence of the equality constraints in (4.2) and (4.6) is lost since $A_{ts} = 0$ for all $s > t$ while

$P_s M_t P_t = P_s = P_s N_t P_t$ for all $s \leq t$. To prove the latter statement, we first use the definition of M_t to deduce the relation

$$P_s M_t P_t \, \xi = P_s \, \mathbb{E}_t(\xi) = P_s \, \xi \quad \mathbb{P}\text{-a.s.} \quad \forall \, s \leq t \, .$$

As the support of $\mathbb{P}$ is assumed to span $\mathbb{R}^k$, the above statement is equivalent to $P_s M_t P_t = P_s$ for all $s \leq t$. Next, we use the definition of N_t and the fact that $P_s = P_s P_t^\top P_t$ for all $s \leq t$ to conclude that

$$P_s N_t P_t = P_s P_t^\top P_t M P_t^\top (P_t M P_t^\top)^{-1} P_t = P_s P_t^\top P_t = P_s \quad \forall \, s \leq t \, .$$

The principal insights gained in this section are summarized in the following theorem. Its proof is omitted since it widely parallels the argumentation in Section 2.

Theorem 3 *Assume that $\mathbb{P}$ has a polyhedral support of the type* (2.1)*, while $\mathbb{E}_t(\xi) = M_t P_t \, \xi$ almost surely for some $M_t \in \mathbb{R}^{k \times k^t}$, $t \in \mathbb{T}$. If $\mathcal{MSP}$ has deterministic constraint matrices and is strictly feasible, then $\mathcal{MSP}^u$ and $\mathcal{MSP}^l$ are equivalent to the linear programs* (4.2) *and* (4.6)*, respectively. The sizes of these linear programs are polynomial in k, l, $m := \sum_{t=1}^T m_t$, and $n := \sum_{t=1}^T n_t$, implying that they are efficiently solvable.*

It is straightforward to generalize the results of this section to decision problems in which some uncertain parameters are *stochastic* in the sense that their distribution is fully known, while all other uncertain parameters are *ambiguous* in the sense that only the support of their distribution is known. In this generalized setting, the objective is to minimize the expected worst-case cost, where the expectation is taken with respect to the stochastic parameters and the worst-case is calculated with respect to the (conditional) support of the ambiguous parameters. In particular, if all uncertain parameters are ambiguous, we end up with a standard robust optimization problem that minimizes the worst-case cost, see e.g. [5,6,9,17]. By means of an epigraph formulation, worst-case optimization problems with a min-max objective can be converted to minimization problems. Thus, the decision rule approximations developed in this section easily carry over to pure robust as well as hybrid stochastic/robust optimization problems. Since these extensions require no new ideas, further details are omitted. However, it is worthwhile to point out that the calculation of the dual bounds requires us to specify a distribution over both the stochastic *and* the ambiguous parameters. This joint distribution can be chosen freely but must be consistent with the given marginal distribution of the stochastic parameters. Each choice of the distribution gives rise to a different dual bound and—in principle—the best (largest) lower bound can be found by maximizing over all admissible distributions. This optimization problem appears to be hard since its objective function is nonconvex, while the feasible set is convex but generically lacks a simple representation in terms of conic constraints. We will return to the problem of choosing an appropriate distribution for the ambiguous parameters in Section 5.

5 Numerical Example

We consider an inventory system proposed by Ben-Tal et al. [4, § 5]. The system consists of a warehouse and I factories, all of which produce a single good. The aim is to satisfy an uncertain demand while minimizing the worst case production costs over a planning horizon of T semimonthly periods. ξ_t represents the demand of the product in period t, while c_{it} denotes the unit production cost, $\overline{x}_{it}$ the production capacity, and x_{it} the actual output of factory i in period t. The cumulative production capacity of factory i over the total planning horizon amounts to $\overline{x}_{\mathrm{tot},i}$.[4] The factories forward all produced goods to the warehouse, which has prescribed maximum and minimum inventory levels $\overline{x}_{\mathrm{wh}}$ and $\underline{x}_{\mathrm{wh}}$,

