Math. Program., Ser. A (2018) 167:395–433
https://doi.org/10.1007/s10107-017-1135-6

FULL LENGTH PAPER

Binary decision rules for multistage adaptive mixed-integer optimization

Dimitris Bertsimas[1] · **Angelos Georghiou**[2]

Received: 2 October 2014 / Accepted: 16 March 2017 / Published online: 24 March 2017

Abstract Decision rules provide a flexible toolbox for solving computationally demanding, multistage adaptive optimization problems. There is a plethora of real-valued decision rules that are highly scalable and achieve good quality solutions. On the other hand, existing binary decision rule structures tend to produce good quality solutions at the expense of limited scalability and are typically confined to worst-case optimization problems. To address these issues, we first propose a linearly parameterised binary decision rule structure and derive the exact reformulation of the decision rule problem. In the cases where the resulting optimization problem grows exponentially with respect to the problem data, we provide a systematic methodology that trades-off scalability and optimality, resulting to practical binary decision rules. We also apply the proposed binary decision rules to the class of problems with random-recourse and show that they share similar complexity as the fixed-recourse problems. Our numerical results demonstrate the effectiveness of the proposed binary decision rules and show that they are (i) highly scalable and (ii) provide high quality solutions.

Keywords Adaptive optimization · Binary decision rules · Mixed-integer optimization

Mathematics Subject Classification 90C15

✉ Dimitris Bertsimas
dbertsim@mit.edu

Angelos Georghiou
angelos.georghiou@mcgill.ca

1 Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

2 Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada

1 Introduction

In this paper, we use robust optimization techniques to develop efficient solution methods for the class of multistage adaptive mixed-integer optimization problems. The one-stage variant of the problem can be described as follows: Given matrices $A \in \mathbb{R}^{m\times k}$, $B \in \mathbb{R}^{m\times q}$, $C \in \mathbb{R}^{n\times k}$, $D \in \mathbb{R}^{q\times k}$, $H \in \mathbb{R}^{m\times k}$ and a probability measure $\mathbb{P}_{\boldsymbol{\xi}}$ supported on set Ξ for the uncertain vector $\boldsymbol{\xi} \in \mathbb{R}^k$, we are interested in choosing n real-valued functions $\boldsymbol{x}(\cdot) \in \mathcal{R}_{k,n}$ and q binary functions $\boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}$ in order to solve:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\Big((C\boldsymbol{\xi})^\top \boldsymbol{x}(\boldsymbol{\xi}) + (D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})\Big) \\ \text{subject to} & \left.\begin{array}{l} \boldsymbol{x}(\cdot) \in \mathcal{R}_{k,n},\ \boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}, \\ A\boldsymbol{x}(\boldsymbol{\xi}) + B\boldsymbol{y}(\boldsymbol{\xi}) \le H\boldsymbol{\xi}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{1.1}$$

Here, $\mathcal{R}_{k,n}$ denotes the space of all real-valued functions from $\mathbb{R}^k$ to $\mathbb{R}^n$, and $\mathcal{B}_{k,q}$ the space of binary functions from $\mathbb{R}^k$ to $\{0, 1\}^q$. Problem (1.1) has a long history both from the theoretical and practical point of view, with applications in many fields such as engineering [22,32], operations management [5,30], and finance [13,14]. From a theoretical point of view, Dyer and Stougie [18] have shown that computing the optimal solution for the class of Problems (1.1) involving only real-valued decisions, is #P-hard, while their multistage variants are believed to be "*computationally intractable already when medium-accuracy solutions are sought*" [31]. In view of these complexity results, there is a need for computationally efficient solution methods that possibly sacrifice optimality for tractability. To quote Shapiro and Nemirovski [31], "*... in actual applications it is better to pose a modest and achievable goal rather than an ambitious goal which we do not know how to achieve*".

A drastic simplification that achieves this goal is to use *decision rules*. This functional approximation restricts the infinite space of the adaptive decisions $\boldsymbol{x}(\cdot)$ and $\boldsymbol{y}(\cdot)$ to admit pre-specified structures, allowing the use of numerical solution methods. Decisions rules for real-valued functions have been used since 1974 [19]. Nevertheless, their potential was not fully exploited until recently when new advances in robust optimization provided the tools to reformulate the decision rule problem as tractable convex optimization problems [4,7]. Robust optimization techniques were first used by Ben-Tal et al. [6] to reformulate the *linear decision rule problem*. In this work, real-valued functions are parameterised as linear function of the random variables, i.e., $x(\boldsymbol{\xi}) = \boldsymbol{x}^\top\boldsymbol{\xi}$ for $\boldsymbol{x} \in \mathbb{R}^k$. The simple structure of linear decision rules offers the scalability needed to tackle multistage adaptive problems. Even though linear decision rules are known to be optimal in a number of problem instances [1,11,23], their simple structure generically sacrifices a significant amount of optimality in return for scalability. To gain back some degree of optimality, attention was focused on the construction of non-linear decision rules. Inspired by the linear decision rule structure, the real-valued adaptive decisions are parameterised as $x(\boldsymbol{\xi}) = \boldsymbol{x}^\top L(\boldsymbol{\xi}),\ \boldsymbol{x} \in \mathbb{R}^{k'}$, where $L : \mathbb{R}^k \to \mathbb{R}^{k'}$, $k' \ge k$, is a non-linear operator defining the structure of the decision rule. This approximation provides significant improvements in solution quality, while retaining in large parts the favourable scalability properties of the linear decision rules.

Table 1 The table summarizes the literature of real-valued decision rule structures for which the resulting semi-infinite optimization problem can be reformulated exactly using robust optimization techniques

Real-valued decision rule structures $x(\boldsymbol{\xi}) = \boldsymbol{x}^\top L(\boldsymbol{\xi})$	Computational burden	Reference
Linear	LOP	[6,22,27,32]
Piecewise linear	LOP	[16,17,20,21]
Multilinear	LOP	[20]
Quadratic	SOCOP	[3,20]
Power, monomial, inverse monomial	SOCOP	[20]
Polynomial	SDOP	[2,12]

The computation burden refers to the structure of the resulting optimization problems for the case where the uncertainty set is described by a polyhedral set, with Linear Optimization Problems denoted by LOP, Second-Order Cone Optimization Problems denoted by SOCOP, and Semi-Definite Optimization Problems denoted by SDOP

Table 1 summarizes non-linear decision rules from the literature that are parameterized as $x(\boldsymbol{\xi}) = \boldsymbol{x}^\top L(\boldsymbol{\xi})$, together with the complexity of the induced convex optimization problems. A notable exception of real-valued decision rules that are alternatively parameterized as $x(\boldsymbol{\xi}) = \max\{\overline{\boldsymbol{x}}_1^\top \boldsymbol{\xi}, \ldots, \overline{\boldsymbol{x}}_P^\top \boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{x}}_1^\top \boldsymbol{\xi}, \ldots, \underline{\boldsymbol{x}}_P^\top \boldsymbol{\xi}\}$, $\overline{\boldsymbol{x}}_p, \underline{\boldsymbol{x}}_p \in \mathbb{R}^k$, is proposed by Bertsimas and Georghiou [10] in the framework of worst-case adaptive optimization. This structure offers near-optimal designs but requires the solution of mixed-integer optimization problems.

In contrast to the plethora of decision rule structures for real-valued decisions, the literature on discrete decision rules is somewhat limited. There are, however, three notable exceptions that deal with the case of worst-case adaptive optimization. The first one is the work of Bertsimas and Caramanis [8], where integer decision rules are parameterized as $y(\boldsymbol{\xi}) = \boldsymbol{y}^\top \lceil \boldsymbol{\xi} \rceil$, $\boldsymbol{y} \in \mathbb{Z}^k$, where $\lceil \cdot \rceil$ is the component-wise ceiling function. In this work, the resulting semi-infinite optimization problem is further approximated and solved using a randomized algorithm [15], providing only probabilistic guarantees on the feasibility of the solution. The second binary decision rule structure was proposed by Bertsimas and Georghiou [10], where the decision rule is parameterized as

$$y(\boldsymbol{\xi}) = \begin{cases} 1, & \max\{\overline{\boldsymbol{y}}_1^\top \boldsymbol{\xi}, \ldots, \overline{\boldsymbol{y}}_P^\top \boldsymbol{\xi}\} - \max\{\underline{\boldsymbol{y}}_1^\top \boldsymbol{\xi}, \ldots, \underline{\boldsymbol{y}}_P^\top \boldsymbol{\xi}\} \leq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{1.2}$$

with $\overline{\boldsymbol{y}}_p, \underline{\boldsymbol{y}}_p \in \mathbb{R}^k$, $p = 1, \ldots, P$. This binary decision rule structure offers near-optimal designs at the expense of scalability. Finally, in the recent work by Hanasusanto et al. [24], the binary decision rule is restricted to the so called "K-adaptable structure", resulting to binary adaptable policies with K contingency plans. This approximation is an adaptation of the work presented in Bertsimas and Caramanis [9], and its use is limited to two-stage adaptive problems.

The goal of this paper is to develop binary decision rule structures that can be used together with the real-valued decision rules listed in Table 1 for solving multistage adaptive mixed-integer optimization problems. The proposed methodology is inspired by the work of Bertsimas and Caramanis [8], and uses the tools developed in Georghiou et al. [20] to robustly reformulate the problem into a finite dimensional mixed-integer linear optimization problem. The motivation of this work is to propose an alternative to the highly flexible but computationally demanding decision rule (1.2). The emphasis of this work is placed on the scalability properties of the solution method, allowing to solve large instances of Problem (1.1) with minimal computational effort. The main contributions of this paper can be summarized as follows.

1. We propose a linear parameterized binary decision rule structure and derive the exact reformulation of the decision rule problem that achieves the best binary decision rule. We prove that for general polyhedral uncertainty sets and arbitrary decision rule structures, the decision rule problem is computationally intractable, resulting in mixed-integer linear optimization problems whose size can grow exponentially with respect to the problem data. To remedy this exponential growth, we use similar ideas as those discussed in Georghiou et al. [20], to provide a systematic trade-off between scalability and optimality, resulting to practical binary decision rules.
2. We apply the proposed binary decision rules to the class of problems with random-recourse, i.e., to problem instances where the technology matrix $B(\boldsymbol{\xi})$ is a given function of the uncertain vector $\boldsymbol{\xi}$. We provide the exact reformulation of the binary decision rule problem, and we show that the resulting mixed-integer linear optimization problem shares similar complexity as in the fixed-recourse case.
3. We demonstrate the effectiveness of the proposed methods in the context of a multistage inventory control problem (fixed-recourse problem), and a multistage knapsack problem (random-recourse problem). We show that for the inventory control problem, we are able to solve problem instances with 50 time stages and 200 binary recourse decisions, while for the knapsack problem we are able to solve problem instances with 25 time stages and 1300 binary recourse decisions, using IBM ILOG CPLEX 12.5 on an Intel Core i5—3570 CPU at 3.40 GHz machine with 8 GB RAM.

The rest of this paper is organized as follows. In Sect. 2, we outline our approach for binary decision rules in the context of one-stage adaptive optimization problems with fixed-recourse, and in Sect. 3 we discuss the extension to random-recourse problems. In Sect. 4, we extend the proposed approach to multistage adaptive optimization problems and in Sect. 5, we present our computational results. Throughout the paper, our work is focused on problem instances involving only binary decision rules for ease of exposition. The extension to problem instances involving both real-valued and binary recourse decisions can be easily extrapolated, and thus omitted.

Notation We denote scalar quantities by lowercase, non-bold face symbols and vector quantities by lowercase, boldface symbols, e.g., $x \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$, respectively. Similarly, scalar and vector valued functions will be denoted by, $x(\cdot) \in \mathbb{R}$ and $\boldsymbol{x}(\cdot) \in \mathbb{R}^n$, respectively. Matrices are denoted by uppercase symbols , e.g., $A \in \mathbb{R}^{n\times m}$. We model uncertainty by a probability space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathbb{P}_{\boldsymbol{\xi}})$ and denote the elements of the sample space $\mathbb{R}^k$ by $\boldsymbol{\xi}$. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ is the set of events that are assigned probabilities by the probability measure $\mathbb{P}_{\boldsymbol{\xi}}$. The support Ξ of $\mathbb{P}_{\boldsymbol{\xi}}$ represents the smallest closed subset of $\mathbb{R}^k$ which has probability 1, and $\mathbb{E}_{\boldsymbol{\xi}}(\cdot)$ denotes the expectation operator with respect to $\mathbb{P}_{\boldsymbol{\xi}}$. $\mathrm{Tr}\,(A)$ denotes the trace of a square matrix $A \in \mathbb{R}^{n\times n}$ and $\mathbf{1}(\cdot)$ is the indicator function. Finally, $\boldsymbol{e}_k$ the kth canonical basis vector, while $\boldsymbol{e}$ denotes the vector whose components are all ones. In both cases, the dimension will be clear from the context.

2 Binary decision rules for fixed-recourse problems

In this section, we present our approach for one-stage adaptive optimization problems with fixed recourse, involving only binary decisions. Given matrices $B \in \mathbb{R}^{m\times q}$, $D \in \mathbb{R}^{q\times k}$, $H \in \mathbb{R}^{m\times k}$ and a probability measure $\mathbb{P}_{\boldsymbol{\xi}}$ supported on set Ξ for the uncertain vector $\boldsymbol{\xi} \in \mathbb{R}^k$, we are interested in choosing binary functions $\boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}$ in order to solve:

$$
\begin{array}{ll}
\text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\Big((D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})\Big) \\
\text{subject to} & \left.\begin{array}{l} \boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}, \\ B\boldsymbol{y}(\boldsymbol{\xi}) \le H\boldsymbol{\xi}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi.
\end{array}
\tag{2.1}
$$

Here, we assume that the uncertainty set Ξ is a non-empty, convex and compact polyhedron

$$
\Xi = \Big\{\boldsymbol{\xi} \in \mathbb{R}^k : \exists \boldsymbol{\zeta} \in \mathbb{R}^v \text{ such that } W\boldsymbol{\xi} + U\boldsymbol{\zeta} \ge \boldsymbol{h},\ \xi_1 = 1\Big\}, \tag{2.2}
$$

where $W \in \mathbb{R}^{l\times k}$, $U \in \mathbb{R}^{l\times v}$ and $\boldsymbol{h} \in \mathbb{R}^l$. Moreover, we assume that Ξ spans $\mathbb{R}^k$. The parameter ξ_1 is set equal to 1 without loss of generality as it allows us to represent affine functions of the non-degenerate outcomes $(\xi_2, \ldots, \xi_k)$ in a compact manner as linear functions of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$.

If the distribution governing the uncertainty $\boldsymbol{\xi}$ is unknown or if the decision maker is very risk-averse, then one might choose to alternatively minimize the worst-case costs with respect to all possible scenarios $\boldsymbol{\xi} \in \Xi$. This can be achieved, by replacing the objective function in Problem (2.1) with the worst-case objective $\max_{\boldsymbol{\xi}\in\Xi}\Big((D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})\Big)$. By introducing the auxiliary variable $\tau \in \mathbb{R}$ and an epigraph formulation, we can equivalently write the worst-case problem as the following adaptive robust optimization problem.

$$
\begin{array}{ll}
\text{minimize} & \tau \\
\text{subject to} & \left.\begin{array}{l} \tau \in \mathbb{R},\ \boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}, \\ (D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) \le \tau, \\ B\boldsymbol{y}(\boldsymbol{\xi}) \le H\boldsymbol{\xi}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi.
\end{array}
\tag{2.3}
$$

Notice that $\boldsymbol{\xi}$ appears in the left hand side of the constraint $(D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) \le \tau$, multiplying the binary decisions $\boldsymbol{y}(\boldsymbol{\xi})$. Therefore, Problem (2.3) is an instance of the class of problems with random-recourse, which will be investigated in Sect. 3.

Problem (2.1) involves a continuum of decision variables and inequality constraints. Therefore, in order to make the problem amenable to numerical solutions, there is a need for suitable functional approximations for $\boldsymbol{y}(\cdot)$.

2.1 Structure of binary decision rules

In this section, we present the structure of the binary decision rules. We restrict the feasible region of the binary functions $\boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}$ to admit the following piecewise constant structure:

$$\left.\begin{array}{l} \boldsymbol{y}(\boldsymbol{\xi}) = YG(\boldsymbol{\xi}),\ Y \in \mathbb{Z}^{q\times g}, \\ 0 \leq YG(\boldsymbol{\xi}) \leq \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi, \tag{2.4a}$$

where $G : \mathbb{R}^k \to \{0, 1\}^g$ is a piecewise constant function:

$$G_1(\boldsymbol{\xi}) := 1, \quad G_i(\boldsymbol{\xi}) := \mathbf{1}\left(\boldsymbol{\alpha}_i^\top \boldsymbol{\xi} \geq \beta_i\right), \quad i = 2, \dots, g, \tag{2.4b}$$

for given $\boldsymbol{\alpha}_i \in \mathbb{R}^k$ and $\beta_i \in \mathbb{R}$, $i = 2, \dots, g$. Notice, that since $G(\cdot)$ is a piecewise constant function mapping to $\{0, 1\}$ and the entries of matrix Y can take any integer values, then $\boldsymbol{y}(\boldsymbol{\xi}) = YG(\boldsymbol{\xi})$ gives rise to integer decision rules. By imposing the additional constraint $0 \leq YG(\boldsymbol{\xi}) \leq \boldsymbol{e}$, for all $\boldsymbol{\xi} \in \Xi$, $\boldsymbol{y}(\cdot)$ is restricted to the space of binary decision rules.

We require the pairs $(\boldsymbol{\alpha}_i, \beta_i)$ defining $G(\cdot)$ to satisfy the following conditions. First, we choose $(\boldsymbol{\alpha}_i, \beta_i)$ which result in unique hyperplanes $\boldsymbol{\alpha}_i^\top \boldsymbol{\xi} - \beta_i = 0$ for all $i \in \{2, \dots, g\}$, i.e., we avoid creating both $\mathbf{1}(\xi \geq 0.5)$ and $\mathbf{1}(\xi \leq 0.5)$, or, both $\mathbf{1}(\xi \leq 0.5)$ and $\mathbf{1}(2\xi \leq 1)$. Second, using $(\boldsymbol{\alpha}_i, \beta_i)$ we partition the uncertainty set Ξ into P polyhedra, where Ξ_p is defined as follows:

$$\Xi_p = \left\{ \begin{array}{ll} \boldsymbol{\xi} \in \Xi \ :\ \boldsymbol{\alpha}_i^\top \boldsymbol{\xi} \geq \beta_i, & i \in \mathcal{G}_p \subseteq \{2, \dots, g\}, \\ \qquad\quad\ \ \boldsymbol{\alpha}_i^\top \boldsymbol{\xi} \leq \beta_i, & i \in \{2, \dots, g\} \backslash \mathcal{G}_p \end{array} \right\}, \quad p = 1, \dots, P. \tag{2.5}$$

Here, given $(\boldsymbol{\alpha}_i, \beta_i)$ the sets $\mathcal{G}_p$ are chosen such that the number of non-empty partitions P is maximized. We require that $(\boldsymbol{\alpha}_i, \beta_i)$, $i = 2, \dots, g$ are chosen such that Ξ_p have (i) non-empty relative interior, (ii) the sets Ξ_p span $\mathbb{R}^k$, (iii) any partition pair Ξ_i and Ξ_j, $i \neq j$, can overlap only on one of their facets, and (iv), $\Xi = \bigcup_{p=1}^P \Xi_p$. The number of partitions P that satisfy these conditions is upper bounded by 2^g. We remark that in condition (ii) due to the definition of Ξ in (2.2), Ξ_p spans $\mathbb{R}^k$ if and only if it has dimension $k - 1$.

