Multi-stage Adjustable Robust Mixed-Integer Optimization via Iterative Splitting of the Uncertainty set (Revision of CentER Discussion Paper 2014-056)

Postek, K.S.; den Hertog, D.

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Postek, K. S., & den Hertog, D. (2016). *Multi-stage Adjustable Robust Mixed-Integer Optimization via Iterative Splitting of the Uncertainty set (Revision of CentER Discussion Paper 2014-056)*. (CentER Discussion Paper; Vol. 2016-006). Operations research.

CentER

Discussion Paper

No. 2016-006

MULTI-STAGE ADJUSTABLE ROBUST MIXED-INTEGER OPTIMIZATION VIA ITERATIVE SPLITTING OF THE UNCERTAINTY SET

By

Krzysztof Postek, Dick den Hertog

1 February 2016

This is a revised version of CentER Discussion Paper No. 2014-056

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set

Krzysztof Postek, Dick den Hertog
CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, k.postek@tilburguniversity.edu, d.denhertog@tilburguniversity.edu

In this paper we propose a methodology for constructing decision rules for integer and continuous decision variables in multiperiod robust linear optimization problems. This type of problems finds application in, for example, inventory management, lot sizing, and manpower management. We show that by iteratively splitting the uncertainty set into subsets one can differentiate the later-period decisions based on the revealed uncertain parameters. At the same time, the problem's computational complexity stays at the same level as for the static robust problem. This holds also in the non-fixed recourse situation. In the fixed recourse situation our approach can be combined with linear decision rules for the continuous decision variables. We provide theoretical results how to split the uncertainty set by identifying sets of uncertain parameter scenarios to be divided for an improvement in the worst-case objective value. Based on this theory, we propose several splitting heuristics. Numerical examples entailing a capital budgeting and a lot sizing problem illustrate the advantages of the proposed approach.

Key words: adjustable, decision rules, integer, multi-stage, robust optimization

1. Introduction

Robust optimization (RO, see Ben-Tal et al. (2009)) has become one of the main approaches to optimization under uncertainty. One of its applications are multiperiod problems where, period after period, values of the uncertain parameters are revealed and new decisions are implemented. Adjustable Robust Optimization (ARO, see Ben-Tal et al. (2004)) addresses such problems by formulating the decision variables as functions of the revealed uncertain parameters. Ben-Tal et al. (2004) prove that without any functional restrictions on the form of adjustability, the resulting problem is NP-hard. For that reason, several functional forms of the decision rules have been proposed, with the most popular being the affinely adjustable decision rules. However, only for a limited class of problems do they yield problems that can be reformulated to a computationally tractable form (see Ben-Tal et al. (2009)). In particular, for problems without fixed recourse, where the later-period problem parameters depend also on the uncertain parameters from earlier periods, it is nontrivial to construct tractable decision rules. The difficulty level grows even more when the adjustable variables are binary or integer. Addressing this problem is the topic of our paper. We propose a simple and intuitive method to construct adjustable decision rules, applicable also to problems with integer adjustable variables and to problems without fixed recourse. For problems with fixed recourse our methodology can be combined with linear decision rules for the continuous decision variables.

The contribution of our paper is twofold. First, we propose a methodology of iterative splitting of the uncertainty set into subsets, for each of which a scalar later-period decision shall be determined. A given decision is implemented in the next period if the revealed uncertain parameter belongs to the corresponding subset. Using scalar decisions per subset ensures that the resulting problem has the same complexity as the static robust problem. This approach provides an upper bound on the optimal value of the adjustable robust problem. Next to that, we propose a method of obtaining lower bounds, being a generalization of the approach of Hadjiyiannis et al. (2011).

