Generalized Decision Rule Approximations for Stochastic Programming via Liftings

Angelos Georghiou[1], Wolfram Wiesemann[2], and Daniel Kuhn[3]

[1]Process Systems Engineering Laboratory, Massachusetts Institute of Technology, USA

[2]Imperial College Business School, Imperial College London, UK

[3]Department of Computing, Imperial College London, UK

February 4, 2014

Abstract

Stochastic programming provides a versatile framework for decision-making under uncertainty, but the resulting optimization problems can be computationally demanding. It has recently been shown that primal and dual linear decision rule approximations can yield tractable upper and lower bounds on the optimal value of a stochastic program. Unfortunately, linear decision rules often provide crude approximations that result in loose bounds. To address this problem, we propose a lifting technique that maps a given stochastic program to an equivalent problem on a higher-dimensional probability space. We prove that solving the lifted problem in primal and dual linear decision rules provides tighter bounds than those obtained from applying linear decision rules to the original problem. We also show that there is a one-to-one correspondence between linear decision rules in the lifted problem and families of nonlinear decision rules in the original problem. Finally, we identify structured liftings that give rise to highly flexible piecewise linear and nonlinear decision rules, and we assess their performance in the context of a dynamic production planning problem.

1 Introduction

Stochastic programming studies models and algorithms for optimal decision making under uncertainty. A salient feature of many stochastic programming problems is their dynamic nature: some of the uncertain parameters are revealed sequentially as time progresses, and thus future decisions must be modeled as functions of the observable data. These adaptive functional decisions are often referred to as *decision rules*, and their presence severely complicates numerical solution procedures. Indeed, when exact solutions are sought, already two-stage stochastic programs whose random parameters obey independent uniform distributions are computationally intractable [23]. Multistage stochastic programs (with at least two adaptive decision stages) remain intractable even if one searches only for approximate solutions of medium accuracy [40].

Over many decades, substantial efforts have been devoted on designing solution schemes that discretize the distribution of the uncertain model parameters [13, 32]. These discretization techniques typically achieve any desired level of accuracy at the expense of significant computational overheads. An alternative solution scheme is obtained by restricting the set of feasible adaptive decisions to those possessing a simple functional form, such as linear or piecewise linear decision rules. This decision rule approach has been neglected for a long time due to a lack of efficient algorithms and due to the somewhat disappointing result that the class of piecewise linear decision rules is not rich enough to contain the optimal solutions of generic linear stochastic programs with more than two stages [26, p. 123]. Only in 2004 it was realized that the best linear decision rules for linear stochastic and robust optimization problems can be computed in polynomial time [7]. After this breakthrough, similar results emerged for piecewise linear [29] and polynomial [12] decision rules. Even though linear decision rules are known to be optimal for the linear quadratic regulator problem [1], certain classes of robust vehicle routing problems [30] and some one-dimensional robust control problems [11], decision rule approximations generically sacrifice a significant amount of optimality in return for scalability. In fact, the worst-case approximation ratio of linear decision rules when applied to two-stage robust optimization problems with m linear constraints is $\mathcal{O}(\sqrt{m})$ [10].

The goal of this paper is to develop and analyze decision rules that provide more flexibility than crude linear decision rules but preserve their favorable scalability properties. The idea is to map the original stochastic program to an equivalent lifted stochastic program on a higher-dimensional probability space. The relation between the uncertain parameters in the original and the lifted problems is determined through a *lifting operator* which will be defined axiomatically. We will show that there is a one-to-one correspondence between linear decision rules in the lifted problem and families of nonlinear decision rules in the original problem that result from linear combinations of the components of the lifting operator. Thus, solving the lifted stochastic program in linear decision rules, which can be done efficiently, is tantamount to solving the original problem with respect to a class of nonlinear decision rules.

The trade-off between optimality and scalability is controlled by the richness of the lifting operator, that is, by the number of its component mappings and their structure. In order to tailor the lifting operator to a given problem instance, it is crucial that the corresponding approximation quality can be estimated efficiently. In this paper we will measure the approximation quality of a lifting by solving the primal as well as the dual of the lifted stochastic program in linear decision rules, thereby obtaining an upper as well as a lower bound on the (exact) optimal value of the original problem. The difference between these bounds provides an efficiently computable measure for the approximation quality offered by the lifting at hand. This primal-dual approach generalizes a method that was first used to estimate the degree of suboptimality of naive linear decision rules, see [31, 35].

Our axiomatic lifting approach provides a unifying framework for several decision rule approximations proposed in the recent literature. Indeed, piecewise linear [5], segregated linear [20, 21, 29], as well as algebraic and trigonometric polynomial decision rules [5, 12] can be seen as special cases of our approach if the lifting operator is suitably defined. To the best or our knowledge, no efficient a posteriori procedure has yet been reported for measuring the approximation quality of these decision rules—the label 'a posteriori' meaning that the resulting quality measure is specific for each problem instance.

Even though decision rule approximations have gained broader attention only since 2004 [7], they have already found successful use in a variety of application areas ranging from supply chain management [6], logistics [30] and portfolio optimization [19] to network design problems [3], project scheduling [28], electricity procurement optimization [38] and automatic control [41]. The lifting techniques developed in this paper enable the modeler to actively control the trade-off between optimality and scalability and may therefore stimulate the exploration of additional application areas.

The main contributions of this paper may be summarized as follows.

1. We axiomatically introduce lifting operators that allow us to map a given stochastic program to an equivalent problem on a higher-dimensional probability space. We prove that solving the lifted problem in primal and dual linear decision rules results in upper and lower bounds on the original problem that are tighter than the bounds obtained by solving the original problem in linear decision rules. Moreover, we demonstrate that there is a one-to-one relation between linear decision rules in the lifted problem and families of nonlinear decision rules in the original problem that correspond to linear combinations of the components of the lifting operator.

2. We define a class of separable lifting operators that give rise to piecewise linear continuous decision rules with an axial segmentation. These are closely related to the segregated linear decision rules developed in [29]. We prove that the resulting lifted problems in primal and dual linear decision rules are intractable. We then identify tractable special cases and construct tractable approximations for the generic case. Next, we propose a class of liftings that result in tractable piecewise linear continuous decision rules with a general segmentation. We show that these decision rules can offer a substantially better approximation quality than the decision rules with axial segmentation.

3. We introduce a class of nonlinear convex liftings, which includes quadratic liftings, power liftings, monomial liftings and inverse monomial liftings as special cases. These liftings can offer additional flexibility when piecewise linear decision rules perform poorly. Under mild assumptions, the re-

sulting nonlinear decision rule problems are equivalent to tractable second-order cone programs. We also define multilinear liftings, which display excellent approximation properties in numerical tests. Maybe surprisingly, the resulting decision rule problems reduce to linear programs.

The decision rule methods developed in this paper offer similar guarantees as the classical bounding methods of stochastic programming. The most popular bounding methods are those based on functional approximations of the recourse costs [14, 16, 42, 43] and those approximating the true distribution of the uncertain parameters by the (discrete) solution of a generalized moment problem [15, 22, 25, 27, 33]. For a general overview, see [13, Chapter 11] or the survey paper [24]. The bounds based on functional approximations often rely on restrictive assumptions about the problem, such as simple or complete recourse, independence of the uncertain parameters, discrete probability space and/or deterministic objective function coefficients. The moment-based bounding approximations tend to be more flexible and can sometimes be interpreted as multilinear decision rule approximations [34]. However, they exhibit exponential complexity in the number of decision stages. In contrast, all the decision rule approximations developed in this paper offer polynomial complexity. We will compare the tightness and the scalability properties of classical bounds and decision rule approximations in Section 7.

Another approach that has been successfully used for solving large-scale stochastic dynamic programs is approximate dynamic programming (ADP); see e.g. [9, 37]. ADP shares similarities with the decision rule approach discussed in this paper in that both methods use linear combinations of prescribed basis functions to approximate more complex nonlinear functions. While ADP applies these parametric approximations to the cost-to-go functions, however, the decision rule approach applies them to the future adaptive decisions. ADP techniques enjoy great flexibility and can be applied even to nonconvex problems and problems involving integer decisions. The decision rule approach presented here is only applicable to convex problems but offers explicit and easily computable error bounds. Similar bounds are often missing in descriptions of ADP methods. A notable exception are the popular performance bounds based on information relaxations [18], which admit an interpretation as restricted dual stochastic programs. We further remark that the decision rule approach can even find near-optimal solutions for stochastic programs *without* relatively complete recourse [17]. In contrast, most ADP methods enforce feasibility only on sample paths and may therefore fail to account for induced constraints.

The rest of this paper is organized as follows. Section 2 reviews recent results on primal and dual linear decision rules, highlighting the conditions needed to ensure tractability of the resulting optimization problems. In Section 3 we introduce our axiomatic lifting approach for one-stage stochastic programs. We show that if the convex hull of the support of the lifted uncertain parameters has a tractable representation (or outer approximation) in terms of conic inequalities, then the resulting lifted problems can be solved (or approximated) efficiently in primal and dual linear decision rules. Two versatile classes of piecewise linear liftings that ensure this tractability condition are discussed in Section 4, while nonlinear convex liftings and multilinear liftings are studied in Section 5. We generalize the proposed lifting techniques to the multistage case in Section 6, and we assess the performance of the new primal and dual nonlinear decision rules in the context of a dynamic production planning problem in Section 7. The electronic companion to this paper contains some of the proofs as well as further auxiliary results.

Notation We model uncertainty by a probability space $\left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right), \mathbb{P}_\xi\right)$ and denote the elements of the sample space $\mathbb{R}^k$ by ξ. The Borel σ-algebra $\mathcal{B}\left(\mathbb{R}^k\right)$ is the set of events that are assigned probabilities by the probability measure $\mathbb{P}_\xi$. The support Ξ of $\mathbb{P}_\xi$ represents the smallest closed subset of $\mathbb{R}^k$ which has probability 1, and $\mathbb{E}_\xi(\cdot)$ denotes the expectation operator with respect to $\mathbb{P}_\xi$. For any $m, n \in \mathbb{N}$, we let $\mathcal{L}_{m,n}$ be the space of all measurable functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ that are bounded on compact sets. As usual, $\operatorname{Tr}(A)$ denotes the trace of a square matrix $A \in \mathbb{R}^{n\times n}$, while $\mathbb{I}_n$ represents the identity matrix in $\mathbb{R}^{n\times n}$. By slight abuse of notation, the relations $A \leq B$ and $A \geq B$ denote component-wise inequalities for $A, B \in \mathbb{R}^{m\times n}$. For a proper cone $\mathcal{K}$ (i.e., a closed, convex and pointed cone with nonempty interior), the relation $x \preceq_{\mathcal{K}} y$ indicates that $y - x \in \mathcal{K}$. Finally, we denote by e_k the kth canonical basis vector, while e denotes the vector whose components are all ones. In both cases, the dimension will usually be clear from the context.

2 Primal and Dual Linear Decision Rules

In the first part of the paper we study one-stage stochastic programs of the following type. A decision maker first observes an element ξ of the sample space $\mathbb{R}^k$ and then selects a decision $x(\xi) \in \mathbb{R}^n$ subject to the constraints $Ax(\xi) \leq b(\xi)$ and at a cost $c(\xi)^\top x(\xi)$. In the framework of stochastic programming, the aim of the decision maker is to find a function $x \in \mathcal{L}_{k,n}$ which minimizes the expected cost. This decision problem can be formalized as the following one-stage <u>*s*</u>*tochastic* <u>*p*</u>*rogram*.

$$
\begin{array}{lll}
\text{minimize} & \mathbb{E}_\xi\Big(c(\xi)^\top x(\xi)\Big) & \\
\text{subject to} & x \in \mathcal{L}_{k,n} & \\
& Ax(\xi) \leq b(\xi) \qquad \mathbb{P}_\xi\text{-a.s.} &
\end{array}
\tag{$\mathcal{SP}$}
$$

Since the matrix $A \in \mathbb{R}^{m\times n}$ does not depend on the uncertain parameters, we say that $\mathcal{SP}$ has *fixed recourse*. By convention, the function x is referred to as a *decision rule*, *strategy* or *policy*. To ensure that $\mathcal{SP}$ is well-defined, we always assume that it satisfies the following regularity conditions.

(S1) Ξ is a compact subset of the hyperplane $\left\{\xi \in \mathbb{R}^k : \xi_1 = 1\right\}$, and its linear hull spans $\mathbb{R}^k$.

(S2) The objective function coefficients and the right hand sides in $\mathcal{SP}$ depend linearly on the uncertain parameters, that is, $c(\xi) = C\xi$ and $b(\xi) = B\xi$ for some $C \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{m\times k}$.

(S3) $\mathcal{SP}$ is strictly feasible, that is, there exists $\delta > 0$ and a policy $x \in \mathcal{L}_{k,n}$ which satisfies the inequality constraint in $\mathcal{SP}$ with $b(\xi)$ replaced by $b(\xi) - \delta \mathrm{e}$.

Condition **(S1)** ensures that $\xi_1 = 1$ almost surely with respect to $\mathbb{P}_\xi$. This non-restrictive assumption will simplify notation, as it allows us to represent affine functions of the non-degenerate uncertain parameters $(\xi_2, \ldots, \xi_k)$ in a compact way as linear functions of $\xi = (\xi_1, \ldots, \xi_k)^\top$. The assumption about the linear hull of Ξ ensures that the second order moment matrix $\mathbb{E}_\xi\left(\xi\,\xi^\top\right)$ of the uncertain parameters is invertible, see [35]. This assumption is also generic as it can always be enforced by reducing the dimension of ξ if necessary. Condition **(S2)** is non-restrictive as we are free to redefine ξ to contain $c(\xi)$ and $b(\xi)$ as subvectors. Finally, the unrestrictive condition **(S3)** is standard in stochastic programming.

$\mathcal{SP}$ is #P-hard even if $\mathbb{P}_\xi$ is the uniform distribution on the unit cube in $\mathbb{R}^k$, see [23]. Hence, there is no efficient algorithm to determine the optimal value of $\mathcal{SP}$ exactly unless P = NP. A convenient way to obtain a tractable approximation for $\mathcal{SP}$ is to restrict the space of feasible policies to those exhibiting a linear dependency on the uncertain parameters. Thus, we focus on *linear decision rules* that satisfy $x(\xi) = X\xi$ for some $X \in \mathbb{R}^{n\times k}$. Under this restriction, we obtain the following approximate problem.

$$
\begin{array}{lll}
\text{minimize} & \mathbb{E}_\xi\left(c(\xi)^\top X\xi\right) & \\
\text{subject to} & X \in \mathbb{R}^{n\times k} & \\
 & AX\xi \le b(\xi) & \mathbb{P}_\xi\text{-a.s.}
\end{array}
\tag{$\mathcal{UB}$}
$$

This problem is of semi-infinite type and provides a conservative approximation for the original stochastic program because we have reduced the underlying feasible set. Thus, the optimal value of $\mathcal{UB}$ constitutes an <u>u</u>*pper* <u>b</u>*ound* on the optimal value of $\mathcal{SP}$.

We can bound the optimal value of $\mathcal{SP}$ from below if we dualize $\mathcal{SP}$ and afterwards restrict the decision rules corresponding to the dual variables to be linear functions of the uncertain data, see [35]. This dual approximation gives rise to the stochastic program

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

which is easily recognized as a relaxation of the original problem. Therefore, its optimal value provides a *lower bound* on the optimal value of $\mathcal{SP}$. For the derivation of problem $\mathcal{LB}$, we refer to Section 2.3 in [35]. Note that $\mathcal{LB}$ involves only finitely many equality constraints. However, $\mathcal{LB}$ still appears to be intractable as it involves a continuum of decision variables and non-negativity constraints.

Although the semi-infinite bounding problems $\mathcal{UB}$ and $\mathcal{LB}$ look intractable, they can be shown to be equivalent to tractable conic problems under the following assumption about the convex hull of Ξ.

(S4) The convex hull of the support Ξ of $\mathbb{P}_\xi$ is a compact set of the form

$$\text{conv}\ \Xi = \left\{\xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h\right\}, \tag{1}$$

where $W \in \mathbb{R}^{l\times k}$, $V \in \mathbb{R}^{l\times p}$, $h \in \mathbb{R}^l$ and $\mathcal{K} \subseteq \mathbb{R}^l$ is a proper cone.

Theorem 2.1 *If $\mathcal{SP}$ satisfies the regularity conditions **(S1)**, **(S2)** and **(S4)**, then $\mathcal{UB}$ is equivalent to*

$$\begin{array}{ll} \text{minimize} & \text{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, \Lambda \in \mathcal{K}_\star^m \\ & AX + \Lambda W = B \\ & \Lambda V = 0,\ \ \Lambda h \geq 0. \end{array} \tag{$\mathcal{UB}^*$}$$

*If $\mathcal{SP}$ additionally satisfies the regularity condition **(S3)**, then $\mathcal{LB}$ is equivalent to*

$$\begin{array}{ll} \text{minimize} & \text{Tr}\left(MC^\top X\right) \\ \text{subject to} & X \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k}, \Gamma \in \mathbb{R}^{p\times m} \\ & AX + S = B \\ & (W - h\mathrm{e}_1^\top)MS^\top + V\Gamma \succeq_{\mathcal{K}^m} 0. \end{array} \tag{$\mathcal{LB}^*$}$$

In both formulations, $M := \mathbb{E}_\xi\left(\xi\,\xi^\top\right)$ is the second order moment matrix of the uncertain parameters, while $\mathcal{K}_\star$ denotes the dual cone of $\mathcal{K}$. The sizes of the conic problems $\mathcal{UB}^*$ and $\mathcal{LB}^*$ are polynomial in k, l, m, n and p, implying that they are tractable.