Applying decision rules (2.4) to Problem (2.1) yields the following semi-infinite problem, which involves a finite number of decision variables $Y \in \mathbb{Z}^{q\times g}$, and an infinite number of constraints:

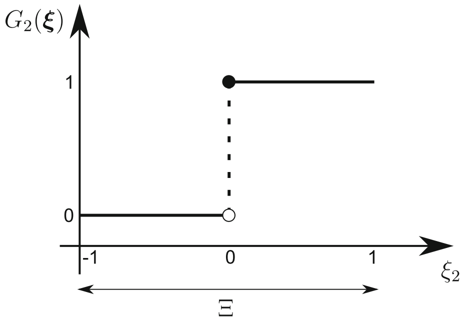

Fig. 1 Plot of piecewise constant function $G_2(\boldsymbol{\xi}) = \mathbf{1}(\xi_2 \geq 0)$. This structure of $G(\cdot)$ will give rise to piecewise constant decision rules $y(\boldsymbol{\xi}) = \boldsymbol{y}^\top G(\boldsymbol{\xi})$, for some $\boldsymbol{y} \in \mathbb{R}^2$. If one imposes constraints $\boldsymbol{y} \in \mathbb{Z}^2$ and $0 \leq \boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1$, $y(\cdot)$ is restricted to the space of binary decision rules induced by G

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\Big((D\boldsymbol{\xi})^\top Y G(\boldsymbol{\xi})\Big) \\ \text{subject to} & \left.\begin{array}{l} Y \in \mathbb{Z}^{q \times g}, \\ B\,Y G(\boldsymbol{\xi}) \leq H\boldsymbol{\xi}, \\ 0 \leq Y G(\boldsymbol{\xi}) \leq \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{2.6}$$

Notice that all decision variables appear linearly in Problem (2.6). Nevertheless, the objective and constraints are non-linear functions of $\boldsymbol{\xi}$.

The following example illustrates the use of the binary decision rule (2.6) in a simple instance of Problem (2.1), briefly discussing the choice of vectors $\boldsymbol{\alpha}_i \in \mathbb{R}^k$ and $\beta_i \in \mathbb{R}$, $i = 2, \ldots, g$, and showing the relationship between Problems (2.1) and (2.6).

Example 1 Consider the following instance of Problem (2.1):

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\big(y(\boldsymbol{\xi})\big) \\ \text{subject to} & \left.\begin{array}{l} y(\cdot) \in \mathcal{B}_{2,1}, \\ y(\boldsymbol{\xi}) \geq \xi_2, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi, \end{array} \tag{2.7}$$

where $\mathbb{P}_{\boldsymbol{\xi}}$ is a uniform distribution supported on $\Xi = \big\{(\xi_1, \xi_2) \in \mathbb{R}^2 \,:\, \xi_1 = 1,\ \xi_2 \in [-1, 1]\big\}$. The optimal solution of Problem (2.7) is $y^*(\boldsymbol{\xi}) = \mathbf{1}(\xi_2 \geq 0)$ achieving an optimal value of $\frac{1}{2}$. One can attain the same solution by solving the following semi-infinite problem where $G(\cdot)$ is defined to be $G_1(\boldsymbol{\xi}) = 1$, $G_2(\boldsymbol{\xi}) = \mathbf{1}(\xi_2 \geq 0)$, i.e., $\boldsymbol{\alpha}_2 = (0, 1)^\top$, $\beta_2 = 0$ and $g = 2$, see Fig. 1.

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\big(\boldsymbol{y}^\top G(\boldsymbol{\xi})\big) \\ \text{subject to} & \left.\begin{array}{l} \boldsymbol{y} \in \mathbb{Z}^2, \\ \boldsymbol{y}^\top G(\boldsymbol{\xi}) \geq \xi_2, \\ 0 \leq \boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{2.8}$$

The optimal solution of Problem (2.8) is $\boldsymbol{y}^* = (0, 1)^\top$ achieving an optimal value of $\frac{1}{2}$, and is equivalent to the optimal binary decision rule in Problem (2.7).

In the following section, we present robust reformulations of the semi-infinite Problem (2.6).

2.2 Computing the best binary decision rule

In this section, we present an exact reformulation of Problem (2.6). The solution of the reformulated problem will achieve the best binary decision rule associated with structure (2.4). The main idea used in this section is to map Problem (2.6) to an equivalent lifted adaptive optimization problem on a higher-dimensional probability space. This will allow us to represent the non-convex constraints with respect to $\boldsymbol{\xi}$ in Problem (2.6), as linear constraints in the lifted adaptive optimization problem. The relation between the uncertain parameters in the original and the lifted problems is determined through the piecewise constant operator $G(\cdot)$. By taking advantage of the linear structure of the lifted problem, we will employ linear duality arguments to reformulate the semi-infinite structure of the lifted problem into a finite dimensional mixed-integer linear optimization problem. The work presented in this section uses the tools developed by Georghiou et al. [20, Section 3].

We define the non-linear operator

$$L : \mathbb{R}^k \to \mathbb{R}^{k'}, \qquad L(\boldsymbol{\xi}) = \begin{pmatrix} \boldsymbol{\xi} \\ G(\boldsymbol{\xi}) \end{pmatrix}, \tag{2.9}$$

where $k' = k + g$, and define the projection matrices $R_{\xi} \in \mathbb{R}^{k \times k'}$, $R_G \in \mathbb{R}^{g \times k'}$ such that

$$R_{\xi} L(\boldsymbol{\xi}) = \boldsymbol{\xi}, \qquad R_G L(\boldsymbol{\xi}) = G(\boldsymbol{\xi}).$$

We will refer to $L(\cdot)$ as the *lifting operator*. Using $L(\cdot)$ and R_{ξ}, R_G, we can rewrite Problem (2.6) into the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\xi}\Big((D R_{\xi} L(\boldsymbol{\xi}))^{\top} Y R_G L(\boldsymbol{\xi})\Big) \\ \text{subject to} & \left.\begin{array}{l} Y \in \mathbb{Z}^{q \times g}, \\ B\, Y R_G L(\boldsymbol{\xi}) \le H R_{\xi} L(\boldsymbol{\xi}), \\ 0 \le Y R_G L(\boldsymbol{\xi}) \le \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{2.10}$$

Notice that this simple reformulation still has infinite number of constraints. Nevertheless, due to the structure of R_{ξ} and R_G, the constraints are now linear with respect to $L(\boldsymbol{\xi})$. The objective function of Problem (2.10) can be further reformulated to

$$\mathbb{E}_{\xi}\Big((D R_{\xi} L(\boldsymbol{\xi}))^{\top} Y R_G L(\boldsymbol{\xi})\Big) = \mathrm{Tr}\Big(M R_{\xi}^{\top} D^{\top} Y R_G\Big),$$

where $M \in \mathbb{R}^{k' \times k'}$, $M = \mathbb{E}_{\xi}(L(\boldsymbol{\xi}) L(\boldsymbol{\xi})^{\top})$ is the second order moment matrix of the uncertain vector $L(\boldsymbol{\xi})$. In some special cases where the $\mathbb{P}_{\xi}$ has a simple structure, e.g., $\mathbb{P}_{\xi}$ is a uniform distribution, and $G(\cdot)$ is not too complicated, M can be calculated analytically. If this is not the case, an arbitrarily good approximation of M can be calculated using Monte Carlo simulations. This is not computationally demanding as it does not does not involve an optimization problem, and can be done offline.

We now define the uncertain vector $\boldsymbol{\xi}' = (\boldsymbol{\xi}_1'^{\top}, \boldsymbol{\xi}_2'^{\top})^{\top} \in \mathbb{R}^{k'}$, such that $\boldsymbol{\xi}' := L(\boldsymbol{\xi}) = (\boldsymbol{\xi}^{\top}, G(\boldsymbol{\xi})^{\top})^{\top}$. The random vector $\boldsymbol{\xi}'$ has probability measure $\mathbb{P}_{\xi'}$ which is defined on the space $\left(\mathbb{R}^{k'}, \mathcal{B}(\mathbb{R}^{k'})\right)$ and is completely determined by the probability measure $\mathbb{P}_{\boldsymbol{\xi}}$ through the relation

$$\mathbb{P}_{\boldsymbol{\xi}'}\left(B'\right) := \mathbb{P}_{\boldsymbol{\xi}}\left(\{\boldsymbol{\xi} \in \mathbb{R}^k : \ L(\boldsymbol{\xi}) \in B'\}\right) \quad \forall B' \in \mathcal{B}(\mathbb{R}^{k'}). \tag{2.11}$$

We also introduce the probability measure support

$$\begin{aligned} \Xi' := L(\Xi) &= \left\{\boldsymbol{\xi}' \in \mathbb{R}^{k'} : \ \exists \boldsymbol{\xi} \in \Xi \text{ such that } L(\boldsymbol{\xi}) = \boldsymbol{\xi}'\right\}, \\ &= \left\{\boldsymbol{\xi}' \in \mathbb{R}^{k'} : \ R_{\boldsymbol{\xi}}\boldsymbol{\xi}' \in \Xi, \quad L(R_{\boldsymbol{\xi}}\boldsymbol{\xi}') = \boldsymbol{\xi}'\right\}, \end{aligned} \tag{2.12}$$

and the expectation operator $\mathbb{E}_{\boldsymbol{\xi}'}(\cdot)$ with respect to the probability measure $\mathbb{P}_{\boldsymbol{\xi}'}$. We will refer to $\mathbb{P}_{\boldsymbol{\xi}'}$ and Ξ' as the *lifted* probability measure and uncertainty set, respectively. Notice that although Ξ is a convex polyhedron, Ξ' can be highly non-convex due to the non-linear nature of $L(\cdot)$. Using the definition of $\boldsymbol{\xi}'$ we introduce the following *lifted adaptive optimization problem*:

$$\begin{aligned} &\text{minimize} && \operatorname{Tr}\left(M' R_{\boldsymbol{\xi}}^\top D^\top Y R_G\right) \\ &\text{subject to} && \left.\begin{aligned} & Y \in \mathbb{Z}^{q\times g}, \\ & B\,Y R_G \boldsymbol{\xi}' \le H R_{\boldsymbol{\xi}} \boldsymbol{\xi}', \\ & 0 \le Y R_G \boldsymbol{\xi}' \le \boldsymbol{e}, \end{aligned}\right\} \quad \forall \boldsymbol{\xi}' \in \Xi', \end{aligned} \tag{2.13}$$

where $M' \in \mathbb{R}^{k'\times k'}$, $M' = \mathbb{E}_{\boldsymbol{\xi}'}(\boldsymbol{\xi}'\boldsymbol{\xi}'^\top)$ is the second order moment matrix associated with $\boldsymbol{\xi}'$. Since there is a one-to-one mapping between $\mathbb{P}_{\boldsymbol{\xi}}$ and $\mathbb{P}_{\boldsymbol{\xi}'}$, $M = M'$.

Proposition 1 *Problems* (2.10) *and* (2.13) *are equivalent in the following sense: both problems have the same optimal value, and there is a one-to-one mapping between feasible and optimal solutions in both problems.*

Proof See [20, Proposition 3.6(i)]. □

The lifted uncertainty set Ξ' is an open set due to the discontinuous nature of $G(\cdot)$. This can be problematic in an optimization framework. We now define $\overline{\Xi}' := \mathrm{cl}(\Xi')$ to be the closure of set Ξ', and introduce the following variant of Problem (2.13).

$$\begin{aligned} &\text{minimize} && \operatorname{Tr}\left(M' R_{\boldsymbol{\xi}}^\top D^\top Y R_G\right) \\ &\text{subject to} && \left.\begin{aligned} & Y \in \mathbb{Z}^{q\times g}, \\ & B\,Y R_G \boldsymbol{\xi}' \le H R_{\boldsymbol{\xi}} \boldsymbol{\xi}', \\ & 0 \le Y R_G \boldsymbol{\xi}' \le \boldsymbol{e}, \end{aligned}\right\} \quad \forall \boldsymbol{\xi}' \in \overline{\Xi}', \end{aligned} \tag{2.14}$$

The following proposition demonstrates that Problems (2.13) and (2.14) are in fact equivalent.

Proposition 2 *Problems* (2.13) *and* (2.14) *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 To prove the assertion, it is sufficient to show that Problems (2.13) and (2.14) have the same feasible region. Notice, that since the constraints are linear in $\boldsymbol{\xi}'$, the semi-infinite constraints in Problems (2.13) can be equivalently written as

$$\begin{aligned}
&(HR_{\boldsymbol{\xi}} - BYR_G) \in (\mathrm{cone}(\Xi')^*)^m,\\
&(YR_G) \in (\mathrm{cone}(\Xi')^*)^q,\\
&(\boldsymbol{e}\boldsymbol{e}_1^\top - YR_G) \in (\mathrm{cone}(\Xi')^*)^q,
\end{aligned}$$

where $\mathrm{cone}(\Xi')^*$ is the dual cone of $\mathrm{cone}(\Xi')$. Similarly, the semi-infinite constraints in Problems (2.14) can be equivalently written as

$$\begin{aligned}
&(HR_{\boldsymbol{\xi}} - BYR_G) \in (\mathrm{cone}(\overline{\Xi}')^*)^m,\\
&(YR_G) \in (\mathrm{cone}(\overline{\Xi}')^*)^q,\\
&(\boldsymbol{e}\boldsymbol{e}_1^\top - YR_G) \in (\mathrm{cone}(\overline{\Xi}')^*)^q.
\end{aligned}$$

From [29, Corollary 6.21], we have that $\mathrm{cone}(\Xi')^* = \mathrm{cone}(\overline{\Xi}')^*$, and therefore, we can conclude that the feasible regions of Problems (2.13) and (2.14) are equivalent. □

Problem (2.14) is linear in both the decision variables Y and the uncertain vector $\boldsymbol{\xi}'$. Despite this nice bilinear structure, in the following we demonstrate that Problems (2.10) and (2.14) are generically intractable for decision rules of type (2.4).

Theorem 1 *Problems* (2.10) *and* (2.14) *defined through* $L(\cdot)$ *in* (2.9) *and* $G(\cdot)$ *in* (2.4b)*, are* NP*-hard even when* $Y \in \mathbb{Z}^{q\times g}$ *is relaxed to* $Y \in \mathbb{R}^{q\times g}$.

Proof See Appendix. □

Theorem 1 provides a rather disappointing result on the complexity of Problems (2.10) and (2.14). Therefore, unless P = NP, there is no algorithm that solves generic problems of type (2.10) and (2.14) in polynomial time. This difficulty stems from the generic structure of $G(\cdot)$ in (2.4b), combined with a generic polyhedral uncertainty set Ξ, resulting in highly non-convex sets $\overline{\Xi}'$. Nevertheless, in the following we derive the exact polyhedral representation of $\mathrm{conv}(\overline{\Xi}')$, allowing us to use linear duality arguments from [4,7], to reformulate Problem (2.14) into a mixed-integer linear optimization problem. We remark that Theorem 1 also covers decision rule problems involving real-valued, piecewise constant decisions rules constructed using (2.4b), and demonstrates that these problems are also computationally intractable.

We construct the convex hull of $\overline{\Xi}'$ by taking convex combinations of its extreme points, denoted by $\mathrm{ext}(\overline{\Xi}')$. These points are either the extreme points of Ξ', or the limit points $\overline{\Xi}'\backslash\Xi'$. We construct these points using the definition of $G(\cdot)$ and the P partitions Ξ_p given in (2.5). Then, for all extreme points $\boldsymbol{\xi} \in \mathrm{ext}(\Xi_p)$, we calculate the one-side limit point at $L(\boldsymbol{\xi})$. The set of all one-side limit points will coincide with the set of extreme points of $\overline{\Xi}'$, see Fig. 2.

We define V and $V(p)$ such that

$$V := \bigcup_{p=1}^{P} V(p), \qquad V(p) := \mathrm{ext}(\Xi_p), \quad p = 1,\ldots,P. \tag{2.15}$$

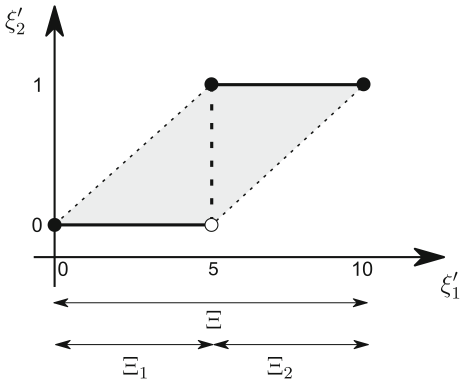

Fig. 2 Convex hull representation of $\overline{\Xi}'$ induced by lifting $\boldsymbol{\xi}' = (\xi'_1, \xi'_2) = L(\xi) = (\xi, G(\xi))^\top$, where $G(\xi) = \mathbf{1}(\xi \geq 5)$ and $\xi \in \Xi = [0, 10]$. Here, $V = \{\text{ext}(\Xi_1), \text{ext}(\Xi_2)\} = \{0, 5, 10\}$. The convex hull is constructed by taking convex combinations of the points $L(0)$, $L(5)$, $L(10)$ (*black dots*), and point $(5, \widehat{G}_1(5))^\top$ (*white dot*), where $\widehat{G}_1(5)$ is the one-side limit point at $\xi = 5$ using partition Ξ_1

Due to the discontinuity of $G(\cdot)$, the set of points $\{\boldsymbol{\xi}' \in \mathbb{R}^{k'} : \boldsymbol{\xi}' = L(\boldsymbol{\xi}), \boldsymbol{\xi} \in V\}$ does not contain the points $\overline{\Xi}' \backslash \Xi'$. To construct these points, we now define the one-sided limit points $\widehat{L}_p(\boldsymbol{\xi}) = (\boldsymbol{\xi}^\top, \widehat{G}_p(\boldsymbol{\xi})^\top)^\top$ for all $\boldsymbol{\xi} \in V(p)$ and each partition $p = 1, \ldots, P$. Here, $\widehat{G}_p(\boldsymbol{\xi}) : \mathbb{R}^k \to \mathbb{R}^g$ is given by:

$$\widehat{G}_p(\boldsymbol{\xi}) := \lim_{\boldsymbol{u} \in \Xi_p, \boldsymbol{u} \to \boldsymbol{\xi}} G(\boldsymbol{u}), \quad \forall \boldsymbol{\xi} \in V(p), \ p = 1, \ldots, P. \tag{2.16}$$

From definition (2.4b), for each partition p, $G(\boldsymbol{\xi})$ is constant for all $\boldsymbol{\xi}$ in the relative interior of Ξ_p, which we denote by relint(Ξ_p). Therefore, for each $\tilde{\boldsymbol{\xi}} \in V(p)$, the one-side limit $\widehat{G}_p(\tilde{\boldsymbol{\xi}})$ is equal to $G(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in$ relint(Ξ_p). Furthermore, since Ξ_p spans $\mathbb{R}^k$ and thus has dimension $k - 1$, then there exists at least k vertices in Ξ_p, and therefore, at least k one-sided limit points $\widehat{G}_p(\boldsymbol{\xi})$. From the definition (2.16), we have that $\widehat{L}_p(\boldsymbol{\xi})$ are the extreme points of $\overline{\Xi}'$ for all $\boldsymbol{\xi} \in V(p)$ and $p = 1, \ldots, P$, see Fig. 2.

The following proposition gives the polyhedral representation of the convex hull of $\overline{\Xi}'$.