As a second contribution, we provide theoretical results supporting the decision of how to split the uncertainty set into smaller subsets for problems with continuous decision variables. The theory identifies sets of scenarios for the uncertain parameters that have to be divided. On the basis of these results, we propose set-splitting heuristics for problems including also integer decision variables. As a side result, we prove the reverse of the result of Gorissen et al. (2014). Namely, we show that the optimal KKT vector of the tractable robust counterpart of a linear robust problem, obtained using the results of Ben-Tal et al. (2014), yields an optimal solution to the optimistic dual (see Beck and Ben-Tal (2009)) of the original problem.

ARO was developed to (approximately) solve problems with continuous variables. Ben-Tal et al. (2004) introduce the concept of using affinely adjustable decision rules and show how to apply such rules to obtain (approximate) optimal solutions to multiperiod problems. Affinely adjustable decisions turn out to be very effective for the inventory management example, which shall be also visible in the results of our paper. Their approach has been later extended to other function classes by Chen et al. (2007), Chen and Zhang (2009), Ben-Tal et al. (2009) and Bertsimas et al. (2011b). Bertsimas et al. (2010) prove that for a specific class of multiperiod control problems the affinely adjustable decision rules result in optimal adjustable solution. Bertsimas and Goyal (2010) show that the static robust solutions perform well in Stochastic Programming problems. Bertsimas et al. (2014b) study cases where static decisions are worst-case optimal in two-period problems and give a tight approximation bound on the performance of static solutions, related to a measure of non-convexity of a transformation of the uncertainty set. Goyal and Lu (2014) study the performance of static solutions in problems with constraint-wise and column-wise uncertainty and provide theoretical bounds on the adaptivity gap between static and optimal adjustable solutions in such a setting.

Later, developments have been made allowing ARO to (approximately) solve problems involving adjustable integer variables. Bertsimas and Caramanis (2007) propose a sampling method for constructing adjustable robust decision rules ensuring, under certain conditions, that the robust constraints are satisfied with high probability. Bertsimas and Caramanis (2010) introduce the term of finite adaptability in two-period problems, with a fixed number of possible second-period decisions. They also show that finding the best values for these variables is NP-hard. In a later paper, Bertsimas et al. (2011a) characterize the geometric conditions for the uncertainty sets under which finite adaptability provides good approximations of the adjustable robust solutions.

Vayanos (2011) split the uncertainty set into hyper-rectangles, assigning to each of them the corresponding later-period adjustable linear and binary variables. Contrary to this, our method does not impose any geometrical form of the uncertainty subsets. Bertsimas and Georghiou (2015) propose to use piecewise linear decision rules, both for the continuous and the binary variables (for the binary variables, value 0 is implemented if the piecewise linear decision rule is positive). They use a cutting plane approach that gradually increases the fraction of the uncertainty set that the solution is robust to, reaching complete robustness when their approach terminates. In our approach, the decision rules proposed ensure full robustness after each of the so-called splitting rounds, and the more splitting rounds, the better the value of the objective function. In a recent paper, Bertsimas and Georghiou (2014a) propose a different type of decision rules for binary variables. Since the resulting problems are exponential in the size of the original formulation, authors propose their conservative approximations, giving a systematic tradeoff between computational tractability and level of conservatism. In our approach, instead of imposing a functional form of the decision rules, we focus on splitting the uncertainty set into subsets with different decisions. Also, we ensure robustness precisely against the specified uncertainty set and allow non-binary integer variables.

Hanasusanto et al. (2014) apply finite adaptability to two-period decision problems with binary variables. In this setting, the decision maker can choose out of K possible decisions in the second period when the uncertain parameter value is known. For each outcome of the uncertain parameter, one of the time-2 decisions must yield a feasible solution. The optimization variables are the here-and-now decisions taken at period 1, and the set of K decisions for period 2. The resulting problems can be transformed to MILP problems of size exponential in the number K of possible decisions (in case of uncertainty in both the objective function and the constraints - for problems with uncertainty only in the objective the reformulation is polynomial). They also study the approximation quality provided by such reformulations and complexity issues. Our approach applies to general multi-period problems and allows also explicitly non-binary integer variables.