Proof This is a straightforward generalization of the results from [35] to conic support sets Ξ. ∎

Theorem 2.1 requires a description of the convex hull of Ξ in terms of conic inequalities, which may not be available or difficult to obtain. In such situations, it may be possible to construct a tractable outer approximation $\widehat{\Xi}$ for the convex hull of Ξ which satisfies the following condition.

($\widehat{\text{S4}}$) There is a compact set $\widehat{\Xi} \supseteq \text{conv}\ \Xi$ of the form $\widehat{\Xi} = \left\{\xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h\right\}$, where W, V, h and $\mathcal{K}$ are defined as in **(S4)**.

Under the relaxed assumption $(\widehat{\mathbf{S4}})$, we can still bound the optimal value of $\mathcal{SP}$.

Corollary 2.2 *If $\mathcal{SP}$ satisfies the regularity conditions **(S1)**, **(S2)** and $(\widehat{\mathbf{S4}})$, then $\mathcal{UB}^*$ provides a conservative approximation (i.e., a restriction) for $\mathcal{UB}$. If $\mathcal{SP}$ additionally satisfies the regularity condition **(S3)**, then $\mathcal{LB}^*$ provides a progressive approximation (i.e., a relaxation) for $\mathcal{LB}$.*

Remark 2.3 *We remark that there is a duality symmetry between the problems $\mathcal{UB}^*$ and $\mathcal{LB}^*$. Indeed, one can show that the dual approximation $\mathcal{LB}^*$ may be obtained by applying the primal approximation $\mathcal{UB}^*$ to the dual of $\mathcal{SP}$ and dualizing the resulting conic program to recover a minimization problem.*

3 Lifted Stochastic Programs

The bounds provided by Theorem 2.1 and Corollary 2.2 can be calculated efficiently by solving tractable conic problems. However, the gap between these bounds can be large if the optimal primal and dual decision rules for the original problem $\mathcal{SP}$ exhibit significant nonlinearities. In this section we elaborate a systematic approach for tightening the bounds that preserves (to some extent) the desirable scalability of the linear decision rule approximations. The basic idea is to lift $\mathcal{SP}$ to a higher-dimensional space and to then apply the linear decision rule approximations to the lifted problem. In this section we axiomatically define the concept of lifting and prove that the application of Theorem 2.1 and Corollary 2.2 to the lifted problem leads to improved bounds on the original problem.

To this end, we introduce a generic *lifting operator* $L : \mathbb{R}^k \to \mathbb{R}^{k'}$, $\xi \mapsto \xi'$, as well as a corresponding *retraction operator* $R : \mathbb{R}^{k'} \to \mathbb{R}^k$, $\xi' \mapsto \xi$. By convention, we will refer to $\mathbb{R}^{k'}$ as the *lifted space*. The operators L and R are assumed to satisfy the following axioms:

(A1) L is continuous and satisfies $\mathrm{e}_1^\top L(\xi) = 1$ for all $\xi \in \Xi$;

(A2) R is linear;

(A3) $R \circ L = \mathbb{I}_k$;

(A4) The component mappings of L are linearly independent, that is, for each $v \in \mathbb{R}^{k'}$, we have

$$L(\xi)^\top v = 0 \quad \mathbb{P}_\xi\text{-a.s.} \quad \Longrightarrow \quad v = 0.$$

Axiom **(A3)** implies that L is an injective operator, which in turn implies that $k' \geq k$.

The following proposition illuminates the relationship between L and R.

Proposition 3.1 *$L \circ R$ is the projection on the range of L along the null space of R.*

Proof By axiom **(A3)** we have $L \circ R \circ L \circ R = L \circ R$, which implies that $L \circ R$ is a projection. Axiom **(A3)** further implies that $L \circ R \circ L = L$, that is, $L \circ R$ is the identity on the range of L. Finally, we have

$$R\left(\xi' - L \circ R\left(\xi'\right)\right) \;=\; R\left(\xi'\right) - R \circ L \circ R\left(\xi'\right) \;=\; 0,$$

where the first and second identity follow from **(A2)** and **(A3)**, respectively. Hence, $\xi' - L \circ R\left(\xi'\right)$ is an element of the null space of R for any $\xi' \in \mathbb{R}^{k'}$, which concludes the proof. ∎

We illustrate the axioms **(A1)**–**(A4)** and Proposition 3.1 with an example.

Example 3.2 *Assume that the dimensions of the original and the lifted space are* $k = 2$ *and* $k' = 3$*, respectively. We define the lifting* L *through* $L((\xi_1, \xi_2)^\top) := \left(\xi_1,\ \xi_2,\ \xi_2^2\right)^\top$*. Similarly, the retraction* R *is given by* $R\left(\xi_1',\ \xi_2', \xi_3'\right)^\top := \left(\xi_1',\ \xi_2'\right)^\top$*. One readily verifies that* L *and* R *satisfy the axioms* ***(A1)**–**(A4)**. Figure 1 illustrates both operators. The lifting* L *maps* $\widehat{\xi}$ *to* $\widehat{\xi}'$*, and the retraction* R *maps any point on the dashed line through* $\widehat{\xi}'$ *to* $\widehat{\xi}$*. The dashed line is given by* $\widehat{\xi}' + \mathrm{kernel}(R)$*, where* $\mathrm{kernel}(R) = \{(0,\ 0,\ \alpha)^\top : \alpha \in \mathbb{R}\}$ *denotes the null space of* R*.*

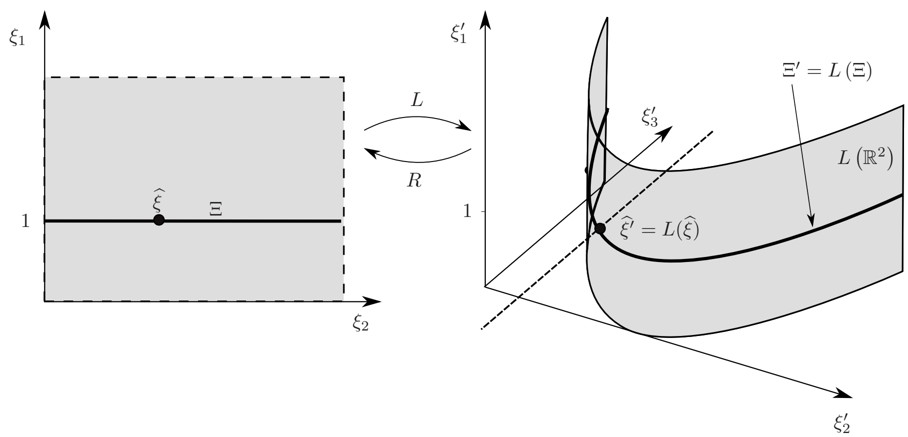

Figure 1: *Illustration of* L *and* R*. The left and right diagram show the original and the lifted space* $\mathbb{R}^k$ *and* $\mathbb{R}^{k'}$*, respectively. The shaded areas and thick solid lines represent* $\mathbb{R}^k$ *and* Ξ *in the left diagram and their lifted counterparts* $L\left(\mathbb{R}^k\right)$ *and* $\Xi' = L(\Xi)$ *in the right diagram.*

We define the probability measure $\mathbb{P}_{\xi'}$ on the lifted space $\left(\mathbb{R}^{k'}, \mathcal{B}\left(\mathbb{R}^{k'}\right)\right)$ in terms of the probability measure $\mathbb{P}_\xi$ on the original space through the relation

$$\mathbb{P}_{\xi'}\left(B'\right) := \mathbb{P}_\xi\left(\{\xi \in \mathbb{R}^k :\ L(\xi) \in B'\}\right) \qquad \forall B' \in \mathcal{B}(\mathbb{R}^{k'}).$$

We also introduce the expectation operator $\mathbb{E}_{\xi'}(\cdot)$ and the support $\Xi' := L(\Xi)$ with respect to the probability measure $\mathbb{P}_{\xi'}$. The following proposition explains the relation between expectations and constraints in the original and lifted space.

Proposition 3.3 *For two measurable functions* $f : \left(\mathbb{R}^{k'}, \mathcal{B}\big(\mathbb{R}^{k'}\big)\right) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *and* $g : \left(\mathbb{R}^{k}, \mathcal{B}\left(\mathbb{R}^{k}\right)\right) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, we have*

(i) $\mathbb{E}_{\xi}\left(f\left(L(\xi)\right)\right) = \mathbb{E}_{\xi'}\left(f\left(\xi'\right)\right)$

(ii) $\mathbb{E}_{\xi}\left(g\left(\xi\right)\right) = \mathbb{E}_{\xi'}\left(g\left(R\xi'\right)\right)$

(iii) $f\left(L\left(\xi\right)\right) \leq 0 \quad \mathbb{P}_{\xi}\text{-a.s.} \quad \Longleftrightarrow \quad f\left(\xi'\right) \leq 0 \quad \mathbb{P}_{\xi'}\text{-a.s.}$

(iv) $g\left(\xi\right) \leq 0 \quad \mathbb{P}_{\xi}\text{-a.s.} \quad \Longleftrightarrow \quad g\left(R\xi'\right) \leq 0 \quad \mathbb{P}_{\xi'}\text{-a.s.}$

Proof Statement (i) follows immediately from [2, Theorem 1.6.12]. In view of (ii), we observe that

$$\mathbb{E}_{\xi}\left(g\left(\xi\right)\right) \;=\; \mathbb{E}_{\xi}\left(g\left(R\circ L\left(\xi\right)\right)\right) \;=\; \mathbb{E}_{\xi'}\left(g\left(R\xi'\right)\right),$$

where the first equality follows from **(A3)** and the second one from statement (i). As for (iii), we have

$$\begin{aligned} f\left(L\left(\xi\right)\right) \leq 0 \quad \mathbb{P}_{\xi}\text{-a.s.} \quad &\Longleftrightarrow \quad \mathbb{E}_{\xi}\left(\max\{0, f\left(L\left(\xi\right)\right)\}\right) = 0 \\ &\Longleftrightarrow \quad \mathbb{E}_{\xi'}\left(\max\{0, f\left(\xi'\right)\}\right) = 0 \\ &\Longleftrightarrow \quad f\left(\xi'\right) \leq 0 \quad \mathbb{P}_{\xi'}\text{-a.s.} \end{aligned}$$

Here, the second equivalence follows from statement *(i)*, while the first and the last equivalences follow from [2, Theorem 1.6.6(b)]. Statement *(iv)* can be shown in a similar manner. ∎

We now consider a variant of the one-stage stochastic program $\mathcal{SP}$ on the lifted probability space.

$$\begin{array}{lll} \text{minimize} & \mathbb{E}_{\xi'}\left(c\left(R\xi'\right)^{\top} x(\xi')\right) & \\ \text{subject to} & x \in \mathcal{L}_{k',n} & \\ & Ax(\xi') \leq b\left(R\xi'\right) \qquad \mathbb{P}_{\xi'}\text{-a.s.} & \end{array} \qquad (\mathcal{LSP})$$

The following proposition shows that the *lifted stochastic program* $\mathcal{LSP}$ is equivalent to $\mathcal{SP}$.

Proposition 3.4 $\mathcal{SP}$ *and* $\mathcal{LSP}$ *are equivalent in the following sense: both problems have the same optimal value, and there is a one-to-one mapping between feasible and optimal solutions in both problems.*

Proof See electronic companion. ∎

Remark 3.5 *If two pairs of lifting and retraction operators* $L^1 : \mathbb{R}^k \rightarrow \mathbb{R}^{k'}$*,* $R^1 : \mathbb{R}^{k'} \rightarrow \mathbb{R}^k$ *and* $L^2 : \mathbb{R}^{k'} \rightarrow \mathbb{R}^{k''}$*,* $R^2 : \mathbb{R}^{k''} \rightarrow \mathbb{R}^{k'}$ *satisfy* ***(A1)****–****(A4)****, then the combined operators* $L := L^2 \circ L^1$*,* $R := R^1 \circ R^2$ *also satisfy* ***(A1)****–****(A4)****. This means that lifted stochastic programs can be constructed iteratively, and all of these lifted programs are equivalent to the original problem* $\mathcal{SP}$*.*

Since $\mathcal{SP}$ and $\mathcal{LSP}$ are equivalent, an upper (lower) bound on the optimal value of $\mathcal{LSP}$ also constitutes an upper (lower) bound on the optimal value of $\mathcal{SP}$. It is therefore useful to investigate the *lifted upper bound* $\mathcal{LUB}$ and the *lifted lower bound* $\mathcal{LLB}$ obtained by applying the primal and dual linear decision rules from the previous section to $\mathcal{LSP}$ instead of $\mathcal{SP}$. In fact, it will turn out that $\mathcal{LUB}$ and $\mathcal{LLB}$ provide a tighter approximation than $\mathcal{UB}$ and $\mathcal{LB}$, which are obtained by applying the linear decision rule approximations directly to $\mathcal{SP}$.

The linear decision rule approximations $\mathcal{LUB}$ and $\mathcal{LLB}$ in the lifted space $\mathbb{R}^{k'}$ correspond to nonlinear decision rule approximations in the original space $\mathbb{R}^{k}$. To show this, we write the lifting operator as $L = (L_1, \ldots, L_{k'})$, where $L_i : \mathbb{R}^k \to \mathbb{R}$ denotes the i^{th} coordinate mapping. These coordinate mappings can be viewed as basis functions for constructing nonlinear decision rules in the original space. To this end, we consider a conservative approximation of $\mathcal{SP}$ that restricts the set of primal decision rules to linear combinations of the coordinate mappings of L, that is, to $x \in \mathcal{L}_{k,n}$ that satisfy $x(\xi) = X'L(\xi)$ for some $X' \in \mathbb{R}^{n \times k'}$. We are thus led to the following *nonlinear upper bound* on $\mathcal{SP}$.

$$
\begin{array}{lll}
\text{minimize} & \mathbb{E}_\xi\left(c(\xi)^\top X'L(\xi)\right) & \\
\text{subject to} & X' \in \mathbb{R}^{n\times k'} & \\
& AX'L(\xi) \le b(\xi) & \mathbb{P}_\xi\text{-a.s.}
\end{array}
\qquad (\mathcal{NUB})
$$

Similarly, we obtain a lower bound on $\mathcal{SP}$ by restricting the set of dual decisions $y \in \mathcal{L}_{k,m}$ in Section 2 to those that can be represented as $y(\xi) = Y'L(\xi)$ for some $Y' \in \mathbb{R}^{m\times k'}$. By using similar arguments as in Section 2, we obtain the following *nonlinear lower bound* on $\mathcal{SP}$.

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

We now show that optimizing over the linear decision rules in the lifted space is indeed equivalent to optimizing over those decision rules in the original space that result from linear combinations of the basis functions $L_1, \ldots, L_{k'}$.

Proposition 3.6 *The nonlinear stochastic programs $\mathcal{NUB}$, $\mathcal{NLB}$ and the linear lifted stochastic programs $\mathcal{LUB}$, $\mathcal{LLB}$ satisfy the following equivalences.*

(i) $\mathcal{NUB}$ and $\mathcal{LUB}$ are equivalent.

(ii) $\mathcal{NLB}$ and $\mathcal{LLB}$ are equivalent.

Equivalent problems attain the same optimal value, and there is a one-to-one mapping between feasible and optimal solutions to equivalent problems.

Proof See electronic companion. ∎

Example 3.7 *In Example 3.2, the lifted linear decision rule* $X'\xi'$ *with* $X' = (1,1,1)$ *corresponds to the nonlinear decision rule* $x(\xi) = \xi_1 + \xi_2 + \xi_2^2$ *in the original space* $\mathbb{R}^k$.

We now show that the linear decision rule approximations in the lifted space $\mathbb{R}^{k'}$ lead to tighter bounds on the optimal value of $\mathcal{SP}$ than the linear decision rule approximations in the original space $\mathbb{R}^k$.

Theorem 3.8 *The optimal values of the approximate problems* $\mathcal{UB}$, $\mathcal{LUB}$, $\mathcal{LB}$ *and* $\mathcal{LLB}$ *satisfy the following chain of inequalities.*

$$\inf \mathcal{LB} \;\le\; \inf \mathcal{LLB} \;\le\; \inf \mathcal{SP} \;=\; \inf \mathcal{LSP} \;\le\; \inf \mathcal{LUB} \;\le\; \inf \mathcal{UB} \tag{2}$$

Proof See electronic companion. ∎

We have shown that the primal and dual linear decision rule approximations to $\mathcal{LSP}$ may result in improved bounds on $\mathcal{SP}$. We now prove that $\mathcal{LSP}$ satisfies the conditions **(S1)**–**(S4)**, which are necessary to obtain tractable reformulations for the approximate lifted problems via Theorem 2.1 and Corollary 2.2.