Proposition 3 *Let $L(\cdot)$ be given by (2.9) and $G(\cdot)$ being defined in (2.4b). Then, the exact representation of the convex hull of $\overline{\Xi}'$ is given by the following polyhedron:*

$$\begin{aligned} conv(\overline{\Xi}') = \Big\{ \boldsymbol{\xi}' = (\boldsymbol{\xi}_1'^\top, \boldsymbol{\xi}_2'^\top)^\top \in \mathbb{R}^{k'} : \exists \zeta_p(\boldsymbol{v}) \in \mathbb{R}_+, \forall \boldsymbol{v} \in V(p), \ p = 1, \ldots, P, \ \textit{such that} \\ \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) = 1, \\ \boldsymbol{\xi}'_1 = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \boldsymbol{v}, \\ \boldsymbol{\xi}'_2 = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \widehat{G}_p(\boldsymbol{v}) \Big\}. \end{aligned} \tag{2.17}$$

Proof The proof is split in two parts: First, we prove that $L(\Xi) \subseteq \text{conv}\,(\overline{\Xi}')$ holds, and then we show that $\text{conv}\,(\text{cl}(L(\Xi))) \supseteq \text{conv}\,(\overline{\Xi}')$ is true as well, concluding that $\text{conv}\,(\text{cl}(L(\Xi))) = \text{conv}\,(\overline{\Xi}')$.

To prove assertion $L(\Xi) \subseteq \text{conv}\,(\overline{\Xi}')$, pick any $\boldsymbol{\xi} \in \Xi$. By construction $\boldsymbol{\xi}$ belongs to some partition of Ξ, say Ξ_p, for which $L(\boldsymbol{\xi})$ is an element of Ξ'. There are two possibilities: (a) $\boldsymbol{\xi}$ lies on a facet of Ξ_p, or (b) $\boldsymbol{\xi}$ lies in the relative interior of Ξ_p. If $\boldsymbol{\xi}$ lies on a facet, and since Ξ_p spans $\mathbb{R}^k$, then Carathéodory's Theorem implies that there are $k-1$ extreme points on the facet, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1} \in V(p)$, such that for some $\delta(\boldsymbol{v}) \in \mathbb{R}^{k-1}_+$ with $\sum_{i=1}^{k-1} \delta(\boldsymbol{v}_i) = 1$, we have $\boldsymbol{\xi} = \sum_{i=1}^{k-1} \delta(\boldsymbol{v}_i)\boldsymbol{v}_i$. Since $G(\cdot)$ is constant on each facet, then the $k-1$ one-side limit points $\widehat{G}_p(\boldsymbol{v}_i)$, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1} \in V(p)$ attain the same value as $G(\boldsymbol{\xi})$. Therefore, we have

$$G\left(\sum_{i=1}^{k-1} \delta(\boldsymbol{v}_i)\boldsymbol{v}_i\right) = \sum_{i=1}^{k-1} \delta(\boldsymbol{v}_i)\widehat{G}_p(\boldsymbol{v}_i).$$

Hence, setting $\zeta_p(\boldsymbol{v}_i) = \delta(\boldsymbol{v}_i)$, $i = 1, \ldots, k-1$, in the definition of the convex hull with the rest of the $\zeta(\boldsymbol{v})$ equal to zero proves part (a) of the assertion. If $\boldsymbol{\xi}$ lies in the relative interior of Ξ_p, then again by Carathéodory's Theorem there are k extreme points of Ξ_p such that for some $\delta(\boldsymbol{v}) \in \mathbb{R}^k_+$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in V(p)$ with $\sum_{i=1}^{k} \delta(\boldsymbol{v}_i) = 1$, we have $\boldsymbol{\xi} = \sum_{i=1}^{k} \delta(\boldsymbol{v}_i)\boldsymbol{v}_i$. By construction, there also exists k one-sided limit in the collection $\{\widehat{G}_p(\boldsymbol{v}) : \boldsymbol{v} \in V(p)\}$ that attain the same value as $G(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \text{relint}(\Xi_p)$. Thus, we have that

$$G\left(\sum_{i=1}^{k} \delta(\boldsymbol{v}_i)\boldsymbol{v}_i\right) = \sum_{i=1}^{k} \delta(\boldsymbol{v}_i)\widehat{G}_p(\boldsymbol{v}_i).$$

Setting $\zeta_p(\boldsymbol{v}_i) = \delta(\boldsymbol{v}_i)$, $i = 1, \ldots, k$, with the rest of the $\zeta(\boldsymbol{v})$ equal to zero proves part (b) of the assertion.

To prove the second part of assertion, $\text{conv}\,(\text{cl}(L(\Xi))) \supseteq \text{conv}\,(\overline{\Xi}')$, fix $\boldsymbol{\xi}' \in \overline{\Xi}'$. By construction, there exists $\zeta_p(v) \in \mathbb{R}_+$, $p = 1, \ldots, P$, $\boldsymbol{v} \in V$, such that

$$\sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) = 1, \quad \boldsymbol{\xi}'_1 = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v})\boldsymbol{v}, \quad \boldsymbol{\xi}'_2 = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v})\widehat{G}_p(\boldsymbol{v}).$$

This implies that

$$\begin{pmatrix} \boldsymbol{\xi}'_1 \\ \boldsymbol{\xi}'_2 \end{pmatrix} = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \begin{pmatrix} \boldsymbol{v} \\ \widehat{G}_p(\boldsymbol{v}) \end{pmatrix},$$

that is, $\boldsymbol{\xi}'$ is a convex combination of either corner points or limit points of $L(\Xi)$. Therefore, $\text{conv}\,(\overline{\Xi}')$ is equal to $\text{conv}\,(\text{cl}(L(\Xi)))$ almost surely. This concludes the proof. $\square$

The convex hull (2.17) is a closed and bounded polyhedral set described by a finite set of linear inequalities. Therefore, one can rewrite (2.17) as the following polyhedron

$$\operatorname{conv}(\overline{\Xi}') = \left\{ \boldsymbol{\xi}' \in \mathbb{R}^{k'} : \exists \boldsymbol{\zeta}' \in \mathbb{R}^{v'} \text{ such that } W'\boldsymbol{\xi}' + U'\boldsymbol{\zeta}' \geq \boldsymbol{h}' \right\}, \tag{2.18}$$

where $W' \in \mathbb{R}^{l' \times k'}$, $U' \in \mathbb{R}^{l' \times v'}$ and $\boldsymbol{h}' \in \mathbb{R}^{l'}$ are completely determined through Propositions 3.

The following proposition captures the essence of robust optimization and provides the tools for reformulating the infinite number of constraints in Problem (2.14), see [4,7]. The proof is repeated here to keep the paper self-contained.

Proposition 4 *For any* $Z \in \mathbb{R}^{m \times k'}$ *and* $\operatorname{conv}(\overline{\Xi}')$ *given by* (2.18)*, the following statements are equivalent.*

(i) $Z\boldsymbol{\xi}' \geq 0$ *for all* $\boldsymbol{\xi}' \in \operatorname{conv}(\overline{\Xi}')$,
(ii) $\exists\, \Lambda \in \mathbb{R}_+^{m \times l'}$ *with* $\Lambda W' = Z,\ \ \Lambda U' = 0,\ \ \Lambda \boldsymbol{h}' \geq 0.$

Proof We denote by $Z_\mu^\top$ the μth row of the matrix Z. Then, statement (i) is equivalent to

$$\left.\begin{array}{ll} & Z\boldsymbol{\xi}' \geq 0 \text{ for all } \boldsymbol{\xi}' \in \operatorname{conv}(\Xi'), \\ \Longleftrightarrow & 0 \leq \min\limits_{\boldsymbol{\xi}'} \left\{ Z_\mu^\top \xi : \exists \boldsymbol{\zeta}' \in \mathbb{R}^{v'},\ \ W'\boldsymbol{\xi}' + U'\boldsymbol{\zeta}' \geq \boldsymbol{h}' \right\}, \qquad \forall\, \mu = 1, \ldots, m \\ \Longleftrightarrow & 0 \leq \max\limits_{\Lambda_\mu \in \mathbb{R}^{l'}} \left\{ \boldsymbol{h}'^\top \Lambda_\mu : W'^\top \Lambda_\mu = Z_\mu,\ U'^\top \Lambda_\mu = 0,\ \Lambda_\mu \geq 0 \right\}, \qquad \forall\, \mu = 1, \ldots, m \\ \Longleftrightarrow & \exists\, \Lambda_\mu \in \mathbb{R}^{l'} \text{ with } W'^\top \Lambda_\mu = Z_\mu,\ U'^\top \Lambda_\mu = 0,\ \boldsymbol{h}'^\top \Lambda_\mu \geq 0,\ \Lambda_\mu \geq 0,\ \forall\, \mu = 1, \ldots, m \end{array}\right\} \tag{2.19}$$

The equivalence in the third line follows from linear duality. Interpreting $\Lambda_\mu^\top$ as the μth row of a new matrix $\Lambda \in \mathbb{R}^{m \times l}$ shows that the last line in (2.19) is equivalent to assertion (ii). Thus, the claim follows. □

Using Proposition 4 together with the polyhedron (2.18), we can now reformulate Problem (2.14) into the following mixed-integer linear optimization problem.

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}\left(M' R_{\boldsymbol{\xi}}^\top D^\top Y R_G \right) \\ \text{subject to} & Y \in \mathbb{Z}^{q \times g},\ \ \Lambda \in \mathbb{R}_+^{m \times l'},\ \ \Gamma \in \mathbb{R}_+^{q \times l'},\ \ \Theta \in \mathbb{R}_+^{q \times l'} \\ & BYR_G + \Lambda W' = HR_{\boldsymbol{\xi}},\ \ \Lambda U' = 0,\ \ \Lambda \boldsymbol{h}' \geq 0, \\ & YR_G = \Gamma W',\ \ \Gamma U' = 0,\ \ \Gamma \boldsymbol{h}' \geq 0, \\ & \boldsymbol{e}\boldsymbol{e}_1^\top - YR_G = \Theta W',\ \ \Theta U' = 0,\ \ \Theta \boldsymbol{h}' \geq 0. \end{array} \tag{2.20}$$

Here, $\Lambda \in \mathbb{R}_+^{m \times l'}$, $\Gamma \in \mathbb{R}_+^{q \times l'}$ and $\Theta \in \mathbb{R}_+^{q \times l'}$ are the auxiliary variables associated with constraints $BYR_G\boldsymbol{\xi}' \leq HR_{\boldsymbol{\xi}}\boldsymbol{\xi}'$, $0 \leq YR_G\boldsymbol{\xi}'$ and $YR_G\boldsymbol{\xi}' \leq \boldsymbol{e}$, respectively. We emphasize that Problem (2.20) is the exact reformulation of Problem (2.14), since (2.17) is the exact representation of the convex hull of $\overline{\Xi}'$. Therefore, *the solution of Problem* (2.20) *achieves the best binary decision rule associated with* $G(\cdot)$ *in* (2.4b).

Notice that the size of the Problem (2.20) grows quadratically with respect to the number q of binary decisions, and l' the number of constraints of $\operatorname{conv}(\overline{\Xi}')$. However, l' can be very large as it depends on the cardinality of V, which is constructed using the extreme points of the P partitions Ξ_p. Therefore, the size of Problem (2.20) can grow exponentially with respect to the description of Ξ and the complexity of the binary decision rule, g, defined in (2.4).

In the following, we present instances of Ξ and $G(\cdot)$, for which the proposed solution method will result to mixed-integer optimization problems whose size grow only polynomially with respect to the input parameters.

2.3 Scalable binary decision rule structures

In this section, we derive a scalable mixed-integer linear optimization reformulation of Problem (2.6) by considering simplified instances of the uncertainty set Ξ and binary decision rules $\boldsymbol{y}(\cdot)$. We will show that for the structure of Ξ and $\boldsymbol{y}(\cdot)$ considered in this section, the size of the resulting mixed-integer linear optimization problem grows polynomially in the size of the original Problem (2.1) as well as the description of the binary decision rules. We consider two cases: (i) uncertainty sets that can be described as hyperrectangles and (ii) general polyhedral uncertainty sets.

We now consider case (i), and define the uncertainty set of Problem (2.1) to have the following hyperrectangular structure:

$$\Xi = \left\{ \boldsymbol{\xi} \in \mathbb{R}^k \;:\; \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u},\; \xi_1 = 1 \right\}, \tag{2.21}$$

for given $\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^k$. We restrict the feasible region of the binary functions $\boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}$ to admit the piecewise constant structure (2.4a), but now $G : \mathbb{R}^k \to \{0, 1\}^g$ is written in the form

$$G(\cdot) = (G_1(\cdot), G_{2,1}(\cdot), \ldots, G_{2,r}(\cdot), \ldots, G_{k,r}(\cdot))^\top,$$

with $g = 1 + (k-1)r$. Here, $G_1 : \mathbb{R}^k \to \{0, 1\}$ and $G_{i,j} : \mathbb{R}^k \to \{0, 1\}$, for $j = 1, \ldots, r,\; i = 2, \ldots, k$, are piecewise constant functions given by

$$G_1(\boldsymbol{\xi}) := 1, \qquad G_{i,j}(\boldsymbol{\xi}) := \mathbf{1}\left(\xi_i \geq \beta_{i,j}\right), \quad j = 1, \ldots, r,\; i = 2, \ldots, k, \tag{2.22}$$

for fixed $\beta_{i,j} \in \mathbb{R},\; j = 1, \ldots, r,\; i = 2, \ldots, k$. By construction, we assume that $l_i < \beta_{i,1} < \ldots, \beta_{i,r} < u_i$, for all $i = 2, \ldots, k$. Structure (2.22) gives rise to binary decision rules that are discontinuous along the axis of the uncertainty set, and have r discontinuities in each direction. Notice that (2.22) is a special case of (2.4b), where vector $\boldsymbol{\alpha}$ are replaced by $\boldsymbol{e}_i$. One can easily modify (2.22) such that the number of discontinuities r is different for each direction ξ_i. We will refer to $\beta_{i,1}, \ldots, \beta_{i,r}$ as the breakpoints in direction ξ_i.

Problem (2.6) still retains the same semi-infinite structure using (2.21) and (2.22). In the following, we use the same arguments as in Sect. 2.2, to (a) redefine Problem (2.6) into a higher dimensional space and (b) construct the convex hull of the lifted uncertainty set and use it together with Proposition 4 to reformulate the infinite number of constraints of the lifted problem.

We define the non-linear lifting operator $L : \mathbb{R}^k \to \mathbb{R}^{k'}$ to be $L(\cdot) = (L_1(\cdot)^\top, \ldots, L_k(\cdot)^\top)^\top$ such that

$$\begin{aligned} &L_1 : \mathbb{R}^k \to \mathbb{R}^2, \quad L_1(\boldsymbol{\xi}) = (\xi_1, G_1(\boldsymbol{\xi}))^\top, \\ &L_i : \mathbb{R}^k \to \mathbb{R}^{r+1}, \ L_i(\boldsymbol{\xi}) = (\xi_i, G_i(\boldsymbol{\xi})^\top)^\top, \quad i = 2, \ldots, k, \end{aligned} \tag{2.23}$$

with $k' = 2 + (k-1)(r+1)$. Here, with slight abuse of notation, $G_i(\cdot) = (G_{i,1}(\cdot), \ldots, G_{i,r})^\top$ for $i = 2, \ldots, k$. Notice that $L(\cdot)$ is separable in each component $L_i(\cdot)$, since $G_i(\cdot)$ only involves ξ_i. We also introduce matrices $R_{\boldsymbol{\xi}} = (R_{\boldsymbol{\xi},1}, \ldots, R_{\boldsymbol{\xi},k}) \in \mathbb{R}^{k \times k'}$ with $R_{\boldsymbol{\xi},1} \in \mathbb{R}^{k \times 2}$, $R_{\boldsymbol{\xi},i} \in \mathbb{R}^{k \times (r+1)}$, $i = 2, \ldots, k$, such that

$$R_{\boldsymbol{\xi}} L(\boldsymbol{\xi}) = \boldsymbol{\xi}, \qquad R_{\boldsymbol{\xi},i} L_i(\boldsymbol{\xi}) = \boldsymbol{e}_i \xi_i, \quad i = 1, \ldots, k, \tag{2.24a}$$

and $R_G = (R_{G,1}, \ldots, R_{G,k}) \in \mathbb{R}^{g \times k'}$ with $R_{G,1} \in \mathbb{R}^{g \times 2}$, $R_{G,i} \in \mathbb{R}^{g \times (r+1)}$, $i = 2, \ldots, k$, such that

$$R_G L(\boldsymbol{\xi}) = G(\boldsymbol{\xi}), \qquad R_{G,i} L_i(\boldsymbol{\xi}) = (0, \ldots, G_i(\boldsymbol{\xi})^\top, \ldots, 0)^\top, \quad i = 1, \ldots, k. \tag{2.24b}$$

As in Sect. 2.2, using the definition of $L(\cdot)$ and $R_{\boldsymbol{\xi}}$, R_G in (2.23) and (2.24), respectively, we can rewrite Problem (2.6) into Problem (2.10).

We now define the uncertain vector $\boldsymbol{\xi}' = (\boldsymbol{\xi}_1'^\top, \ldots, \boldsymbol{\xi}_k'^\top)^\top \in \mathbb{R}^{k'}$ such that $\boldsymbol{\xi}' = L(\boldsymbol{\xi})$ and $\boldsymbol{\xi}_i' = L_i(\boldsymbol{\xi})$ for $i = 1, \ldots, k$. $\boldsymbol{\xi}'$ has probability measure $\mathbb{P}_{\boldsymbol{\xi}'}$ defined using (2.11), and corresponding probability measure support is given by

$$\Xi' = L(\Xi) = \left\{ \boldsymbol{\xi}' \in \mathbb{R}^{k'} : \boldsymbol{l} \le R_{\boldsymbol{\xi}} \boldsymbol{\xi}' \le \boldsymbol{u}, \ \ L(R_{\boldsymbol{\xi}} \boldsymbol{\xi}') = \boldsymbol{\xi}' \right\}. \tag{2.25}$$

Notice that since both Ξ and $L(\cdot)$ are separable in each component ξ_i, Ξ' can be written in the following equivalent form:

$$\Xi' = L(\Xi) = \left\{ (\boldsymbol{\xi}_1'^\top, \ldots, \boldsymbol{\xi}_k'^\top)^\top \in \mathbb{R}^{k'} : \boldsymbol{\xi}_i \in \Xi_i', \ i = 1, \ldots, k \right\},$$

where $\Xi_i' = L_i(\Xi)$. Therefore, we can express $\mathrm{conv}(\mathrm{cl}(\Xi'))$ as

$$\begin{aligned} \mathrm{conv}(\mathrm{cl}(\Xi')) &= \left\{ (\boldsymbol{\xi}_1'^\top, \ldots, \boldsymbol{\xi}_k'^\top)^\top \in \mathbb{R}^{k'} : \boldsymbol{\xi}_i' \in \mathrm{conv}(\mathrm{cl}(\Xi_i')), \ i = 1, \ldots, k \right\}, \\ &= \left\{ (\boldsymbol{\xi}_1'^\top, \ldots, \boldsymbol{\xi}_k'^\top)^\top \in \mathbb{R}^{k'} : \boldsymbol{\xi}_i' \in \mathrm{conv}(\overline{\Xi}_i'), \ i = 1, \ldots, k \right\}. \end{aligned} \tag{2.26}$$

It is thus sufficient to derive a closed-form representation for the marginal convex hulls $\mathrm{conv}(\overline{\Xi}_i')$.

