COMISEF Computational Optimization Methods in Statistics, Econometrics and Finance

- Marie Curie Research and Training Network funded by the EU Commission through MRTN-CT-2006-034270 -

COMISEF WORKING PAPERS SERIES

WPS-043 24/08/2010

Generalized Decision Rule Approximations for Stochastic Programming via Liftings

A.Georghiou
W. Wiesemann
D. Kuhn

Generalized Decision Rule Approximations for Stochastic Programming via Liftings

Angelos Georghiou, Wolfram Wiesemann and Daniel Kuhn

Department of Computing, Imperial College London

180 Queen's Gate, London SW7 2AZ, United Kingdom.

August 9, 2010

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 non-linear decision rules in the original problem. Finally, we identify structured liftings that give rise to highly flexible piecewise linear decision rules and assess their performance in the context of a stylized investment 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 [14]. Multistage stochastic programs (with at least two adaptive decision stages) remain intractable even if one searches only for approximate solutions of medium accuracy [25].

Over the past decades, research has focused on developing solution schemes that discretize the distribution of the uncertain model parameters [10, 20, 26]. These discretization approaches theoretically achieve any desired level of accuracy at the cost of significant computational overheads. Recently, an alternative solution paradigm has emerged which preserves the exact distribution of the uncertain parameters but restricts the set of feasible adaptive decisions to those possessing a simple functional form, such as linear, piecewise linear or polynomial decision rules [6, 8, 17]. An attractive feature of these decision rule approaches is that they typically lead to polynomial-time solution schemes. Even though linear decision rules are known to be optimal for the linear quadratic regulator problem [1] and some one-dimensional robust control problems [9], 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 $\mathcal{O}(\sqrt{m})$ [7].

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 non-linear 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 non-linear 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 [19, 22].

Our axiomatic lifting approach provides a unifying framework for several decision rule approximations proposed in the recent literature. Indeed, piecewise linear [4], segregated linear [12, 13, 17], as well as algebraic and trigonometric polynomial decision rules [4, 8] 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 [6], they have already found successful use in a variety of application areas ranging from supply chain management [5] and portfolio optimization [11] to network design problems [3], project scheduling [16] and electricity procurement optimization [23]. 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 non-linear 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 [17]. We prove that the corresponding lifted problems in primal and dual linear decision rules are generically intractable. We then identify tractable special cases and construct tractable approximations for the generic case.

3. We propose a class of separable liftings that result in tractable piecewise linear continuous decision rules with a general segmentation. We show that these highly flexible decision rules can offer a substantially better approximation quality than the decision rules with axial segmentation.

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 linear 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. Section 5 generalizes the proposed lifting techniques to the multistage case, and Section 6 assesses the performance of the new non-linear primal and dual decision rules in the context of a stylized investment planning problem.

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, $\mathrm{Tr}(A)$ 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 $A, B \in \mathbb{R}^{m\times n}$. 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 *stochastic program*.

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

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 [22]. 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 [14]. 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.

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

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. For this purpose, it is more convenient to rewrite $\mathcal{SP}$ as

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_\xi\left(c(\xi)^\top x(\xi)\right) \\ \text{subject to} & x \in \mathcal{L}_{k,n}\,,\ s \in \mathcal{L}_{k,m} \\ & \left.\begin{array}{l} Ax(\xi) + s(\xi) = b(\xi) \\ s(\xi) \ge 0 \end{array}\right\} \quad \mathbb{P}_\xi\text{-a.s.}, \end{array} \tag{1}$$

where we have converted the inequality constraints to equality constraints by introducing slack variables $s \in \mathcal{L}_{k,m}$. We then proceed by establishing a min-max reformulation for problem (1).

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

Here, we have dualized the equality constraints by multiplying them with dual decisions $y \in \mathcal{L}_{k,m}$ and moving them to the objective function. It can be shown that (1) and (2) are equivalent, see [27]. Note that the maximization over the dual decisions in (2) imposes an infinite penalty on all primal decisions (x, s) that violate the equality constraints $Ax(\xi) + s(\xi) = b(\xi)$ on a set of strictly positive probability.

In the following, we use the shorthand notation '$\inf_{x,s}$' to denote the infimum over all $x \in \mathcal{L}_{k,n}$ and over all $s \in \mathcal{L}_{k,m}$ that are almost surely nonnegative. Similarly, '$\sup_y$' and '$\sup_Y$' represent the suprema over all $y \in \mathcal{L}_{k,m}$ and $Y \in \mathbb{R}^{m\times k}$, respectively. Using the equivalence of (1) and (2), we obtain

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

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

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

$\mathcal{LB}$ represents a relaxation of $\mathcal{SP}$, and therefore its optimal value provides a *lower bound* on the optimal value of $\mathcal{SP}$. 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 linear programs 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 polyhedron of the form

$$\mathrm{conv}\ \Xi = \left\{ \xi \in \mathbb{R}^k : W\xi \geq h \right\}, \tag{3}$$

where $W \in \mathbb{R}^{l\times k}$ and $h \in \mathbb{R}^l$ satisfy $W = (\mathrm{e}_1, -\mathrm{e}_1, \widehat{W}^\top)^\top$ and $h = (1, -1, 0, \ldots, 0)^\top$ for some matrix $\widehat{W} \in \mathbb{R}^{(l-2)\times k}$.

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

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

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

$$\begin{aligned} \text{minimize} \quad & \operatorname{Tr}\left(MC^\top X\right) \\ \text{subject to} \quad & X \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k} \\ & AX + S = B \\ & \left(W - h\,\mathrm{e}_1^\top\right) MS^\top \geq 0, \end{aligned} \tag{$\mathcal{LB}^*$}$$

where $M := \mathbb{E}_\xi\left(\xi\,\xi^\top\right)$ denotes the second order moment matrix of the uncertain parameters. The sizes of the linear programs $\mathcal{UB}^*$ and $\mathcal{LB}^*$ are polynomial in k, l, m and n, implying that they are tractable.

Proof See [22]. ∎

Theorem 2.1 requires a description of the convex hull of Ξ in terms of linear 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.

$(\widehat{\textbf{S4}})$ There is a compact polyhedron $\widehat{\Xi} \supseteq \operatorname{conv}\Xi$ of the form $\widehat{\Xi} = \left\{\xi \in \mathbb{R}^k : W\xi \geq h\right\}$, where W and h are defined as in **(S4)**.