Proposition 3.9 *If* $\mathcal{SP}$ *satisfies* ***(S1)**–**(S3)**, then* $\mathcal{LSP}$ *satisfies these conditions as well.*

Proof The support Ξ' of $\mathbb{P}_{\xi'}$ is compact as it is the image of a compact set under the continuous mapping L, see axiom **(A1)**. Axiom **(A1)** also guarantees that L maps Ξ to a subset of the hyperplane $\{\xi \in \mathbb{R}^{k'} : \xi'_1 = 1\}$. We now show that Ξ' spans $\mathbb{R}^{k'}$. Assume to the contrary that Ξ' does not span $\mathbb{R}^{k'}$. Then there is $v \in \mathbb{R}^{k'}$, $v \neq 0$, such that

$$\xi'^{\top} v = 0 \quad \mathbb{P}_{\xi'}\text{-a.s.} \qquad \Longleftrightarrow \qquad L(\xi)^{\top} v = 0 \quad \mathbb{P}_{\xi}\text{-a.s.},$$

where the equivalence follows from Proposition 3.3 *(iii)*. By axiom **(A4)** we conclude that $v = 0$. This is a contradiction, and hence the claim follows. In summary, we have shown that $\mathcal{LSP}$ satisfies **(S1)**.

Axiom **(A2)** ensures that the retraction operator R is linear. Hence, the objective and right hand side coefficients of $\mathcal{LSP}$ are linear in the uncertain parameter ξ', and thus $\mathcal{LSP}$ satisfies **(S2)**.

To show that $\mathcal{LSP}$ satisfies **(S3)**, we will use a similar argument as in Proposition 3.4. Suppose that $x \in \mathcal{L}_{k,n}$ is strictly feasible in $\mathcal{SP}$. We define the function $x' \in \mathcal{L}_{k',n}$ through

$$x'(\xi') := x(R\xi') \qquad \forall \xi' \in \mathbb{R}^{k'}.$$

The strict feasibility of x in $\mathcal{SP}$ implies that there exists $\delta > 0$ such that

$$\begin{aligned} & Ax(\xi) \le b(\xi) - \delta \mathrm{e} && \mathbb{P}_\xi\text{-a.s.} \\ \iff \quad & Ax(R\xi') \le b(R\xi') - \delta \mathrm{e} && \mathbb{P}_{\xi'}\text{-a.s.} \\ \iff \quad & Ax'(\xi') \le b(R\xi') - \delta \mathrm{e} && \mathbb{P}_{\xi'}\text{-a.s.}, \end{aligned}$$

where the equivalences follow from Proposition 3.3 *(iv)* and the definition of x', respectively. Therefore, x' is strictly feasible in $\mathcal{LSP}$, and thus $\mathcal{LSP}$ satisfies **(S3)**. ∎

In order to apply Theorem 2.1 and Corollary 2.2 to $\mathcal{LUB}$ and $\mathcal{LLB}$, we also need an exact representation or an outer approximation of the convex hull of Ξ' in terms of conic inequalities, see conditions **(S4)** and **($\widehat{\text{S4}}$)**. In the following sections we will show that these conditions hold in a number of relevant special cases. We close this section with an explicit description of Ξ' in terms of Ξ and L.

Proposition 3.10 *The support Ξ' of the probability measure $\mathbb{P}_{\xi'}$ on the lifted space is given by*

$$\Xi' = \left\{ \xi' \in \mathbb{R}^{k'} : R\xi' \in \Xi, \quad L \circ R(\xi') = \xi' \right\}.$$

Proof The support of $\mathbb{P}_{\xi'}$ can be expressed as

$$\begin{aligned} \Xi' \;=\; L(\Xi) \quad &= \quad \left\{ \xi' \in \mathbb{R}^{k'} : \exists \xi \in \mathbb{R}^k \text{ with } \xi \in \Xi \text{ and } L(\xi) = \xi' \right\} \\ &= \quad \left\{ \xi' \in \mathbb{R}^{k'} : R\xi' \in \Xi, \quad L \circ R(\xi') = \xi' \right\}, \end{aligned}$$

where the identity in the second line follows from Proposition 3.1. ∎

4 Piecewise Linear Continuous Decision Rules

In this section we propose a class of supports Ξ and piecewise linear lifting operators L that satisfy the axioms **(A1)**–**(A4)** and that ensure that the convex hull of $\Xi' = L(\Xi)$ has a tractable representation or outer approximation. We show that the sizes of the corresponding approximate problems $\mathcal{LUB}$ and $\mathcal{LLB}$ are polynomial in the size of the original problem $\mathcal{SP}$ as well as the description of L. We can then invoke Theorem 2.1 and Corollary 2.2 to conclude that $\mathcal{LUB}$ and $\mathcal{LLB}$ can be solved efficiently.

4.1 Piecewise Linear Continuous Decision Rules with Axial Segmentation

The first step towards defining our nonlinear lifting is to select a set of breakpoints for each coordinate axis in $\mathbb{R}^k$. These breakpoints will define the structure of the lifted space, and they are denoted by

$$z_1^i < z_2^i < \ldots < z_{r_i-1}^i \qquad \text{for } i = 2, \ldots, k,$$

where $r_i - 1$ denotes the number of breakpoints along the ξ_i axis. We allow the case $r_i = 1$, where there are no breakpoints along the ξ_i axis. Due to the degenerate nature of the first uncertain parameter ξ_1, we always set $r_1 = 1$. Without loss of generality, we assume that all breakpoints $\{z_j^i\}_{j=1}^{r_i-1}$ are located in the interior of the marginal support of ξ_i. In the remainder of this section we will work with a lifted space whose dimension is given by $k' := \sum_{i=1}^k r_i$. The vectors in the lifted space $\mathbb{R}^{k'}$ can be written as

$$\xi' = \left(\xi'_{1,1},\ \xi'_{2,1}, \ldots, \xi'_{2,r_2},\ \xi'_{3,1}, \ldots, \xi'_{3,r_3},\ \ldots,\ \xi'_{k,1}, \ldots, \xi'_{k,r_k}\right)^\top.$$

Next, we use the breakpoints to define the lifting operator $L = (L_{1,1}, \ldots, L_{k,r_k})$, where the coordinate mapping $L_{i,j}$ corresponds to the $\xi'_{i,j}$ axis in the lifted space and is defined through

$$L_{i,j}(\xi) := \begin{cases} \xi_i & \text{if } r_i = 1, \\ \min\left\{\xi_i, z_1^i\right\} & \text{if } r_i > 1,\ j = 1, \\ \max\left\{\min\left\{\xi_i, z_j^i\right\} - z_{j-1}^i, 0\right\} & \text{if } r_i > 1,\ j = 2, \ldots, r_i - 1, \\ \max\left\{\xi_i - z_{j-1}^i, 0\right\} & \text{if } r_i > 1,\ j = r_i. \end{cases} \tag{3}$$

By construction, $L_{i,j}$ is continuous and piecewise linear with respect to ξ_i and constant in all of its other arguments, see Figure 2. If $r_i = 1$ for all $i = 1, \ldots, k$, then L reduces to the identity mapping on $\mathbb{R}^k$. The linear retraction operator corresponding to L is denoted by $R = (R_1, \ldots, R_k)$, where the coordinate mapping R_i corresponds to the ξ_i axis in the original space and is defined through

$$R_i(\xi') := \sum_{j=1}^{r_i} \xi'_{i,j}. \tag{4}$$

We now show that L and R satisfy the axioms **(A1)**–**(A4)**.

Proposition 4.1 *The operators L and R defined in* (3) *and* (4) *satisfy the axioms **(A1)**–**(A4)**.*

Proof The axioms **(A1)** and **(A2)** are satisfied by construction. Axiom **(A3)** is satisfied if

$$R_i(L(\xi)) = \sum_{j=1}^{r_i} L_{i,j}(\xi) = \xi_i \qquad \forall i = 1, \ldots, k.$$

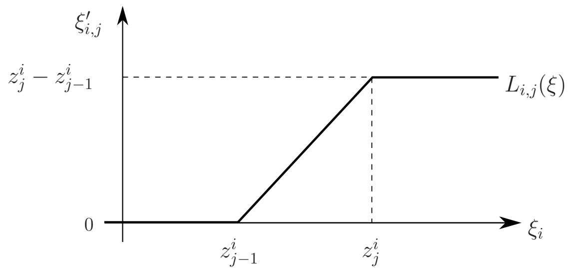

Figure 2: *Graph of the coordinate mapping $L_{i,j}$ for $1 < i \leq k$ and $1 < j < r_i$.*

For $r_i = 1$ this condition is trivially satisfied. For $r_i > 1$ we distinguish the following two cases.

(i) If $\xi_i \leq z^i_1$, then $L_{i,1}(\xi) = \xi_i$ and $L_{i,j}(\xi) = 0$ for all $j = 2, \ldots, r_i$. Thus, $\sum_{j=1}^{r_i} L_{i,j}(\xi) = \xi_i$.

(ii) If $\xi^i > z^i_1$, then set $j^* := \max\{j \in \{1, \ldots, r_i - 1\} : z^i_j \leq \xi_i\}$ so that

$$L_{i,j}(\xi) = \begin{cases} z^i_j & \text{if } j = 1 \\ z^i_j - z^i_{j-1} & \text{if } j = 2, \ldots, j^* - 1 \\ \xi_i - z^i_{j-1} & \text{if } j = j^* \\ 0 & \text{if } j > j^* \end{cases}$$

and thus

$$\sum_{j=1}^{r_i} L_{i,j}(\xi) = z^i_1 + \sum_{j=2}^{j^*-1} \left(z^i_j - z^i_{j-1}\right) + \xi_i - z^i_{j^*-1} = \xi_i.$$

The above arguments apply for each $i = 1, \ldots, k$, and thus **(A3)** follows. Axiom **(A4)** is also satisfied since $L_{i,1}, \ldots, L_{i,r_i}$ are non-constant on disjoint subsets of $\mathbb{R}^k$, each of which has a non-empty intersection with Ξ. ∎

As in Section 3, we use the lifting operator L to define the probability measure $\mathbb{P}_{\xi'}$ on the lifted space and denote the support of $\mathbb{P}_{\xi'}$ by Ξ'. The lifted problems $\mathcal{LSP}$, $\mathcal{LUB}$ and $\mathcal{LLB}$, as well as the problems $\mathcal{NUB}$ and $\mathcal{NLB}$ involving nonlinear decision rules, are defined in the usual way. One can show that the space of nonlinear decision rules induced by the lifting (3) corresponds exactly to the space of piecewise linear and continuous functions with kinks at the breakpoints. A proof is provided in the electronic companion. We now demonstrate that the problems $\mathcal{LUB}$ and $\mathcal{LLB}$ are generically intractable for liftings of the type (3).

Theorem 4.2 *The approximate problems $\mathcal{LUB}$ and $\mathcal{LLB}$ defined through a lifting operator L of the type* (3) *are* NP*-hard even if there is only one breakpoint per coordinate axis.*

Proof See electronic companion. ∎

Theorem 4.2 implies that we cannot solve $\mathcal{LUB}$ and $\mathcal{LLB}$ efficiently for generic liftings of the type (3), even though these problems arise from a linear decision rule approximation. However, Theorem 2.1 ensures that $\mathcal{LUB}$ and $\mathcal{LLB}$ can be solved efficiently if conv Ξ' has a tractable representation of the type (1). We now show that if Ξ constitutes a hyperrectangle within $\{\xi \in \mathbb{R}^k : \mathrm{e}_1^\top \xi = 1\}$, then there exists such a tractable representation for liftings of the type (3). Afterwards, we construct a tractable outer approximation for conv Ξ' in generic situations.

Let Ξ be a hyperrectangle of the type

$$\Xi = \left\{\xi \in \mathbb{R}^k : \ell \leq \xi \leq u\right\}, \tag{5}$$

where $\ell_1 = u_1 = 1$. By Proposition 3.10, the support Ξ' of the lifted probability measure $\mathbb{P}_{\xi'}$ induced by L is given by

$$\Xi' = \left\{\xi' \in \mathbb{R}^{k'} : L \circ R\left(\xi'\right) = \xi',\ \xi'_1 = 1,\ \ell \leq R(\xi') \leq u\right\}$$

and constitutes a non-convex, connected and compact set, see (3). In order to calculate its convex hull, we exploit a separability property of Ξ' that originates from the rectangularity of Ξ. For the further argumentation, we define the *partial lifting operators*

$$L_i := \begin{cases} \mathbb{R}^k \to \mathbb{R}^{r_i} \\ \xi \mapsto \xi'_i := \left(L_{i,1}\left(\xi\right), \ldots, L_{i,r_i}\left(\xi\right)\right)^\top \end{cases} \tag{6}$$

for $i = 1, \ldots, k$. Note that due to (3) the vector-valued function L_i is piecewise affine in ξ_i and constant in its other arguments. By the rectangularity of Ξ we conclude that

$$\Xi' \;=\; L\left(\Xi\right) \;=\; \bigtimes_{i=1}^{k} L_i\left(\Xi\right) \;=\; \bigtimes_{i=1}^{k} \Xi'_i, \tag{7}$$

where $\Xi'_i := L_i\left(\Xi\right)$. The marginal supports Ξ'_i inherit the non-convexity, connectedness and compactness from Ξ'. Note that (7) implies

$$\text{conv}\ \Xi' \;=\; \bigtimes_{i=1}^{k} \text{conv}\ \Xi'_i,$$

and therefore it is sufficient to derive a closed-form representation for the marginal convex hulls conv Ξ'_i. Recall that $\ell_i < z_1^i$ and $z_{r_i-1}^i < u_i$ for $i = 2, \ldots, k$, that is, all breakpoints along the ξ_i-axis in $\mathbb{R}^k$ lie in the interior of the marginal support $[\ell_i, u_i]$.

Lemma 4.3 *The convex hull of Ξ'_i, $i = 2, \ldots, k$, is given by*

$$\text{conv}\ \Xi'_i = \left\{\xi'_i \in \mathbb{R}^{r_i} : V_i^{-1}(1, \xi_i'^\top)^\top \geq 0\right\},$$

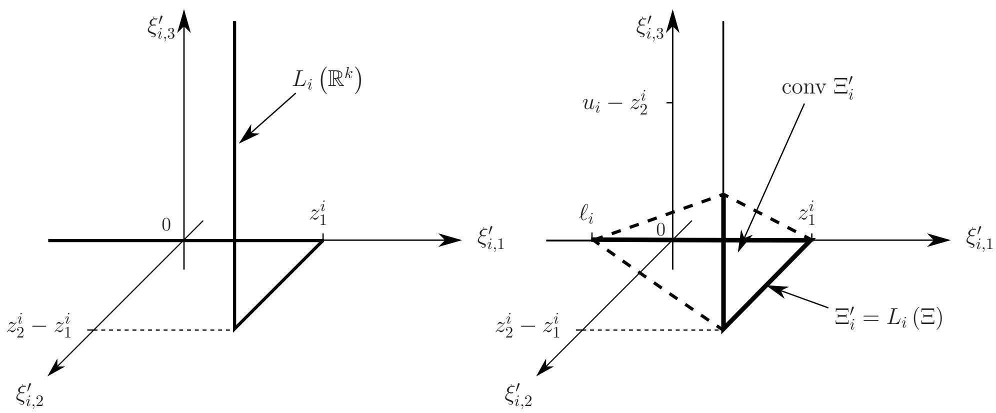

Figure 3: *The left diagram illustrates the range of the partial lifting L_i, which consists of three perpendicular line segments. Here, we assume that there are only two breakpoints at z_1^i and z_2^i along the ξ_i direction (i.e., $r_i = 3$). The right diagram shows the marginal support Ξ'_i (thick line) as well as its convex hull, which is given by a simplex (thick and dashed lines).*

where

$$V_i^{-1} := \begin{pmatrix} \frac{z_1^i}{z_1^i-\ell_i} & -\frac{1}{z_1^i-\ell_i} & & & & \\ -\frac{\ell_i}{z_1^i-\ell_i} & \frac{1}{z_1^i-\ell_i} & -\frac{1}{z_2^i-z_1^i} & & & \\ & & \frac{1}{z_2^i-z_1^i} & \ddots & & \\ & & & \ddots & -\frac{1}{z_{r_i-1}^i-z_{r_i-2}^i} & \\ & & & & \frac{1}{z_{r_i-1}^i-z_{r_i-2}^i} & -\frac{1}{u_i-z_{r_i-1}^i} \\ & & & & & \frac{1}{u_i-z_{r_i-1}^i} \end{pmatrix}.$$