We now construct the polyhedral representation of $\mathrm{conv}(\overline{\Xi}_i')$. As before, $\mathrm{conv}(\overline{\Xi}_i')$ will be constructed by taking convex combinations of its extreme points. Set $\mathrm{conv}(\overline{\Xi}_1')$ has the trivial representation $\mathrm{conv}(\overline{\Xi}_1') = \{\boldsymbol{\xi}_1' \in \mathbb{R}^2 : \boldsymbol{\xi}_1' = \boldsymbol{e}\}$. We define the sets

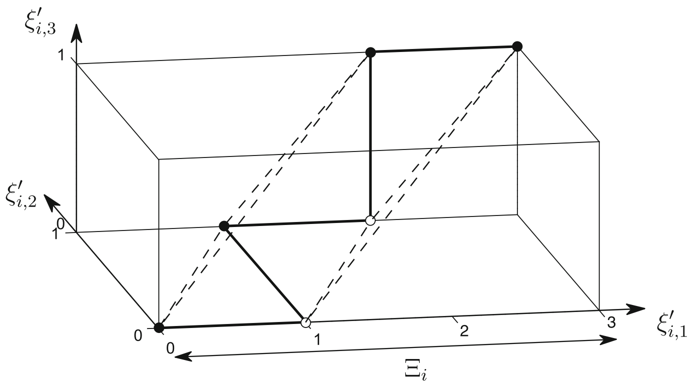

Fig. 3 Convex hull representation of $\overline{\Xi}'_i$ induced by lifting $\boldsymbol{\xi}'_i = L_i(\xi) = (\xi_i, G_i(\xi)^\top)^\top$, where $G_i(\xi) = (\mathbf{1}(\xi_i \geq 1), \mathbf{1}(\xi_i \geq 2))^\top$ and $\xi_i \in [0, 3]$. Here, $V_i = \{0, 1, 2, 3\}$. The convex hull is constructed by taking convex combinations of the points $L(0)$, $L(1)$, $L(2)$, $L(3)$ (*black dots*), and points $(1, \widehat{G}_{1,1}(1))^\top$, $(2, \widehat{G}_{1,2}(2))^\top$ (*white dot*), where $\widehat{G}_{1,1}(1)$ and $\widehat{G}_{1,2}(2)$ are the one-side limit point at points $\xi_i = 1$ and $\xi_i = 2$, respectively

$V_i = \{l_i, \beta_{i,1}, \ldots, \beta_{i,r}, u_i\}$, $i = 2, \ldots, k$, and introduce the following partitions for each dimension of Ξ.

$$\left.\begin{aligned} &\Xi_{i,1} := \{\xi_i \in \mathbb{R} \;:\; l_i \leq \xi_i \leq \beta_{i,1}\}, \\ &\Xi_{i,p} := \{\xi_i \in \mathbb{R} \;:\; \beta_{i,p-1} \leq \xi_i \leq \beta_{i,p}\}, \quad p = 2, \ldots, r, \\ &\Xi_{i,r+1} := \{\xi_i \in \mathbb{R} \;:\; \beta_{i,r} \leq \xi_i \leq u_i\}, \end{aligned}\right\} i = 2, \ldots, k. \tag{2.27}$$

Therefore, $V_i(1) = \{l_i, \beta_{i,1}\}$, $V_i(p) = \{\beta_{i,p-1}, \beta_{i,p}\}$, $p = 2, \ldots, r$, $V_i(r+1) = \{\beta_{i,r}, u_i\}$. It is easy to see that points $L(\boldsymbol{e}_i\xi_i)$, $\xi_i \in V_i$, are extreme points of $\mathrm{conv}(\overline{\Xi}'_i)$, see Fig. 3. Notice that $G_i(\boldsymbol{e}_i\xi_i)$ is constant for all ξ_i in the interior of $\Xi_{i,p}$, i.e., $\xi_i \in \mathrm{int}(\Xi_{i,p})$, $p = 1, \ldots, r+1$, but can attain different values on the boundaries of the partitions. To this end, for each partition and $\xi_i \in V_i(p)$, we define the one-sided limit points $\widehat{G}_{i,p}(\boldsymbol{e}_i\xi_i) \in \mathbb{R}^{r+1}$ such that

$$\widehat{G}_{i,p}(\boldsymbol{e}_i\xi_i) = \lim_{u \in \Xi_{i,p},\, u \to \xi_i} G_i(\boldsymbol{e}_i u), \quad \forall \xi_i \in V_i(p), \;\; p = 1, \ldots, r+1, \tag{2.28}$$

for all $i = 2, \ldots, k$. $G_i(\boldsymbol{e}_i\xi_i)$ is constant for all $\xi_i \in \mathrm{int}(\Xi_{i,p})$, and each partition p. Therefore, for each $\tilde{\xi}_i \in V_i(p)$, the one-side limit $\widehat{G}_{i,p}(\boldsymbol{e}_i\tilde{\xi}_i)$ is equal to $G_i(\boldsymbol{e}_i\xi_i)$ for all $\xi_i \in \mathrm{int}(\Xi_{i,p})$.

The following proposition gives the polyhedral representation of each $\mathrm{conv}(\overline{\Xi}'_i)$, $i = 2, \ldots, k$. A visual representation of $\mathrm{conv}(\overline{\Xi}'_i)$ is depicted in Fig. 3.

Proposition 5 *For each $i = 2 \ldots, k$, let $L_i(\cdot)$ be given by (2.23) and $G_i(\cdot)$ being defined in (2.22). Then, the exact representation of the convex hull of each $\overline{\Xi}'_i$ is given by the following polyhedron:*

$$conv\,(\overline{\Xi}'_i) = \Big\{ \boldsymbol{\xi}'_i = (\xi'^{\top}_{i,1}, \boldsymbol{\xi}'^{\top}_{i,2})^{\top} \in \mathbb{R}^{r+1} : \exists \zeta_{i,p}(v) \in \mathbb{R}_+,\ \forall v \in V_i(p),\ p = 1, \ldots, r+1,\ \textit{such that}$$
$$\sum_{p=1}^{r+1} \sum_{\boldsymbol{v} \in V_i(p)} \zeta_{i,p}(v) = 1,$$
$$\xi'_{i,1} = \sum_{p=1}^{r+1} \sum_{\boldsymbol{v} \in V_i(p)} \zeta_{i,p}(v) v, \tag{2.29}$$
$$\boldsymbol{\xi}'_{i,2} = \sum_{p=1}^{r+1} \sum_{v \in V_i(p)} \zeta_{i,p}(v) \widehat{G}_p(\boldsymbol{e}_i v) \Big\}.$$

Proof Proposition 5 can be proven using similar arguments as in Proposition 3. □

Set $\text{conv}(\overline{\Xi}'_i)$ has $2r + 2$ extreme points. Using the extreme point representation (2.29), $\text{conv}(\overline{\Xi}'_i)$ can be represented through $4r+4$ constraints. Therefore, $\text{conv}(\overline{\Xi}') = \times_{i=1}^{k} \text{conv}(\overline{\Xi}'_i)$ can be represented using a total of $2+k(4r+4)$ constraints, i.e., its size grows quadratically in the dimension of Ξ, k, and the complexity of the binary decision rule r. One can now use the polyhedral description (2.29) together with Proposition 4, to reformulate Problem (2.14) into a mixed-integer linear optimization problem which can be expressed in the form (2.20). The size of the mixed-integer linear optimization problem will be polynomial in q, m, k and r. We emphasize that since the convex hull representation (2.29) is exact, *the solution of Problem* (2.20) *achieves the best binary decision rule associated with* $G(\cdot)$ *defined in* (2.22). For $G(\cdot)$ and Ξ defined in (2.22) and (2.21), respectively, the following proposition allows to instead of optimizing over integers in Problem (2.20), to restrict the integer variables to $Y \in \{-1, 0, 1\}^{q \times g}$ without introducing an additional approximation. This will significantly reduce the computational time needed to solve instances of Problem (2.20).

Proposition 6 *Let* $G(\cdot)$ *being defined in* (2.22) *and* Ξ *being a box uncertainty set defined in* (2.21). *Then, constraints* (2.4a) *imply that we can restrict* $Y \in \mathbb{Z}^{q \times g}$ *to* $Y \in \{-1, 0, 1\}^{q \times g}$ *without loss of generality.*

Proof First notice that by definition of $G(\cdot)$ in (2.22), the break points $\beta_{i,j}$ are unique in each direction $i \in \{2, \ldots, k\}$, and therefore the components of $G(\boldsymbol{\xi})$ are linearly independent for all $\boldsymbol{\xi} \in \Xi$, i.e,

$$\boldsymbol{v}^{\top} G(\boldsymbol{\xi}) = 0, \quad \forall \boldsymbol{\xi} \in \Xi \quad \Longrightarrow \quad \boldsymbol{v} = 0.$$

First we discuss problem instances involving one uncertain parameter, i.e., $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, l \le \xi_2 \le u\}$, and $y(\boldsymbol{\xi}) = \boldsymbol{y}^{\top} G(\boldsymbol{\xi})$, $\boldsymbol{y} \in \mathbb{Z}^{r+1}$ is one of the decision rules satisfying (2.4a). We will now prove by induction that constraint (2.4a) implies $y_i \in \{-1, 0, 1\}$ for all $i = 1, \ldots, r+1$. We will prove the basis step by contradiction. Assume that $y_{r+1} > 1$, then either $y_1 = 0, \ldots, y_r = 0$, which implies that

$$\boldsymbol{y}^{\top} G(\boldsymbol{\xi}) > 1, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1,\ \xi_2 \in [\beta_{2,r}, u]\},$$
$$\boldsymbol{y}^{\top} G(\boldsymbol{\xi}) = 0, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1,\ \xi_2 \in [l, \beta_{2,r})\},$$

or $\sum_{j=1}^{r} y_j \leq 1 - y_{r+1} < 0$ which implies that there exists a subset $\Xi_r \subseteq \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [l, \beta_{2,r})\}$ such that

$$\begin{aligned} &\boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [\beta_{2,r}, u]\}, \\ &\boldsymbol{y}^\top G(\boldsymbol{\xi}) < 0, \quad \forall \boldsymbol{\xi} \in \Xi_r. \end{aligned} \tag{2.30}$$

In both cases, the constraint $0 \leq \boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1$ for all $\boldsymbol{\xi} \in \Xi$, is violated, and therefore it must be the case that $y_{r+1} \leq 1$. Demonstrating the case that $y_{r+1} < -1$ is also infeasible can be shown in a similar way, thus concluding that $y_{r+1} \in \{-1, 0, 1\}$.

We will also prove the inductive step by contradiction. For some $i \in \{1, \ldots, r\}$, assume that $y_{i+1} \in \{-1, 0, 1\}, \ldots, y_{r+1} \in \{-1, 0, 1\}$, and assume that $y_i > 1$. Notice that $\boldsymbol{y}^\top G(\boldsymbol{\xi})$ can influence its value on $\boldsymbol{\xi} \in [\beta_{2,i}, \beta_{2,i+1})$, by only taking combinations of basis functions $G_1(\cdot), \ldots, G_i(\cdot)$. Then, either $y_1 = 0, \ldots, y_{i-1} = 0$, which implies that

$$\begin{aligned} &\boldsymbol{y}^\top G(\boldsymbol{\xi}) > 1, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [\beta_{2,i}, \beta_{2,i+1})\}, \\ &\boldsymbol{y}^\top G(\boldsymbol{\xi}) = 0, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [l, \beta_{2,i})\}, \end{aligned}$$

or $\sum_{j=1}^{i-1} y_j \leq 1 - y_i < 0$ which implies that there exists a subset $\Xi_i \subseteq \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [l, \beta_{2,i})\}$ such that

$$\begin{aligned} &\boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1, \quad \forall \boldsymbol{\xi} \in \{\boldsymbol{\xi} \in \mathbb{R}^2 : \xi_1 = 1, \xi_2 \in [\beta_{2,i}, \beta_{2,i+1})\}, \\ &\boldsymbol{y}^\top G(\boldsymbol{\xi}) < 0, \quad \forall \boldsymbol{\xi} \in \Xi_i. \end{aligned} \tag{2.31}$$

In both cases, the constraint $0 \leq \boldsymbol{y}^\top G(\boldsymbol{\xi}) \leq 1$, for all $\boldsymbol{\xi} \in \Xi$ is violated, and therefore it must be the case that $y_i \leq 1$. Again, demonstrating the case that $y_i < -1$ is also infeasible can be shown in a similar way, thus concluding that $\boldsymbol{y} \in \{-1, 0, 1\}^{r+1}$.

If the uncertainty set involves more than one uncertain parameter, since the components of $G(\boldsymbol{\xi})$ are linearly independent for all $\boldsymbol{\xi} \in \Xi$, it is easy to see that if for some $p \in \{2, \ldots, k\}$ and $j \in \{1, \ldots, r\}$, $y_{p,j} \in \{-1, 1\}$, i.e., non-zero, then $y_{i,j} = 0$ for all $i \in \{2 \ldots, k\}$, $i \neq p$ and $j \in \{1, \ldots, r\}$. Therefore, we conclude that if $G(\cdot)$ defined in (2.22) and Ξ in (2.21), we can restrict $Y \in \mathbb{Z}^{q \times g}$ to $Y \in \{-1, 0, 1\}^{q \times g}$ without loss of generality. □

We now consider case (ii), and we assume that Ξ is a generic set of the type (2.2) and $G(\cdot)$ is given by (2.22). From Sect. 2, we know that the number of constraints needed to describe the polyhedral representation of $\text{conv}(\overline{\Xi}')$ grows exponentially with the description of Ξ and complexity of the decision rule. In the following, we present a systematic way to construct a tractable outer approximation for $\text{conv}(\overline{\Xi}')$. Using this outer approximation to reformulate the lifted problem (2.10), will not yield the best binary decision rule structure associated with function $G(\cdot)$ but rather a conservative approximation. Nevertheless, the size of the resulting mixed-integer linear optimization problem will only grow polynomially with respect to the constrains of Ξ and complexity of the binary decision rule.

To this end, let $\{\boldsymbol{\xi} \in \mathbb{R}^k : \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u}\}$ be the smallest hyperrectangle containing Ξ. We have

$$\begin{aligned} \Xi &= \left\{\boldsymbol{\xi} \in \mathbb{R}^k : \exists \boldsymbol{\zeta} \in \mathbb{R}^v \text{ such that } W\boldsymbol{\xi} + U\boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi_1 = 1\right\}, \\ &= \left\{\boldsymbol{\xi} \in \mathbb{R}^k : \exists \boldsymbol{\zeta} \in \mathbb{R}^v \text{ such that } W\boldsymbol{\xi} + U\boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi_1 = 1,\ \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u}\right\}, \end{aligned} \tag{2.32}$$

which implies that the lifted uncertainty set $\Xi' = L(\Xi)$ can be expressed as $\Xi' = \Xi_1' \cap \Xi_2'$, where

$$\begin{aligned} \Xi_1' &:= \left\{\boldsymbol{\xi}' \in \mathbb{R}^{k'} : \exists \boldsymbol{\zeta} \in \mathbb{R}^v \text{ such that } W R_{\boldsymbol{\xi}}\boldsymbol{\xi}' + U\boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi_1' = 1\right\} \\ \Xi_2' &:= \left\{\boldsymbol{\xi}' \in \mathbb{R}^{k'} : \boldsymbol{l} \leq R_{\boldsymbol{\xi}}\boldsymbol{\xi}' \leq \boldsymbol{u},\ L(R_{\boldsymbol{\xi}}\boldsymbol{\xi}') = \boldsymbol{\xi}'\right\}. \end{aligned}$$

Notice that Ξ_2' has exactly the same structure as (2.25). We thus conclude that

$$\widehat{\Xi}' := \{\Xi_1' \cap \operatorname{conv}(\overline{\Xi}_2')\} \quad \supseteq \quad \operatorname{conv}(\overline{\Xi}'), \tag{2.33}$$

and therefore, $\widehat{\Xi}'$ can be used as an outer approximation of $\operatorname{conv}(\overline{\Xi}')$. Since $\operatorname{conv}(\overline{\Xi}_2')$ can be written as the polyhedron induced by Proposition 5, set $\widehat{\Xi}'$ has a total of $(l+1)+(2+k(4r+4))$ constraints. As before, one can now use the polyhedral description of $\widehat{\Xi}'$ together with Proposition 4, to reformulate Problem (2.14) into a mixed-integer linear optimization problem which can be expressed in the form (2.20). The size of the mixed-integer linear optimization problem will be polynomial in $q,\ m,\ k,\ l$ and r. Since $\widehat{\Xi}' \supseteq \operatorname{conv}(\overline{\Xi}')$, the *solution of Problem* (2.20) *might not achieve the best binary decision rule induced by* (2.22), *but rather a conservative approximation.* Nevertheless, using this systematic decomposition of the uncertainty set, we can efficiently apply binary decision rules to problems where the uncertainty set has arbitrary convex polyhedral structure.

One can use $G(\cdot)$ given by (2.22) together with this systematic decomposition of Ξ to achieve the same binary decision rule structures as those offered by $G(\cdot)$ defined in (2.4b). We will illustrate this through the following example. We assume that the problem in hand requires a generic Ξ of type (2.2), and we want to parameterize the binary decision rule to be a linear function of $G(\boldsymbol{\xi}) = \mathbf{1}\left(\boldsymbol{\alpha}^\top\boldsymbol{\xi} \geq \beta\right)$ for some $\boldsymbol{\alpha} \in \mathbb{R}^k$ and $\beta \in \mathbb{R}$. We can introduce an additional random variable $\tilde{\xi}$ in Ξ such that $\tilde{\xi} = \boldsymbol{\alpha}^\top\boldsymbol{\xi}$. Here $\tilde{\xi}$ is completely determined by the random variables already presented in Ξ. The modified support can now be written as follows,

$$\begin{aligned} \Xi &= \left\{(\boldsymbol{\xi}, \tilde{\xi}) \in \mathbb{R}^{k+1} : \exists \boldsymbol{\zeta} \in \mathbb{R}^v,\ W\boldsymbol{\xi} + U\boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi_1 = 1,\ \tilde{\xi} = \boldsymbol{\alpha}^\top\boldsymbol{\xi}\right\}, \\ &= \left\{(\boldsymbol{\xi}, \tilde{\xi}) \in \mathbb{R}^{k+1} : \exists \boldsymbol{\zeta} \in \mathbb{R}^v,\ W\boldsymbol{\xi} + U\boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi_1 = 1,\ \tilde{\xi} = \boldsymbol{\alpha}^\top\boldsymbol{\xi},\ \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u},\ \tilde{l} \leq \tilde{\xi} \leq \tilde{u}\right\}, \end{aligned} \tag{2.34}$$

where $\{(\boldsymbol{\xi}, \tilde{\xi}) \in \mathbb{R}^{k+1} : \boldsymbol{l} \leq \boldsymbol{\xi} \leq \boldsymbol{u},\ \tilde{l} \leq \tilde{\xi} \leq \tilde{u}\}$ is the smallest hyperrectangle containing Ξ. We can now use

$$G(\boldsymbol{\xi}, \tilde{\xi}) = \mathbf{1}\left(\tilde{\xi} \geq \beta\right), \tag{2.35}$$

which is an instance of (2.22), to achieve the same structure as $G(\boldsymbol{\xi}) = \mathbf{1}\left(\boldsymbol{\alpha}^\top \boldsymbol{\xi} \geq \beta\right)$. Defining appropriate $L(\cdot)$ and $R_{\boldsymbol{\xi}}$, R_G, we can use the following decomposition of the lifted uncertainty set $\Xi' = \Xi'_1 \cap \Xi'_2$ such that

$$\begin{aligned} \Xi'_1 &:= \left\{ (\boldsymbol{\xi}', \tilde{\xi}') \in \mathbb{R}^{k'+1} : \exists \boldsymbol{\zeta} \in \mathbb{R}^v \text{ such that } W R_{\boldsymbol{\xi}} \boldsymbol{\xi}' + U \boldsymbol{\zeta} \geq \boldsymbol{h},\ \xi'_1 = 1,\ \tilde{\xi}' = \boldsymbol{\alpha}^\top R_{\boldsymbol{\xi}} \boldsymbol{\xi}' \right\}, \\ \Xi'_2 &:= \left\{ (\boldsymbol{\xi}', \tilde{\xi}') \in \mathbb{R}^{k'+1} : \boldsymbol{l} \leq R_{\boldsymbol{\xi}} \boldsymbol{\xi}' \leq \boldsymbol{u},\ \tilde{l} \leq \tilde{\xi}' \leq \tilde{u},\ L(R_{\boldsymbol{\xi}}(\boldsymbol{\xi}', \tilde{\xi}')) = (\boldsymbol{\xi}', \tilde{\xi}') \right\}. \end{aligned}$$

By defining, $\widehat{\Xi}' := \{\Xi'_1 \cap \text{conv}(\Xi'_2)\} \supseteq \text{conv}(\overline{\Xi}')$ we can reformulate Problem (2.14) into a mixed-integer linear optimization problem with polynomial number of decision variable and constraints with respect to the the input data. Once more, we emphasize that this systematic decomposition might not achieve the best binary decision rule induced by $G(\cdot)$ but rather a conservative approximation. Nevertheless, we can now achieve the same flexibility for the binary decision rules as those offered by (2.4b) without the exponential growth in the size of the resulting problem.