We test our methodology on problem instances from Bertsimas and Georghiou (2015) and Hanasusanto et al. (2014). The experiments reveal that our methodology performs worse on problems with uncertainty only in the objective function and on small instances, where the ‘more exact’ approaches of other authors can be solved fast to optimality. However, as the problems grow in size, it is able to provide comparable or better results after a significantly shorter computation.

The idea of partitioning the support of random variables in order to improve approximations of the objective function has been subject of intensive study in Stochastic Programming (SP). There, the use of partitions is to apply some bounds on the expectation of a function of a random variable on a per-partition basis, obtaining tighter bounds in this way. Examples of such partitions are given in Birge and Wets (1986) and Frauendorfer and Kall (1988). In some cases, similarly to our methodology, these methods use dual information to decide about the positioning of the partitions (see Birge and Wets (1986)). For an overview of bounds and their partition-based refinements used in SP we refer the reader to Birge and Louveaux (2011). Despite these similarities with our approach, our method is different for its focus on the worst-case outcomes without assuming distributional information.

The composition of the remainder of the paper is as follows. Section 2 introduces the set-splitting methodology for the case of two-period problems with adjustable continuous variables. Section 3 extends the approach to multiperiod problems, and Section 4 extends the multiperiod case to problems with integer decision variables. Section 5 proposes heuristics to be used as a part of the method. Section 6 gives two numerical examples, showing that the methodology of our paper offers substantial gains in terms of the worst-case objective function improvement. Section 7 concludes and lists the potential directions for future research.

2. Two-period problems

For ease of exposition we first introduce our methodology on the case of two-period problems with continuous decision variables only. The extension to multi-period problems is given in Section 3, and the extension to problems with integer variables is given in Section 4.

2.1. Description

Consider the following two-period optimization problem:

$$\begin{aligned} \min_{\mathbf{x}_1,\mathbf{x}_2} \ & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2 \\ \text{s.t.} \ & \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\mathbf{x}_2 \leq \mathbf{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{aligned} \tag{1}$$

where $\mathbf{c}_1 \in \mathbb{R}^{d_1}, \mathbf{c}_2 \in \mathbb{R}^{d_2}$, $\mathbf{b} \in \mathbb{R}^m$ are fixed parameters, $\boldsymbol{\zeta} \in \mathbb{R}^L$ is the uncertain parameter and $\mathcal{Z} \subset \mathbb{R}^L$ is a compact and convex uncertainty set. Vector $\mathbf{x}_1 \in \mathbb{R}^{d_1}$ is the decision implemented at time 1 before the value of $\boldsymbol{\zeta}$ is known, and $\mathbf{x}_2 \in \mathbb{R}^{d_2}$ is the decision vector implemented at time 2, after the value of $\boldsymbol{\zeta}$ is known. It is assumed that the functions $\mathbf{A}_1 : \mathbb{R}^L \to \mathbb{R}^{m\times d_1}, \mathbf{A}_2 : \mathbb{R}^L \to \mathbb{R}^{m\times d_2}$ are linear. We refer to the rows of matrix $\mathbf{A}_1$ and $\mathbf{A}_2$ as $\mathbf{a}_{1,i}^T(\boldsymbol{\zeta})$ and $\mathbf{a}_{2,i}^T(\boldsymbol{\zeta})$ respectively, with $\mathbf{a}_{1,i}(\boldsymbol{\zeta}) = \mathbf{P}_{1,i}\boldsymbol{\zeta}$ and $\mathbf{a}_{2,i}(\boldsymbol{\zeta}) = \mathbf{P}_{2,i}\boldsymbol{\zeta}$, where $\mathbf{P}_{1,i} \in \mathbb{R}^{d_1\times L}, \mathbf{P}_{2,i} \in \mathbb{R}^{d_2\times L}$ (uncertain parameter can contain a single fixed component, which would result in the intercepts of the affine transformations $\mathbf{A}_1(\boldsymbol{\zeta}), \mathbf{A}_2(\boldsymbol{\zeta})$.