[4] It is assumed that $\overline{x}_{\mathrm{tot},i}$ is smaller than $\sum_{t=1}^T \overline{x}_{it}$.

respectively. The initial inventory level amounts to x^0_{wh}. We assume that $\xi = (\xi_1, \dots, \xi_T)$ can take any value within a rectangle of the form

$$\Xi := \times_{t=1}^{T} [(1-\theta)\,\xi^*\,\varsigma_t, (1+\theta)\,\xi^*\,\varsigma_t]\,,$$

where ξ^* represents the nominal demand, θ determines the demand variability, and the seasonality factor

$$\varsigma_t := 1 + \frac{1}{2}\sin\left[\frac{\pi(t-1)}{12}\right]$$

reflects our expectation that demands are highest in spring. In order to minimize the worst-case cost over all possible demand scenarios, we solve the following *robust* inventory management problem.

$$\begin{array}{ll} \text{minimize} & x_{\mathrm{obj}} \\ \text{subject to} & x_{\mathrm{obj}} \in \mathbb{R}, \quad x_{it} \in \mathcal{L}^2_{t,1} \quad \forall i = 1\dots, I,\, t \in \mathbb{T} \\ & \left.\begin{array}{l} \displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I} c_{it}\, x_{it}(\xi^t) \le x_{\mathrm{obj}} \\ 0 \le x_{it}(\xi^t) \le \overline{x}_{it} \quad \forall i = 1\dots, I,\, t \in \mathbb{T} \\ \displaystyle\sum_{t=1}^{T} x_{it}(\xi^t) \le \overline{x}_{\mathrm{tot},i} \quad \forall i = 1, \dots, I \\ \underline{x}_{\mathrm{wh}} \le x^0_{\mathrm{wh}} + \displaystyle\sum_{s=1}^{t}\sum_{i=1}^{I} x_{is}(\xi^s) - \sum_{s=1}^{t}\xi_s \le \overline{x}_{\mathrm{wh}} \quad \forall t \in \mathbb{T} \end{array}\right\} \mathbb{P}\text{-a.s.} \end{array} \tag{5.8}$$

Note that $\mathbb{P}$ may be chosen freely by the modeller as long as its support is given by Ξ. This choice does not affect the exact solution of (5.8), but—as manifested by our numerical results below—may have a substantial impact on the quality of the dual linear decision rule approximation.

Replacing the first constraint in (5.8) by

$$\mathbb{E}\left(\sum_{t=1}^{T}\sum_{i=1}^{I} c_{it}\, x_{it}(\xi^t)\right) \le x_{\mathrm{obj}} \tag{5.9}$$

transforms (5.8) to a *stochastic* inventory management problem that minimizes the expected value of the production costs instead of their worst-case realization. In this setting, we assume that $\mathbb{P}$ is the uniform distribution on Ξ. Both the robust as well as the stochastic inventory management problems are readily recognizable as instances of problem $\mathcal{MSP}$ in Section 4.

For our computational experiments, we use the same parameters as in [4]. We assume that there are $I = 3$ factories, while the maximum instantaneous production capacity is set to $\overline{x}_{it} = 567$ units uniformly over all factories and time periods. The cumulative production capacity amounts to $\overline{x}_{\mathrm{tot},i} = 13,600\,T/24$ units for each factory (thus scaling linearly with T), and the maximum and minimum inventory levels are set to $\overline{x}_{\mathrm{wh}} = 2,000$ and $\underline{x}_{\mathrm{wh}} = 500$, respectively. Like the demand, the production costs are subject to seasonal changes, that is, we set $c_{it} = \alpha_i\,\varsigma_t$, where $\alpha_1 = 1$, $\alpha_2 = 1.5$, and $\alpha_3 = 2$. Finally, we fix $\xi^* = 1,000$ units and impose a significant demand uncertainty by setting $\theta = 30\%$.