3 Binary decision rules for random-recourse problems

In this section, we present our approach for one-stage adaptive optimization problems with random recourse. The solution method of this class of problems is an adaptation of the solution method presented in Sect. 2.2. Given function $B : \mathbb{R}^k \rightarrow \mathbb{R}^{m \times q}$ and matrices $D \in \mathbb{R}^{q \times k}$, $H \in \mathbb{R}^{m \times k}$ and a probability measure $\mathbb{P}_{\boldsymbol{\xi}}$ supported on set Ξ for the uncertain vector $\boldsymbol{\xi} \in \mathbb{R}^k$, we are interested in choosing binary functions $\boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}$ in order to solve:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\left((D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})\right) \\ \text{subject to} & \left.\begin{array}{l} \boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}, \\ B(\boldsymbol{\xi})\boldsymbol{y}(\boldsymbol{\xi}) \leq H\boldsymbol{\xi}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi, \end{array} \tag{3.1}$$

where Ξ is a generic polyhedron of type (2.2). We assume that the recourse matrix $B(\boldsymbol{\xi})$ depends linearly on the uncertain parameters. Therefore, the μth row of $B(\boldsymbol{\xi})$ is representable as $\boldsymbol{\xi}^\top B_\mu$ for matrices $B_\mu \in \mathbb{R}^{k \times q}$ with $\mu = 1, \ldots, m$. Problem (3.1) can therefore be written as the following problem:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\left((D\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})\right) \\ \text{subject to} & \left.\begin{array}{l} \boldsymbol{y}(\cdot) \in \mathcal{B}_{k,q}, \\ \boldsymbol{\xi}^\top B_\mu \boldsymbol{y}(\boldsymbol{\xi}) \leq H_\mu^\top \boldsymbol{\xi}, \quad \mu = 1, \ldots, m \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi, \end{array} \tag{3.2}$$

where $H_\mu^\top$ denotes the μth row of matrix H. Problem (3.2) involves a continuum of decision variables and inequality constraints. Therefore, in order to make the problem amenable to numerical solutions, we restrict $\boldsymbol{y}(\cdot)$ to admit structure (2.4). Applying the binary decision rules (2.4) to Problem (3.2) yields the following semi-infinite problem, which involves a finite number of decision variables $Y \in \mathbb{Z}^{q \times g}$, and an infinite number of constraints:

$$\left.\begin{array}{ll}\text{minimize} & \mathbb{E}_{\boldsymbol{\xi}}\Big(\boldsymbol{\xi}^\top D^\top Y G(\boldsymbol{\xi})\Big) \\ \text{subject to} & Y \in \mathbb{Z}^{q\times g}, \\ & \boldsymbol{\xi}^\top B_\mu Y G(\boldsymbol{\xi}) \le H_\mu^\top \boldsymbol{\xi}, \quad \mu = 1,\dots,m, \\ & 0 \le Y G(\boldsymbol{\xi}) \le \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \varXi. \tag{3.3}$$

Notice that all decision variables appear linearly in Problem (2.6). Nevertheless, the objective and constraints are non-linear functions of $\boldsymbol{\xi}$.

We now define the non-linear lifting operator $L : \mathbb{R}^k \to \mathbb{R}^{k'}$ such that

$$L : \mathbb{R}^k \to \mathbb{R}^{k'}, \qquad L(\boldsymbol{\xi}) = \begin{pmatrix} \boldsymbol{\xi} \\ G(\boldsymbol{\xi}) \\ G(\boldsymbol{\xi})\xi_1 \\ \vdots \\ G(\boldsymbol{\xi})\xi_k \end{pmatrix}, \tag{3.4}$$

$k' = k + g + gk$. Note that including both components $G(\boldsymbol{\xi})$ and $G(\boldsymbol{\xi})\xi_1$ in $L(\boldsymbol{\xi})$ is redundant as by construction $\xi_1 = 1$. Nevertheless, in the following we adopt formulation (3.4) for ease of exposition. We also define matrices $R_{\boldsymbol{\xi}} \in \mathbb{R}^{k\times k'}$, and $R_G \in \mathbb{R}^{q\times k'}$ such that

$$R_{\boldsymbol{\xi}} L(\boldsymbol{\xi}) = \boldsymbol{\xi}, \qquad R_G L(\boldsymbol{\xi}) = G(\boldsymbol{\xi}). \tag{3.5}$$

In addition, we define $F : \mathbb{R}^{k\times g} \to \mathbb{R}^{k'}$ that expresses the quadratic polynomials $\boldsymbol{\xi}^\top B_\mu Y G(\boldsymbol{\xi})$, as linear functions of $L(\boldsymbol{\xi})$ through the following constraints:

$$\mathbf{y}'_\mu = F(B_\mu Y), \quad \mathbf{y}'^\top_\mu L(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top B_\mu Y G(\boldsymbol{\xi}), \quad \mathbf{y}'_\mu \in \mathbb{R}^{k'}, \quad \mu = 1,\dots,m, \quad \forall \boldsymbol{\xi} \in \varXi. \tag{3.6}$$

Constraints $\mathbf{y}'_\mu = F(B_\mu Y)$, $\mu = 1,\dots,m$, are linear, since $F(\cdot)$ effectively reorders the entries of matrix $B_\mu Y$ into the vector $\mathbf{y}'_\mu$. Using (3.4), (3.5) and (3.6), Problem (3.3) can now be rewritten into the following optimization problem.

$$\left.\begin{array}{ll}\text{minimize} & \mathrm{Tr}\Big(M R_{\boldsymbol{\xi}}^\top D^\top Y R_G\Big) \\ \text{subject to} & Y \in \mathbb{Z}^{q\times g}, \\ & \mathbf{y}'^\top_\mu L(\boldsymbol{\xi}) \le H_\mu^\top R_{\boldsymbol{\xi}} L(\boldsymbol{\xi}), \quad \mu = 1,\dots,m, \\ & \mathbf{y}'_\mu = F(B_\mu Y),\ \mathbf{y}'_\mu \in \mathbb{R}^{k'},\ \mu = 1,\dots,m, \\ & 0 \le Y R_G L(\boldsymbol{\xi}) \le \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \varXi. \tag{3.7}$$

Problem (3.7) has similar structure as Problem (2.10), i.e., both Y and $L(\boldsymbol{\xi})$ appear linearly in the constraints. Therefore, in the following we redefine Problem (3.7) in a higher dimensional space and apply Proposition 4 to reformulate the problem into a mixed-integer optimization problem.

We define the uncertain vector $\boldsymbol{\xi}' = (\boldsymbol{\xi}'^\top_{1,1}, \boldsymbol{\xi}'^\top_{1,2}, \boldsymbol{\xi}'^\top_{2,1}, \dots, \boldsymbol{\xi}'^\top_{2,k})^\top \in \mathbb{R}^{k'}$ such that $\boldsymbol{\xi}' = L(\boldsymbol{\xi})$ and $\boldsymbol{\xi}'_{1,1} = \boldsymbol{\xi}$, $\boldsymbol{\xi}'_{1,2} = G(\boldsymbol{\xi})$, and $\boldsymbol{\xi}'_{2,i} = G(\boldsymbol{\xi})\xi_i$ for $i = 1,\dots,k$. $\boldsymbol{\xi}'$ has

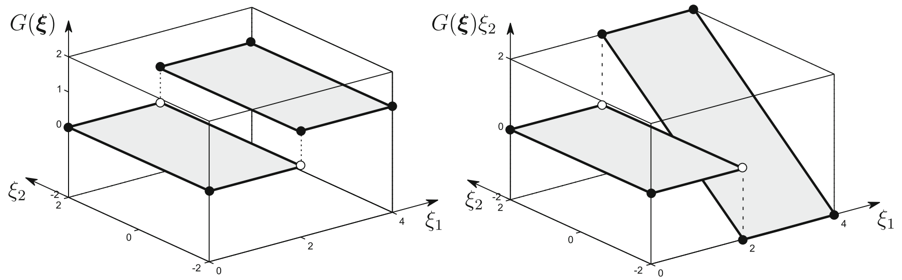

Fig. 4 Visualization of $L(\boldsymbol{\xi}) = (\xi_1, \xi_2, G(\boldsymbol{\xi}))$ (*left*) and $\widetilde{L}(\boldsymbol{\xi}) = (\xi_1, \xi_2, G(\boldsymbol{\xi})\xi_2)$ (*right*) where $G(\boldsymbol{\xi}) = \mathbf{1}(\xi_1 \geq 2)$, for $\xi_1 \in [0, 4]$ and $\xi_2 \in [-2, 2]$. Here, $V = \{(0, -2), (0, 2), (2, -2), (2, 2), (4, -2), (4, 2)\}$. The convex hull of $L(\boldsymbol{\xi})$ is constructed by taken convex combinations of $L(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in V$ and $(2, -2, \widehat{G}(2, -2)))^\top$, $(2, 2, \widehat{G}((2, 2)))^\top$, and the convex hull of $\widetilde{L}(\boldsymbol{\xi})$ is constructed by taken convex combinations of $\widetilde{L}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in V$ and $(2, -2, \widehat{G}(2, -2)(-2))^\top$, $(2, 2, \widehat{G}(2, 2)2)^\top$

probability measure $\mathbb{P}_{\boldsymbol{\xi}'}$ defined as in (2.11) and support $\Xi' = L(\Xi)$. Problem (3.7) can now be written as the equivalent semi-infinite problem defined on the lifted space $\boldsymbol{\xi}'$.

$$
\begin{array}{ll}
\text{minimize} & \operatorname{Tr}\left(M' R_{\boldsymbol{\xi}}^\top D^\top Y R_G\right) \\
\text{subject to} & Y \in \mathbb{Z}^{q \times g}, \\
& \left.\begin{array}{ll}
\mathbf{y}_\mu'^\top \boldsymbol{\xi}' \leq H_\mu^\top R_{\boldsymbol{\xi}}\, \boldsymbol{\xi}', & \mu = 1, \ldots, m, \\
\mathbf{y}_\mu' = F(B_\mu Y),\ \mathbf{y}_\mu' \in \mathbb{R}^{k'}, & \mu = 1, \ldots, m, \\
0 \leq Y R_G \boldsymbol{\xi}' \leq \boldsymbol{e},
\end{array}\right\} \quad \forall \boldsymbol{\xi}' \in \Xi',
\end{array}
\tag{3.8}
$$

We will construct the convex hull of $\overline{\Xi}'$ in the same way as in Sect. 2.2. Notice that $\operatorname{conv}(\overline{\Xi}')$ will have the same number of extreme points as the convex hull defined through the lifting $L(\boldsymbol{\xi}) = (\boldsymbol{\xi}^\top, G(\boldsymbol{\xi})^\top)^\top$, see Fig. 4.

By defining the partitions Ξ_p of Ξ as in (2.5), and $\widehat{G}_p(\boldsymbol{\xi})$ as in (2.16), it is easy to see that $\operatorname{conv}(\overline{\Xi}')$ can be described as a convex combination of the following points.

$$
\begin{pmatrix} \boldsymbol{\xi} \\ \widehat{G}_p(\boldsymbol{\xi}) \\ \widehat{G}_p(\boldsymbol{\xi})\xi_1 \\ \vdots \\ \widehat{G}_p(\boldsymbol{\xi})\xi_k \end{pmatrix}, \qquad \boldsymbol{\xi} \in V(p),\ p = 1, \ldots, P.
$$

The following proposition gives the polyhedral representation of the convex hull of $\overline{\Xi}'$.

Proposition 7 *Let $L(\cdot)$ being defined in* (3.4) *and $G(\cdot)$ being defined in* (2.4b). *Then, the exact representation of the convex hull of $\overline{\Xi}'$ is given by the following polyhedron:*

$$
\begin{aligned}
conv\,(\overline{\Xi}') = \Big\{ & \boldsymbol{\xi}' = (\boldsymbol{\xi}_{1,1}'^{\top}, \boldsymbol{\xi}_{1,2}'^{\top}, \boldsymbol{\xi}_{2,1}'^{\top}, \ldots, \boldsymbol{\xi}_{2,k}'^{\top})^{\top} \in \mathbb{R}^{k'} : \exists \zeta_p(\boldsymbol{v}) \in \mathbb{R}_+,\ \forall \boldsymbol{v} \in V(p),\ p = 1, \ldots, P, \\
& \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) = 1, \\
& \boldsymbol{\xi}_{1,1}' = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \boldsymbol{v}, \\
& \boldsymbol{\xi}_{1,2}' = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \widehat{G}_p(\boldsymbol{v}), \\
& \boldsymbol{\xi}_{2,i}' = \sum_{p=1}^{P} \sum_{\boldsymbol{v} \in V(p)} \zeta_p(\boldsymbol{v}) \widehat{G}_p(\boldsymbol{v}) v_i,\ i = 1, \ldots, k \Big\}.
\end{aligned} \tag{3.9}
$$

Proof Proposition 7 can be proven using similar arguments as in Proposition 3. □

The description of polyhedron (3.9) inherits the same exponential complexity as (2.17), since the number of constraints in (3.9) depends on the cardinality of V which is constructed using the extreme points of the partitions Ξ_p. Nevertheless, the description (3.9) provides the exact representation of the convex hull for conv $(\overline{\Xi}')$.

Using Proposition 4 together with the polyhedron (3.9), we can now reformulate Problem (3.8) into the following mixed-integer linear optimization problem.

$$
\begin{aligned}
\text{minimize} \quad & \mathrm{Tr}\Big(M' R_{\boldsymbol{\xi}}^{\top} D^{\top} Y R_G \Big) \\
\text{subject to} \quad & Y \in \mathbb{Z}^{q \times g},\ \Gamma \in \mathbb{R}_+^{q \times l'},\ \Theta \in \mathbb{R}_+^{q \times l'}, \\
& \left. \begin{aligned}
& \mathbf{y}_\mu' \in \mathbb{R}^{k'},\ \boldsymbol{\lambda}_\mu \in \mathbb{R}_+^{m \times l'}, \\
& \mathbf{y}_\mu' + \boldsymbol{\lambda}_\mu W' = R_{\boldsymbol{\xi}}^{\top} H_\mu,\ \boldsymbol{\lambda}_\mu U' = 0,\ \boldsymbol{\lambda}_\mu \boldsymbol{h}' \geq 0 \\
& \mathbf{y}_\mu' = F(B_\mu Y),\ \mathbf{y}_\mu' \in \mathbb{R}^{k'},
\end{aligned} \right\} \quad \mu = 1, \ldots, m, \\
& Y R_G = \Gamma W',\ \Gamma U' = 0,\ \Gamma \boldsymbol{h}' \geq 0, \\
& \boldsymbol{e}\boldsymbol{e}_1^{\top} - Y R_G = \Theta W',\ \Theta U' = 0,\ \Theta \boldsymbol{h}' \geq 0.
\end{aligned} \tag{3.10}
$$

Here, the auxiliary variables $\boldsymbol{\lambda}_\mu \in \mathbb{R}_+^{m \times l'}$ are associated with constraints $\mathbf{y}_\mu'^{\top} \boldsymbol{\xi}' \leq H_\mu^{\top} R_{\boldsymbol{\xi}}\, \boldsymbol{\xi}'$, for $\mu = 1, \ldots, m$, and $\Gamma \in \mathbb{R}_+^{q \times l'}$ and $\Theta \in \mathbb{R}_+^{q \times l'}$ with constraints $0 \leq Y R_G \boldsymbol{\xi}'$ and $Y R_G \boldsymbol{\xi}' \leq \boldsymbol{e}$, respectively. Problem (3.10) is the exact reformulation of Problem (3.8), since (3.9) is the exact representation of the convex hull of $\overline{\Xi}'$, and thus *the solution of Problem* (3.10) *achieves the best binary decision rule associated with* $G(\cdot)$ *in* (2.4b). Nevertheless, the size of Problem (3.10) is affected by the exponential growth of the constraints in conv $(\overline{\Xi}')$. One can mitigate this exponential growth by considering simplified structures of Ξ and $G(\cdot)$, following similar guidelines as those those discussed in Sect. 2.3.

4 Binary decision rules for multistage problems

In this section, we extend the methodology presented in Sect. 2 to cover multistage adaptive optimization problems with fixed-recourse. The mathematical formulations presented can easily be adapted to random-recourse problems by following the guidelines of Sect. 3.

The dynamic decision process considered can be described as follows: A decision maker first observes an uncertain parameter $\boldsymbol{\xi}_1 \in \mathbb{R}^{k_1}$ and then takes a binary decision $\boldsymbol{y}_1(\boldsymbol{\xi}_1) \in \{0,1\}^{q_1}$. Subsequently, a second uncertain parameter $\boldsymbol{\xi}_2 \in \mathbb{R}^{k_2}$ is revealed, in response to which the decision maker takes a second decision $\boldsymbol{y}_2(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in \{0,1\}^{q_2}$. This sequence of alternating observations and decisions extends over T time stages. To enforce the non-anticipative structure of the decisions, a decision taken at stage t can only depend on the observed parameters up to and including stage t, i.e., $\boldsymbol{y}_t(\boldsymbol{\xi}^t)$ where $\boldsymbol{\xi}^t = (\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_t^\top)^\top \in \mathbb{R}^{k^t}$, with $k^t = \sum_{s=1}^t k_s$. For consistency with the previous sections, and with slight abuse of notation, we assume that $k_1 = 1$ and $\boldsymbol{\xi}_1 = 1$. As before, setting $\boldsymbol{\xi}_1 = 1$ is a non-restrictive assumption which allows to represent affine functions of the non-degenerate outcomes $(\boldsymbol{\xi}_2^\top, \ldots, \boldsymbol{\xi}_t^\top)^\top$ in a compact manner as linear functions of $(\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_t^\top)^\top$. We denote by $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_T^\top)^\top \in \mathbb{R}^k$ the vector of all uncertain parameters, where $k = k^T$. Finally, we denote by $\mathcal{B}_{k^t, q_t}$ the space of binary functions from $\mathbb{R}^{k^t}$ to $\{0,1\}^{q_t}$.