The static robust problem (1) where the decision vector $\mathbf{x}_2$ is independent from the value of $\boldsymbol{\zeta}$ makes no use of the fact that $\mathbf{x}_2$ can adjust to the revealed $\boldsymbol{\zeta}$. The adjustable version of problem (1) is:

$$\begin{aligned}\min_{\mathbf{x}_1,\mathbf{x}_2(\boldsymbol{\zeta}),z} \quad & z \\ \text{s.t.} \quad & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2(\boldsymbol{\zeta}) \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\mathbf{x}_2(\boldsymbol{\zeta}) \leq \mathbf{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}.\end{aligned} \tag{2}$$

Since this problem is NP-hard (see Ben-Tal et al. (2009)), the concept of linear decision rules has been proposed. Then, the time 2 decision vector is defined as $\mathbf{x}_2 = \mathbf{v} + \mathbf{V}\boldsymbol{\zeta}$, where $\mathbf{v} \in \mathbb{R}^{d_2}$, $\mathbf{V} \in \mathbb{R}^{d_2 \times L}$ (see Ben-Tal et al. (2009)) and the problem is:

$$\begin{aligned}\min_{\mathbf{x}_1,\mathbf{v},\mathbf{V}} \quad & z \\ \text{s.t.} \quad & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\left(\mathbf{v} + \mathbf{V}\boldsymbol{\zeta}\right) \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\left(v + V\boldsymbol{\zeta}\right) \leq \mathbf{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}.\end{aligned} \tag{3}$$

In the general case such constraints are quadratic in $\boldsymbol{\zeta}$, because of the term $\mathbf{A}_2(\boldsymbol{\zeta})\left(\mathbf{v} + \mathbf{V}\boldsymbol{\zeta}\right)$. Only for special cases the constraint system can be rewritten as a computationally tractable system of inequalities. Moreover, linear decision rules cannot be used if (part of) the decision vector $\mathbf{x}_2$ is required to be integer.

We propose a different approach. Before introducing it, we need to introduce the term of *splitting* a set. By *splitting* a set $\mathcal{Z}$ it is understood such a partition $\mathcal{Z} = \mathcal{Z}^+ \cup \mathcal{Z}^-$ that there exist $\boldsymbol{\zeta}^+ \in \mathcal{Z}^+$ and $\boldsymbol{\zeta}^- \in \mathcal{Z}^-$ such that:

$$\boldsymbol{\zeta}^+ \in \mathcal{Z}^+ \setminus \mathcal{Z}^-, \; \boldsymbol{\zeta}^- \in \mathcal{Z}^- \setminus \mathcal{Z}^+.$$

Our idea lies in splitting the set $\mathcal{Z}$ into a collection of subsets $\mathcal{Z}_{r,s}$ where $s \in \mathcal{N}_r$ and $\cup_{s \in \mathcal{N}_r} \mathcal{Z}_{r,s} = \mathcal{Z}$ (r denotes the index of the splitting round and s denotes the set index). For each $\mathcal{Z}_{r,s}$ a different, fixed time 2 decision shall be determined. We split the set $\mathcal{Z}$ in rounds into smaller and smaller subsets using hyperplanes. In this way, all the uncertainty subsets remain convex, which is a typical assumption for RO problems. The following example illustrates this idea.