Proof The set Ξ'_i is a union of r_i connected finite line segments, see Figure 3. Its extreme points are $v_0 = (\ell_i, 0, \cdots, 0)^\top \in \mathbb{R}^{k'}$, $v_1 = (z_1^i, 0, \cdots, 0)^\top \in \mathbb{R}^{k'}$, $v_2 = (z_1^i, z_2^i - z_1^i, 0, \cdots, 0)^\top \in \mathbb{R}^{k'}, \ldots,$ $v_{r_i} = (z_1^i, z_2^i - z_1^i, \cdots, z_{r_i-1}^i - z_{r_i-2}^i, u_i - z_{r_i-1}^i)^\top \in \mathbb{R}^{k'}$. Since $\ell_i < z_1^i < \ldots < z_{r_i-1}^i < u_i$, the difference vectors $\{v_j - v_0\}_{j=1}^{r_i}$ are linearly independent. The convex hull of Ξ'_i is thus given by the non-degenerate simplex with vertices $\{v_j\}_{j=0}^{r_i}$, that is, it coincides with the set of all points representable as

$$\xi'_i = \sum_{j=0}^{r_i} \lambda_j v_j, \qquad \sum_{j=0}^{r_i} \lambda_j = 1, \quad \lambda_0, \ldots, \lambda_{r_i} \geq 0.$$

Therefore, we have

$$\begin{aligned} \text{conv } \Xi'_i &= \left\{ \xi'_i \in \mathbb{R}^{r_i} : \exists \lambda \in \mathbb{R}^{r_i+1} \text{ with } V_i \lambda = (1, \xi_i'^\top)^\top, \ \lambda \geq 0 \right\} \\ &= \left\{ \xi'_i \in \mathbb{R}^{r_i} : V_i^{-1} (1, \xi_i'^\top)^\top \geq 0 \right\}, \end{aligned}$$

where

$$V_i := \begin{pmatrix} 1 & \cdots & 1 \\ v_0 & \cdots & v_{r_i} \end{pmatrix} \in \mathbb{R}^{(r_i+1)\times(r_i+1)},$$

and V_i^{-1} is shown in the assertion of this lemma. ∎

Lemma 4.3 allows us to write the convex hull of Ξ' as

$$\begin{aligned} \operatorname{conv} \Xi' &= \bigtimes_{i=1}^{k} \operatorname{conv} \Xi'_i \\ &= \left\{ \xi' = (\xi'_1, \ldots, \xi'_k) \in \bigtimes_{i=1}^{k} \mathbb{R}^{r_i} : \xi'_1 = 1,\ V_i^{-1}(1, \xi_i'^{\top})^{\top} \geq 0 \ \ \forall i = 2, \ldots, k \right\}. \end{aligned} \tag{8}$$

Note that $\operatorname{conv} \Xi'$ is of the form (1) and therefore satisfies condition **(S4)**. This implies that Theorem 2.1 is applicable, which ensures that $\mathcal{LUB}$ and $\mathcal{LLB}$ are equivalent to the conic problems $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ that result from applying the upper and lower bound formulations from Section 2 to the lifted stochastic program $\mathcal{LSP}$. Moreover, since $\operatorname{conv} \Xi'$ is described by $O(k')$ linear inequalities, the sizes of $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are polynomial in k, l, m, n and the total number k' of breakpoints.

Assume now that Ξ is a generic set of the type (1). Then the convex hull of Ξ' has no tractable representation. However, we can systematically construct a tractable outer approximation for $\operatorname{conv} \Xi'$. To this end, let $\{\xi \in \mathbb{R}^k : \ell \leq \xi \leq u\}$ be the smallest hyperrectangle containing Ξ. We have

$$\begin{aligned} \Xi &= \left\{ \xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h \right\} \\ &= \left\{ \xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h,\ \ell \leq \xi \leq u \right\}, \end{aligned} \tag{9}$$

which implies that $\Xi' = \Xi'_1 \cap \Xi'_2$, where

$$\begin{aligned} \Xi'_1 &:= \left\{ \xi' \in \mathbb{R}^{k'} : \exists \zeta \in \mathbb{R}^p \text{ with } WR\xi' + V\zeta \succeq_{\mathcal{K}} h \right\} \\ \Xi'_2 &:= \left\{ \xi' \in \mathbb{R}^{k'} : L \circ R\,(\xi') = \xi',\ \ell \leq R(\xi') \leq u \right\}. \end{aligned}$$

We thus conclude that

$$\widehat{\Xi}' := \left\{ \xi' \in \mathbb{R}^{k'} : \exists \zeta \in \mathbb{R}^p \text{ with } WR\xi' + V\zeta \succeq_{\mathcal{K}} h,\ V_i^{-1}(1, \xi_i'^{\top})^{\top} \geq 0 \ \ \forall i = 2, \ldots, k \right\} \ \supseteq \ \operatorname{conv} \Xi' \tag{10}$$

since $\widehat{\Xi}' = \Xi'_1 \cap \operatorname{conv} \Xi'_2$ and $\Xi'_1 = \operatorname{conv} \Xi'_1$, see (8). Note that $\widehat{\Xi}'$ is of the form (1) and therefore satisfies condition $\widehat{\textbf{(S4)}}$. This implies that Corollary 2.2 is applicable, which ensures that $\mathcal{LUB}$ is conservatively approximated by $\mathcal{LUB}^*$, while $\mathcal{LLB}$ is progressively approximated by $\mathcal{LLB}^*$. Moreover, the sizes of $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are polynomial in k, l, m, n, p and the dimension k' of the lifted space.

The main results of this subsection can be summarized in the following theorem.

Theorem 4.4 *Assume that the original problem $\mathcal{SP}$ satisfies **(S1)–(S4)** and consider any lifting of the type* (3)*. Then the following hold.*

*(i) The lifted problem $\mathcal{LSP}$ satisfies **(S1)–(S3)** and $\widehat{\textbf{(S4)}}$.*

(ii) If Ξ is a hyperrectangle of the type (5)*, then $\mathcal{LSP}$ satisfies the stronger conditions **(S1)–(S4)**.*

(iii) The sizes of the bounding problems $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are polynomial in k, l, m, n, p and k', implying that they are efficiently solvable.

We emphasize the sizes of $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are not only polynomial in the problem dimensions but also in the number of breakpoints. Thus, it is relatively cheap to introduce enough breakpoints along each coordinate direction until the bounds saturate. In contrast, determining the best positions of a fixed number of breakpoints would require the solution of a non-convex global optimization problem.

4.2 Piecewise Linear Continuous Decision Rules with General Segmentation

Even though the liftings considered in Section 4.1 provide considerable flexibility in tailoring piecewise linear decision rules, all pieces of linearity are rectangular and aligned with the coordinate axes in $\mathbb{R}^k$. It is easy to construct problems for which such an axial segmentation results in infeasible or severely suboptimal decisions.

Example 4.5 *Consider the stochastic program*

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_\xi\left(x\left(\xi\right)\right) \\ \text{subject to} & x \in \mathcal{L}_{3,1} \\ & x\left(\xi\right) \geq \max\{|\xi_2|, |\xi_3|\} \quad \mathbb{P}_\xi\text{-a.s.}, \end{array}$$

where ξ_2 and ξ_3 are independent and uniformly distributed on $[-1, 1]$. The optimal solution $x(\xi) = \max\{|\xi_2|, |\xi_3|\}$ is kinked along the main diagonals in the (ξ_2, ξ_3)-plane, and the corresponding optimal value amounts to $2/3$. The best piecewise linear decision rule with axial segmentation (which is also the best affine decision rule) is $x(\xi) = 1$ and achieves the suboptimal objective value 1.

Example 4.5 motivates us to investigate piecewise linear decision rules with generic segmentations that are not necessarily aligned with the coordinate axes. Our aim is to construct piecewise linear decision rules whose kinks are perpendicular to prescribed folding directions. In the following, we demonstrate that such versatile decision rules can be constructed by generalizing the liftings discussed in Section 4.1.

Select finitely many *folding directions* $f_i \in \mathbb{R}^k,\ i = 1, \ldots, k_\eta$, which span $\mathbb{R}^k$ (thus, we have $k_\eta \geq k$). Moreover, for each folding direction f_i select finitely many breakpoints

$$z_1^i < z_2^i < \ldots < z_{r_i-1}^i. \tag{11}$$

For technical reasons, we always set $f_1 = \mathrm{e}_1$ and $r_1 = 1$. We now construct piecewise linear decision rules with kinks along hyperplanes that are perpendicular to f_i and at a distance $z_j^i/\|f_i\|$ from the origin. The general idea is to apply a lifting of the type (3) to the augmented random vector $\eta := F\xi$ instead of ξ, where $F := (f_1, \ldots, f_{k_\eta})^\top$ is the rank-k matrix whose rows correspond to the folding directions.

Define now the piecewise linear lifting operator $L^\eta : \mathbb{R}^{k_\eta} \to \mathbb{R}^{k'_\eta},\ \eta \mapsto \eta'$, and the corresponding retraction operator $R^\eta : \mathbb{R}^{k'_\eta} \to \mathbb{R}^{k_\eta},\ \eta' \mapsto \eta$, as in (3) and (4) by using the breakpoints (11). We set $k'_\eta := \sum_{i=1}^{k_\eta} r_i$. One can show that the k'_η component mappings of the combined lifting $L^\eta \circ F$ span the space of all piecewise linear continuous functions in $\mathbb{R}^k$ which are non-smooth on the hyperplanes $\{\xi \in \mathbb{R}^k : f_i^\top \xi = z_j^i\}$. However, $L^\eta \circ F$ is not a valid lifting if $k_\eta > k$, that is, if the number of folding directions strictly exceeds the dimension of ξ, since then it violates axiom **(A4)**. Indeed, for $k_\eta > k$ the kernel of $F^\top$ is not a singleton. Therefore, there exists $\eta \in \mathrm{kernel}(F^\top),\ \eta \neq 0$, such that by setting $v := (R^\eta)^\top \eta$ we observe that $v \neq 0$ since $v^\top L^\eta(\eta) = \eta^\top \eta \neq 0$ by axiom **(A3)**, see Proposition 4.1. Nevertheless, we have

$$v^\top L^\eta \circ F(\xi) = \eta^\top F(\xi) = 0 \quad \mathbb{P}_\xi\text{-a.s.},$$

and thus $L^\eta \circ F$ fails to satisfy axiom **(A4)**.

To remedy this shortcoming, we define E as the linear hull of $L^\eta \circ F(\Xi)$ and let $g_i \in \mathbb{R}^{k'_\eta},\ i = 1, \ldots, k'$, be a basis for E. For technical reasons, we always set $g_1 = \mathrm{e}_1$. Note that $k' \leq k'_\eta$ since E is a subspace of $\mathbb{R}^{k'_\eta}$. We now define the lifting $L : \mathbb{R}^k \to \mathbb{R}^{k'}$ through

$$L := G \circ L^\eta \circ F, \tag{12}$$

where $G := (g_1, \ldots, g_{k'})^\top \in \mathbb{R}^{k' \times k'_\eta}$ is the rank-k' matrix whose rows correspond to the basis vectors of E. The purpose of G in (12) is to remove all linear dependencies among the component mappings of $L^\eta \circ F$. The corresponding retraction $R : \mathbb{R}^{k'} \to \mathbb{R}^k$ is defined through

$$R := F^+ \circ R^\eta \circ G^+, \tag{13}$$

where $F^+ := (F^\top F)^{-1} F^\top \in \mathbb{R}^{k \times k_\eta}$ and $G^+ := G^\top (G G^\top)^{-1} \in \mathbb{R}^{k'_\eta \times k'}$ are the left and right inverses of F and G, respectively.

Proposition 4.6 *The operators L and R defined in* (12) *and* (13) *satisfy* ***(A1)–(A4)***.

Proof Axioms **(A1)** and **(A2)** are satisfied by construction. Axiom **(A3)** is satisfied if

$$R \circ L \;=\; F^{+} \circ R^{\eta} \circ G^{+} \circ G \circ L^{\eta} \circ F \;=\; \mathbb{I}_k. \tag{14}$$

We have $F^{+} \circ R^{\eta} \circ L^{\eta} \circ F = \mathbb{I}_k$ since $F^{+}F = \mathbb{I}_k$ by the definition of the left inverse and since L^{η} and R^{η} satisfy axiom **(A3)**, see Proposition 4.1. Thus, (14) follows if we can show that $G^{+}G$ acts as the identity on the range of $L^{\eta} \circ F$. As the columns of $G^{\top}$ constitute a basis for E, we conclude that for any $\eta' \in E$ there exists $v \in \mathbb{R}^{k'}$ such that $G^{\top} v = \eta'$. This implies that

$$\begin{aligned} G^{+}G\eta' &= G^{\top}\left(GG^{\top}\right)^{-1} G\eta' \\ &= G^{\top}\left(GG^{\top}\right)^{-1} GG^{\top} v \\ &= G^{\top} v \;=\; \eta' \qquad \forall \eta' \in E. \end{aligned}$$

Thus $G^{+}G$ acts as the identity on the range of $L^{\eta} \circ F$, and therefore **(A3)** follows from (14).

To prove axiom **(A4)**, we first show that the orthogonal complement of E satisfies

$$E^{\perp} \subseteq \left\{ (R^{\eta})^{\top}\eta : \; \eta \in \mathrm{kernel}(F^{\top}) \right\}. \tag{15}$$

This holds if $L^{\eta} \circ F(\xi)$ is orthogonal to $(R^{\eta})^{\top}\eta$ for all $\xi \in \Xi$ and $\eta \in \mathrm{kernel}(F^{\top})$. Indeed, we have

$$\eta^{\top} R^{\eta} \circ L^{\eta} \circ F(\xi) \;=\; \xi^{\top}(F^{\top}\eta) \;=\; 0 \quad \forall \xi \in \Xi,\; \eta \in \mathrm{kernel}(F^{\top}),$$

and thus (15) follows. Next, choose $v \in \mathbb{R}^{k'}$, $v \neq 0$, and observe that $G^{\top} v \in E$ since the row space of G coincides with E. This implies that $G^{\top} v \notin E^{\perp}$, and thus

$$\exists \eta' \in E \,:\, v^{\top} G \eta' \neq 0 \quad \Longrightarrow \quad \exists \xi \in \Xi \,:\, v^{\top} G \circ L^{\eta} \circ F(\xi) = v^{\top} L(\xi) \neq 0.$$

Since L is continuous, $v^{\top} L(\xi)$ cannot vanish $\mathbb{P}_{\xi}$-almost surely. This implies axiom **(A4)**. ∎

The liftings of type (12) provide much flexibility in designing piecewise linear decision rules. In particular, they cover the class of liftings considered in Section 4.1 if we set F and G to $\mathbb{I}_k$ and $\mathbb{I}_{k'_{\eta}}$, respectively. This implies that the lifted approximate problems $\mathcal{LUB}$ and $\mathcal{LLB}$ are computationally intractable for generic liftings of the type (12) even if there is only one breakpoint per folding direction, see Theorem 4.2. As in Section 4.1 we need to construct a tractable representation or outer approximation for the convex hull of $\Xi' = L(\Xi)$ in order to invoke Theorem 2.1 or Corollary 2.2. In the following, we develop an outer approximation for the convex hull of hyperrectangular sets Ξ.

The convex hull of Ξ' is given by

$$\begin{aligned} \text{conv}\ \Xi' &= \text{conv}\ L(\Xi) = \text{conv}\ G \circ L^{\eta} \circ F(\Xi) \\ &= G\left(\text{conv}\ L^{\eta} \circ F(\Xi)\right), \end{aligned}$$

where the last equality holds since the linear operator G preserves convexity, see [39, Proposition 2.21]. Therefore, our problem reduces to finding an outer approximation for conv $L^{\eta} \circ F(\Xi)$. To this end, let $\{\eta \in \mathbb{R}^{k_\eta} : \ell \leq \eta \leq u\}$ be the smallest hypercube that encloses $\Theta := F(\Xi)$. In analogy to Proposition 3.10, one can show that

$$\begin{aligned} \Theta &= \{\eta \in \mathbb{R}^{k_\eta} : \exists \xi \in \Xi \ \text{ with } \ F\xi = \eta\} \\ &= \{\eta \in \mathbb{R}^{k_\eta} : \exists \zeta \in \mathbb{R}^p \text{ with } WF^{+}\eta + V\zeta \succeq_{\mathcal{K}} h,\ FF^{+}\eta = \eta\} \\ &= \{\eta \in \mathbb{R}^{k_\eta} : \exists \zeta \in \mathbb{R}^p \text{ with } WF^{+}\eta + V\zeta \succeq_{\mathcal{K}} h,\ FF^{+}\eta = \eta,\ \ell \leq \eta \leq u\}, \end{aligned}$$

where the second equality holds since FF^{+} is the orthogonal projection onto the range of F and since $\xi = F^{+}\eta$ by definition of F^{+} and η. Note that Θ has the same structure as Ξ in (9) in the sense that it involves a set of generic conic constraints as well as box constraints. Thus, an outer approximation for the convex hull of $L^{\eta}(\Theta)$ is given by

$$\widehat{\Theta}' := \left\{ \eta' := (\eta'_1, \ldots, \eta'_{k_\eta}) \in \bigtimes_{i=1}^{k_\eta} \mathbb{R}^{r_i} : \exists \zeta \in \mathbb{R}^p \text{ with } WF^{+} \circ R^{\eta}(\eta') + V\zeta \succeq_{\mathcal{K}} h,\ V_i^{-1}(1, \eta_i'^{\top})^{\top} \geq 0 \right\},$$

see (10), where the matrices V_i^{-1} are defined as in Lemma 4.3. Thus the resulting outer approximation for conv Ξ' is given by $\widehat{\Xi}' := G(\widehat{\Theta}')$, which satisfies condition **($\widehat{\text{S4}}$)**. This implies that Corollary 2.2 is applicable, which ensures that $\mathcal{LUB}$ is conservatively approximated by $\mathcal{LUB}^*$, while $\mathcal{LLB}$ is progressively approximated by $\mathcal{LLB}^*$.