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

Corollary 2.2 *If $\mathcal{SP}$ satisfies the regularity conditions **(S1)**, **(S2)** and $(\widehat{\textbf{S4}})$, 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 $\mathcal{LB}$.*

3 Lifted Stochastic Programs

The bounds provided by Theorem 2.1 and Corollary 2.2 can be calculated efficiently by solving tractable linear programs. 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 non-linearities. 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*

$$L : \mathbb{R}^k \to \mathbb{R}^{k'}, \qquad \xi \mapsto \xi', \tag{4a}$$

as well as a corresponding *retraction operator*

$$R : \mathbb{R}^{k'} \to \mathbb{R}^k, \qquad \xi' \mapsto \xi. \tag{4b}$$

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

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

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

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

where the first and second identity follow from **(A2)** and **(A3)**, respectively. Hence, $\xi' - L \circ R(\xi')$ 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.*

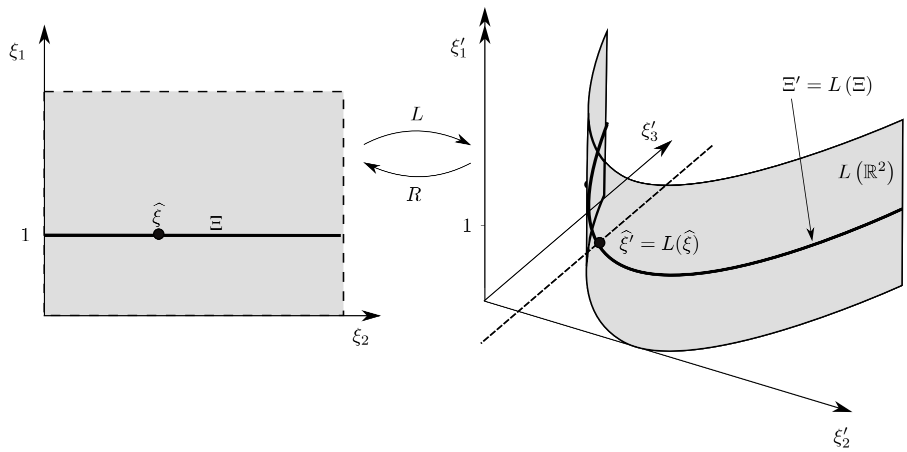

Figure 1: *Illustration of L and R. The left and right diagram show the original and the lifted space $\mathbb{R}^k$ and $\mathbb{R}^{k'}$, respectively. The shaded areas and thick solid lines represent $\mathbb{R}^k$ and Ξ in the left diagram and their lifted counterparts $L\left(\mathbb{R}^k\right)$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

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

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

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 $f : \left(\mathbb{R}^{k'}, \mathcal{B}(\mathbb{R}^{k'})\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $g : \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 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) $f\left(L\left(\xi\right)\right) \leq 0 \quad \mathbb{P}_\xi\text{-a.s.} \iff f\left(\xi'\right) \leq 0 \quad \mathbb{P}_{\xi'}\text{-a.s.}$

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

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.

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

The following proposition shows that the *l̲ifted s̲tochastic p̲rogram* $\mathcal{LSP}$ 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 We show that any feasible solution in $\mathcal{SP}$ corresponds to a feasible solution in $\mathcal{LSP}$ with the same objective value and vice versa. Suppose that $x \in \mathcal{L}_{k,n}$ is feasible in $\mathcal{SP}$, and consider the decision $x' \in \mathcal{L}_{k',n}$ defined through

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

The feasibility of x in $\mathcal{SP}$ implies that

$$\begin{aligned} & Ax\left(\xi\right) \leq b\left(\xi\right) && \mathbb{P}_{\xi}\text{-a.s.} \\ \Longleftrightarrow \quad & Ax\left(R\xi'\right) \leq b\left(R\xi'\right) && \mathbb{P}_{\xi'}\text{-a.s.} \\ \Longleftrightarrow \quad & Ax'\left(\xi'\right) \leq b\left(R\xi'\right) && \mathbb{P}_{\xi'}\text{-a.s.} \end{aligned}$$

Here, the first and second equivalence follow from Proposition 3.3 *(iv)* and the definition of x', respectively. Therefore, x' is feasible in $\mathcal{LSP}$. Moreover, by Proposition 3.3 *(ii)* we have

$$\mathbb{E}_{\xi}\left(c\left(\xi\right)^{\top} x\left(\xi\right)\right) \;=\; \mathbb{E}_{\xi'}\left(c\left(R\xi'\right)^{\top} x\left(R\xi'\right)\right) \;=\; \mathbb{E}_{\xi'}\left(c\left(R\xi'\right)^{\top} x'\left(\xi'\right)\right),$$

which implies that x in $\mathcal{SP}$ and x' in $\mathcal{LSP}$ share the same objective value.

Suppose now that $x' \in \mathcal{L}_{k',n}$ is feasible in $\mathcal{LSP}$. We define the function $x \in \mathcal{L}_{k,n}$ through

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

The feasibility of x' in $\mathcal{LSP}$ implies that

$$\begin{aligned} & Ax'(\xi') \leq b(R\xi') && \mathbb{P}_{\xi'}\text{-a.s.} \\ \iff \quad & Ax'(L(\xi)) \leq b(R \circ L(\xi)) && \mathbb{P}_{\xi}\text{-a.s.} \\ \iff \quad & Ax(\xi) \leq b(\xi) && \mathbb{P}_{\xi}\text{-a.s.} \end{aligned}$$

Here, the first equivalence follows from Proposition 3.3 (iii), while the second equivalence is due to the definition of x and **(A3)**. Hence, x is feasible in $\mathcal{SP}$. Proposition 3.3 (i) and **(A3)** also imply that

$$\mathbb{E}_{\xi'}\left(c(R\xi')^\top x'(\xi')\right) \quad = \quad \mathbb{E}_{\xi}\left(c(R \circ L(\xi))^\top x'(L(\xi))\right) \quad = \quad \mathbb{E}_{\xi}\left(c(\xi)^\top x(\xi)\right),$$

which guarantees that x' in $\mathcal{LSP}$ and x in $\mathcal{SP}$ share the same objective value. ∎

Remark 3.5 *If two pairs of lifting and retraction operators $L^1 : \mathbb{R}^k \to \mathbb{R}^{k'}$, $R^1 : \mathbb{R}^{k'} \to \mathbb{R}^k$ and $L^2 : \mathbb{R}^{k'} \to \mathbb{R}^{k''}$, $R^2 : \mathbb{R}^{k''} \to \mathbb{R}^{k'}$ 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 non-linear 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 non-linear 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 *n̲on-linear u̲pper b̲ound* on $\mathcal{SP}$.

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

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 $y\left(\xi\right) = Y'L\left(\xi\right)$ for some $Y' \in \mathbb{R}^{m\times k'}$. By using similar arguments as in Section 2, we obtain the following *n̲on-linear l̲ower b̲ound* on $\mathcal{SP}$.