Example 1. We split the uncertainty set $\mathcal{Z}$ with a hyperplane $\mathbf{g}^T\boldsymbol{\zeta} = h$ into the following two sets:

$$\mathcal{Z}_{1,1} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \mathbf{g}^T\boldsymbol{\zeta} \leq h\right\} \quad \text{and} \quad \mathcal{Z}_{1,2} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \mathbf{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

At time 2 the following decision is implemented:

$$\mathbf{x}_2 = \begin{cases} \mathbf{x}_2^{(1,1)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \\ \mathbf{x}_2^{(1,2)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,2} \\ \mathbf{x}_2^{(1,1)} \text{ or } \mathbf{x}_2^{(1,2)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \cap \mathcal{Z}_{1,2}. \end{cases}$$

The splitting is illustrated in Figure 1. Now, the following constraints have to be satisfied:

$$\begin{cases} \mathbf{A}_1\left(\boldsymbol{\zeta}\right)\mathbf{x}_1 + \mathbf{A}_2\left(\boldsymbol{\zeta}\right)\mathbf{x}_2^{(1,1)} \leq b, & \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \\ \mathbf{A}_1\left(\boldsymbol{\zeta}\right)\mathbf{x}_1 + \mathbf{A}_2\left(\boldsymbol{\zeta}\right)\mathbf{x}_2^{(1,2)} \leq b, & \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,2}. \end{cases}$$

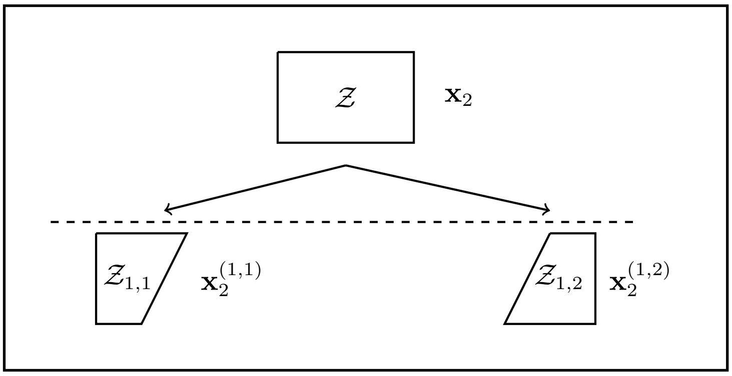

Figure 1 Scheme of the first splitting.

Since there are two values for the decision at time 2, there are also two 'objective function' values: $\mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(1,1)}$ and $\mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(1,2)}$. The worst-case value is:

$$z = \max\left\{\mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(1,1)}, \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(1,2)}\right\}.$$

After splitting $\mathcal{Z}$ into two subsets, one is solving the following problem:

$$\begin{aligned} \min\ & z^{(1)} \\ \text{s.t. } & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(1,s)} \le z^{(1)}, \quad s = 1,2 \\ & \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\mathbf{x}_2^{(1,s)} \le \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,s}, \quad s = 1,2. \end{aligned} \tag{4}$$

Since for each s the constraint system is less restrictive than in (1), an improvement in the optimal value can be expected. Also, the average-case performance is expected to be better than in the case of (1), due to the variety of time 2 decision variants. □

The splitting process can be continued so that the already existing sets $\mathcal{Z}_{r,s}$ are split with hyperplanes. This is illustrated by the continuation of our example.

Example 2. Figure 2 illustrates the second splitting round, where the set $\mathcal{Z}_{1,1}$ is left not split, but the set $\mathcal{Z}_{1,2}$ is split with a new hyperplane into two new subsets $\mathcal{Z}_{2,2}$ and $\mathcal{Z}_{2,3}$. Then, a problem results with three uncertainty subsets and three decision variants $\mathbf{x}_2^{(2,s)}$ for time 2. □

In general, after the r-th splitting round there are N_r uncertainty subsets $\mathcal{Z}_{r,s}$ and N_r decision variants $\mathbf{x}_2^{(r,s)}$. The problem is then:

$$\begin{aligned} \min\ & z^{(r)} \\ \text{s.t. } & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(r,s)} \le z^{(r)}, \quad s \in \mathcal{N}_r \\ & \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\mathbf{x}_2^{(r,s)} \le \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \qquad s \in \mathcal{N}_r = \{1, ..., N_r\}. \end{aligned} \tag{5}$$