The insights of this subsection are summarized in the following theorem.

Theorem 4.7 *Assume that the original problem $\mathcal{SP}$ satisfies* ***(S1)–(S4)*** *and consider any lifting of the type* (12)*. Then the following hold.*

(i) The lifted problem $\mathcal{LSP}$ satisfies ***(S1)–(S3)*** *and* ***($\widehat{S4}$)****.*

(iii) The sizes of the bounding problems $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are polynomial in k, l, m, n, p and k', implying that they are efficiently solvable.

We emphasize that the sizes of $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are not only polynomial in the problem dimensions but also in the number of folding directions and breakpoints. Thus, it is relatively cheap to add enough breakpoints along each folding direction until the bounds saturate. More care needs to be taken when choosing the folding directions. In the absence of structural knowledge about the optimal solution, we propose to use the folding directions $e_i + e_j$ and $e_i - e_j$ for $1 \le i < j \le k$ and potentially similar combinations of more than two basis vectors. For a practical example we refer to Section 7.

Example 4.8 *Consider again the stylized stochastic program of Example 4.5. As the optimal solution is kinked along the main diagonals in the (ξ_2, ξ_3)-plane, we define a lifting of the type (12) with a single breakpoint along each (nontrivial) folding direction $f_2 = (1,1)$ and $f_3 = (1,-1)$. Thus, we choose*

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad L^\eta(\eta) = \begin{pmatrix} \eta_1 \\ \min\{\eta_2, 0\} \\ \max\{\eta_2, 0\} \\ \min\{\eta_3, 0\} \\ \max\{\eta_3, 0\} \end{pmatrix}, \quad G = \mathbb{I}_5 \quad \Longrightarrow \quad L(\xi) = \begin{pmatrix} \xi_1 \\ \min\{\xi_2 + \xi_3, 0\} \\ \max\{\xi_2 + \xi_3, 0\} \\ \min\{\xi_2 - \xi_3, 0\} \\ \max\{\xi_2 - \xi_3, 0\} \end{pmatrix}.$$

Note that G can be set to the identity matrix as the number of folding directions matches the dimension of ξ, which implies that the components of $L^\eta(F\xi)$ constitute linearly independent functions on Ξ. It is easy to verify directly that the linear retraction operator corresponding to L is of the form

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

This pair of lifting and retraction operators gives rise to the following instance of $\mathcal{LUB}$,

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\xi'}\left(X'\xi'\right) \\ \text{subject to} & X' \in \mathbb{R}^{1\times 5} \\ & X'\xi' \ge \frac{1}{2}\max\{|\xi_2' + \xi_3' + \xi_4' + \xi_5'|, |\xi_2' + \xi_3' - \xi_4' - \xi_5'|\} \quad \forall \xi' \in \widehat{\Xi}', \end{array}$$

where

$$\widehat{\Xi}' = \Big\{\xi' \in \mathbb{R}^5 \;\;:\;\; \xi_1 = 1, \;\; -1 \le \tfrac{1}{2}(\xi_2' + \xi_3' + \xi_4' + \xi_5') \le 1, \;\; -1 \le \tfrac{1}{2}(\xi_2' + \xi_3' - \xi_4' - \xi_5') \le 1,$$
$$\xi_2' \le 0, \; \xi_3' \ge 0, \; \xi_4' \le 0, \; \xi_5' \ge 0, \; \xi_3' - \xi_2' \le 2, \; \xi_5' - \xi_4' \le 2\Big\}.$$

The first line in the definition of $\widehat{\Xi}'$ encodes the requirement $R\xi' \in \Xi$, while the second line ensures that ξ' lies in the convex hull of $L^\eta\{\eta \in \mathbb{R}^5 : \ell \le \eta \le u\}$ for $\ell = (-2, 0, -2, 0)$ and $u = (0, 2, 0, 2)$. The optimal solution of $\mathcal{LUB}$ is found to be $X' = (0, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})$. In fact, this solution remains optimal even if *the support of* ξ' *is extended to* $\{\xi' \in \mathbb{R}^5 : \xi_1 = 1, \xi'_2 \leq 0, \xi'_3 \geq 0, \xi'_4 \leq 0, \xi'_5 \geq 0\}$. *The corresponding objective value coincides with the minimum of the original stochastic program (i.e.,* $2/3$*). Unfortunately, for the given lifting the optimal value of the lower bounding problem* $\mathcal{LLB}$ *amounts to* 0.542 *and is therefore strictly smaller than the true minimum. The reason for this is that the optimal dual solution is discontinuous along the main diagonals in the* (ξ_2, ξ_3)*-plane. However, a slightly more flexible lifting with two breakpoints at* $\pm\epsilon$ *along each of the folding directions* $f_2 = (1,1)$ *and* $f_3 = (1,-1)$ *ensures that the optimal values of* $\mathcal{LUP}$ *and* $\mathcal{LLB}$ *both converge to* $2/3$ *as* ϵ *tends to zero.*

5 Nonlinear Continuous Decision Rules

In this section we depart from the assumption that the lifting operator L is piecewise linear. In particular, we investigate different types of nonlinear convex liftings in Section 5.1, and we propose a class of nonconvex multilinear liftings in Section 5.2. These liftings may offer additional flexibility when piecewise linear decision rules perform poorly, see Example 5.1. Moreover, they may prove useful for the approximation of linear multistage stochastic programs, whose optimal solutions generically fail to be piecewise linear [26, p. 123].

Example 5.1 *Consider the stochastic program*

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_\xi\left(-x\left(\xi\right)\right) \\ \text{subject to} \quad & x \in \mathcal{L}_{3,1} \\ & x(\xi) \geq 0,\ x\left(\xi\right) \leq \xi_2,\ x(\xi) \leq \xi_3 \quad \mathbb{P}_\xi\text{-a.s.,} \end{aligned}$$

where ξ_2 *and* ξ_3 *are independent and uniformly distributed on* $[0,1]$. *The optimal solution* $x(\xi) = \min\{\xi_2, \xi_3\}$ *with objective value* $-1/3$ *is kinked along the main diagonal of the* (ξ_2, ξ_3)*-plane. As expected, the best piecewise linear decision rule with axial segmentation is* $x(\xi) = 0$ *with objective value* 0. *Maybe surprisingly, the best piecewise linear decision rule with general segmentation and folding directions* $f_1 = (1,0,0)$ *and* $f_2 = (0,1,1)$ *is also* $x(\xi) = 0$ *(irrespective of the number of breakpoints).*

In the following, we represent $\xi \in \Xi$ as $\xi = (\xi_1, \ldots, \xi_T)$, where the subvectors $\xi_t \in \mathbb{R}^{k_t}$ satisfy $k_1 = 1$ and $\sum_{t=1}^T k_t = k$. We assume that the lifted random vector $\xi' := (\xi'_1, \ldots, \xi'_T)$ has a similar structure as ξ, and that the subvectors $\xi'_t \in \mathbb{R}^{k'_t}$ satisfy $k'_1 = 1$, $\xi'_1 = 1$ $\mathbb{P}_{\xi'}$-a.s. and $\sum_{t=1}^T k'_t = k'$. The computational tractability of our nonlinear decision rules is intimately related to a seperability property of the support and the lifting operator.

Definition 5.2 (Separability) *The support* Ξ *and the lifting operator* L *are called* separable *if*

(i) the support Ξ *satisfies* $\Xi = \times_{t=1}^T \Xi_t$, *where* $\Xi_t = \left\{\xi_t \in \mathbb{R}^{k_t} : \xi \in \Xi\right\}$, *and*

(ii) the lifting L satisfies $L = (L_1, \ldots, L_T)$, where $L_t : \mathbb{R}^{k_t} \to \mathbb{R}^{k'_t}$, $\xi_t \mapsto \xi'_t$, depends only on ξ_t.

This definition is non-restrictive. In fact, any support Ξ and lifting operator $L : \mathbb{R}^k \to \mathbb{R}^{k'}$ are separable if we choose $T = 2$, $k_2 = k-1$ and $k'_2 = k'-1$. As we will see below, however, the computational complexity of the bounding problems $\mathcal{LLB}$ and $\mathcal{LUB}$ will depend on $\max_t k_t$, the maximum dimension over all subvectors ξ_t of ξ. The decision rules in this section are therefore particularly suited for problem instances where each of the subvectors ξ_t is of small dimension (e.g. below 10). This is the case, for example, if the optimal decision can be approximated well through the combination of a 'global' decision $x \in \mathcal{L}_{k,n}$ that is linear in ξ, as well as several 'local' decisions $x_t \in \mathcal{L}_{k_t,n}$ that may be highly nonlinear in the components of each subvector ξ_t. This assumption is naturally satisfied for multi-stage stochastic programs with stagewise independent random vectors, see Section 6. It may also be justified in operations management, where manufacturing and distribution decisions can exhibit strong nonlinearities in the random demands for related products or in adjacent regions, whereas the decisions relating to different products or distant regions can be almost independent.

As before, we require the lifting and retraction operators to satisfy **(A1)**–**(A4)**. In view of **(A2)** and **(A3)**, this implies that ξ'_t must contain ξ_t (or an invertible linear transformation thereof) as a subvector.

5.1 Nonlinear Convex Liftings

In this section we assume that Ξ is polyhedral and that Ξ and L are separable in the sense of Definition 5.2. We also assume that the lifting operator L satisfies $L_1(\xi_1) = 1$ and $L_t(\xi_t) = (\xi'_{t,1}, \xi'_{t,2})$, $t = 2, \ldots, T$, where $\xi'_{t,1} = \xi_t$ and $\xi'_{t,2} = f_t(\xi_t)$ for functions $f_t : \mathbb{R}^{k_t} \to \mathbb{R}$ that have conic representable epigraphs. A function f_t has a conic representable epigraph if the set $\{(x, y) \in \Xi_t \times \mathbb{R} : f_t(x) \le y\}$ can be expressed through conic inequalities that may involve additional auxiliary variables.

Proposition 5.3 *The convex hull of $L(\Xi)$ is described through the conic representable set*

$$\Xi' = \Bigg\{\xi' \in \mathbb{R}^{k'} : \xi'_1 = 1, \ \exists \lambda_t(v) \in \mathbb{R}_+, t = 2, \ldots, T \text{ and } v \in V_t, \text{ such that } \sum_{v \in V_t} \lambda_t(v) = 1,$$
$$\xi'_{t,1} = \sum_{v \in V_t} \lambda_t(v) v \ \text{ and } \ \xi'_{t,2} \in \left[f_t(\xi'_{t,1}), \sum_{v \in V_t} \lambda_t(v) f_t(v) \right] \ \ \forall t = 2, \ldots, T \Bigg\},$$

where $V_t = \operatorname{ext} \Xi_t$ denotes the set of extreme points defining Ξ_t.

Proof Note that for any two sets A and B, we have $\operatorname{conv}(A \times B) = \operatorname{conv}(A) \times \operatorname{conv}(B)$. Due to the separability of Ξ and L, as well as the fact that by construction, $\Xi' = \times_{t=1}^{T} \Xi'_t$ for $\Xi'_t = \left\{\xi'_t \in \mathbb{R}^{k'_t} : \xi' \in \Xi'\right\}$, we can therefore restrict ourselves to the case $T = 2$.

We first show that $\operatorname{conv} L(\Xi) \subseteq \Xi'$. To this end, fix any $\xi' \in \operatorname{conv} L(\Xi)$. Carathéodory's Theorem implies that there is $\delta \in \mathbb{R}_+^{k+1}$ and $u_1, \ldots, u_{k+1} \in \Xi_2$ such that $\mathrm{e}^\top \delta = 1$, $\xi'_1 = 1$, $\xi'_{2,1} = \sum_{i=1}^{k+1} \delta_i \, u_i$ and $\xi'_{2,2} = \sum_{i=1}^{k+1} \delta_i \, f_2(u_i)$. Since Ξ_2 is convex, we have $\Xi_2 = \operatorname{conv} \operatorname{ext} \Xi_2$, and another application of Carathéodory's Theorem implies that for each u_i there is $\kappa_i(v) \in \mathbb{R}_+$, $v \in V_2$, such that $\sum_{v \in V_2} \kappa_i(v) = 1$ and $u_i = \sum_{v \in V_2} \kappa_i(v) v$. We now set $\lambda_2(v) = \sum_{i=1}^{k+1} \delta_i \, \kappa_i(v)$ for all $v \in V_2$. By construction, we have $\lambda_2(v) \geq 0$, $v \in V_2$, $\sum_{v \in V_2} \lambda_2(v) = 1$ and $\xi'_{2,1} = \sum_{v \in V_2} \lambda_2(v) v$. Moreover, one readily verifies that

$$\begin{aligned} f_2(\xi'_{2,1}) \;=\; f_2\left(\sum_{i=1}^{k+1} \delta_i \, u_i \right) \;=\; f_2\left(\sum_{i=1}^{k+1} \delta_i \sum_{v \in V_2} \kappa_i(v) v \right) \;&\leq\; \sum_{i=1}^{k+1} \delta_i \, f_2\left(\sum_{v \in V_2} \kappa_i(v) v \right) \\ &\leq\; \sum_{i=1}^{k+1} \delta_i \sum_{v \in V_2} \kappa_i(v) f_2(v) \;=\; \sum_{v \in V_2} \lambda_2(v) f_2(v), \end{aligned}$$

where both inequalities follow from the convexity of f_2. Since the last expression in the first row equals $\sum_{i=1}^{k+1} \delta_i \, f_2(u_i) = \xi'_{2,2}$, we have $\xi'_{2,2} \in \left[f_2(\xi'_{2,1}), \, \sum_{v \in V_2} \lambda_2(v) f_2(v) \right]$, and the assertion follows.

To show that $\operatorname{conv} L(\Xi) \supseteq \Xi'$, fix any $\xi' \in \Xi'$. By construction, there is $\lambda_2(v) \in \mathbb{R}_+$, $v \in V_2$, and $\delta \in [0, 1]$ such that $\xi'_1 = 1$, $\sum_{v \in V_2} \lambda_2(v) = 1$, $\xi'_{2,1} = \sum_{v \in V_2} \lambda_2(v) v$ and $\xi'_{2,2} = \delta f_2(\xi'_{2,1}) + (1 - \delta) \sum_{v \in V_2} \lambda_2(v) f_2(v)$. This implies that

$$\begin{pmatrix} \xi'_{2,1} \\ \xi'_{2,2} \end{pmatrix} \;=\; \sum_{v \in V_2} (1 - \delta) \lambda_2(v) \begin{pmatrix} v \\ f_2(v) \end{pmatrix} + \delta \begin{pmatrix} \xi'_{2,1} \\ f_2(\xi'_{2,1}) \end{pmatrix},$$

that is, ξ' is a convex combination of points contained in $L(\Xi)$. This concludes the proof. ∎

The number of auxiliary variables $\lambda_t(v)$ in Ξ' is proportional to $\max_t |\operatorname{ext} \Xi_t|$, which in general will be exponential in $\max_t k_t$. We are therefore primarily interested in liftings where the dimensions of the subvectors ξ_t are fairly small. Despite this limitation, Proposition 5.3 provides us with remarkable flexibility in defining nonlinear decision rules. In the following, we present a few immediate applications of the result. The involved epigraph formulations of f_t are standard, see e.g. [8].

Example 5.4 (Quadratic Liftings) *Consider the component lifting $L_t(\xi_t) = (\xi_t, \, \xi_t^\top Q \, \xi_t)$, where Q is positive semidefinite. Then the epigraph of $f_t(\xi_t) = \xi_t^\top Q \, \xi_t$ has a conic quadratic representation as*

$$\left\{ (x, y) \in \mathbb{R}^{k_t} \times \mathbb{R} \, : \, \left\| \begin{pmatrix} Q^{1/2} x \\ (y-1)/2 \end{pmatrix} \right\|_2 \leq (y+1)/2 \right\}.$$

Example 5.4 allows us to optimize over quadratic decision rules $x(\xi_1, \ldots, \xi_T) = \sum_{t=2}^{T} x_{1t} (\xi_t^\top Q_t \, \xi_t) + x_2^\top \xi$ that are parametrized in $x_1 \in \mathbb{R}^T$ and $x_2 \in \mathbb{R}^k$. The resulting bounding problems $\mathcal{LLB}$ and $\mathcal{LUB}$

are tractable conic quadratic programs if each subvector ξ_t is of modest dimension.