Given matrices $B_{ts} \in \mathbb{R}^{m_t \times q_s}$, $D_t \in \mathbb{R}^{q_t \times k^t}$ and $H_t \in \mathbb{R}^{m_t \times k^t}$, and a probability measure $\mathbb{P}_{\boldsymbol{\xi}}$ supported on set Ξ given by (2.2), for the uncertain vector $\boldsymbol{\xi} \in \mathbb{R}^k$, we are interested in choosing binary functions $\boldsymbol{y}_t(\cdot) \in \mathcal{B}_{k^t, q_t}$ in order to solve:

$$
\begin{array}{ll}
\text{minimize} & \mathbb{E}_{\boldsymbol{\xi}} \left(\displaystyle\sum_{t=1}^{T} (D_t \boldsymbol{\xi}^t)^\top \boldsymbol{y}_t(\boldsymbol{\xi}^t) \right) \\
\text{subject to} & \left. \begin{array}{l} \boldsymbol{y}_t(\cdot) \in \mathcal{B}_{k^t, q_t}, \\ \displaystyle\sum_{s=1}^{t} B_{ts} \boldsymbol{y}_s(\boldsymbol{\xi}^s) \leq H_t \boldsymbol{\xi}^t, \end{array} \right\} \quad t = 1, \ldots, T, \quad \forall \boldsymbol{\xi} \in \Xi.
\end{array}
\tag{4.1}
$$

Problem (4.1) involves a continuum of decision variables and inequality constraints. Therefore, in order to make the problem amenable to numerical solutions, we restrict the feasible region of the binary functions $\boldsymbol{y}_t(\cdot) \in \mathcal{B}_{k^t, q_t}$ to admit the following piecewise constant structure:

$$
\left. \begin{array}{l} \boldsymbol{y}_t(\boldsymbol{\xi}) = Y_t G^t(\boldsymbol{\xi}), \ Y \in \mathbb{Z}^{q_t \times g^t}, \\ 0 \leq Y_t G^t(\boldsymbol{\xi}) \leq \boldsymbol{e}, \end{array} \right\} \quad t = 1, \ldots, T, \quad \forall \boldsymbol{\xi} \in \Xi, \tag{4.2a}
$$

where $G^t : \mathbb{R}^k \to \{0,1\}^{g^t}$ can be expressed by $G^t(\cdot) = (G_1(\cdot)^\top, \ldots, G_t(\cdot)^\top)$, and $G_t : \mathbb{R}^{k^t} \to \{0,1\}^{g_t}$ are the piecewise constant functions:

$$
G_1(\boldsymbol{\xi}_1) := 1, \quad G_{t,i}(\boldsymbol{\xi}) := \mathbf{1}\left(\boldsymbol{\alpha}_{t,i}^\top \boldsymbol{\xi}^t \geq \beta_{t,i} \right), \quad i = 1, \ldots, g_t, \ t = 2, \ldots, T, \tag{4.2b}
$$

for given $\boldsymbol{\alpha}_{t,i} \in \mathbb{R}^{k^t}$ and $\beta_{t,i} \in \mathbb{R}$, $i = 1, \ldots, g$, $t = 2, \ldots, T$. The dimension of $G^t(\cdot)$ is g^t, with $g^t = \sum_{s=1}^t g_s$. The non-anticipativity of $\boldsymbol{y}_t(\cdot)$ is ensured by restricting $G^t(\cdot)$ to depend only on random parameters up to and including stage t.

The pairs $(\boldsymbol{\alpha}_{t,i}, \beta_{t,i})$ defining the $G(\cdot)$ need to satisfy the same requirements as those presented in Sect. (2.4), namely, $(\boldsymbol{\alpha}_{t,i}, \beta_{t,i})$ are chosen such that they result to unique hyperplanes $\boldsymbol{\alpha}_{t,i}^{\top}\boldsymbol{\xi}^{t} - \beta_{t,i} = 0$ for all $i \in \{2,\dots,g\}$ and $t \in \{2,\dots,T\}$, and the partitions Ξ_p given by (2.5) that result from the choices of $(\boldsymbol{\alpha}_{t,i}, \beta_{t,i})$, must have (i) non-empty relative interior, (ii) Ξ_p spans $\mathbb{R}^k$, (iii) any partition pair Ξ_i and Ξ_j, $i \neq j$, can overlap only on one of their facets, and (iv) $\Xi = \bigcup_{p=1}^{P} \Xi_p$.

Applying decision rules (4.2) to Problem (4.1) yields the following semi-infinite problem.

$$\begin{aligned} &\text{minimize} \quad \mathbb{E}_{\boldsymbol{\xi}}\left(\sum_{t=1}^{T}(D_t\boldsymbol{\xi}^t)^{\top}Y_tG^t(\boldsymbol{\xi})\right) \\ &\text{subject to} \quad \left.\begin{aligned} &Y_t \in \mathbb{Z}^{q_t\times g^t}, \\ &\sum_{s=1}^{t} B_{ts}Y_sG^s(\boldsymbol{\xi}) \le H_t\boldsymbol{\xi}^t, \\ &0 \le Y_tG^t(\boldsymbol{\xi}) \le \boldsymbol{e} \end{aligned}\right\} \quad t=1,\dots,T, \quad \forall \boldsymbol{\xi} \in \Xi. \end{aligned} \tag{4.3}$$

We can now use the lifting techniques to first express Problem (4.3) in a higher dimensional space, compute the convex hull associated with the non-convex uncertainty set of the lifted problem, and finally reformulate the semi-infinite structure using Proposition 4. We define lifting $L^t : \mathbb{R}^k \rightarrow \mathbb{R}^{k'^t}$, such that $L^t(\cdot) = (L_1(\cdot)^{\top},\dots,L_t(\cdot)^{\top})^{\top}$, where

$$L_t : \mathbb{R}^{k_t} \rightarrow \mathbb{R}^{k'_t}, \qquad L_t(\boldsymbol{\xi}) = (\boldsymbol{\xi}_t^{\top}, G_t(\boldsymbol{\xi})^{\top})^{\top},\ t=1,\dots,T, \tag{4.4a}$$

Here, $k'_t = k_t + g_t$. By convention $L(\cdot) = (L_1(\cdot)^{\top},\dots,L_T(\cdot)^{\top})^{\top}$ and $k'^t = \sum_{s=1}^{t} k'_s$ with $k' = k'^T$. In addition, we define matrices, $R_{\boldsymbol{\xi},t} \in \mathbb{R}^{k^t\times k'}$ and $R_{G,t} \in \mathbb{R}^{g^t\times k'}$ such that

$$R_{\boldsymbol{\xi},t}L(\boldsymbol{\xi}) = \boldsymbol{\xi}^t, \qquad R_{G,t}L(\boldsymbol{\xi}) = G(\boldsymbol{\xi})^t, \quad t=1,\dots,T. \tag{4.4b}$$

Using (4.4), Problem (4.3) can now be rewritten in the following form:

$$\begin{aligned} &\text{minimize} \quad \mathbb{E}_{\boldsymbol{\xi}}\left(\sum_{t=1}^{T}(D_tR_{\boldsymbol{\xi},t}L(\boldsymbol{\xi}))^{\top}Y_tR_{G,t}L(\boldsymbol{\xi})\right) \\ &\text{subject to} \quad \left.\begin{aligned} &Y_t \in \mathbb{Z}^{q_t\times g^t}, \\ &\sum_{s=1}^{t} B_{ts}Y_sR_{G,s}L(\boldsymbol{\xi}) \le H_tR_{\boldsymbol{\xi},t}L(\boldsymbol{\xi}), \\ &0 \le Y_tR_{G,t}L(\boldsymbol{\xi}) \le \boldsymbol{e} \end{aligned}\right\} \quad t=1,\dots,T, \quad \forall \boldsymbol{\xi} \in \Xi. \end{aligned} \tag{4.5}$$

Problem (4.5) has similar structure as Problem (2.10), i.e., both Y_t and $L(\boldsymbol{\xi})$ appear linearly in the constraints. Therefore, in the following we redefine Problem (4.5) in a higher dimensional space and apply Proposition 4 to reformulate the problem into a mixed-integer optimization problem.

We now define the uncertain vector $\boldsymbol{\xi}' = (\boldsymbol{\xi}_1'^{\top},\dots,\boldsymbol{\xi}_T'^{\top})^{\top} \in \mathbb{R}^{k'}$ such that $\boldsymbol{\xi}' = L(\boldsymbol{\xi})$ and $\boldsymbol{\xi}'_t = L_t(\boldsymbol{\xi})$ for $t=1,\dots,T$. $\boldsymbol{\xi}'$ has probability measure $\mathbb{P}_{\boldsymbol{\xi}'}$ defined using (2.11), and probability measure support $\Xi' = L(\Xi)$. Using the definition of $\boldsymbol{\xi}'$, the lifted semi-infinite problem is given as follows:

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\boldsymbol{\xi}'}\left(\sum_{t=1}^{T}(D_t R_{\boldsymbol{\xi},t}\boldsymbol{\xi}')^\top Y_t R_{G,t}\boldsymbol{\xi}'\right) \\ \text{subject to} & \left.\begin{array}{l} Y_t \in \mathbb{Z}^{q_t \times g^t}, \\ \sum_{s=1}^{t} B_{ts} Y_s R_{G,s}\boldsymbol{\xi}' \leq H_t R_{\boldsymbol{\xi},t}\boldsymbol{\xi}', \\ 0 \leq Y_t R_{G,t}\boldsymbol{\xi}' \leq \boldsymbol{e} \end{array}\right\} \quad t = 1, \ldots, T, \quad \forall \boldsymbol{\xi}' \in \Xi'. \end{array} \tag{4.6}$$

The definition of $L(\cdot)$ in (4.4a) share almost identical structure as $L(\cdot)$ in (2.9). Therefore, the convex hull of $\overline{\Xi}'$ can be expressed as a slight variant of the polyhedron given in (2.17), and will constitute the exact representation of the convex hull of $\text{conv}(\overline{\Xi}')$. Using Proposition 4 together with the polyhedral representation of $\text{conv}(\overline{\Xi}')$, we can now reformulate Problem (3.8) into the following mixed-integer linear optimization problem.

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T} \text{Tr}\Big(M' R_{\boldsymbol{\xi},t}^\top D_t^\top Y_t R_{G,t}\Big) \\ \text{subject to} & \left.\begin{array}{l} Y_t \in \mathbb{Z}^{q_t \times g^t}, \;\; \Lambda \in \mathbb{R}_+^{m_t \times l'}, \;\; \Gamma \in \mathbb{R}_+^{q^t \times l'}, \;\; \Theta \in \mathbb{R}_+^{q^t \times l'} \\ \sum_{s=1}^{t} B_{ts} Y R_{G,s} + \Lambda_t W' = H_t R_{\boldsymbol{\xi},t}, \;\; \Lambda_t U' = 0, \;\; \Lambda_t \boldsymbol{h}' \geq 0, \\ Y_t R_{G,t} = \Gamma_t W', \;\; \Gamma_t U' = 0, \;\; \Gamma_t \boldsymbol{h}' \geq 0, \\ \boldsymbol{e}\boldsymbol{e}_1^\top - Y_t R_{G,t} = \Theta_t W', \;\; \Theta_t U' = 0, \;\; \Theta_t \boldsymbol{h}' \geq 0. \end{array}\right\} t = 1, \ldots, T, \end{array} \tag{4.7}$$

where $M' \in \mathbb{R}^{k' \times k'}$, $M' = \mathbb{E}_{\boldsymbol{\xi}'}(\boldsymbol{\xi}'\boldsymbol{\xi}'^\top)$. Once more, we emphasize that Problem (4.7) is the exact reformulation of Problem (4.6). Therefore, *the solution of Problem* (4.7) *achieves the best binary decision rule associated with* $G(\cdot)$ *in* (2.4b) at the cost of the exponential growth of its size with respect to the description of Ξ and the complexity of the binary decision rules (4.2). The exponential growth in the time horizon T is implicit through the description of Ξ and the structure of the binary decision rules. One can mitigate this exponential growth by considering simplified structures of Ξ and $G(\cdot)$, following similar guidelines as those discussed in Sect. 2.3.

5 Computational results

In this section, we apply the proposed binary decision rules to a multistage inventory control problem and to a multistage knapsack problem. In both problems, the structure of the binary decision rules is constructed using $G(\cdot)$ defined in (2.22), where the breakpoints $\beta_{i,1}, \ldots, \beta_{i,r}$ are placed equidistantly within the marginal support of each random parameter ξ_i. To improve the scalability of the solution method, we further restrict the structure of the binary decisions rules to be

$$\left.\begin{array}{l} \boldsymbol{y}(\boldsymbol{\xi}) = YG(\boldsymbol{\xi}),\ Y \in \{-1, 0, 1\}^{q\times g}, \\ 0 \le YG(\boldsymbol{\xi}) \le \boldsymbol{e}, \end{array}\right\} \quad \forall \boldsymbol{\xi} \in \Xi. \tag{5.1}$$

As shown in Proposition 6, for problem instances where $G(\cdot)$ and Ξ are defined in (2.22) and (2.21), respectively, this restriction does not introduce an additional conservatism. All of our numerical results are carried out using the IBM ILOG CPLEX 12.5 optimization package on a Intel Core i5 − 3570 CPU at 3.40GHz machine with 8GB RAM [26].

5.1 Multistage inventory control

In this case study, we consider a single item inventory control problem where the ordering decisions are discrete. This problem can be formulated as an instance of the multistage adaptive optimization problem with fixed-recourse, and can be described as follows. At the beginning of each time period $t \in \mathcal{T} := \{2, \ldots, T\}$, the decision maker observes the product demand ξ_t that needs to be satisfied robustly. This demand can be served in two ways: (i) by pre-ordering at stage 1 a maximum of N lots, each delivering a fixed quantity q_z at the beginning of period t, for a unit cost of c_z; (ii) or by placing an order for a maximum of N lots, each delivering immediately a fixed quantity q_y, for a unit cost of c_y, with $c_z < c_y$. For each lot $n = 1, \ldots, N$, the pre-ordering binary decisions delivered at stage t are denoted by $z_{n,t} \in \{0, 1\}$, and the recourse binary ordering decisions are denoted by $y_{n,t}(\cdot) \in \mathcal{B}_{k^t,1}$. If the ordered quantity is greater than the demand, the excess units are stored in a warehouse, incurring a unit holding cost c_h, and can be used to serve future demand. The level of available inventory at each period is given by $I_t(\cdot) \in \mathcal{R}_{k^t,1}$. In addition, the cumulative volume of pre-orders $\sum_{s=1}^{t}\sum_{n=1}^{N} z_{n,s}$ must not exceed the ordering budget $\overline{B}_{\text{tot},t}$. The decision maker wishes to determine the orders $z_{n,t}$ and $y_{n,t}(\cdot)$ that minimize the total ordering and holding costs associated with the worst-case demand realization over the planning horizon T. The problem can be formulated as the following multistage adaptive optimization problem.

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\boldsymbol{\xi}\in\Xi} \left(\sum_{t\in\mathcal{T}} \sum_{n=1}^{N} c_z\, q_z z_{n,t} + c_y\, q_y y_{n,t}(\boldsymbol{\xi}^t) + c_h\, I_t(\boldsymbol{\xi}^t) \right) \\ \text{subject to} & \left.\begin{array}{l} I_t(\cdot) \in \mathcal{R}_{k^t,1}, \\ z_{n,t} \in \{0,1\},\ y_{n,t}(\cdot) \in \mathcal{B}_{k^t,1}, \quad \forall n = 1, \ldots, N, \\ \displaystyle I_t(\boldsymbol{\xi}^t) = I_{t-1}(\boldsymbol{\xi}^{t-1}) + \sum_{n=1}^{N} q_z z_{n,t} + q_y y_{n,t}(\boldsymbol{\xi}^t) - \xi_t, \\ I_t(\boldsymbol{\xi}^t) \ge 0, \\ \displaystyle\sum_{s=1}^{t}\sum_{n=1}^{N} q_z z_{n,s} \le \overline{B}_{\text{tot},t}, \end{array}\right\} \forall t \in \mathcal{T},\ \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{5.2}$$

The uncertainty set Ξ is given by,

$$\Xi := \left\{ \boldsymbol{\xi} \in \mathbb{R}^T : \ \xi_1 = 1,\ \boldsymbol{l} \le \boldsymbol{\xi} \le \boldsymbol{u} \right\}.$$

Table 2 Comparison of average improvement of adaptive versus static binary decision rules using the performance measure $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$, (top), and average solution time of binary decision rules (bottom)

	Global optimality (%)			1% optimality (%)			5% optimality (%)		
	$r=1$	$r=5$	$r=9$	$r=1$	$r=5$	$r=9$	$r=1$	$r=5$	$r=9$
Average objective value improvements, $N=2$									
$T=10$	15	17	17	14	17	17	14	16	16
$T=15$	16	18	18	16	18	18	16	17	17
$T=20$	18	19	20	18	19	20	17	18	19
	Global optimality (s)			1% optimality (s)			5% optimality (s)		
	$r=1$	$r=5$	$r=9$	$r=1$	$r=5$	$r=9$	$r=1$	$r=5$	$r=9$
Average solution time, $N=2$									
$T=10$	<1	4	42	<1	4	15	<1	1	8
$T=15$	<1	166	983	<1	16	83	<1	3	10
$T=20$	1	688	5580	<1	46	710	<1	9	92

We emphasize that since we are interested in minimizing the total costs associated with the worst-case demand realization, the model does not require to specify a distribution for our uncertainty demand. Moreover, notice that the real-valued inventory decision $I_t(\boldsymbol{\xi}^t)$ can be eliminated from formulation (5.2) by substituting $I_t(\boldsymbol{\xi}^t) = I_1 + \sum_{s=1}^{t}\left(\sum_{i=1}^{N} q_z z_{n,s} + q_y y_{n,s}(\boldsymbol{\xi}^s) - \xi_s\right)$ in both the objective and constraints, thus eliminating the need to further approximate $I_t(\cdot)$ using decision rules.

For our computational experiments we randomly generated 30 instances of Problem (5.2). The parameters are randomly chosen using a uniform distribution from the following sets: Advanced and instant ordering costs are chosen from $c_z \in [0, 5]$ and $c_y \in [0, 10]$, respectively, such that $c_z < c_y$, and holding costs are elements of $c_h \in [0, 5]$. The bounds for each random parameter are chosen from $l_i \in [0, 5]$ and $u_i \in [10, 15]$ for $i = 2, \ldots, T$. The cumulative ordering budget equals to $\overline{B}_{\text{tot},t} = 10(t-1)$, for $t = 2, \ldots, T$. We also assume that $q_z = q_y = 15/N$, and the initial inventory level equals to zero, i.e., $I_1 = 0$.

In the first test series, we compared the performance of the binary decision rules versus the non-adaptive binary decisions for the randomly generated instances. We consider binary decision rules for $y_{n,t}(\cdot)$ that have $r \in \{1, 5, 9\}$ breakpoints in each direction ξ_i, and planning horizons $T \in \{10, 15, 20\}$.