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

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 It follows from Proposition 3.3 that $\mathcal{NUB}$ is equivalent to

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_{\xi'} \left(c\left(R\xi'\right)^\top X'\xi' \right) \\ \text{subject to} \quad & X' \in \mathbb{R}^{n\times k'} \\ & AX'\xi' \leq b\left(R\xi'\right) \qquad \mathbb{P}_{\xi'}\text{-a.s.}, \end{aligned}$$

which can readily be identified as $\mathcal{LUB}$. Thus assertion *(i)* follows. By using similar arguments as in

Proposition 3.4, one can further show that $\mathcal{NLB}$ is equivalent to

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

which we recognize as $\mathcal{LLB}$. This observation establishes assertion (ii). ∎

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 $\mathcal{SP}$ that 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.*

$$\inf \mathcal{LB} \;\leq\; \inf \mathcal{LLB} \;\leq\; \inf \mathcal{SP} \;=\; \inf \mathcal{LSP} \;\leq\; \inf \mathcal{LUB} \;\leq\; \inf \mathcal{UB} \tag{5}$$

Proof In Section 2 we have already seen that $\inf \mathcal{LB} \leq \inf \mathcal{SP} \leq \inf \mathcal{UB}$. Proposition 3.4 implies that $\inf \mathcal{SP} = \inf \mathcal{LSP}$, and from Proposition 3.6 we conclude that $\inf \mathcal{LLB} \leq \inf \mathcal{LSP} \leq \inf \mathcal{LUB}$. Thus, it only remains to be shown that $\inf \mathcal{LUB} \leq \inf \mathcal{UB}$ and $\inf \mathcal{LB} \leq \inf \mathcal{LLB}$.

As for the first inequality, let X be feasible in $\mathcal{UB}$ and set $X' := XR$. Then X' is feasible in $\mathcal{LUB}$ since

$$\begin{array}{lll} & AX\xi \leq b(\xi) & \mathbb{P}_\xi\text{-a.s.} \\ \iff & AXR \circ L(\xi) \leq b(R \circ L(\xi)) & \mathbb{P}_\xi\text{-a.s.} \\ \iff & AX'L(\xi) \leq b(R \circ L(\xi)) & \mathbb{P}_\xi\text{-a.s.} \\ \iff & AX'\xi' \leq b(R\xi') & \mathbb{P}_{\xi'}\text{-a.s.,} \end{array}$$

where the equivalences follow from axiom **(A3)**, the definition of X' and Proposition 3.3 *(iii)*, respectively. Moreover, X in $\mathcal{UB}$ and X' in $\mathcal{LUB}$ share the same objective value since

$$\begin{aligned} \mathbb{E}_\xi\left(c(\xi)^\top X\xi\right) &= \mathbb{E}_\xi\left(c(R \circ L(\xi))^\top XR \circ L(\xi)\right) \\ &= \mathbb{E}_\xi\left(c(R \circ L(\xi))^\top X'L(\xi)\right) \\ &= \mathbb{E}_{\xi'}\left(c(R\xi')^\top X'\xi'\right), \end{aligned}$$

where the identities follow from axiom **(A3)**, the definition of X' and Proposition 3.3 *(i)*, respectively.

On the other hand, for a generic X' feasible in $\mathcal{LUB}$ there may be no X feasible in $\mathcal{UB}$ with the same objective value. Therefore, we have $\inf \mathcal{LUB} \leq \inf \mathcal{UB}$.

Next, let (x, s) be feasible in $\mathcal{NLB}$, which is equivalent to $\mathcal{LLB}$ due to Proposition 3.6 *(ii)*. Then (x, s) is feasible in $\mathcal{LB}$ since

$$\begin{aligned} 0 &= \mathbb{E}_\xi \left([Ax(\xi) + s(\xi) - b(\xi)] \, L(\xi)^\top \right) \\ \implies 0 &= \mathbb{E}_\xi \left([Ax(\xi) + s(\xi) - b(\xi)] \, L(\xi)^\top \right) R^\top = \mathbb{E}_\xi \left([Ax(\xi) + s(\xi) - b(\xi)] \, \xi^\top \right). \end{aligned}$$

Here, the identities follow from the feasibility of (x, s) in $\mathcal{NLB}$ and axiom **(A3)**. As $\mathcal{LB}$ and $\mathcal{NLB}$ have the same objective function, we conclude that $\inf \mathcal{LB} \leq \inf \mathcal{NLB} = \inf \mathcal{LLB}$. ∎

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

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

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

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

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 linear inequalities, see conditions **(S4)** and **($\widehat{\textbf{S4}}$)**. 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 non-linear 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

$$\xi' = \left(\xi'_{11},\, \xi'_{21}, \ldots, \xi'_{2r_2},\, \xi'_{31}, \ldots, \xi'_{3r_3},\, \ldots,\, \xi'_{k1}, \ldots, \xi'_{kr_k}\right)^\top.$$

Next, we use the breakpoints to define the lifting operator $L = (L_{11}, \ldots, L_{kr_k})$, where the coordinate mapping L_{ij} corresponds to the ξ'_{ij} axis in the lifted space and is defined through

$$L_{ij}(\xi) := \begin{cases} \xi_i & \text{if } r_i = 1, \\ \min\left\{\xi_i, z_1^i\right\} & \text{if } r_i > 1,\ j = 1, \\ \max\left\{\min\left\{\xi_i, z_j^i\right\} - z_{j-1}^i, 0\right\} & \text{if } r_i > 1,\ j = 2, \ldots, r_i - 1, \\ \max\left\{\xi_i - z_{j-1}^i, 0\right\} & \text{if } r_i > 1,\ j = r_i. \end{cases} \tag{6}$$

By construction, L_{ij} 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

$$R_i(\xi') := \sum_{j=1}^{r_i} \xi'_{ij}. \tag{7}$$

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

Proposition 4.1 *The operators L and R defined in* (6) *and* (7) *satisfy the axioms **(A1)**–**(A4)**.*

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

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

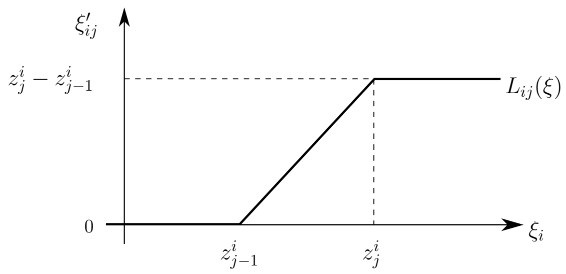

Figure 2: *Graph of the coordinate mapping L_{ij} for $1 < i \leq k$ and $1 < j < r_i$.*

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 $L_{i1}(\xi) = \xi_i$ and $L_{ij}(\xi) = 0$ for all $j = 2, \ldots, r_i$. Thus, $\sum_{j=1}^{r_i} L_{ij}(\xi) = \xi_i$.

(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

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

and thus

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

The above arguments apply for each $i = 1, \ldots, k$, and thus **(A3)** follows. Axiom **(A4)** is also satisfied since $L_{i1}, \ldots, L_{ir_i}$ 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 non-linear decision rules, are defined in the usual way. We now give a precise characterization of the decision rules that can be represented as linear combinations of the coordinate mappings (6) of the lifting L. To this end, we need the following definition.