Example 5.5 (Power Liftings) *Consider the component lifting* $L_t(\xi_t) = (\xi_t, g(\xi_t)^{p/q})$*, where* $g : \mathbb{R}^{k_t} \to \mathbb{R}_+$ *has a conic representable epigraph and* $p, q \in \mathbb{N}$ *with* $p > q$*. Then the epigraph of* $f_t(\xi_t) = g(\xi_t)^{p/q}$ *has a conic representation as*

$$\left\{ (x,y) \in \mathbb{R}^{k_t} \times \mathbb{R}_+ : \begin{array}{c} \exists w_{r,s} \in \mathbb{R}_+, r = 1, \ldots, Q \text{ and } s = 1, \ldots, 2^{Q-r} \text{ such that} \\ w_{r,s} \leq \sqrt{w_{r-1,2s-1} w_{r-1,2s}} \;\; \forall r = 1, \ldots, Q,\, s = 1, \ldots, 2^{Q-r}, \;\; g(x) \leq w_{Q,1} \end{array} \right\},$$

where we use the notational shorthands $Q = \lceil \log_2 q \rceil$ *and* $w_{0,s} = w_{Q,1}$ *for* $s = 1, \ldots, 2^Q - p$*;* $= y$ *for* $s = 2^Q - p + 1, \ldots, 2^Q - p + q$*;* $= 1$ *otherwise. In particular, the epigraph of* f_t *has a conic quadratic representation whenever the function* g *has one.*

Example 5.5 allows us to formulate conic quadratic bounding problems $\mathcal{LLB}$ and $\mathcal{LUB}$ that optimize over decision rules such as $x(\xi_1, \ldots, \xi_6) = x_1 \xi_2^2 + x_2 |\xi_3|^{3/2} + x_3(\xi_3 - \xi_4)^4 + x_4 [\xi_5 - \xi_6]_+^{5/2} + x_5^\top \xi$, which is parametrized in $x_1, \ldots, x_4 \in \mathbb{R}$ and $x_5 \in \mathbb{R}^6$, since the mappings $x \mapsto |x|$ and $y \mapsto \left[g^\top y\right]_+$ have nonnegative polyhedral epigraphs.

Example 5.6 (Monomial Liftings) *Consider the component lifting* $L_t(\xi_t) = (\xi_t, -\prod_{l=1}^{\ell} g_l(\xi_t)^{p_l/q})$*, where* $g_l : \mathbb{R}^{k_t} \to \mathbb{R}_+$*,* $l = 1, \ldots, \ell$*, have conic representable hypographs and* $p_l, q \in \mathbb{N}$ *satisfy* $\sum_{l=1}^{\ell} \frac{p_l}{q} \leq 1$*. Then the epigraph of* $f_t(\xi_t) = -\prod_{l=1}^{\ell} g_l(\xi_t)^{p_l/q}$ *has a conic representation as*

$$\left\{ (x,y) \in \mathbb{R}^{k_t} \times \mathbb{R} : \begin{array}{c} \exists w_{r,s} \in \mathbb{R}_+, r = 1, \ldots, Q, s = 1, \ldots, 2^{Q-r} \text{ and } \tau \in \mathbb{R}_+^{\ell} \text{ such that} \\ w_{r,s} \leq \sqrt{w_{r-1,2s-1} w_{r-1,2s}} \;\; \forall r = 1, \ldots, Q,\, s = 1, \ldots, 2^{Q-r}, \;\; \tau \leq g(x), \;\; y + w_{Q,1} \geq 0 \end{array} \right\},$$

where we use the notational shorthands $Q = \lceil \log_2 q \rceil$ *and* $w_{0,s} = \tau_l$ *for* $s = \sum_{l'=1}^{l-1} p_{l'} + 1, \ldots, \sum_{l'=1}^{l} p_{l'}$ *and* $l \in \{1, \ldots, \ell\}$*;* $= w_{Q,1}$ *for* $s = \sum_{l=1}^{\ell} p_l + 1, \ldots, \sum_{l=1}^{\ell} p_l + 2^Q - q$*;* $= 1$ *otherwise. In particular, the epigraph of* f_t *has a conic quadratic representation whenever the function* g *has one.*

Using Example 5.6 and the fact that $x \mapsto |x|$ and $y \mapsto [y]_+$ have polyhedral epigraphs, we can formulate conic quadratic bounding problems that optimize over decision rules such as $x(\xi_1, \ldots, \xi_5) = x_1 \sqrt{M_1 - (\xi_2 - \xi_3)^2} + x_2 \sqrt{\left(M_2 - [\xi_4]_+\right)\left(M_3 - [\xi_5]_+\right)} + x_3^\top \xi$, which is parametrized in $x_1, x_2 \in \mathbb{R}$ and $x_3 \in \mathbb{R}^5$. Here, M_1, M_2, M_3 are constants that ensure nonnegativity of the terms inside the square roots.

Example 5.7 (Inverse Monomial Liftings) *Consider the lifting* $L_t(\xi_t) = (\xi_t, \prod_{l=1}^{\ell} g_l(\xi_t)^{-p_l/q})$*, where* $g_l : \mathbb{R}^{k_t} \to \mathbb{R}_+$*,* $l = 1, \ldots, \ell$*, are strictly positive functions with conic representable hypographs and*

$p_l, q \in \mathbb{N}$. *Then the epigraph of* $f_t(\xi_t) = \prod_{l=1}^{\ell} g_l(\xi_t)^{-p_l/q}$ *has a conic representation as*

$$\left\{ (x,y) \in \mathbb{R}^{k_t} \times \mathbb{R} \; : \; \begin{array}{l} \exists w_{r,s} \in \mathbb{R}_+, r = 1,\dots,Q-1, s = 1,\dots,2^{Q-r} \text{ and } \tau \in \mathbb{R}^{\ell}_+ \text{ such that} \\ w_{r,s} \le \sqrt{w_{r-1,2s-1} w_{r-1,2s}} \;\; \forall r = 1,\dots,Q, \; s = 1,\dots,2^{Q-r}, \;\; \tau \le g(x) \end{array} \right\},$$

where we use the notational shorthands $Q = \left\lceil \log_2 \left(\sum_{l=1}^{\ell} p_l + q \right) \right\rceil$, $w_{Q,1} = 1$ *and* $w_{0,s} = \tau_l$ *for* $s = \sum_{l'=1}^{l-1} p_{l'} + 1, \dots, \sum_{l'=1}^{l} p_{l'}$ *and* $l \in \{1,\dots,\ell\}$*;* $= y$ *for* $s = \sum_{l=1}^{\ell} p_l + 1, \dots, \sum_{l=1}^{\ell} p_l + q$*;* $= 1$ *otherwise. In particular, the epigraph of* f_t *has a conic quadratic representation whenever the function* g *has one.*

Assuming that $\Xi \subseteq \operatorname{int} \mathbb{R}^k_+$, Example 5.7 allows us to formulate bounding problems that optimize over decision rules such as $x(\xi) = x_1 / \prod_{i=2}^{k} \xi_i + x_2^\top \xi$, which is parametrized in $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}^k$, as well as $x(\xi_1,\dots,\xi_4) = x_1/\xi_2 + x_2/\xi_3^2 + x_3/\sqrt{\xi_4} + x_4^\top \xi$, which is parametrized in $x_1, x_2, x_3 \in \mathbb{R}$ and $x_4 \in \mathbb{R}^4$. We can also model inverse power liftings of the form $L_t(\xi_t) = (\xi_t, g(\xi_t)^{-p/q})$ if we set $\ell = 1$ in Example 5.7.

We remark that the component liftings in this section can be combined with each other and with the piecewise linear liftings from Section 4 as long as Ξ and L remain separable in the sense of Definition 5.2.

5.2 Multilinear Liftings

We now assume that the support of ξ is described by the hypercube $\Xi = \left\{ \xi \in \mathbb{R}^k : \ell \le \xi \le u \right\}$. The regularity condition **(S1)** then implies that $\ell_1 = u_1 = 1$ and $\ell_i < u_i$ for all $i = 2,\dots,k$. As in the previous section, we require Ξ and L to be separable in the sense of Definition 5.2. We also assume that the lifting operator L satisfies $L_1(\xi_1) = 1$ and $L_t(\xi_t) = (\xi'_{t,1}, \xi'_{t,2})$, $t = 2,\dots,T$, where $\xi'_{t,1} = \xi_t$ and $\xi'_{t,2} = f_t(\xi_t)$ for multilinear functions $f_t : \mathbb{R}^{k_t} \mapsto \mathbb{R}^{s_t}$ of the form

$$f_t(\xi_t) = \begin{pmatrix} f_{t,1}(\xi_t) \\ \vdots \\ f_{t,s_t}(\xi_t) \end{pmatrix} = \begin{pmatrix} \prod_{i \in I_{t,1}} \xi_{t,i} \\ \vdots \\ \prod_{i \in I_{t,s_t}} \xi_{t,i} \end{pmatrix}, \qquad \begin{array}{r} \text{where } I_{t,s} \subseteq \{1,\dots,k_t\} \text{ and } |I_{t,s}| > 1, \\ t = 2,\dots,T \text{ and } s = 1,\dots,s_t. \end{array}$$

Consistency then requires that $k'_1 = 1$ and $k'_t = k_t + s_t$ for all $t = 2,\dots,T$.

Proposition 5.8 *The convex hull of $L(\Xi)$ is described through the polyhedron*

$$\Xi' = \Bigg\{ \xi' \in \mathbb{R}^{k'} : \xi'_1 = 1, \ \exists \lambda_t(v) \in \mathbb{R}_+, t = 2, \ldots, T \text{ and } v \in V_t, \text{ such that}$$

$$\sum_{v \in V_t} \lambda_t(v) = 1 \text{ and } \xi'_t = \sum_{v \in V_t} \lambda_t(v) \begin{pmatrix} v \\ f_t(v) \end{pmatrix} \quad \forall t = 2, \ldots, T \Bigg\},$$

where $V_t = \text{ext } \Xi_t$ denotes the set of extreme points defining Ξ_t.

Proof In analogy to the proof of Proposition 5.3, it is sufficient to consider the case $T = 2$. By construction, any element in Ξ' is a convex combination of the points $(1, v, f_2(v))$, $v \in V_2$. Since all of these points are elements of $L(\Xi)$, we conclude that $\Xi' \subseteq \text{conv } L(\Xi)$.

We now show that $\text{conv } L(\Xi) \subseteq \Xi'$. Since Ξ' is convex, it suffices to show that $L(\Xi) \subseteq \Xi'$. Fix $\xi' \in L(\Xi)$ and choose the weights $\lambda_2(v) = \prod_{i=1}^{k_2} \phi_i(v_i, \xi'_{2,1,i})$, $v \in V_2$, where $\phi_i(v_i, \xi'_{2,1,i}) = (u_{i+1} - \xi'_{2,1,i})/(u_{i+1} - l_{i+1})$ if $v_i = \ell_{i+1}$ and $\phi_i(v_i, \xi'_{2,1,i}) = (\xi'_{2,1,i} - \ell_{i+1})/(u_{i+1} - l_{i+1})$ if $v_i = u_{i+1}$. By construction, $\phi_i(v_i, \xi'_{2,1,i}) \geq 0$ for all i and v, which implies that $\lambda_2(v) \geq 0$ for all $v \in V_2$. Moreover, we have that

$$\sum_{v \in V_2} \lambda_2(v) = \sum_{v \in V_2} \prod_{i=1}^{k_2} \phi_i(v_i, \xi'_{2,1,i}) = \prod_{i=1}^{k_2} \left[\phi_i(\ell_{i+1}, \xi'_{2,1,i}) + \phi_i(u_{i+1}, \xi'_{2,1,i}) \right] = 1,$$

and for all $i = 1, \ldots, k_2$, we have that

$$\begin{aligned}
\sum_{v \in V_2} \lambda_2(v) v_i &= \sum_{v \in V_2} \left[\prod_{j=1}^{k_2} \phi_j(v_j, \xi'_{2,1,j}) \right] v_i \\
&= \ell_{i+1} \sum_{\substack{v \in V_2: \\ v_i = \ell_i}} \left[\prod_{j=1}^{k_2} \phi_j(v_j, \xi'_{2,1,j}) \right] + u_{i+1} \sum_{\substack{v \in V_2: \\ v_i = u_i}} \left[\prod_{j=1}^{k_2} \phi_j(v_j, \xi'_{2,1,j}) \right] \\
&= \left[\ell_{i+1} \, \phi_i(\ell_{i+1}, \xi'_{2,1,i}) + u_{i+1} \, \phi_i(u_{i+1}, \xi'_{2,1,i}) \right] \prod_{\substack{j=1,\ldots,k_2 \\ j \neq i}} \left[\phi_j(\ell_{j+1}, \xi'_{2,1,j}) + \phi_i(u_{j+1}, \xi'_{2,1,j}) \right] \\
&= \ell_{i+1} \phi_i(\ell_{i+1}, \xi'_{2,1,i}) + u_{i+1} \phi_i(u_{i+1}, \xi'_{2,1,i}) = \xi'_{2,1,i}.
\end{aligned}$$

One similarly shows that $\sum_{v \in V_2} \lambda_2(v) f_2(v) = \xi'_{2,2}$. We thus have $\xi' \in \Xi'$, which concludes the proof. ∎

In analogy to Proposition 5.3, the number of auxiliary variables $\lambda_t(v)$ in Proposition 5.8 is proportional to $\max_t |\text{ext } \Xi_t| = \max_t 2^{k_t}$. For practical applications, the dimensions of the subvectors ξ_t should therefore be sufficiently small. Proposition 5.8 allows us to formulate bounding problems $\mathcal{LLB}$ and $\mathcal{LUB}$ that optimize over decision rules such as $x(\xi_1, \ldots, \xi_4) = x_1 \xi_2 \xi_3 + x_2 \xi_2 \xi_4 + x_3 \xi_3 \xi_4 + x_4^\top \xi$, which is parametrized in $x_1, x_2, x_3 \in \mathbb{R}$ and $x_4 \in \mathbb{R}^4$. It is worth noting that the resulting bounding problems constitute linear programs, despite the fact that they optimize over nonlinear decision rules.

We now consider a special case of Proposition 5.8 that allows us to derive bounding problems $\mathcal{LLB}$ and $\mathcal{LUB}$ which are tractable irrespective of the sizes k_t of the subvectors ξ_t of ξ.

Proposition 5.9 *Assume that in addition to the assumptions of Proposition 5.8, we have that*

(i) the support Ξ satisfies either $\ell_i = 0$ for all $i = 2, \ldots, k$ or $\ell_i = -u_i$ for all $i = 2, \ldots, k$, and

(ii) the functions f_t in the lifting L satisfy $s_t = 1$ for all $t = 2, \ldots, T$.

Then the convex hull of $L(\Xi)$ is described through the polyhedron

$$\begin{aligned}\Xi' = \big\{\xi' \in \mathbb{R}^{k'} \,:\, \xi'_1 = 1,\ \exists \theta_{t,i}, t = 2, \ldots, T \text{ and } i = 1, \ldots, n_t, \text{ such that}\\ \theta_{t,i} \geq \kappa_{t,i}\, \xi'_{t,1,i_{t,i}} + u_{i_{t,i}} \theta_{t,i+1} - \kappa_{i,j}\, u_{i_{t,i}},\ \ \theta_{t,i} \geq \lambda_{t,i}\, \xi'_{t,1,i_{t,i}} + \ell_{i_{t,i}} \theta_{t,i+1} - \lambda_{i,j}\, \ell_{i_{t,i}},\\ \theta_{t,i} \leq \kappa_{t,i}\, \xi'_{t,1,i_{t,i}} + \ell_{i_{t,i}} \theta_{t,i+1} - \kappa_{i,j}\, \ell_{i_{t,i}},\ \ \theta_{t,i} \leq \lambda_{t,i}\, \xi'_{t,1,i_{t,i}} + u_{i_{t,i}} \theta_{t,i+1} - \lambda_{i,j}\, u_{i_{t,i}},\\ \xi'_{t,1} \in \Xi_t,\ \ \xi'_{t,2} = \theta_{t,n_t} \qquad \forall t = 2, \ldots, T,\ \forall i = 1, \ldots, n_t - 1\big\},\end{aligned}$$

where $I_{t,1} = \{i_{t,1}, \ldots, i_{t,n_t}\}$, $t = 2, \ldots, T$, $\lambda_{t,n_t-1} = \ell_{i_{t,n_t}}$ and $\kappa_{t,n_t-1} = u_{i_{t,n_t}}$, and $\lambda_{t,i}$ and $\kappa_{t,i}$ denote the minimum and maximum of $\big\{\kappa_{t,i+1} u_{i_{t,i+1}},\ \kappa_{t,i+1}\ell_{i_{t,i+1}},\ \lambda_{t,i+1} u_{i_{t,i+1}},\ \lambda_{t,i+1}\ell_{i_{t,i+1}}\big\}$ for all $t = 2, \ldots, T$ and $i = 1, \ldots, n_t - 2$, respectively.