In a first test series, we solve the bounding problems $\mathcal{MSP}^u$ and $\mathcal{MSP}^l$ corresponding to the *stochastic* inventory problem for planning horizons between 1 and 72 time units (i.e., three years), see Figure 1, left. The relative gap between the bounds is consistently below 5%, which demonstrates that the linear decision rule approximation achieves a high degree of precision in this example. CPLEX 11.2 can solve the two largest bounding problems (each with a remarkable 72 decision stages) within 3,519 seconds on a 2.4 GHz machine. In Table 1 we compare the linear decision rule approximation (LDR) with a sample average

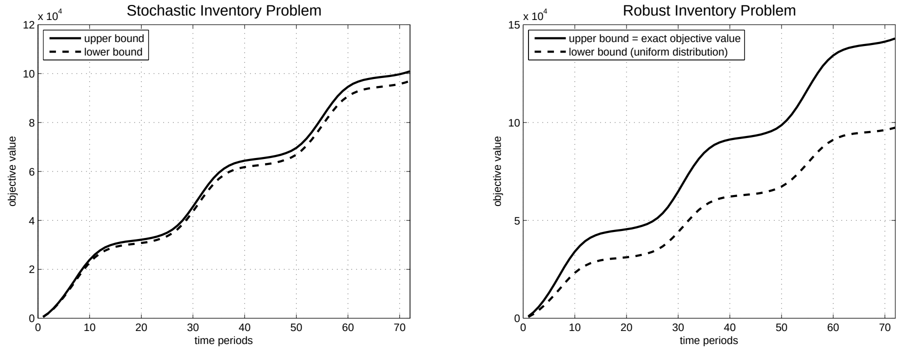

Figure 1: Linear decision rule-based bounds for an inventory problem

approximation (SAA) that replaces the true demand distribution by a discrete scenario tree obtained via conditional sampling, see e.g. Shapiro [25] for details. Due to run time restrictions, the *branching factor* of the randomized scenario tree is fixed to 2 (SAA2), 3 (SAA3), or 4 (SAA4) branches per node, while the planning horizon ranges from 1 to 10 decision stages. Each SAA problem is solved for 50 statistically independent scenario trees. Table 1 reports the 10%, 50%, and 90% quantiles of the resulting optimal objective values. Missing entries (n/a) indicate that the corresponding SAA problems could not be solved since CPLEX ran out of memory (>4GB). Table 1 also lists the lower (lb) and upper (ub) bounds associated with the linear decision rule approximation. We observe that the SAA approximation is consistent with the linear decision rule approximation and achieves a similar degree of accuracy—at least for the low branching factors under consideration. However, the linear decision rule approach exhibits superior scalability, as manifested by the average run times (see columns labeled 'CPU'). For the linear decision rule approximation we report the average run time required to calculate the upper and lower bounds, and for the SAA approximation we disclose the average run time per scenario tree. All run times are quoted in seconds on a 2.4 GHz machine using CPLEX 11.2.

Next, we solve the *robust* inventory problem, which offers some flexibility in choosing the probability measure $\mathbb{P}$. As neither the objective function nor the constraints of problem (5.8) contain expectations, the upper bounding problem $\mathcal{MSP}^u$ associated with (5.8) depends on $\mathbb{P}$ only through its support. However, the lower bounding problem $\mathcal{MSP}^l$ is also affected by the second-order moments of $\mathbb{P}$. Thus, any choice of $\mathbb{P}$ may result in a different lower bound. If one naively sets $\mathbb{P}$ to the uniform distribution $\mathbb{P}^{\mathrm{u}}$ on Ξ, the resulting lower bound is weak, see Figure 1, right. We expect to obtain stronger lower bounds by choosing measures that concentrate probability mass on particularly unfavorable scenarios within Ξ. This reasoning motivates us to solve the lower bounding problem under a probability measure $\mathbb{P}^{\mathrm{wc}}$ that places unit mass[5] on the maximum demand scenario ξ^{wc}. This scenario is defined through $\xi^{\mathrm{wc}}_t := (1+\theta)\xi^*\varsigma_t$, $t \in \mathbb{T}$, and may be expected to result in particularly high production costs. It turns out that the lower bound associated with $\mathbb{P}^{\mathrm{wc}}$ collapses with the upper bound (which is independent of the choice of $\mathbb{P}$), implying that the robust problem (5.8) is solved *exactly* by linear decision rules.