Moreover, we consider three different termination criteria for the mixed-integer linear optimization problems: (i) global optimality, (ii) 1% optimality and (iii) 5% optimality. All non-adaptive problems were solved to global optimality. Table 2 presents the results from the case where $N = 2$, in terms of the average improvement of the binary decision rules versus the static binary decisions. The improvement is defined as quantity $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$, where Non-Adapt. and Adapt. are the objective values of the non-adaptive and adaptive binary decision rule problems, respectively. The first observation is the significant improvement over the non-adaptive decisions, and

Table 3 Comparison between the binary decision rules discussed in Bertsimas and Georghiou [10] (Design), given by (1.2) with $P = 1$, and the proposed decision rules (Adaptive)

	Average performance (%)				Solution time (s)				
	$r = 0$	$r = 1$	$r = 5$	$r = 9$	Design	$r = 0$	$r = 1$	$r = 5$	$r = 9$
Binary decision rule structure (1.2) v.s. (2.4), $N = 2$									
$T = 2$	19	3	2	1	<1	<1	<1	<1	<1
$T = 4$	29	16	12	10	3	<1	<1	<1	<1
$T = 6$	52	34	27	26	607	<1	<1	<1	<1
$T = 8$	60	40	29	28	1628	<1	<1	1	1
$T = 10$	47	38	31	30	16,613	<1	<1	1	5

The performance is defined as quantity $\frac{(\text{Adapt.}-\text{Design})}{\text{Design}}$. The proposed decision rule problems have great scalability properties but do not perform as well compared to binary decision rules of type (1.2)

the improvement in solution quality as the complexity of the binary decision rules increase. As expected, solving the problems to global optimality achieves a slightly better objective value compared to the relaxed termination criteria cases. Nevertheless, the computational burden is substantially larger, rendering the decision rules impractical for $T > 20$ and $r > 9$. Using the problem instances of this test series, we also compared the binary decision rule structure given in (2.4a) and (5.1), i.e., the decision variable in the mixed-integer linear optimization problems were $Y \in \mathbb{Z}^{q\times g}$ and $Y \in \{-1, 0, 1\}^{q\times g}$, respectively. In both cases, the optimal solution was identical, but the solution time for the case $Y \in \mathbb{Z}^{q\times g}$ was substantially longer, rendering the decision rules impractical for $T > 15$ and $r > 5$.

In our second test series, we investigate the relative performance between the proposed decision rules and the binary decision rules discussed in Bertsimas and Georghiou [10]. In particular, we used decision rules of type (1.2) with $P = 1$, and the corresponding decision rule problem was solved using the semi-infinite procedure defined by Algorithm 2 with parameters $\delta = 0.01$, $\beta = 10^{-4}$, $\widehat{\epsilon} = 0.01$ and $\widehat{\beta} = 10^{-4}$, see [10, Section 2.3] for more details. The termination criteria for all mixed-integer linear optimization problems was fixed to 5%. The results are presented Table 3, and constructed using the randomly generated instances for the cases where $N = 2$, $T \in \{2, \ldots, 10\}$ and $r \in \{0, 1, 5, 9\}$. The performance is defined as quantity $\frac{(\text{Adapt.}-\text{Design})}{\text{Design}}$, where Adapt. and Design are the objective values of the proposed binary decision rule structure and the decision rule structure given by (1.2), respectively. As expected, structure (1.2) is much more flexible compared to the proposed structure, producing high quality designs. This is the case as the discontinuity of the binary decision rule is decided endogenously by the optimization problem. However, this flexibility comes at a high computational cost, rendering decision rules (1.2) impractical for problem instances with $T > 10$. On the other hand, the proposed decision rules are highly scalable, with all instances being solved in under 5 S.

The quality of the proposed decision rules can be further improved by enriching the choice of $G(\cdot)$. Using definition (2.35), one can introduce additional piecewise constant components along directions $\boldsymbol{\alpha}_i \in \mathbb{R}^k$ that are not aligned with the coordinate axis of $\boldsymbol{\xi}$.

Table 4 Improvement in the proposed design of decision rules by using the flexible design of (1.2)

$T = 2$		$T = 4$	
Enriching the decision rule structure			
"Enriched"	"Simple"	"Enriched"	"Simple"
1.4%	7.1%	9.8%	14.5%

If the choice of $\boldsymbol{\alpha}_i$ cannot be motivated by the structure of the problem, then one can envision a hybrid method that first solves a variant of Problem (5.2) with a short horizon using the binary decision rules in [10], and extracting $\boldsymbol{\alpha}_i$ by analyzing the piecewise structure of the decision rules. Subsequently, Problem (5.2) with a longer horizon is solved using the enriched basis functions. To demonstrate this heuristic on a small scale example, we solve instances of Problem (5.2) with $N = 1$, $q_z = 7$, $q_y = 10$ and $T = 3$ using the binary decision rules given in (1.2) with $P = 1$. For each instance, we denote the optimal second and third stage decisions by

$$y_2^*(\xi_2) = \begin{cases} 1, & \alpha_{2,2}^* \xi_2 \geq \beta_2^*, \\ 0, & \text{otherwise}, \end{cases}$$
$$y_3^*(\xi_2, \xi_3) = \begin{cases} 1, & \alpha_{3,2}^* \xi_2 + \alpha_{3,3}^* \xi_3 \geq \beta_3^*, \\ 0, & \text{otherwise}, \end{cases}$$

and use the optimal $(\boldsymbol{\alpha}_2^*, \beta_2^*)$ and $(\boldsymbol{\alpha}_3^*, \beta_3^*)$ to create additional basis functions $G(\cdot)$ as described in equations (2.34) and (2.35). In addition to the single break point defined by the values β_2^* and β_3^*, 5 uniformly placed breakpoints are added along the directions $\boldsymbol{\alpha}_2^*$ and $\boldsymbol{\alpha}_3^*$ enriching further the basis $G(\boldsymbol{\xi})$. We then proceed to resolve the same instances with $T = 3$ and $T = 4$ with the proposed "enriched" decision rules and the "simpler" decision rules used in the previous experiment with $r = 10$. For problem instances with $T = 4$, we use the same data as the corresponding instance of $T = 3$, plus additional data to form the fourth stage. We compare the performance of the two approximations to the decision rule structure given by (1.2) using again the metric $\frac{(\text{Adapt.}-\text{Design})}{\text{Design}}$. The average results over 30 instances are presented in Table 4. We see that for $T = 3$, the performance of the "enriched" structure almost matches the performance of the flexible decision rules (1.2), while for $T = 4$, since we have only utilized the optimal policy from the second and third stage, the solution deteriorates but still performs better compared to the "simpler" decision rules. We note that for $T = 3$, the "enriched" structure does not exactly match the performance of the decision rule structure given by (1.2), since for the latter approximation the constraints are satisfied with high probability $(1 - \delta)$, while for the proposed decision rules the constraints are robustly satisfied. Nevertheless, we can still utilize the additional information and improve upon the "simpler" decision rule structure. Of course, the knowledge of finding good choices of $\boldsymbol{\alpha}_i$ comes with the price of solving computationally expensive problems and can restrict the applicability of this heuristic to small problem instances.

In our fourth test series, we investigate the scalability properties of the proposed binary decision rules and their relative performance. In this test series, the termination criterion of the mixed-integer linear optimization problems was fixed to 5%. The results

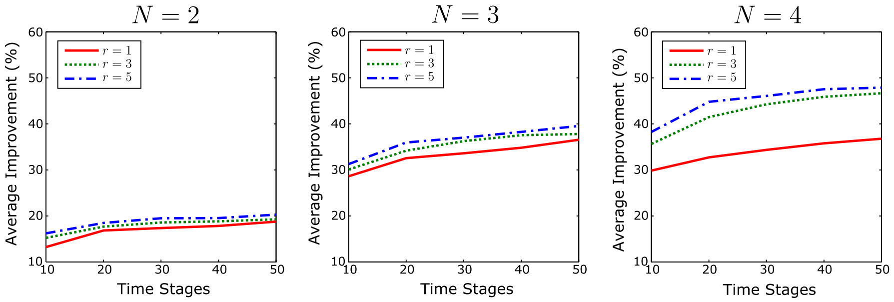

Fig. 5 Comparison of average performance of adaptive versus static binary decision rules for $N \in \{2, 3, 4\}$ and $r \in \{1, 3, 5\}$ using the performance measure $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$

Table 5 Average solution time for $N \in \{2, 3, 4\}$ and $r \in \{1, 3, 5\}$

	$N = 2$ (s)			$N = 3$ (s)			$N = 4$ (s)		
	$r = 1$	$r = 3$	$r = 5$	$r = 1$	$r = 3$	$r = 5$	$r = 1$	$r = 3$	$r = 5$
Average solution time									
$T = 10$	<1	<1	1	<1	<1	2	<1	<1	4
$T = 20$	<1	1	9	<1	2	48	<1	5	57
$T = 30$	<1	5	63	<1	7	224	<1	31	380
$T = 40$	<1	9	114	1	12	453	1	95	4615
$T = 50$	1	12	225	1	36	1951	1	357	19,298

are presented in Fig. 5 and Table 5, and constructed using the randomly generated instances for the cases where $N \in \{2, 3, 4\}$, $T \in \{10, \ldots, 50\}$ and $r \in \{1, 3, 5\}$. As expected, solution quality increases as the decision rules become more flexible. It is interesting to see that these improvements saturate for $r > 3$, getting diminishing returns for the computational effort required.

5.2 Multistage knapsack problem

In this case study, we consider a multistage knapsack problem, which can be formulated as an instance of the multistage adaptive optimization problem with random recourse and binary recourse decisions. In the classic one-stage knapsack problem, the decision maker is given N items, each weighing w_n, $n = 1, \ldots, N$, and a knapsack that can carry a maximum weight $\overline{w}$. The decision maker wants to maximize the number of items that fit into the knapsack, by choosing binary decisions $y_n \in \{0, 1\}$ for each item n, while keeping the total weight of the items in the knapsack $\sum_{n=1}^{N} w_n y_n$ less than $\overline{w}$. In the multistage variant of the problem, at the beginning of each time period $t \in \mathcal{T} := \{1, \ldots, T\}$, the decision maker is given N items, each weighing $w_{n,t}(\xi_t)$, $n = 1, \ldots, N$, and an empty knapsack with capacity $\overline{w}$. The weights $w_{n,t}(\xi_t)$ are assume to be linear functions of the random parameter ξ_t that realizes at stage t. Upon observing the weight of each item, the decision maker then decides the subset of the N items that can fit in the knapsack. The items left behind can be taken at later stages. We denote by $y_{n,t,s}(\cdot) \in \mathcal{B}_{k^s,1}$ the binary decision corresponding to item with weight $w_{n,t}(\xi_t)$, taking value 1 if the item is fitted into the knapsack at stage s, $s \geq t$, and zero otherwise. The decision maker wishes to maximize the expected number of items that can fit into the knapsacks over the planning horizon T by determining the collection of items $y_{n,t,s}(\cdot)$. The problem can be formulated as the following multistage optimization problem with random recourse.

$$\begin{array}{ll} \text{maximize} & \mathbb{E}_{\boldsymbol{\xi}}\left(\sum_{n=1}^{N}\sum_{t\in\mathcal{T}}\sum_{s=t}^{T} y_{n,t,s}(\boldsymbol{\xi}^s)\right) \\ \text{subject to} & \left.\begin{array}{l} y_{n,t,s}(\cdot) \in \mathcal{B}_{k^s,1}, \quad s = 1\ldots,t,\ n = 1\ldots,N, \\ \sum_{s=1}^{t}\sum_{n=1}^{N} w_{n,s}(\xi_s)\, y_{n,s,t}(\boldsymbol{\xi}^t) \leq \overline{w}, \\ \sum_{s=t}^{T} y_{n,t,s}(\boldsymbol{\xi}^s) \leq 1, \quad n = 1\ldots,N, \end{array}\right\} \quad \forall t \in \mathcal{T},\ \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{5.3}$$

Here, the second set of constraints ensure that at each stage t, the total weight of the knapsack is below its capacity $\overline{w}$, while the last set of constraints ensure that an item can only be carried in the knapsack only once over the time horizon. The random vector $\boldsymbol{\xi}$ follows a beta distribution, Beta(α, β), and the uncertainty set Ξ is given by,

$$\Xi := \left\{\boldsymbol{\xi} \in \mathbb{R}^T : \xi_1 = 1,\ 0 \leq \boldsymbol{\xi} \leq \boldsymbol{e}\right\}.$$

For our computational experiments we randomly generated 30 instances of Problem (5.3). The parameters are randomly chosen using a uniform distribution from the following sets: The weight of item n at stage t is given by $w_{n,t}(\xi_t) = \widehat{w}_{n,t}\xi_t$ where $\widehat{w}_{n,t} \in [0.5, 1]$, and the parameters of the beta distribution are chosen from $\alpha \in (0, 5]$ and $\beta \in (0, 5]$. The capacity of the knapsack is set to $\overline{w} = N/2$.

In the first test series, we compared the performance of the binary decision rules versus the non-adaptive binary decisions for the randomly generated instances. We consider binary decision rules that have $r = \{1, 2, 3\}$ breakpoints in each direction ξ_i, for problem instances with $N \in \{2, 3, 4\}$ and planning horizons $T \in \{3, \ldots, 15\}$. The termination criterion of the mixed-integer linear optimization problems was fixed to 5%. Figure 6 depicts the average improvement in objective value of the binary decisions rules versus the static binary decisions. As before, the improvement in objective value is defined as quantity $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$. The results show a substantial improvement in solution quality for all problem instances $N \in \{2, 3, 4\}$, with the most flexible decision rule structure achieving the best results. For the time horizon and complexity of the binary decision rules considered, all problem instances were solved within 20 minutes. The solution time for instances with $T > 15$ and $r > 3$ was substantially longer.

In the second test series, we improve the scalability of the solution method by restricting the information available to the binary decision rules. To this end, we define

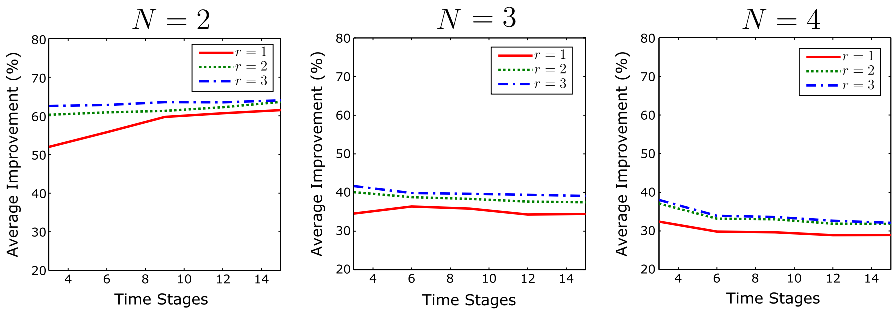

Fig. 6 Comparison of average performance of adaptive versus static binary decision rules for $N \in \{2, 3, 4\}$ and $r \in \{1, 2, 3\}$, using the performance measure $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$. All problems were solved within 20 min

Table 6 Comparison of average performance and average computational times for binary decision rules with full and partial information structure, using the performance measure $\frac{\text{(Non-Adapt.}-\text{Adapt.)}}{\text{Non-Adapt.}}$

T	Full information (%)			Partial information (%)		
	$r = 1$	$r = 2$	$r = 3$	$r = 1$	$r = 2$	$r = 3$
Average objective value improvements, $N = 4$						
$T = 5$	32.5	34.8	35.7	30.5	32.5	34.2
$T = 10$	28.3	32.0	32.1	27.9	30.3	31.5
$T = 15$	28.9	31.5	32.6	26.9	30.1	31.2
$T = 20$	28.3	n/a	n/a	27.6	30.9	31.7
$T = 25$	26.8	n/a	n/a	26.0	31.0	31.9
	Full information (s)			Partial information (s)		
Average solution time, $N = 4$						
$T = 5$	<1	1	1	<1	<1	1
$T = 10$	4	7	25	2	4	6
$T = 15$	27	52	364	13	42	28
$T = 20$	76	n/a	n/a	34	169	431
$T = 25$	396	n/a	n/a	162	449	11,349

$\boldsymbol{\eta}_{s,t} = (\xi_s, \xi_t)^\top$, and we restrict the recourse decisions $y_{n,t,s}(\boldsymbol{\xi}^s)$ to have structure $y_{n,t,s}(\boldsymbol{\eta}_{s,t})$ for all $n = 1, \ldots, N$, $s = t, \ldots, T$ and $t = 1 \ldots, T$. The number of integer decision variables needed to approximate $y_{n,t,s}(\boldsymbol{\xi}^s)$ using the binary decision rule structure (5.1) and (2.22) is $s(r+1)$, while for $y_{n,t,s}(\boldsymbol{\eta}_{s,t})$ the number of integer decision variables required is $2(r+1)$. We will refer to $\boldsymbol{\eta}_{s,t} = (\xi_s, \xi_t)^\top$ as the partial information available to the decision rule at stage s. Table 6 compares the solution quality and computation time, of the full information and partial information structures for problem instances with $N = 4$, $r \in \{1, 2, 3\}$ and $T \in \{5, \ldots, 25\}$. The solution quality of the full information structure is slightly better, but the solution method suffers with respect to scalability. On the other hand, using the partial information structure we can solve much bigger problem instances within very reasonable amount of time. We note that the problem instance with $N = 4$ and $T = 25$ involves a total of 1300 binary decision rules.

6 Conclusions

In this paper, we present linearly parameterised binary decision rule structures that can be used in conjunction with real-valued decision rules appearing in the literature, for solving multistage adaptive mixed-integer optimization problems. We provide a systematic way to reformulate the binary decision rule problem into a finite dimensional mixed-integer linear optimization problem, and we identify instances where the size of this problem grows polynomially with respect to the input data. The theory presented covers both fixed-recourse and random-recourse problems.

Particular emphasis of the numerical section is put on the comparison between the proposed decision rule structures and the decision rules presented in [10]. We show that decision rules parameterized as in (1.2) can provide better policy designs at the expense of scalability. This is the case as both the structure and shape of the decision rule is decided endogenously through the solution of a sequence of mixed-integer linear optimization problems. This necessitates the repeated use of non-convex optimization algorithms in order to achieve robust feasibility, reducing the scope of these decision rules to relatively small problem instances (horizons ≤ 10). In contrast, the proposed decision rule structures are highly scalable (horizons ≤ 50, $\sim$5 min), since the structure of the decision rule is dictated by the a priori choice of the non-linear operator $G(\cdot)$, and only the shape of the decision rule is decided through the solution of a single mixed-integer linear optimization problem. We show that the proposed decision rules can provide significant improvements compared to non-adaptive policies, making the method particularly attractive for practical, large scale problems.

Acknowledgements We like to thank Dr. Wolfram Wiesemann for very helpful discussions, and the Associate Editor and referees of the paper for their suggestions which improved the paper significantly.

7 Appendix: Technical proofs

In this section, we provide the proof for Theorem 1. To do this, we need the following auxiliary results given in Lemmas 1, 2 and 3. Lemma 1 gives the complexity of the epsilon integer feasibility problem, which will be used in Lemmas 2 and 3 to prove that checking feasibility of a semi-infinite constraint involving indicator functions is NP-hard.

Lemma 1 (Epsilon integer feasibility problem) *The following decision problem is* NP-*hard:*

$$\begin{array}{ll}\textsc{Instance.} & W \in \mathbb{R}^{l\times k},\ \boldsymbol{h} \in \mathbb{R}^l \text{ and } N \in \mathbb{Z}_+ \text{ with } N < k \text{ and}\\ & 0 < \epsilon < \min\left\{\min_{i\in\{1,\ldots,l\}}\{(\sum_{j=1}^k |W_{i,j}|)^{-1}\}, \frac{1}{k}, \frac{1}{2}\right\} \text{ that satisfy:}\\ & (A1)\ \text{The set } \Xi = \{\boldsymbol{\xi} \in [0,1]^k : W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N\} \text{ spans } \mathbb{R}^k;\\ & (A2)\ \text{For each } i = 1,\ldots,k, \text{ the set } \{\xi_i \in [0,1] : W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N\} \supseteq [1-\epsilon, 1].\\ \textsc{Question.} & \text{Is there } \boldsymbol{\xi} \in \{[0,\epsilon],[1-\epsilon,1]\}^k \text{ such that } W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N?\end{array} \tag{7.1}$$

The condition $\boldsymbol{\xi} \in \{[0,\epsilon],[1-\epsilon,1]\}^k$ implies that each component of $\boldsymbol{\xi}$ can take values either in the set $[0,\epsilon]$ or in the set $[1-\epsilon,1]$. Note that $(A1)$ is a standard assumption, while $(A2)$ is non-restrictive since if the projection of a component of $\boldsymbol{\xi}$ is strictly less than $1-\epsilon$, (or strictly greater than ϵ) then that component of $\boldsymbol{\xi}$ can be a priori fixed to an element of $[0, \epsilon]$ (or $[1-\epsilon]$), or in the case where $\epsilon < \xi_i < 1-\epsilon$ the dimension of Ξ can be reduced without changing the problem structure. Assumptions $(A1)$ and $(A2)$ are not necessary for the proofs of Lemmas 1, 2 and 3, but will be used in the proof of Theorem 1.