Definition 4.2 *Let $\mathcal{F}_L$ be the linear space of all continuous functions $f : \mathbb{R}^k \to \mathbb{R}$ which vanish at the origin and are affine on the hyperrectangles*

$$\bigtimes_{i=1}^{k} \left[z^i_{j_i-1} - z^i_{j_i}\right], \qquad \forall j_i = 1, \ldots, r_i,\; i = 1, \ldots, k. \tag{8}$$

Here, we use the convention $z^1_0 := -\infty$ and $z^i_{r_i} := +\infty$. We will refer to $\mathcal{F}_L$ as the space of all piecewise *linear continuous decision rules induced by L.*

Proposition 4.3 *Without loss of generality, assume that $z_1^i = 0$ for all $i = 2, \ldots, k$. Then $\mathcal{F}_L$ coincides with the space of all functions $f : \mathbb{R}^k \to \mathbb{R}$ of the form $f(\xi) = v^\top L(\xi)$ for some $v \in \mathbb{R}^{k'}$, that is, all functions that can be represented as linear combinations of the coordinate mappings corresponding to L.*

Proof Choose $f \in \mathcal{F}_L$. We first show that there exist piecewise constant functions $\phi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, k$, such that

$$\frac{\partial f(\xi)}{\partial \xi_i} = \phi_i(\xi_i) \qquad \forall \xi_i \notin \left\{z_1^i, \ldots, z_{r_i-1}^i\right\}, \tag{9}$$

where ϕ_i is constant on (z_{j-1}^i, z_j^i) for all $j = 1, \ldots, r_i$. Since f is piecewise linear, $\partial f(\xi)/\partial \xi_i$ is constant on the interior of each hyperrectangle in (8). It remains to be shown that $\partial f(\xi)/\partial \xi_i$ is constant in $\xi_{-i} := (\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_k)$. To this end, choose $\xi_i \notin \{z_1^i, \ldots, z_{r_i-1}^i\}$ and consider the function

$$\xi_{-i} \mapsto \frac{\partial f(\xi_i, \xi_{-i})}{\partial \xi_i}. \tag{10}$$

Since f is piecewise linear, (10) is locally constant at any ξ_{-i} whose components do not coincide with any breakpoint. Also, (10) inherits continuity from f and is therefore globally constant. This establishes (9).

Since $f \in \mathcal{F}_L$, f vanishes at the origin, and we can recover f from its partial derivatives through

$$f(\xi) = \sum_{i=1}^{k} \int_0^{\xi_i} \phi_i(\hat{\xi}_i) \, \mathrm{d}\hat{\xi}_i.$$

Since ϕ_i is a piecewise constant function, we conclude that $F_i(\xi_i) := \int_0^{\xi_i} \phi_i(\hat{\xi}_i) \, \mathrm{d}\hat{\xi}_\mathrm{i}$ is continuous and piecewise linear on the intervals $[z_{j-1}^i, z_j^i]$ for all $i = 1, \ldots, k$, $j = 1, \ldots, r_i$. Note that $L_{ij}(0) = 0$ for all $i = 1, \ldots, k$, $j = 1, \ldots, r_i$. Thus, (6) implies that there are unique coefficients $v_{i1}, \ldots, v_{ir_i} \in \mathbb{R}$ such that

$$F_i(\xi_i) = \sum_{j=1}^{r_i} v_{ij} L_{ij}(\xi) \qquad \forall \xi \in \mathbb{R}^k$$

for each $i = 1, \ldots, k$. We therefore have

$$f(\xi) \;=\; \sum_{i=1}^{k} F_i(\xi_i) \;=\; \sum_{i=1}^{k} \sum_{j=1}^{r_i} v_{ij} L_{ij}(\xi).$$

Thus, f is equivalent to a linear combination of the coordinate mappings corresponding to L. Conversely, since each coordinate mapping of L is contained in $\mathcal{F}_L$, it is clear that $v^\top L \in \mathcal{F}_L$ for all $v \in \mathbb{R}^{k'}$. ∎

Proposition 4.3 implies that the approximate problems $\mathcal{NUB}$ and $\mathcal{NLB}$ optimize over all piecewise linear continuous decision rules that are induced by L. We now demonstrate that these problems are generically intractable for liftings of the type (6). To this end, we need the following auxiliary result.

Lemma 4.4 *The following decision problem is* NP*-hard:*

> INSTANCE. A convex polytope $\Xi \subset \mathbb{R}^k$ of the form (3) and $\tau \in \mathbb{R}$.
>
> QUESTION. Do all $\xi \in \Xi$ satisfy $\sum_{i=1}^{k} |\xi_i| \le \tau$?

Proof See [18, Lemma 3.2]. ∎

Theorem 4.5 *The approximate problems $\mathcal{LUB}$ and $\mathcal{LLB}$ defined through a lifting operator L of the type* (6) *are* NP*-hard even if there is only one breakpoint per coordinate axis.*

Proof Let $\Xi \subset \mathbb{R}^k$ be a convex polytope of the type (3) and denote by $\mathbb{P}_\xi$ the uniform distribution on Ξ. For a fixed scalar $\tau \in \mathbb{R}$, we define the following instance of $\mathcal{SP}$.

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & x \in \mathcal{L}_{k,k} \\ & \left.\begin{array}{l} -x_i(\xi) \le \xi_i \le x_i(\xi)\,,\ i = 1,\ldots,k \\ \displaystyle\sum_{i=1}^{k} x_i(\xi) \le \tau \end{array}\right\} \ \mathbb{P}_\xi\text{-a.s.} \end{array} \tag{$\mathcal{SP}'$}$$

$\mathcal{SP}'$ is optimized by $x^* \in \mathcal{L}_{k,k}$ defined through $x_i^*(\xi) := |\xi_i|$, $i = 1,\ldots,k$. We thus have

$$\sum_{i=1}^{k} |\xi_i| \le \tau \quad \forall \xi \in \Xi \qquad \Longleftrightarrow \qquad \inf \mathcal{SP}' \le 0.$$

Hence, Lemma 4.4 implies that checking the feasibility of $\mathcal{SP}'$ is NP-hard. We now set $r_i := 2$ and $z_1^i := 0$ for all $i = 2,\ldots,k$, and we define the lifting operator $L : \mathbb{R}^k \to \mathbb{R}^{k'}$ as in (6) with $k' := 2k-1$. By construction, there exists $X' \in \mathbb{R}^{k \times k'}$ such that $x^*(\xi) = X' L(\xi)$, and thus we have

$$\inf \mathcal{SP}' \;=\; \inf \mathcal{NUB}' \;=\; \inf \mathcal{LUB}'.$$

The above arguments allow us to conclude that

$$\sum_{i=1}^{k} |\xi_i| \le \tau \quad \forall \xi \in \Xi \qquad \Longleftrightarrow \qquad \inf \mathcal{LUB}' \le 0.$$

By Lemma 4.4, $\mathcal{LUB}'$ is thus NP-hard. Hence, generic problems of the type $\mathcal{LUB}$ are NP-hard as well.