This example demonstrates that the choice of $\mathbb{P}$ is critical for the quality of the dual bounds when the underlying decision problem optimizes a worst case objective. We emphasize that finding the worst case scenarios of generic multistage robust optimization problems is NP-hard [4]. However, a good understanding of a particular decision problem

[5] Note that the support of $\mathbb{P}^{\mathrm{wc}}$ is given by $\{\xi^{\mathrm{wc}}\}$ instead of Ξ, and thus we are—strictly speaking—not allowed to set $\mathbb{P} = \mathbb{P}^{\mathrm{wc}}$. This technical problem can be circumvented by considering the sequence of measures $\{(1-\frac{1}{\nu})\mathbb{P}^{\mathrm{wc}} + \frac{1}{\nu}\mathbb{P}^{\mathrm{u}}\}_{\nu\in\mathbb{N}}$, which converges weakly to $\mathbb{P}^{\mathrm{wc}}$.

under consideration will often allow the modeller to guess unfavourable scenarios that can be used to construct probability measures suitable for the dual bound calculations.

References

1. Ang, M., Chou, M., Sim, M., and So, K. A robust optimization framework for analyzing distribution systems with transshipments. Working paper, National University of Singapore, Singapore, 2008.
2. Ben-Tal, A., Boyd, S., and Nemirovski, A. Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems. *Math. Program. 107*, 1-2, Ser. B (2006), 63–89.
3. Ben-Tal, A., Golany, B., Nemirovski, A., and Vial, J. Supplier-retailer flexible commitments contracts: A robust optimization approach. *Manufacturing Service Oper. Managment 7*, 3 (2005), 248–273.
4. Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. Adjustable robust solutions of uncertain linear programs. *Math. Program. 99*, 2, Ser. A (2004), 351–376.
5. Ben-Tal, A., and Nemirovski, A. Robust convex optimization. *Math. Oper. Res. 23*, 4 (1998), 769–805.
6. Ben-Tal, A., and Nemirovski, A. Robust solutions of uncertain linear programs. *Oper. Res. Lett. 25*, 1 (1999), 1–13.
7. Ben-Tal, A., Nemirovski, A., and Roos, C. Robust solutions of uncertain quadratic and conic quadratic problems. *SIAM J. Optim. 13*, 2 (2002), 535–560 (electronic).
8. Bertsekas, D. *Dynamic Programming and Optimal Control, Volumes I and II.* Athena Scientific, Belmont, MA, 2001.
9. Bertsimas, D., and Sim, M. The price of robustness. *Oper. Res. 52*, 1 (2004), 35–53.
10. Billingsley, P. *Convergence of Probability Measures.* John Wiley, New York, NY, 1968.
11. Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. *Linear matrix inequalities in system and control theory*, vol. 15 of *SIAM Studies in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.
12. Calafiore, G. Multi-period portfolio optimization with linear control policies. *Automatica 44*, 10 (2008), 2463–2473.
13. Chen, X., Sim, M., and Sun, P. A robust optimization perspective on stochastic programming. *Oper. Res. 55*, 6 (2007), 1058–1071.
14. Chen, X., Sim, M., Sun, P., and Zhang, J. A linear decision-based approximation approach to stochastic programming. *Oper. Res. 56*, 2 (2008), 344–357.
15. Chen, X., and Zhang, Y. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* (2009), opre.1080.0605.
16. Dyer, M., and Stougie, L. Computational complexity of stochastic programming problems. *Math. Program. A 106*, 3 (2006), 423–432.
17. El Ghaoui, L., Oustry, F., and Lebret, H. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim. 9*, 1 (1998), 33–52.
18. Garstka, S. J., and Wets, R. J.-B. On decision rules in stochastic programming. *Math. Programming 7* (1974), 117–143.
19. Klein Haneveld, W. *Duality in Stochastic Linear and Dynamic Programming*, vol. 274 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 1985.
20. Kuhn, D. An information-based approximation scheme for stochastic optimization problems in continuous time. *Math. Oper. Res. 34*, 2 (2009), 428–444.
21. Ordóñez, F., and Zhao, J. Robust capacity expansion of network flows. *Netw. 50*, 2 (2007), 136–145.
22. Rockafellar, R., and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk 2*, 3 (2000), 21–41.
23. Rockafellar, R., and Wets, R.-B. *Variational Analysis*, vol. 317 of *A Series of Comprehensive Studies in Mathematics.* Springer-Verlag, New York, 1998.
24. Shapiro, A. On duality theory of conic linear problems. In *Semi-Infinite Programming* (2001), Kluwer Academic Publishers, pp. 135–165.
25. Shapiro, A. Inference of statistical bounds for multistage stochastic programming problems. *Math. Methods Oper. Res. 58*, 1 (2003), 57–68.
26. Shapiro, A., and Nemirovski, A. On complexity of stochastic programming problems. In *Continuous Optimization: Current Trends and Applications* (2005), V. Jeyakumar and A. Rubinov, Eds., Springer, pp. 111–144.
27. Vandenberghe, L., and Boyd, S. Semidefinite programming. *SIAM Rev. 38*, 1 (1996), 49–95.
28. Wright, S. Primal-dual aggregation and disaggregation for stochastic linear programs. *Math. Oper. Res. 19*, 4 (1994), 893–908.
29. Ye, Y. *Interior point algorithms: theory and analysis.* John Wiley & Sons, Inc., New York, NY, USA, 1997.