The finer the splitting of the uncertainty set, the lower optimal value one may expect. In the limiting case, as the maximum diameter of the uncertainty subsets for a given r converges to 0 as $r \to +\infty$, it should hold that the optimal value of (5) converges to $\overline{z}_{\text{adj}}$ - the optimal value of (2). In Bertsimas

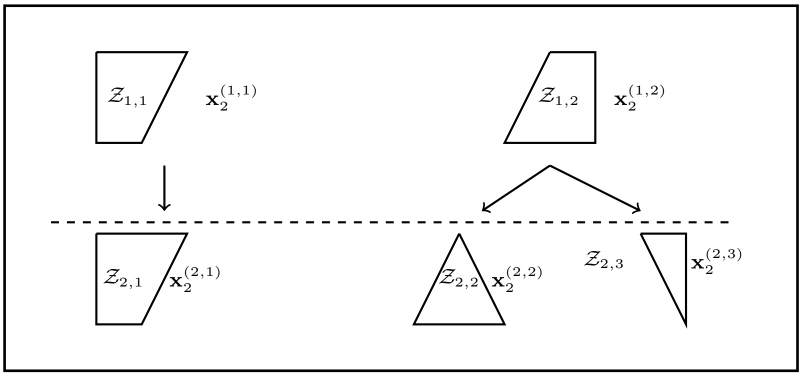

Figure 2 An example of second split for the two-period case.

et al. (2010) authors study the question of finding the optimal K time 2 decision variants, and prove under several regularity assumptions that as the number K of variants tends to $+\infty$, the optimal solution to the k-adaptable problem converges to $\overline{z}_{\text{adj}}$.

Determining whether further splitting is needed and finding the proper hyperplanes is crucial for an improvement in the worst-case objective value to occur. The next two subsections provide theory for determining (1) how far the current optimum is from the best possible value, (2) what are the conditions for the split to bring an improvement in the objective function value.

2.2. Lower bounds

As the problem becomes larger with subsequent splitting rounds, it is important to know how far the current optimal value is from $\overline{z}_{\text{adj}}$ or its lower bound. We use a lower bounding idea proposed for two-period robust problems in Hadjiyiannis et al. (2011), and used also in Bertsimas and Georghiou (2015).

Let $\overline{\mathcal{Z}} = \left\{ \boldsymbol{\zeta}^{(1)}, \ldots, \boldsymbol{\zeta}^{(|\overline{\mathcal{Z}}|)} \right\} \subset \mathcal{Z}$ be a finite set of scenarios for the uncertain parameter. Consider the problem

$$\begin{aligned} \min_{w, \mathbf{x}_1, \mathbf{x}_2^{(i)}, i=1,\ldots,|\overline{\mathcal{Z}}|} \quad & w \\ \text{s.t.} \quad & \mathbf{c}_1^T \mathbf{x}_1 + \mathbf{c}_2^T \mathbf{x}_2^{(i)} \leq w, \quad i = 1, ..., |\overline{\mathcal{Z}}| \\ & \mathbf{A}_1 \left(\boldsymbol{\zeta}^{(i)} \right) \mathbf{x}_1 + \mathbf{A}_2 \left(\boldsymbol{\zeta}^{(i)} \right) \mathbf{x}_2^{(i)} \leq \mathbf{b}, \quad i = 1, ..., |\overline{\mathcal{Z}}|, \end{aligned} \tag{6}$$

where each $\mathbf{x}_1 \in \mathbb{R}^{d_1}$ and $\mathbf{x}_2^{(i)} \in \mathbb{R}^{d_2}$, for all i. Then, the optimal value of (6) is a lower bound for $\overline{z}_{\text{adj}}$, the optimal value of (2) and hence, to any problem (5).