To prove NP-hardness of $\mathcal{LLB}$, we consider the following instance of $\mathcal{SP}$.

$$
\begin{array}{lll}
\text{minimize} & \mathbb{E}_\xi\Big(\Big(\tau - \sum_{i=1}^{k} \xi_i\Big) x_0(\xi) + 2\sum_{i=1}^{k} \xi_i\, x_i(\xi)\Big) & \\
\text{subject to} & x = (x_0, \ldots, x_k) \in \mathcal{L}_{k,k+1},\ s^1 \in \mathcal{L}_{k,k+1},\ s^2 \in \mathcal{L}_{k,k+1} & \\
& \left.\begin{array}{l} x_i(\xi) - s_i^1(\xi) = 0 \\ x_0(\xi) - x_i(\xi) - s_i^2(\xi) = 0 \\ s_i^1(\xi) \geq 0,\ s_i^2(\xi) \geq 0 \end{array}\right\} i = 0, \ldots, k, \quad \mathbb{P}_\xi\text{-a.s.}
\end{array}
\qquad (\mathcal{SP}'')
$$

Note that we used the equivalent reformulation (1) of $\mathcal{SP}$ to express $\mathcal{SP}''$. The problem $\mathcal{SP}''$ uses the same measure $\mathbb{P}_\xi$ and support Ξ as our previous problem $\mathcal{SP}'$.[1] We can now construct $\mathcal{NLB}''$ in the usual way by using the same lifting operator L as in the first part of this proof.

$$
\begin{array}{lll}
\text{minimize} & \mathbb{E}_\xi\Big(\Big(\tau - \sum_{i=1}^{k} \xi_i\Big) x_0(\xi) + 2\sum_{i=1}^{k} \xi_i\, x_i(\xi)\Big) & \\
\text{subject to} & x = (x_0, \ldots, x_k) \in \mathcal{L}_{k,k+1},\ s^1 \in \mathcal{L}_{k,k+1},\ s^2 \in \mathcal{L}_{k,k+1} & \\
& \left.\begin{array}{l} \mathbb{E}_\xi\big(\big[x_i(\xi) - s_i^1(\xi)\big] L(\xi)\big) = 0 \\ \mathbb{E}_\xi\big(\big[x_0(\xi) - x_i(\xi) - s_i^2(\xi)\big] L(\xi)\big) = 0 \\ s_i^1(\xi) \geq 0,\ s_i^2(\xi) \geq 0 \end{array}\right\} i = 0, \ldots, k, \quad \mathbb{P}_\xi\text{-a.s.}
\end{array}
\qquad (\mathcal{NLB}'')
$$

The dual of $\mathcal{NLB}''$ in the sense of [15] is given by

$$
\begin{array}{ll}
\text{maximize} & 0 \\
\text{subject to} & y_i^1 \in \mathbb{R}^{k'},\ y_i^2 \in \mathbb{R}^{k'},\ i = 0, \ldots, k \\
& \left.\begin{array}{l} \tau - \sum_{i=1}^{k} \xi_i + y_0^{1\top} L(\xi) + \sum_{i=1}^{k} y_i^{2\top} L(\xi) = 0 \\ 2\xi_i + y_i^{1\top} L(\xi) - y_i^{2\top} L(\xi) = 0,\ i = 1, \ldots, k \\ y_i^{1\top} L(\xi) \leq 0,\ i = 0, \ldots, k \\ y_i^{2\top} L(\xi) \leq 0,\ i = 0, \ldots, k \end{array}\right\} \mathbb{P}_\xi\text{-a.s.}
\end{array}
\qquad (11)
$$

Proposition 4.1 in [15] implies that strong duality holds. Thus the inequality $\inf \mathcal{NLB}'' \geq 0$ is satisfied if and only if (11) is feasible. Using the second constraint group in (11) to eliminate the variables y_i^2,

[1] We remark that $\mathcal{SP}''$ can be related to the dual of $\mathcal{SP}'$. However, this relation is irrelevant for our argumentation.

$i = 0, \ldots, k$, the constraints in (11) can be equivalently expressed as

$$\left.\begin{array}{l} \tau + \sum_{i=1}^{k} \xi_i + \sum_{i=1}^{k} y_i^{1\top} L(\xi) = -y_0^{1\top} L(\xi) \\ y_i^{1\top} L(\xi) \leq \min\{0, -2\xi_i\},\ i = 1, \ldots, k \\ y_0^{1\top} L(\xi) \leq 0 \end{array}\right\} \quad \mathbb{P}_\xi\text{-a.s.}$$

Next, using the first equation in the above constraint system to eliminate y_0^1, we obtain

$$\left.\begin{array}{l} \tau + \sum_{i=1}^{k} \xi_i + \sum_{i=1}^{k} y_i^{1\top} L(\xi) \geq 0 \\ y_i^{1\top} L(\xi) \leq \min\{0, -2\xi_i\},\ i = 1, \ldots, k \end{array}\right\} \quad \mathbb{P}_\xi\text{-a.s.}$$

By setting $y_i^{1\top} L(\xi) = \min\{0, -2\xi_i\}$, which is possible because $\min\{0, -2\xi_i\}$ is a continuous piecewise linear function with a breakpoint at 0, we find that this last constraint system is feasible if and only if

$$\tau - \sum_{i=1}^{k} |\xi_i| \geq 0 \qquad \mathbb{P}_\xi\text{-a.s.}$$

Proposition 3.6 implies that

$$\inf \mathcal{LLB}'' \geq 0 \iff \inf \mathcal{NLB}'' \geq 0 \iff \sum_{i=1}^{k} |\xi_i| \leq \tau \quad \forall \xi \in \Xi.$$

By Lemma 4.4, $\mathcal{LLB}''$ is thus NP-hard. Hence, generic problems of the type $\mathcal{LLB}$ are NP-hard as well.
■

Theorem 4.5 implies that we cannot solve $\mathcal{LUB}$ and $\mathcal{LLB}$ efficiently for generic liftings of the type (6), 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 (3). 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 (6). Afterwards, we construct a tractable outer approximation for conv Ξ' in generic situations.

Let Ξ be a hyperrectangle of the type

$$\Xi = \left\{\xi \in \mathbb{R}^k : \xi_1 = 1,\ \ell \leq \xi \leq u\right\}. \tag{12}$$

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

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

and constitutes a non-convex, connected and compact set, see (6). 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*

$$L_i := \begin{cases} \mathbb{R}^k \to \mathbb{R}^{r_i} \\ \xi \mapsto \xi'_i := \left(L_{i1}\left(\xi\right), \ldots, L_{ir_i}\left(\xi\right)\right)^\top \end{cases} \tag{13}$$

for $i = 1, \ldots, k$. Note that due to (6) 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

$$\Xi' \;=\; L\left(\Xi\right) \;=\; \mathop{\times}_{i=1}^{k} L_i\left(\Xi\right) \;=\; \mathop{\times}_{i=1}^{k} \Xi'_i, \tag{14}$$

where $\Xi'_i := L_i\left(\Xi\right)$. The marginal supports Ξ'_i inherit the non-convexity, connectedness and compactness from Ξ'. Note that (14) implies

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

and therefore it is sufficient to derive a closed-form representation for the marginal convex hulls conv Ξ'_i. Recall that $\ell_i < z^i_1$ and $z^i_{r_i-1} < u_i$ 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 $[\ell_i, u_i]$.