	LDR			SAA2				SAA3				SAA4			
T	lb	ub	CPU	10%	50%	90%	CPU	10%	50%	90%	CPU	10%	50%	90%	CPU
1	508.3	558.3	0.00	256.0	487.2	635.0	0.0	252.1	499.7	635.1	0.0	320.3	509.4	634.9	0.0
2	1972.7	2032.6	0.00	1555.4	1988.0	2238.1	0.0	1672.6	1944.2	2150.1	0.0	1690.6	1959.2	2176.6	0.0
3	3825.5	4005.3	0.00	3024.9	3854.1	4238.1	0.0	3342.0	3808.0	4004.9	0.0	3350.7	3836.7	4071.7	0.0
4	6090.7	6356.0	0.00	5361.8	6057.9	6643.0	0.0	5319.8	6068.5	6344.9	0.0	5801.3	6125.8	6363.9	0.3
5	8665.4	9064.0	0.00	7818.7	8686.2	9240.1	0.0	8187.3	8689.1	9038.4	0.5	8290.4	8712.9	8976.3	45.6
6	11483.9	12047.5	0.01	10705.5	11433.6	12039.2	0.0	10848.7	11583.7	11985.6	28.8	10753.4	11498.3	11902.2	8476.8
7	14433.5	15182.7	0.02	13947.8	14599.9	15086.9	0.2	13837.6	14499.1	14904.5	1654.0	n/a	n/a	n/a	n/a
8	17434.4	18329.3	0.03	16601.6	17616.6	18117.8	2.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
9	20255.9	21279.0	0.04	19456.0	20412.3	20877.5	26.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
10	22769.3	23869.9	0.05	22025.4	22927.3	23368.5	327.2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 1: Approximate objective values and run times for the stochastic inventory problem.