Remark 5.10 *The polyhedron Ξ' emerges from a recursive application of the classical McCormick representation of the convex hull of bilinear terms. The conditions in the statement guarantee that the resulting outer approximation of* conv $L(\Xi)$ *is tight, which is not the case in general.*

Proof of Proposition 5.9 In analogy to the proof of Propositions 5.3, it is sufficient to consider the case $T = 2$. The results then follow from Theorems 1 and 2, as well as Corollaries 1 and 2 in [36]. ∎

Contrary to Proposition 5.8, the description of conv $L(\Xi)$ provided in Proposition 5.9 involves polynomially many auxiliary variables and therefore scales gracefully with the dimensions k_t of the subvectors ξ_t of ξ. Note that the second condition in Proposition 5.9 is satisfied for the decision rule $x(\xi_1, \ldots, \xi_5) = x_1\xi_2\xi_3 + x_2\xi_4\xi_5 + x_3^\top \xi$, but it is violated by the decision rule $x(\xi_1, \ldots, \xi_4) = x_1\xi_2\xi_3 + x_2\xi_3\xi_4 + x_3^\top \xi$ since the nonlinear terms $\xi_2\xi_3$ and $\xi_3\xi_4$ share a common component of ξ.

Example 5.11 *Consider again the stochastic program of Example 5.1 and define the multilinear lifting $L(\xi) = (\xi_1, \xi_2, \xi_3, \xi_2\xi_3)$ with retraction operator $R(\xi') = (\xi'_1, \xi'_2, \xi'_3)$. The convex hull of $L(\Xi)$ is given by*

$$\Xi' = \big\{\xi' \in \mathbb{R}^4 \,:\, \xi'_1 = 1,\ 0 \leq \xi'_2 \leq 1,\ 0 \leq \xi'_3 \leq 1,\ 0 \leq \xi'_4 \leq 1,\ \xi'_4 \geq \xi'_2,\ \xi'_3 \geq \xi'_4\big\}.$$

This lifting provides an upper bound of $-1/4$*. Recall that all piecewise linear liftings resulted in the trivial upper bound* 0*. The lower bounds associated with the piecewise linear liftings are given by* -0.377 *(axial segmentation with* 10 *equispaced breakpoints) and* -0.336 *(general segmentation with folding directions* $f_1 = (1,0,0)$ *and* $f_2 = (0,1,1)$ *and* 10 *equispaced breakpoints). Interestingly, the multilinear lifting offers a weaker lower bound of only* -0.416*.*

6 Multistage Stochastic Programs

In this section we demonstrate that the lifting techniques developed for the single-stage stochastic program $\mathcal{SP}$ extend to *multistage stochastic programs* of the form

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

Here it is assumed that ξ is 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 $t \in \mathbb{T} := \{1, \ldots, T\}$. Without loss of generality, we assume that $k_1 = 1$ and $\xi_1 = 1$ $\mathbb{P}_\xi$-a.s. The history of observations up to time t is denoted by $\xi^t := (\xi_1, \ldots, \xi_t) \in \mathbb{R}^{k^t}$, where $k^t := \sum_{s=1}^{t} k_s$. Consistency then requires that $\xi^T = \xi$ and $k^T = k$. 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 $x_t \in \mathcal{L}_{k^t, n_t}$, $t \in \mathbb{T}$, which map the available observations to decisions and minimize a linear expected cost function subject to 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. We will henceforth assume that $\mathcal{MSP}$ satisfies the following regularity conditions.

(M1) The support Ξ of the probability measure $\mathbb{P}_\xi$ of ξ is a compact subset of the hyperplane $\{\xi \in \mathbb{R}^k : \xi_1 = 1\}$ and its linear hull spans $\mathbb{R}^k$.

(M2) The objective function coefficients and the right hand sides in $\mathcal{MSP}$ depend linearly on ξ, that is, $c_t(\xi^t) = C_t \xi^t$ and $b_t(\xi^t) = B_t \xi^t$ for some $C_t \in \mathbb{R}^{n_t \times k^t}$ and $B_t \in \mathbb{R}^{m_t \times k^t}$, $t \in \mathbb{T}$.

(M3) $\mathcal{MSP}$ is strictly feasible.

(M4) The random vectors $\{\xi_t\}_{t \in \mathbb{T}}$ are mutually independent.

The conditions **(M1)**–**(M3)** are the multistage equivalents of the conditions **(S1)**–**(S3)** for $\mathcal{SP}$. The additional condition **(M4)** is a widely used standard assumption in multistage stochastic programming. **(M4)** guarantees tractability of the lifted lower bound problem to be developed below.

As in the single-stage case, the intractable problem $\mathcal{MSP}$ can be bounded above and below by two semi-infinite problems $\mathcal{MUB}$ and $\mathcal{MLB}$, which are obtained by requiring the primal and dual decisions in $\mathcal{MSP}$ to be linear in ξ, respectively [35]. These problems turn out to be tractable if the convex hull of Ξ is representable by a finite set of conic inequalities, as stated in the following assumption.

(M5) The convex hull of the support Ξ of $\mathbb{P}_\xi$ is a compact set of the form

$$\text{conv}\,\Xi = \left\{\xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h\right\},$$

where $W \in \mathbb{R}^{l\times k}$, $V \in \mathbb{R}^{l\times p}$, $h \in \mathbb{R}^l$ and $\mathcal{K} \subseteq \mathbb{R}^l$ is a proper cone, see also condition **(S4)**.

Condition **(M5)** is the multistage equivalent of **(S4)**. We can now generalize Theorem 2.1 to $\mathcal{MSP}$.

Theorem 6.1 *If $\mathcal{MSP}$ satisfies the conditions **(M1)**, **(M2)** and **(M5)**, then $\mathcal{MUB}$ is equivalent to*

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{t=1}^{T} \text{Tr}\left(P_t M P_t^\top C_t^\top X_t\right) & \\ \text{subject to} & \left.\begin{array}{l} X_t \in \mathbb{R}^{n_t \times k^t},\ \Lambda_t \in \mathcal{K}_\star^{m_t} \\ \displaystyle\sum_{s=1}^{t} A_{ts} X_s P_s + \Lambda_t W = B_t P_t \\ \Lambda_t V = 0,\ \Lambda_t h \geq 0 \end{array}\right\} \forall t \in \mathbb{T}, \end{array} \qquad (\mathcal{MUB}^*)$$

*where the truncation operators P_t, $t \in \mathbb{T}$, are defined through $P_t : \mathbb{R}^k \to \mathbb{R}^{k^t}$, $\xi \mapsto \xi^t$. If $\mathcal{MSP}$ also satisfies the conditions **(M3)** and **(M4)**, then $\mathcal{MLB}$ is equivalent to*

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{t=1}^{T} \text{Tr}\left(P_t M P_t^\top C_t^\top X_t\right) & \\ \text{subject to} & \left.\begin{array}{l} X_t \in \mathbb{R}^{n_t \times k^t},\ S_t \in \mathbb{R}^{m_t \times k^t},\ \Gamma_t \in \mathbb{R}^{p \times m_t} \\ \displaystyle\sum_{s=1}^{t} A_{ts} X_s P_s + S_t P_t = B_t P_t \\ (W - h\mathrm{e}_1) M P_t^\top S_t^\top + V\Gamma_t \succeq_{\mathcal{K}^{m_t}} 0 \end{array}\right\} \forall t \in \mathbb{T} \end{array} \qquad (\mathcal{MLB}^*)$$

The sizes of the conic problems $\mathcal{MUB}^*$ and $\mathcal{MLB}^*$ are polynomial in $k := \sum_{t=1}^T k_t$, l, $m := \sum_{t=1}^T m_t$, $n := \sum_{t=1}^T n_t$, and p, implying that they are efficiently solvable.

Proof This is a straightforward generalization of the results from [35] to conic support sets Ξ. ∎

If conv Ξ has no tractable representation, it may be possible to construct a tractable outer approximation $\widehat{\Xi}$ for the convex hull of Ξ which satisfies the following condition.

($\widehat{\textbf{M5}}$) There is a compact set $\widehat{\Xi} \supseteq \text{conv } \Xi$ of the form $\widehat{\Xi} = \{\xi \in \mathbb{R}^k : \exists \zeta \in \mathbb{R}^p \text{ with } W\xi + V\zeta \succeq_{\mathcal{K}} h\}$, where $W \in \mathbb{R}^{l \times k}$, $V \in \mathbb{R}^{l \times p}$, $h \in \mathbb{R}^l$ and $\mathcal{K} \subseteq \mathbb{R}^l$ is a proper cone, see also condition **($\widehat{\textbf{S4}}$)**.

If condition **($\widehat{\textbf{M5}}$)** holds, then we can extend Corollary 2.2 to $\mathcal{MSP}$ as follows.

Corollary 6.2 *If $\mathcal{MSP}$ satisfies the conditions **(M1)**, **(M2)** and **($\widehat{\textbf{M5}}$)**, then $\mathcal{MUB}^*$ provides a conservative approximation (i.e., a restriction) for $\mathcal{MUB}$. If $\mathcal{MSP}$ additionally satisfies the conditions **(M3)** and **(M4)**, then $\mathcal{MLB}^*$ provides a progressive approximation (i.e., a relaxation) for $\mathcal{MLB}$.*

We can use lifting techniques to improve the upper and lower bounds on $\mathcal{MSP}$ provided by $\mathcal{MUB}$ and $\mathcal{MLB}$. To this end, we introduce a lifting operator $L : \mathbb{R}^k \to \mathbb{R}^{k'}$, $\xi \mapsto \xi'$, as well as a retraction operator $R : \mathbb{R}^{k'} \to \mathbb{R}^k$, $\xi' \mapsto \xi$. We assume that the lifted random vector $\xi' := (\xi'_1, \ldots, \xi'_T)$ has a similar temporal structure as ξ, where $\xi'_t \in \mathbb{R}^{k'_t}$, $\xi'^t := (\xi'_1, \ldots, \xi'_t) \in \mathbb{R}^{k'^t}$, $k'^t := \sum_{s=1}^t k'_s$, $\xi'^T = \xi'$ and $k'^T = k'$. As in Section 3, admissible pairs of lifting and retraction operators must satisfy the axioms **(A1)**–**(A4)**. Due to the temporal structure inherent in $\mathcal{MSP}$ we need to impose the following additional axiom.

(A5) The lifting L satisfies $L = (L_1, \ldots, L_T)$, where $L_t : \mathbb{R}^{k_t} \to \mathbb{R}^{k'_t}$, $\xi_t \mapsto \xi'_t$, depends only on the observation of ξ at time t. Likewise, the retraction R satisfies $R = (R_1, \ldots, R_T)$, where $R_t : \mathbb{R}^{k'_t} \to \mathbb{R}^{k_t}$, $\xi'_t \mapsto \xi_t$, depends only on the observation of ξ' at time t.

Intuitively, the new axiom **(A5)** guarantees that the lifting L preserves the non-anticipative nature of the decision problem at hand. As before, we use L and R to define the lifted version of $\mathcal{MSP}$:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\xi'}\left(\sum_{t=1}^{T} c_t(P_t R\xi')^\top x'_t(\xi'^t)\right) \\ \text{subject to} & x'_t \in \mathcal{L}_{k'^t, n_t} \ \forall t \in \mathbb{T} \\ & \sum_{s=1}^{t} A_{ts} x'_s(\xi'^s) \leq b_t(P_t R\xi') \quad \mathbb{P}_{\xi'}\text{-a.s. } \forall t \in \mathbb{T}, \end{array} \tag{$\mathcal{LMSP}$}$$

where $\mathbb{P}_{\xi'}$ and P_t are defined in Section 3 and Theorem 6.1, respectively.

Proposition 6.3 *$\mathcal{MSP}$ and $\mathcal{LMSP}$ are equivalent in the following sense: both problems attain the same optimal value, and there is a one-to-one mapping between feasible and optimal solutions in both problems.*

Proof The proof of this proposition widely parallels the proof of Proposition 3.4. The only difference is that axiom **(A5)** is needed to establish a one-to-one correspondence between non-anticipative policies in $\mathcal{MSP}$ and $\mathcal{LMSP}$. ∎

Our goal is to apply Theorem 6.1 and Corollary 6.2 to the lifted problem $\mathcal{LMSP}$ to obtain tighter bounds on the original problem $\mathcal{MSP}$. However, this is only possible if $\mathcal{LMSP}$ satisfies **(M1)**–**(M4)** and a tractable representation or outer approximation of conv Ξ is given by **(M5)** or **$(\widehat{\mathbf{M5}})$**, respectively. In a first step we verify the satisfaction of the conditions **(M1)**–**(M4)**.

Proposition 6.4 *If $\mathcal{MSP}$ satisfies conditions **(M1)**–**(M4)**, then $\mathcal{LMSP}$ also satisfies these conditions.*

Proof The proof that $\mathcal{LMSP}$ satisfies **(M1)**–**(M3)** is largely parallel to the proof of Proposition 3.9 and is thus omitted. To prove that $\mathcal{LMSP}$ satisfies **(M4)**, recall that the random vectors $\{\xi_t\}_{t\in\mathbb{T}}$ are mutually independent, which implies via axiom **(A5)** that $\{L_t(\xi_t)\}_{t\in\mathbb{T}}$ are also mutually independent with respect to $\mathbb{P}_\xi$. By construction of the probability distribution $\mathbb{P}_{\xi'}$ of ξ', the random vectors $\{\xi'_t\}_{t\in\mathbb{T}}$ are therefore also mutually independent with respect to $\mathbb{P}_{\xi'}$. Hence, $\mathcal{LMSP}$ satisfies **(M4)**. ∎

The axioms **(A1)**–**(A5)** are not sufficient to guarantee that $\mathcal{LMSP}$ satisfies condition **(M5)** or **$(\widehat{\mathbf{M5}})$** whenever $\mathcal{MSP}$ does so. However, if each of the stagewise liftings $L_t : \mathbb{R}^{k_t} \to \mathbb{R}^{k'_t}$, $t \in \mathbb{T}$, is constructed like the single-stage liftings in Section 4 or 5, then it is easy to show that $\mathcal{LMSP}$ satisfies either **(M5)** or **$(\widehat{\mathbf{M5}})$** whenever $\mathcal{MSP}$ does so. In this situation, we can solve the approximate linear decision rule problems $\mathcal{LMUB}^*$ and $\mathcal{LMLB}^*$ efficiently.

Remark 6.5 *If we are only interested in the conservative approximation $\mathcal{LMUB}$ and have no intention to solve $\mathcal{LMLB}$, then the assumptions **(M3)** and **(M4)** on the original problem $\mathcal{MSP}$ are not needed. Moreover, axiom **(A5)** can be amended to allow for history-dependent liftings of the form*

$$L_t : \mathbb{R}^{k^t} \to \mathbb{R}^{k'_t}, \quad \xi^t \mapsto \xi'_t.$$

*In this generalized setting, the lifted problem $\mathcal{LMSP}$ can still be shown to be equivalent to $\mathcal{MSP}$ and to satisfy **(M1)** and **(M2)**. Moreover, for the liftings discussed in Sections 4 and 5, $\mathcal{LMSP}$ can be shown to satisfy **(M5)** or **$(\widehat{\mathbf{M5}})$** whenever $\mathcal{MSP}$ does so. Thus, $\mathcal{LMUB}^*$ provides a tractable conservative approximation for the original problem $\mathcal{MSP}$.*

7 Numerical Example

We test different decision rule approximations in the context of a stochastic dynamic inventory control problem with multiple products and backlogging. The objective is to determine a sales and order policy that maximizes the expected profit over a planning horizon of T months. At the beginning of month t, we observe a vector of risk factors ξ_t that explains the uncertainty in the current demand $D_{t,p}(\xi_t)$ and the unit sales price $R_{t,p}(\xi_t)$ of each product $p = 1, \ldots, P$. Having observed ξ_t, we select the quantity

$s_{t,p}$ of product p that is sold in month t at the current price. We also determine the amount $o_{t,p}$ of the product that is ordered to replenish the inventory as well as the amount $b_{t,p}$ of the product that is backlogged to the next month at unit cost CB. We require that the sales $s_{t,p}$ in month t are served from orders placed in month $t-1$ or earlier. The inventory level at the beginning of month t is denoted by I_t. For ease of exposition we assume that one unit of each product occupies the same amount of space and incurs the same monthly inventory holding costs CI. We require that the inventory level remains nonnegative and does not exceed the capacity limit $\bar{\mathrm{I}}$ throughout the planning horizon.