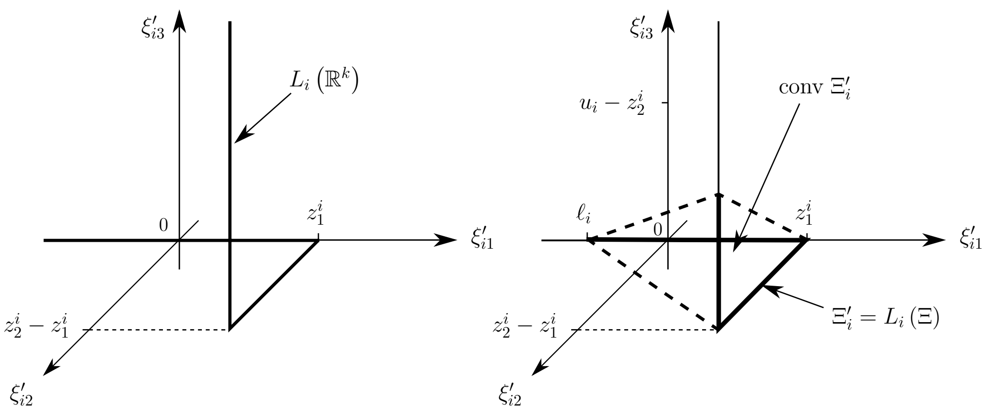

Figure 3: *The left diagram illustrates the range of the partial lifting L_i, which consists of three perpendicular line segments. Here, we assume that there are only two breakpoints at z^i_1 and z^i_2 along the ξ_i direction (i.e., $r_i = 3$). The right diagram shows the marginal support Ξ'_i (thick line) as well as its convex hull, which is given by a simplex (thick and dashed lines).*

Lemma 4.6 *The convex hull of* Ξ'_i*,* $i = 2, \ldots, k$*, is given by*

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

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

$$v_0 = \begin{pmatrix} \ell_i \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, v_1 = \begin{pmatrix} z_1^i \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} z_1^i \\ z_2^i - z_1^i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, v_{r_i} = \begin{pmatrix} z_1^i \\ z_2^i - z_1^i \\ \vdots \\ z_{r_i-1}^i - z_{r_i-2}^i \\ u_i - z_{r_i-1}^i \end{pmatrix}.$$

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 4.6 allows us to write the convex hull of Ξ' as

$$\begin{aligned}\text{conv}\,\Xi' &= \bigtimes_{i=1}^{k} \text{conv}\,\Xi'_i \\ &= \left\{\xi' = (\xi'_1, \ldots, \xi'_k) \in \bigtimes_{i=1}^{k} \mathbb{R}^{r_i} : \xi'_1 = 1,\ V_i^{-1}(1, \xi_i'^{\top})^{\top} \geq 0,\ i = 2, \ldots, k\right\}.\end{aligned} \tag{15}$$

Note that conv Ξ' is of the form (3) 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 linear programs $\mathcal{LUB}^*$ 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 conv Ξ' is described by $O(k')$ 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 polytope of the type (3). Then the convex hull of Ξ' has no tractable representation. However, we can systematically construct a tractable outer approximation for conv Ξ'. To this end, let $\{\xi \in \mathbb{R}^k : \ell \leq \xi \leq u\}$ be the smallest hyperrectangle containing Ξ. We have

$$\begin{aligned}\Xi &= \left\{\xi \in \mathbb{R}^k : W\xi \geq h\right\} \\ &= \left\{\xi \in \mathbb{R}^k : W\xi \geq h,\ \ell \leq \xi \leq u\right\},\end{aligned} \tag{16}$$

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

$$\begin{aligned}\Xi'_1 &:= \left\{\xi' \in \mathbb{R}^{k'} : WR\xi' \geq h\right\} \\ \Xi'_2 &:= \left\{\xi' \in \mathbb{R}^{k'} : L \circ R\,(\xi') = \xi',\ \ell \leq R(\xi') \leq u\right\}.\end{aligned}$$

We thus conclude that

$$\widehat{\Xi}' := \left\{\xi' \in \mathbb{R}^{k'} : WR\xi' \geq h,\ V_i^{-1}(1, \xi_i'^{\top})^{\top} \geq 0,\ i = 2, \ldots, k\right\} \supseteq \text{conv}\,\Xi' \tag{17}$$

since $\widehat{\Xi}' = \Xi'_1 \cap \text{conv}\,\Xi'_2$ and $\Xi'_1 = \text{conv}\,\Xi'_1$, see (15). Note that $\widehat{\Xi}'$ is of the form (3) and therefore satisfies condition $(\widehat{\mathbf{S4}})$. 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, since $\widehat{\Xi}'$ has $O(l + k')$ facets, where l denotes the number of rows in matrix W, the sizes of $\mathcal{LUB}^*$ and $\mathcal{LLB}^*$ are polynomial in k, l, m, n and the dimension k' of the lifted space.

The main results of this subsection can be 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* (6)*. Then the following hold.*

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

(ii) If Ξ is a hyperrectangle of the type (12)*, 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 and k', implying that they are efficiently solvable.

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 4.8 *Consider the stochastic program*

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

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 4.8 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

$$z_1^i < z_2^i < \ldots < z_{r_i-1}^i. \tag{18}$$

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 (6) 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 (6) and (7) by using the breakpoints (18). We set

$k'_\eta := \sum_{i=1}^{k_\eta} r_i$. In analogy to Proposition 4.3, 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 $\eta \in \text{kernel}(F^\top)$, $\eta \neq 0$, 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

$$L := G \circ L^\eta \circ F, \tag{19}$$

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 (19) 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

$$R := F^+ \circ R^\eta \circ G^+, \tag{20}$$

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 4.9 *The operators L and R defined in* (19) *and* (20) *satisfy **(A1)**–**(A4)**.*

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

$$R \circ L \;=\; F^+ \circ R^\eta \circ G^+ \circ G \circ L^\eta \circ F \;=\; \mathbb{I}_k. \tag{21}$$

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, (21) 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 (21).

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

$$E^\perp \subseteq \left\{ (R^\eta)^\top \eta : \; \eta \in \mathrm{kernel}(F^\top) \right\}. \tag{22}$$

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 (22) 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 (19) 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 (19) even if there is only one breakpoint per folding direction, see Theorem 4.5. 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

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

where the last equality holds since the linear operator G preserves convexity, see [24, Proposition 2.21]. Therefore, our problem reduces to finding an outer approximation for $\mathrm{conv}\; L^\eta \circ F\left(\Xi\right)$. 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

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

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 (16) in the sense that it involves a set of generic linear constraints as well as box constraints. Thus, an outer approximation for the convex hull of $L^\eta(\Theta)$ is given by

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

see (17), where the matrices V_i^{-1} are defined as in Lemma 4.6. Thus the resulting outer approximation for conv Ξ' is given by $\widehat{\Xi}' := G(\widehat{\Theta}')$. This set represents a polytope that 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 $\mathcal{LLB}^*$.