Proof The proof is a slight variant of [25, Lemma 2] and references within. □

In particular, the work of Hanasusanto et al. [25] shows that the epsilon integer feasibility problem is in fact equivalent to the integer feasibility problem which is known to be NP-hard, see [28]. More precisely, they prove that the decision problem (7.1) has an affirmative answer if and only if there exists a vector $\boldsymbol{\chi} \in \{0, 1\}^k$ which gives an affirmative answer to the corresponding integer feasibility problem, with $\chi_i = 1$ if $\xi_i > 1-\epsilon$, and $\chi_i = 0$ otherwise.

Lemma 2 *The following decision problem is* NP*-hard:*

INSTANCE. $W \in \mathbb{R}^{l\times k}$, $\boldsymbol{h} \in \mathbb{R}^l$ *and* $N \in \mathbb{Z}_+$ *with* $N < k$ *and*

$$0 < \epsilon < \min\left\{\min_{i\in\{1,\ldots,l\}}\{(\sum_{j=1}^{k}|W_{i,j}|)^{-1}\}, \tfrac{1}{k}, \tfrac{1}{2}\right\} \text{ that satisfy:}$$

$(A1)$ *The set* $\Xi = \{\boldsymbol{\xi} \in [0,1]^k \,:\, W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N\}$ *spans* $\mathbb{R}^k$;

$(A2)$ *For each* $i = 1, \ldots, k$, *the set* $\{\xi_i \in [0,1] \,:\, W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N\} \supseteq [1-\epsilon, 1]$.

QUESTION. *Is there* $\boldsymbol{\xi} \in [0,1]^k$ *such that* $W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N,\ \sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) = N$?

$$(7.2)$$

Proof We will prove the assertion in two steps. In the first step, assume that the decision problem (7.1) has an affirmative answer, i.e., there exists a $\boldsymbol{\xi} \in \{[0,\epsilon], [1-\epsilon, 1]\}^k$ that satisfies (7.1). Then, by [25, Lemma 2] there exists $\boldsymbol{\chi} \in \{0,1\}^k$ with $\chi_i = 1$ if $\xi_i > 1-\epsilon$, and $\chi_i = 0$ otherwise, that satisfies constraints $W\boldsymbol{\chi} \geq \boldsymbol{h},\ \sum_{i=1}^k \chi_i = N$. It is easy to see that this vector $\boldsymbol{\chi}$ will also satisfy $\sum_{i=1}^k \mathbf{1}\left(\chi_i \geq 1 - \frac{\epsilon}{k}\right) = N$ due to the fact that $N < k$, and therefore decision problem (7.2) will have an affirmative answer. For the second step, assume that the decision problem (7.2) has an affirmative answer. We now prove by contradiction that a vector $\boldsymbol{\xi} \in [0,1]^k$ satisfying (7.2) must also satisfy (7.1). Assume that there exists $\boldsymbol{\xi}$ with $\epsilon < \xi_1 < 1-\epsilon$ that satisfies $W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N,\ \sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1-\frac{\epsilon}{k}\right) = N$. Since $\epsilon < \xi_1 < 1-\epsilon$, then $\sum_{i=2}^k \xi_i < N - \epsilon$ and as a result $\sum_{i=2}^k \mathbf{1}(\xi_i \geq 1-\epsilon/k) < N$, which contradicts the assumption that $\sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1-\frac{\epsilon}{k}\right) = N$. The latter follows since if distributed evenly, the value each ξ_i can take is $\xi_i = 1 - \frac{\epsilon}{N}$, for $i = 2, \ldots, N+1$, and since $N < k$ it implies that $\frac{1}{N}(N-\epsilon) = 1 - \frac{\epsilon}{N} < 1 - \frac{\epsilon}{k}$.

Therefore, we conclude that $\boldsymbol{\xi} \in \{[0,\epsilon], [1-\epsilon, 1]\}^k$. Together, the two steps allow us to conclude that (7.2) is NP-hard. □

The following lemma provides the key ingredient for proving Theorem 1.

Lemma 3 *The following decision problem is* NP-*hard:*

INSTANCE. *A convex polytope* $\Xi \subset \mathbb{R}^k$ *with* $W \in \mathbb{R}^{l \times k}$, $\boldsymbol{h} \in \mathbb{R}^l$, $N \in \mathbb{Z}_+$ *with* $N < k$, *and*

$$0 < \epsilon < \min \left\{ \min_{i \in \{1,\ldots,l\}} \{(\sum_{j=1}^{k} |W_{i,j}|)^{-1}\}, \frac{1}{k}, \frac{1}{2} \right\} \text{ that satisfy:}$$

$(A1)$ *The set* $\Xi = \{\boldsymbol{\xi} \in [0,1]^k \,:\, W\boldsymbol{\xi} \geq \boldsymbol{h},\, \sum_{i=1}^k \xi_i = N\}$ *spans* $\mathbb{R}^k$;

$(A2)$ *For each* $i = 1, \ldots, k$, *the set* $\{\xi_i \in [0,1] \,:\, W\boldsymbol{\xi} \geq \boldsymbol{h},\, \sum_{i=1}^k \xi_i = N\} \supseteq [1-\epsilon, 1]$.

QUESTION. *Do all* $\boldsymbol{\xi} \in \Xi$ *satisfy* $\sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) \leq N - 1$?

(7.3)

Proof Notice that the decision problem (7.3) is the negation of the decision problem (7.2). In other words, the decision problem (7.3) evaluates to true if there is no $\boldsymbol{\xi}$ such that

$$\boldsymbol{\xi} \in [0,1]^k \text{ such that } W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N,\ \sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) = N,$$

which corresponds to checking if decision problem (7.2) holds. Notice that there is no $\boldsymbol{\xi}$ such that $\sum_{i=1}^k \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) > N$. This is the case since $\sum_{i=1}^k \xi_i = N$ and $N < k$, implying that $N - N(1 - \epsilon/k) = N\epsilon/k < 1$. The reverse statement holds in a similar manner. □

We now have all the ingredients to prove Theorem 1.

Proof of Theorem 1 Let Ξ be a convex polytope given by

$$\Xi := \left\{ \boldsymbol{\xi} \in [0,1]^k \,:\, \xi_1 = 1,\ W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=2}^k \xi_i = N \right\}, \tag{7.4}$$

with $N \in \mathbb{Z}_+$, $N < k - 1$, and the choice of W and $\boldsymbol{h}$ are such that Ξ spans $\mathbb{R}^k$ and the projection of Ξ on each component ξ_i is a superset of $[1-\epsilon, 1]$ for $i = 2, \ldots, k$, with fixed $\epsilon \in (0, \min\{\min_{i \in \{1,\ldots,l\}}\{(\sum_{j=1}^k |W_{i,j}|)^{-1}\}, \frac{1}{k}, \frac{1}{2}\})$. In other words, the set Ξ satisfies $(A1)$ and $(A2)$ in the decision problem (7.3).

Given the above instance, we can construct the projection of Ξ on each component ξ_i, i.e, for each $i = 2, \ldots, k$, there exists $l_i \in [0, 1-\epsilon]$ such that $\{\xi_i \in [0,1] \,:\, W\boldsymbol{\xi} \geq \boldsymbol{h},\ \sum_{i=1}^k \xi_i = N\} = [l_i, 1]$. Now, let $\boldsymbol{\alpha}_i = \boldsymbol{e}_i$ and $\beta_i = 1 - \frac{\epsilon}{k}$, for $i = 2, \ldots, k$ in the description of $G(\cdot)$ in (2.4b). By construction, the components of $G(\boldsymbol{\xi})$ are linearly independent for all $\boldsymbol{\xi} \in \Xi$, i.e,

$$\boldsymbol{v}^\top G(\boldsymbol{\xi}) = 0, \quad \forall \boldsymbol{\xi} \in \Xi \quad \Longrightarrow \quad \boldsymbol{v} = 0.$$

This is the case since each component of $G(\boldsymbol{\xi})$ is non-constant on disjoint subsets of $\mathbb{R}^k$ and each of these subsets has a non-empty intersection with Ξ, with only $G_1(\boldsymbol{\xi}) = 1$ being constant for all $\boldsymbol{\xi} \in \Xi$. This is satisfied by the construction of Ξ which is a convex set, spans $\mathbb{R}^k$ and the projection of Ξ on each component ξ_i is $[l_i, 1]$ for $i = 2, \ldots, k$, with $l_i \leq 1 - \epsilon < 1 - \frac{\epsilon}{k} < 1$.

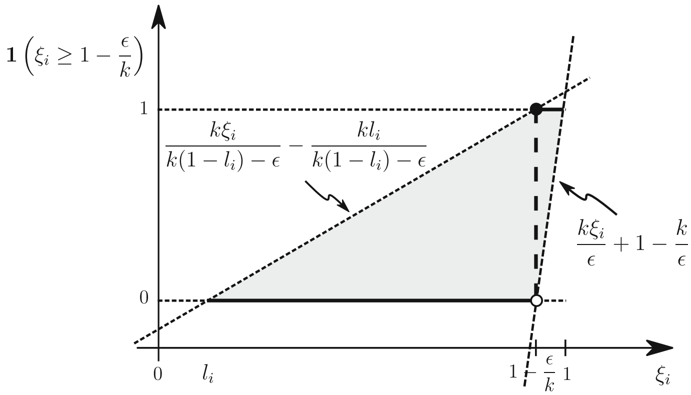

Fig. 7 Visualization of the function $G_i(\boldsymbol{\xi}) = \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right)$ defined by the choice of $\boldsymbol{\alpha}_i = \boldsymbol{e}_i$ and $\beta_i = 1 - \frac{\epsilon}{k}$ in the description of $G(\cdot)$ in (2.4b), together with the visualization of the second and third constraints of Problem (7.5) which create the convex hull of $\mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right)$ on $[l_i, 1]$

We define the following feasibility problem:

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \left.\begin{array}{ll} \boldsymbol{y}_i \in \mathbb{R}^k, & \forall i = 2, \ldots, k, \\ 0 \leq \boldsymbol{y}_i^\top G(\boldsymbol{\xi}) \leq 1, & \forall i = 2, \ldots, k, \\ \frac{k\xi_i}{\epsilon} + 1 - \frac{k}{\epsilon} \leq \boldsymbol{y}_i^\top G(\boldsymbol{\xi}) \leq \frac{k\xi_i}{k(1-l_i)-\epsilon} - \frac{kl_i}{k(1-l_i)-\epsilon}, & \forall i = 2, \ldots, k, \\ \displaystyle\sum_{i=2}^{k} \boldsymbol{y}_i^\top G(\boldsymbol{\xi}) \leq N - 1, & \end{array}\right\} \forall \boldsymbol{\xi} \in \Xi. \end{array} \tag{7.5}$$

It is easy to see that for each $i = 2, \ldots, k$, a feasible vector $\boldsymbol{y}_i$ in the second and third constraints satisfies $\boldsymbol{e}_i \leq \boldsymbol{y}_i \leq \boldsymbol{e}_i + \boldsymbol{v}_i$, for some $\boldsymbol{v}_i \in \mathbb{R}^k$ with $v_{i,i} = 0$. In particular, the solution $\boldsymbol{y}_i = \boldsymbol{e}_i$ for each $i = 2, \ldots, k$ is feasible in the second and third constraint. This is the case since for each i, the second and third constraint create the convex hull of function $\mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right)$ on $\xi_i \in [l_i, 1]$, see Fig. 7.

We will also show that for the smallest N for which the last constraint is feasible, the only feasible solution is $\boldsymbol{y}_i = \boldsymbol{e}_i$ for all $i = 2, \ldots, k$. We will show this by contradiction. Assume that for each $i = 2, \ldots, k$, there exist non-zero vectors $\boldsymbol{v}_i \in \mathbb{R}^k$ such that $\boldsymbol{y}_i = \boldsymbol{e}_i + \boldsymbol{v}_i$ is feasible. Since we have the smallest N, with $N < k - 1$, constraint

$$\sum_{i=2}^{k} (\boldsymbol{e}_i + \boldsymbol{v}_i)^\top G(\boldsymbol{\xi}) \leq N - 1, \quad \forall \xi \in \Xi,$$

implies that $\boldsymbol{v}_i^\top G(\boldsymbol{\xi}) = 0$ for all $\boldsymbol{\xi} \in \Xi$ and $i = 2, \ldots, k$. This contradicts the fact that the components of the components of $G(\boldsymbol{\xi})$ are linearly independent for all $\boldsymbol{\xi} \in \Xi$.

Therefore, we conclude that if Problem (7.5) is feasible, then $\boldsymbol{y}_i = \boldsymbol{e}_i$ for all $i = 2, \ldots, k$ is a feasible solution that satisfies the last constraint, and in addition, if constraint $\sum_{i=2}^{k} \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) \leq N - 1$ is feasible, then Problem (7.5) is feasible, i.e.,

$$\sum_{i=2}^{k} \mathbf{1}\left(\xi_i \geq 1 - \frac{\epsilon}{k}\right) \leq N - 1, \quad \forall \boldsymbol{\xi} \in \Xi \quad \Longleftrightarrow \quad \text{Problem (7.5) is feasible.}$$

Hence, Lemma 3 implies that checking the feasibility of Problem (7.5) is NP-hard. Since the admissible feasible solutions are integers, i.e., $\boldsymbol{y_i} \in \mathbb{Z}^k$ for all $i = 2, \dots, k$, Problem (7.5) can be reduced to an instance of Problem (2.10). Therefore,

$$\begin{aligned} \sum_{i=1}^{k} \mathbf{1}(\xi_i \geq 1) \leq N - 1, \quad \forall \boldsymbol{\xi} \in \Xi \quad &\Longleftrightarrow \quad \text{Problem (2.10) is feasible,} \\ &\underset{\text{Propositions 1 \& 2}}{\Longleftrightarrow} \quad \text{Problem (2.14) is feasible.} \end{aligned}$$

Lemma 3 implies that Problems (2.10) and (2.14) are NP-hard, even when $Y \in \mathbb{Z}^{k \times g}$ is relaxed to $Y \in \mathbb{R}^{k \times g}$. □

References

1. Anderson, B.D.O., Moore, J.B.: Optimal Control: Linear Quadratic Methods. Prentice Hall, Upper Saddle River (1990)
2. Bampou, D., Kuhn, D.: Scenario-free stochastic programming with polynomial decision rules. In: 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, pp. 7806–7812 (December 2011)
3. Ben-Ta, A., den Hertog, D.: Immunizing Conic Quadratic Optimization Problems Against Implementation Errors, vol. 2011-060. Tilburg University, Tilburg (2011)
4. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
5. Ben-Tal, A., Golany, B., Shtern, S.: Robust multi-echelon multi-period inventory control. Eur. J. Oper. Res. **199**(3), 922–935 (2009)
6. Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A.: Adjustable robust solutions of uncertain linear programs. Math. Program. **99**(2), 351–376 (2004)
7. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**(4), 769–805 (1998)
8. Bertsimas, D., Caramanis, C.: Adaptability via sampling. In: 46th IEEE Conference on Decision and Control, New Orleans, LA, USA, pp. 4717–4722 (December 2007)
9. Bertsimas, D., Caramanis, C.: Finite adaptability in multistage linear optimization. IEEE Trans. Autom. Control **55**(12), 2751–2766 (2010)
10. Bertsimas, D., Georghiou, A.: Design of near optimal decision rules in multistage adaptive mixed-integer optimization. Oper. Res. **63**(3), 610–627 (2015)
11. Bertsimas, D., Iancu, D.A., Parrilo, P.A.: Optimality of affine policies in multi-stage robust optimization. Math. Oper. Res. **35**(2), 363–394 (2010)
12. Bertsimas, D., Iancu, D.A., Parrilo, P.A.: A hierarchy of near-optimal policies for multistage adaptive optimization. IEEE Trans. Autom. Control **56**(12), 2809–2824 (2011)
13. Calafiore, G.C.: Multi-period portfolio optimization with linear control policies. Automatica **44**(10), 2463–2473 (2008)
14. Calafiore, G.C.: An affine control method for optimal dynamic asset allocation with transaction costs. SIAM J. Control Optim. **48**(4), 2254–2274 (2009)
15. Caramanis, C.: Adaptable Optimization: Theory and Algorithms. Ph.D. thesis, Massachusetts Institute of Technology (2006)
16. Chen, X., Sim, M., Sun, P., Zhang, J.: A linear decision-based approximation approach to stochastic programming. Oper. Res. **56**(2), 344–357 (2008)
17. Chen, X., Zhang, Y.: Uncertain linear programs: extended affinely adjustable robust counterparts. Oper. Res. **57**(6), 1469–1482 (2009)

18. Dyer, M., Stougie, L.: Computational complexity of stochastic programming problems. Math. Program. **106**(3), 423–432 (2006)
19. Garstka, S.J., Wets, R.J.-B.: On decision rules in stochastic programming. Math. Program. **7**(1), 117–143 (1974)
20. Georghiou, A., Wiesemann, W., Kuhn, D.: Generalized decision rule approximations for stochastic programming via liftings. Math. Program. **152**(1–2), 301–338 (2015)
21. Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Oper. Res. **58**(4), 902–917 (2010)
22. Goulart, P.J., Kerrigan, E.C., Maciejowski, J.M.: Optimization over state feedback policies for robust control with constraints. Automatica **42**(4), 523–533 (2006)
23. Gounaris, C.E., Wiesemann, W., Floudas, C.A.: The robust capacitated vehicle routing problem under demand uncertainty. Oper. Res. **61**(3), 677–693 (2013)
24. Hanasusanto, G.A., Kuhn, D., Wiesemann, W.: K-Adaptability in Two-Stage Robust Binary Programming. Oper Res. **63**(4), 877–891 (2015). doi:10.1287/opre.2015.1392
25. Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W.: Ambiguous joint chance constraints under mean and dispersion information. Submitted for publication, available via Optimization-Online (2016)
26. IBM ILOG CPLEX. http://www.ibm.com/software/integration/optimization/cplex-optimizer
27. Kuhn, D., Wiesemann, W., Georghiou, A.: Primal and dual linear decision rules in stochastic and robust optimization. Math. Program. **130**(1), 177–209 (2011)
28. Papadimitriou, C.H.: On the complexity of integer programming. J. Assoc. Comput. Mach. **28**(4), 765–768 (1981)
29. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Dordrecht (1997)
30. See, C.-T., Sim, M.: Robust approximation to multiperiod inventory management. Oper. Res. **58**(3), 583–594 (2010)
31. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. Appl. Optim. **99**(1), 111–146 (2005)
32. Skaf, J., Boyd, S.P.: Design of affine controllers via convex optimization. IEEE Trans. Autom. Control **55**(11), 2476–2487 (2010)