The inventory control problem described above can be formulated as

$$\begin{array}{lll} \text{maximize} & \mathbb{E}\left[\sum_{t=1}^{T}\sum_{p=1}^{P} R_{t,p}(\xi_t)s_{t,p}(\xi^t) - \mathrm{CB}\cdot b_{t,p}(\xi^t) - \mathrm{CI}\cdot I_{t,p}(\xi^t)\right] & \\ \text{subject to} & I_t(\xi^t) = \mathbb{I}_{[t\neq 1]}I_{t-1}(\xi^{t-1}) + \sum_{p=1}^{P}\mathbb{I}_{[t\neq 1]}o_{t-1,p}(\xi^{t-1}) - s_{t,p}(\xi^t) & \forall t=1,\ldots,T, \\ & \left.\begin{array}{l} b_{t,p}(\xi^t) = \mathbb{I}_{[t\neq 1]}b_{t-1,p}(\xi^{t-1}) + D_{t,p}(\xi_t) - s_{t,p}(\xi^t) \\ o_{t,p}(\xi^t), s_{t,p}(\xi^t), b_{t,p}(\xi^t), I_{t,p}(\xi^t) \geq 0,\;\; I_{t,p}(\xi^t) \leq \bar{\mathrm{I}} \end{array}\right\} & \begin{array}{l}\forall t=1,\ldots,T,\\ \forall p=1,\ldots,P,\end{array} \end{array} \tag{16}$$

where all constraints are assumed to hold with probability 1. We define the product prices as

$$R_{t,p}(\xi_t) = 2 + \gamma_{\mathrm{R}}\left[\alpha_{1,p}\xi_{t,1} + \alpha_{2,p}\xi_{t,2}\right]$$

with factor loadings $\alpha_{1,p}, \alpha_{2,p} \in [-1,1]$ and uncertainty level $\gamma_{\mathrm{R}} \in [0,1]$. Similarly, we set the demands to

$$D_{t,p}(\xi_t) = \begin{cases} 2 + \sin\left(\frac{2\pi(t-1)}{12}\right) + \frac{1}{2}\gamma_{\mathrm{D}}\left[\alpha_{3,p}\xi_{t,3} + \alpha_{4,p}\xi_{t,4}\right] & \text{for } p = 1,\ldots,P/2 \\ 2 + \cos\left(\frac{2\pi(t-1)}{12}\right) + \frac{1}{2}\gamma_{\mathrm{D}}\left[\alpha_{3,p}\xi_{t,3} + \alpha_{4,p}\xi_{t,4}\right] & \text{for } p = P/2+1,\ldots,P \end{cases}$$

with $\alpha_{3p}, \alpha_{4p} \in [-1,1]$ and $\gamma_{\mathrm{D}} \in [0,1]$. The sine (cosine) terms in the above expression encode the stylized fact that the expected demands of the first (last) $P/2$ products are high in spring (winter) and low in autumn (summer). We assume that the vectors of risk factors $\xi_t \in \mathbb{R}^4$, $t = 1,\ldots,T$, are serially independent and uniformly distributed on $[-1,1]^4$. Note that $\xi_{t,1}$ and $\xi_{t,2}$ only impact the prices, while $\xi_{t,3}$ and $\xi_{t,4}$ only impact the demands. This ensures the applicability of the scenario tree-based bounding methods proposed in [25, 33]. We emphasize, however, that none of the decision rule approximations developed in this paper require such a separation of the risk factors.

All numerical experiments are based on 25 randomly generated instances of the inventory control problem with $P = 4$ products, identical backlogging and inventory holding costs $\mathrm{CB} = \mathrm{CI} = 0.2$ and inventory capacity $\bar{\mathrm{I}} = 24$. The uncertainty levels of the prices and demands are set to $\gamma_{\mathrm{P}} = \gamma_{\mathrm{D}} = 1$, and the factor loadings $\alpha_{1,p}$, $\alpha_{2,p}$, $\alpha_{3,p}$ and $\alpha_{4,p}$ are sampled uniformly from the interval $[-1,1]$. Each instance is solved with the decision rule approximations induced by the following liftings: linear (**LDR**), piecewise linear with axial segmentation and 5 equidistant breakpoints (**AS5**), piecewise linear with general segmentation and one breakpoint at 0 (**GS1**), bilinear (**BL**) and trilinear (**TL**). The lifting **GS1** uses the folding directions e_i for $i = 1, \ldots, k$ as well as $\mathrm{e}_i + \mathrm{e}_j$ and $\mathrm{e}_i - \mathrm{e}_j$ for $1 \leq i < j \leq k$. Moreover, the liftings **BL** and **TL** include all possible pairs and triplets of mutually different random parameters of any stage as component functions, respectively. A major benefit of the decision rule techniques developed in this paper is their modularity, which allows us to combine basic liftings to generate more flexible liftings. For instance, we can construct the combined liftings **BL-1** and **TL-1**, which are defined as compositions of a piecewise linear lifting with axial segmentation and one breakpoint at 0 with the bi- and trilinear liftings **BL** and **TL**, respectively (thus resulting in piecewise bi- and trilinear decision rules).

We compare the different decision rule approximations with the scenario tree-based bounding method described in [25, 33] (**SCN**). This method provides both upper and lower bounds that are reminiscent of the classical Jensen and Edmundson-Madansky bounds of stochastic programming [13] and that can be viewed as multilinear decision rule bounds [34]. However, the underlying scenario trees and—*a fortiori*—the computational effort required to compute these bounds grows exponentially with the horizon length T, whereas all decision rule approximations developed in this paper scale gracefully with T.

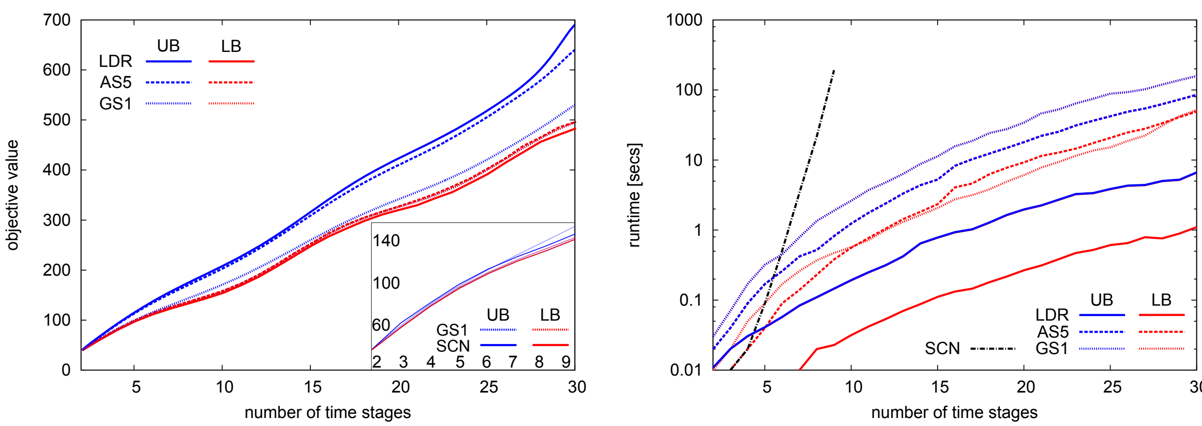

Figure 4: *Upper(UB) and lower(LB) bounds (left) and runtimes (right) of the linear and piecewise linear decision rule approximations and the scenario tree-based approximation.*

All numerical results are obtained using the IBM ILOG CPLEX 12 optimization package on a dual-core 2.4GHz machine with 4GB RAM. Table 1 reports the relative gaps between the bounds obtained from the different approximations and for planning horizons of up to 30 months. All numbers represent averages over 25 randomly generated problem instances. As expected, the **LDR** bounds are the weakest with gaps of up to 43.1%. The **AS5** bounds already provide a noticeable improvement, but the corresponding gaps are still of the order of 25%. A truly substantial improvement is offered by the **GS1** bounds, which collapse the gaps to less than 7.5% uniformly across all $T \leq 30$; see also Figure 4 (left),

	Number of time stages						
	2	5	10	15	20	25	30
LDR	0.9%	28.4%	28.8%	28.8%	32.0%	36.2%	43.1%
AS5	0.5%	24.5%	27.3%	23.9%	27.3%	24.1%	29.1%
GS1	0.6%	5.9%	5.6%	5.7%	6.5%	6.3%	7.5%
BL	0.8%	14.1%	13.7%	15.5%	17.7%	12.6%	18.9%
TL	0.7%	14.1%	13.4%	15.3%	17.4%	12.4%	18.8%
BL-1	0.7%	6.8%	5.6%	5.5%	5.9%	6.1%	7.7%
TL-1	0.7%	5.5%	4.1%	4.4%	5.6%	4.1%	5.7%
SCN	1.1%	4.1%	—	—	—	—	—

Table 1: *Relative gaps between upper and lower bounds of different approximations.*

where the upper **GS1** bound is closer to the lower **LDR** and **AS5** bounds than to the respective upper bounds. The **BL** and **TL** bounds are also noticeably stronger than the **AS5** bounds but do not achieve the high level of precision of the **GS1** bounds. However, multilinear liftings can still be of great value when used in conjunction with piecewise linear liftings. Indeed, the **TL-1** bounds dominate all other bounds in terms of accuracy across all time horizons $T \leq 30$. The scenario tree-based **SCN** bounds are competitive with the best decision rule bounds whenever they are available. However, for $T > 9$ the **SCN** bounds could not be solved within our memory limit of 4GB RAM. Table 2 reports the runtimes for computing the different bounds and clearly illustrates the trade-off between accuracy and computational cost. Note that the runtimes of all new decision rule bounds scale subexponentially with T, which is in stark contrast to the **SCN** bounds, whose runtime grows exponentially; see Figure 4 (right).

Experiments with different parameter settings have shown that the complexity of solving the inventory control problem with decision rules (as measured in terms of relative gap size) is largely independent of the number of products P, the demand uncertainty level γ_{D} and the inventory capacity $\bar{\mathrm{I}}$ but increases with the price uncertainty level γ_{P}. Moreover, the problem is most difficult to solve if the backlogging and inventory holding costs differ substantially from the average sales price $\mathbb{E}_{\xi}(R_{t,p}(\xi_t)) = 2$. The inventory control problem is therefore particularly hard to solve for the specific parameters considered here.

In principle, one could also use polynomial decision rules of a fixed degree $d \in \mathbb{N}$ and sums-of-squares polynomial inequalities to construct approximations for problem (16) [4, 12]. Polynomial decision rules emerge as a special case of the lifting approach discussed in this paper if we define a lifting whose component mappings coincide with the monomials of ξ up to degree d. The resulting primal and dual approximations admit tractable semidefinite restrictions and relaxations, respectively. When applied to problem (16), however, the resulting semidefinite programs would involve $\mathcal{O}(T^{2d})$ decision variables and $\mathcal{O}(T^{2d})$ linear matrix inequalities of dimension $\mathcal{O}(T^{d-1})$ each; see [4, Proposition 2.1]. More precisely, for $T = 4$ and $d = 2$ these semidefinite programs would already accommodate 196,416 decision variables and would thus be beyond the reach of current state-of-the-art semidefinite programming solvers. In

	Number of time stages						
	2	5	10	15	20	25	30
LDR	0.0 sec	0.0 sec	0.0 sec	0.1 sec	0.3 sec	0.6 sec	1.1 sec
	0.0 sec	0.0 sec	0.2 sec	0.8 sec	2.0 sec	3.8 sec	6.6 sec
AS5	0.0 sec	0.0 sec	0.6 sec	2.3 sec	9.3 sec	20.7 sec	49.2 sec
	0.0 sec	0.2 sec	1.2 sec	5.3 sec	18.0 sec	42.0 sec	85.2 sec
GS1	0.0 sec	0.1 sec	0.6 sec	2.1 sec	6.1 sec	15.1 sec	51.5 sec
	0.0 sec	0.3 sec	2.6 sec	11.3 sec	34.1 sec	88.4 sec	158.3 sec
BL	0.0 sec	0.1 sec	0.3 sec	1.0 sec	2.6 sec	7.1 sec	12.5 sec
	0.0 sec	0.3 sec	4.5 sec	20.3 sec	53.7 sec	160.6 sec	271.9 sec
TL	0.0 sec	0.1 sec	0.6 sec	1.5 sec	5.2 sec	13.7 sec	32.3 sec
	0.0 sec	0.4 sec	6.2 sec	25.5 sec	78.9 sec	196.1 sec	281.4 sec
BL-1	0.0 sec	0.1 sec	0.5 sec	1.9 sec	5.3 sec	13.4 sec	29.5 sec
	0.0 sec	0.6 sec	9.2 sec	36.9 sec	99.7 sec	215.75 sec	496.0 sec
TL-1	0.0 sec	0.1 sec	1.1 sec	2.5 sec	11.7 sec	21.3 sec	47.8 sec
	0.0 sec	1.0 sec	13.2 sec	48.6 sec	174.6 sec	326.3 sec	623.0 sec
SCN	0.0 sec	0.1 sec	—	—	—	—	—
	0.0 sec	0.1 sec	—	—	—	—	—

Table 2: *Runtimes of primal (top) and dual (bottom) bounds for different approximations.*

contrast, the new decision rules introduced in this paper result in linear and second-order cone programs that display more attractive scaling properties. Nevertheless, they offer significant flexibility and, as a consequence, high-quality solutions with provably small optimality gaps.

Acknowledgments: The authors thank EPSRC for financial support under grant EP/H0204554/1.

References

[1] B. Anderson and J. Moore. *Optimal Control: Linear Quadratic Methods.* Prentice Hall, 1990.

[2] R. Ash and C. Doléans-Dade. *Probability and Measure Theory.* Academic Press, 2000.

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

[4] D. Bampou and D. Kuhn. Scenario-free stochastic programming with polynomial decision rules. *IEEE Conference on Decision and Control and European Control Conference, Orlando, USA*, December 2011.

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

[6] A. Ben-Tal, B. Golany, A Nemirovski, and J.-P. Vial. Supplier-retailer flexible commitments contracts: A robust optimization approach. *Manufacturing and Service Operations Management*, 7(3):63–89, 2005.

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

[8] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization.* SIAM, 2001.

[9] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, 1996.

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

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

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

[13] J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.

[14] J. Birge and S. Wallace. A separable piecewise linear upper bound for stochastic linear programs. *SIAM Journal on Control and Optimization*, 26(3):725–739, 1988.

[15] J. Birge and R. Wets. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research*, 12(1):149–162, 1987.

[16] J. Birge and R. Wets. Sublinear upper bounds for stochastic programs with recourse. *Mathematical Programming*, 43(1–3):131–149, 1989.

[17] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2nd edition, 2011.

[18] D. B. Brown, J. Smith, and P. Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4):785–801, 2010.

[19] G. Calafiore. Multi-period portfolio optimization with linear control policies. *Automatica*, 44(10):2463–2473, 2008.

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

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

[22] J. Dupačová (as Žáčková). On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky*, 91(4):423–430, 1966.

[23] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming: Series A*, 106(3):423–432, 2006.

[24] C. Edirisinghe. Stochastic programming approximations using limited moment information, with application to asset allocation. In Gerd Infanger, editor, *Stochastic Programming*, volume 150 of *International Series in Operations Research & Management Science*, pages 97–138. Springer, 2011.

[25] K. Frauendorfer. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming*, 75(2):277–293, 1996.

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

[27] H. Gassmann and W. Ziemba. A tight upper bound for the expectation of a convex function of a multivariate random variable. In A. Prékopa and R. Wets, editors, *Stochastic Programming 84 Part I*, volume 27 of *Mathematical Programming Studies*, pages 39–53. Springer, 1986.

[28] J. Goh, N. Hall, and M. Sim. Robust optimization strategies for total cost control in project management. *Submitted for publication*, 2010.

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

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

[31] M. Hadjiyiannis, P. Goulart, and D. Kuhn. An efficient method to estimate the suboptimality of affine controllers. *IEEE Transactions on Automatic Control*, 56(12):2841–2853, 2011.

[32] P. Kall and S. Wallace. *Stochastic Programming*. John Wiley & Sons, 1994.

[33] D. Kuhn. *Generalized Bounds for Convex Multistage Stochastic Programs*, volume 548 of *Lecture Notes in Mathematics and Economical Systems*. Springer, 2004.

[34] D. Kuhn, P. Parpas, and B. Rustem. Bound-based decision rules in multistage stochastic programming. *Kybernetika*, 44(2):134–150, 2008.

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

[36] J. Luedtke, M. Namazifar, and J. Linderoth. Some results on the strength of relaxations of multilinear functions. *Mathematical Programming B*, 136(2):325–351, 2012.

[37] W. B. Powell. *Approximate Dynamic Programming*. John Wiley & Sons, 2nd edition, 2011.

[38] P. Rocha and D. Kuhn. Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *European Journal of Operational Research*, 216(2):397–408, 2012.

[39] R. Rockafellar and R. Wets. *Variational Analysis*. Springer, 1997.

[40] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. *Applied Optimization*, 99(1):111–146, 2005.

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

[42] S. Wallace. A piecewise linear upper bound on the network recourse function. *Mathematical Programming*, 38(2):133–146, 1987.

[43] S. Wallace and T. Yan. Bounding multi-stage stochastic programs from above. *Mathematical Programming*, 61(1–3):111–129, 1993.