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

Theorem 4.10 *Assume that the original problem $\mathcal{SP}$ satisfies **(S1)**–**(S4)** and consider any lifting of the type* (19)*. Then the following hold.*

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

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

5 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

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

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)** is necessary to guarantee 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 [22]. These problems turn out to be tractable if the convex hull of Ξ is representable by a finite set of linear inequalities, as stated in the following assumption.

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

$$\text{conv}\,\Xi = \left\{\xi \in \mathbb{R}^k : W\xi \geq h\right\},$$

where $W \in \mathbb{R}^{l\times k}$ and $h \in \mathbb{R}^l$ are defined in condition **(S4)**, see Section 2.

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

Theorem 5.1 *If $\mathcal{MSP}$ satisfies the conditions **(M1)**, **(M2)** and **(M5)**, then $\mathcal{MUB}$ is equivalent to*

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

*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*

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

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

Proof See [22]. ∎

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.

$(\widehat{\mathbf{M5}})$ There is a compact polyhedron $\widehat{\Xi} \supseteq \operatorname{conv} \Xi$ of the form $\widehat{\Xi} = \left\{\xi \in \mathbb{R}^k : W\xi \geq h\right\}$, where W and h are defined in condition **(S4)**, see Section 2.

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

Corollary 5.2 *If $\mathcal{MSP}$ satisfies the conditions **(M1)**, **(M2)** and $(\widehat{\mathbf{M5}})$, 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 $\mathcal{MSP}$:

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

where $\mathbb{P}_{\xi'}$ and P_t are defined in Section 3 and Theorem 5.1, respectively.

Proposition 5.3 *$\mathcal{MSP}$ 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 5.1 and Corollary 5.2 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 **($\widehat{\textbf{M5}}$)**, respectively. In a first step we verify the satisfaction of the conditions **(M1)**–**(M4)**.

Proposition 5.4 *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 **($\widehat{\textbf{M5}}$)** 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, then it is easy to show that $\mathcal{LMSP}$ satisfies either **(M5)** or $\widehat{\textbf{(M5)}}$ whenever $\mathcal{MSP}$ does so. In this situation, we can solve the approximate linear decision rule problems $\mathcal{LMUB}^*$ and $\mathcal{LMLB}^*$ efficiently.

Remark 5.5 *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 piecewise linear liftings discussed in Section 4, $\mathcal{LMSP}$ can be shown to satisfy **(M5)** or $\widehat{\textbf{(M5)}}$ whenever $\mathcal{MSP}$ does so. Thus, $\mathcal{LMUB}^*$ provides a tractable conservative approximation for the original problem $\mathcal{MSP}$.*

6 Numerical Example

We apply the decision rule approximations of Section 4 to a stylized version of the capacity expansion model discussed in [21]. Consider an electricity system comprising a set $R = \{1, \ldots, \overline{r}\}$ of regions, where the electricity demand in region $r \in R$ is described by the random variable δ_r. All demands have to be satisfied by a set $N = \{1, \ldots, \overline{n}\}$ of power plants, where plant $n \in N$ can produce up to $\overline{g}_n$ units of electricity at uncertain unit costs ζ_n. Each power plant $n \in N$ is located in one of the regions $r \in R$, and we denote the set of plants located in region r by $N(r) \subseteq N$. The regions $r \in R$ are connected by a set $M = \{1, \ldots, \overline{m}\}$ of directed transmission lines, where line $m \in M$ has a capacity of $\overline{f}_m$ units of electricity. We denote by $M_+(r) \subseteq M$ the set of transmission lines that are directed towards region $r \in R$, while $M_-(r) \subseteq M$ represents the set of lines that emanate from region r.

We model the capacity expansion problem as a two-stage stochastic program. In the first stage, the capacity of plant $n \in N$ (line $m \in M$) can be expanded by a fraction $u_n \in [0, 1]$ ($v_m \in [0, 1]$) at linear costs $c_n u_n$ ($d_m v_m$). Then, the uncertain demands $\delta = (\delta_r)_{r \in R}$ and operating costs $\zeta = (\zeta_n)_{n \in N}$ are revealed. In the second stage, the expanded system is put into operation, that is, the amount of electricity g_n produced by plant $n \in N$ and the amount of electricity f_m transmitted on line $m \in M$ are chosen. If $f_m \geq 0$, then $|f_m|$ units of electricity are transmitted along the direction of line $m \in M$, whereas $f_m < 0$ means that $|f_m|$ units of electricity are transmitted in the opposite direction. The goal is to minimize the sum of investment costs and expected operating costs while satisfying all regional demands. The problem can be formulated as the following instance of $\mathcal{MSP}$.

$$
\begin{array}{ll}
\text{minimize} & \sum\limits_{n\in N} c_n u_n + \sum\limits_{m\in M} d_m v_m + \mathbb{E}_\xi \left(\sum\limits_{n\in N} \zeta_n g_n(\xi) \right) \\
\text{subject to} & u \in \mathbb{R}^{\overline{n}},\ v \in \mathbb{R}^{\overline{m}},\ g \in \mathcal{L}^2_{k,\overline{n}},\ f \in \mathcal{L}^2_{k,\overline{m}} \\
& \left.\begin{array}{ll}
0 \le u_n \le 1 & \forall n \in N \\
0 \le v_m \le 1 & \forall m \in M \\
0 \le g_n(\xi) \le \overline{g}_n\,(1+u_n) & \forall n \in N \\
-\overline{f}_m\,(1+v_m) \le f_m(\xi) \le \overline{f}_m\,(1+v_m) & \forall m \in M \\
\sum\limits_{n\in N(r)} g_n(\xi) + \sum\limits_{m\in M_+(r)} f_m(\xi) \ge \sum\limits_{m\in M_-(r)} f_m(\xi) + \delta_r & \forall r \in R
\end{array}\right\} \ \mathbb{P}_\xi\text{-a.s.}
\end{array}
\tag{23}
$$

By a slight abuse of notation, we denote the second stage random variables by $\xi = (\delta, \zeta)$. The first two constraints in (23) limit the expansion potential. The next pair of constraints ensures that the capacities of the generators and transmission lines are obeyed. Finally, the last constraint ensures energy conservation: in any region $r \in R$, the total amount of outflowing electricity must not exceed the total amount of inflowing electricity.

We can find suboptimal but feasible solutions by solving (23) in linear or piecewise linear decision rules. To assess the performance of these approximations, we generate random test instances of problem (23) according to the following procedure. For a given number of $\bar{r}$ regions, we randomly construct a connected electricity network which accommodates, on average, $\bar{r}/2$ power plants and $\bar{r}^2/4$ transmission lines. The regional demands are modeled as independent random variables with uniform distributions, and the initial plant and line capacities are chosen such that the nominal system demand can be served. The uncertain operating costs of the plants are modeled as affine functions of two risk factors (*e.g.*, the prices for oil and gas).

We generate upper and lower bounds on the optimal value of (23) via linear decision rules, piecewise linear continuous decision rules with axial segmentation (hereafter 'axial decision rules'), and piecewise linear continuous decision rules with general segmentation ('general decision rules'). In all experiments the breakpoints are placed uniformly within the marginal supports of the respective random parameters. The axial decision rules are additively separable with respect to the components of $\xi = (\delta, \zeta)$, whereas the general decision rules include an additional term that is piecewise linear in the difference between the two risk factors explaining the operating costs. Intuitively, this extra term allows the system operator to respond to changes in the fuel price structure. Note that other terms depending *e.g.* on demand differences between adjacent regions might also prove beneficial. For the sake of simplicity, we disregard such extensions.

In the first test series, we solve 100 instances of the bounding problems $\mathcal{LMUB}^*$ and $\mathcal{LMLB}^*$ for a network of 10 regions ($\bar{r} = 10$). The aim is to assess the performance of the axial and general decision rules as their complexity (number of breakpoints) increases, see Figure 4, left. We observe that the average relative gap for linear decision rules (zero breakpoints) amounts to 51%, while as the average relative gap for the axial and general decision rules with 9 breakpoints amounts to 17% and 3%, respectively. Our results show that the piecewise linear decision rules outperform the linear decision rules by a significant margin. Moreover, the example demonstrates that the general decision rules can provide the extra flexibility needed for finding a near-optimal solution. Every instance in this test was solved within 2 minutes on a 2.4 GHz machine using CPLEX 12.1.

We now investigate the scalability of the decision rule approximations. To this end, we solve 100 instances of $\mathcal{LMUB}^*$ and $\mathcal{LMLB}^*$ corresponding to networks of up to 30 regions using linear as well as axial and general decision rules with 5 breakpoints. We observe that the average relative gap for all three types of decision rules is constant in the instance size, indicating that the approximation quality of the decision rules is independent of the problem size. The average solution time for an instance with 30 regions amounts to 5, 2600 and 2700 seconds for the linear, axial and general decision rules, respectively.

In the third test series, we examine the approximation quality of the decision rules as the degree of demand and price uncertainty changes. We solve 100 instances of $\mathcal{LMUB}^*$ and $\mathcal{LMLB}^*$ corresponding to networks of 20 regions and vary the size of the support of the random parameters, see Figure 4, right. We observe that the approximation quality of the linear decision rules is highly sensitive to the size of the random parameters' support, while the piecewise linear decision rules (with 5 breakpoints) perform more consistently as the support increases. Again, we observe that the approximation quality of the general decision rules is significantly better than that of the axial decision rules.

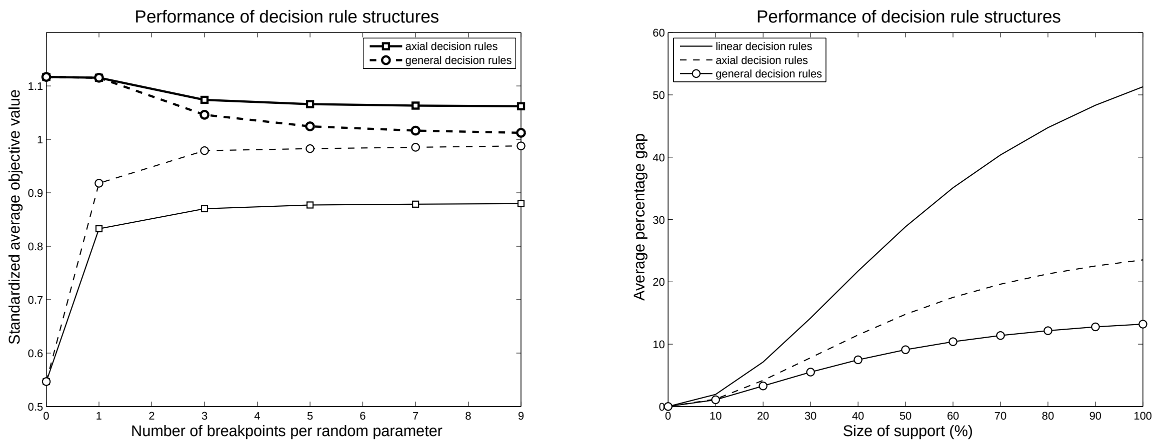

Figure 4: *The left diagram illustrates the performance of the piecewise linear decision rules in dependence of the number of breakpoints. The thick lines indicate the average standardized objective value of $\mathcal{LMUB}^*$, while the thin lines indicate the average standardized objective value of problem $\mathcal{LMLB}^*$. The right diagram illustrates the deterioration of the solution quality for the three different types of decision rules as a function of the size of the random parameters' support.*

Acknowledgements:

The authors would like to thank EPSRC for financial support under grant EP/H0204554/1.

References

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

[2] R.B. Ash. *Measure, Integration, and Functional Analysis.* Academic Press, 1972.

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

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

[5] 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.

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

[7] D. Bertsimas and V. Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Submitted for publication*, 2009.

[8] D. Bertsimas, D.A. Iancu, and P.A. Parrilo. A hierarchy of near-optimal policies for multi-stage adaptive optimization. *Submitted for publication*, 2009.

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

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

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

[12] Xin Chen, Melvyn Sim, Peng Sun, and Jiawei Zhang. A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357, 2008.

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

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

[15] M.J. Eisner and P. Olsen. Duality for stochastic programming interpreted as L.P. in L_p-space. *SIAM Journal on Applied Mathematics*, 28(4):779–792, 1975.

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

[17] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, to appear, 2010.

[18] E. Guslitser. Uncertainty-immunized solutions in linear programming. Master's thesis, Technion-Israel Institute of Technology, 2002.

[19] M.J. Hadjiyiannis, P.J. Goulart, and D. Kuhn. An efficient method to estimate the suboptimality of affine controllers. *Submitted for publication*, 2010.

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

[21] D. Kuhn, P. Parpas, and B. Rustem. Stochastic optimization of investment planning problems in the electric power industry. *Process Systems Engineering*, Volume 5: Energy Systems Engineering, (M. Georgiadis, E. Kikkinides, and E. Pistikopoulos, eds.), Wiley-VCH:215–230, 2008.

[22] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming: Series A, to appear*, 2009.

[23] P. Rocha and D. Kuhn. Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *Submitted for publication*, 2010.

[24] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis.* Springer, 1997.

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

[26] R.J.B. Wets. *Stochastic Programming.* North-Holland, 1989.

[27] S.E. Wright. Primal-dual aggregation and disaggregation for stochastic linear programs. *Mathematics of Operations Research*, 19(4):893–908, 1994.