Since each scenario in $\overline{\mathcal{Z}}$ increases the size of the problem to solve, it is essential to include a possibly small number of scenarios determining the current optimal value of problem (5). The next section indicates a special class of scenarios and based on this, in Section 5 we propose heuristic techniques to construct $\overline{\mathcal{Z}}$.

2.3. How to split

In this section, we introduce the key results related to the way in which the uncertainty sets $\mathcal{Z}_{r,s}$ should be split. The main idea behind splitting the sets is as follows. For each $\mathcal{Z}_{r,s}$ we identify a finite set $\overline{\mathcal{Z}}_{r,s} \subset \mathcal{Z}_{r,s}$ of *critical scenarios*. If $\overline{\mathcal{Z}}_{r,s}$ contains more than one element, a hyperplane is constructed such that at least two elements of $\overline{\mathcal{Z}}_{r,s}$ are on different sides of the hyperplane. We call this process *dividing* the set $\overline{\mathcal{Z}}_{r,s}$. This hyperplane becomes also the splitting hyperplane of $\mathcal{Z}_{r,s}$. To avoid confusion, we use the term *split* in relation to continuous uncertainty sets $\mathcal{Z}_{r,s}$ and the term *divide* in relation to the finite sets $\overline{\mathcal{Z}}_{r,s}$ of critical scenarios.

2.3.1. General theorem To obtain results supporting the decision about splitting the subsets $\mathcal{Z}_{r,s}$, we study the dual of problem (5). We assume that (5) satisfies Slater's condition. By result of Ben-Tal and Beck (2009) the dual of (5) is equivalent to:

$$
\begin{aligned}
\max_{\left\{\boldsymbol{\lambda}^{(r,s)}\right\}_{s=1}^{N_r},\boldsymbol{\mu}^{(r)},\left\{\left\{\boldsymbol{\zeta}^{(r,s,i)}\right\}_{s=1}^{N_r}\right\}_{i=1}^{m}} \quad & -\sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}b_i \\
\text{s.t.} \quad & \sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}\mathbf{a}_{1,i}\left(\boldsymbol{\zeta}^{(r,s,i)}\right)+\sum_{s\in\mathcal{N}_r}\mu_s^{(r)}\mathbf{c}_1=0 \\
& \sum_{i=1}^{m}\lambda_i^{(r,s)}\mathbf{a}_{2,i}\left(\boldsymbol{\zeta}^{(r,s,i)}\right)+\mu_s^{(r)}\mathbf{c}_2=0, \quad \forall s\in\mathcal{N}_r \\
& \sum_{s\in\mathcal{N}_r}\mu_s^{(r)}=1 \\
& \boldsymbol{\lambda}^{(r,s)}\geq 0, s\in\mathcal{N}_r \\
& \boldsymbol{\mu}^{(r)},\boldsymbol{\lambda}^{(r)}\geq 0 \\
& \boldsymbol{\zeta}^{(r,s,i)}\in\mathcal{Z}_{r,s}, \quad \forall s\in\mathcal{N}_r, \quad \forall 1\leq i\leq m.
\end{aligned} \tag{7}
$$

Interestingly, problem (7) is nonconvex in the decision variables, which is not the case for duals of nonrobust problems. This phenomenom has been noted already in Beck and Ben-Tal (2009). Because Slater's condition holds, strong duality holds, and for an optimal $\overline{x}^{(r)}$ to problem (5), with objective value $\overline{z}^{(r)}$, there exist $\overline{\boldsymbol{\lambda}}^{(r)},\overline{\boldsymbol{\mu}}^{(r)},\overline{\boldsymbol{\zeta}}^{(r)}$ such that the dual optimal value is attained and equal to $\overline{z}^{(r)}$. In the following, we use a shorthand notation:

$$
\mathbf{x}^{(r)}=\left(\mathbf{x}_1,\left\{\mathbf{x}_2^{(r,s)}\right\}_{s=1}^{N_r}\right),\boldsymbol{\lambda}^{(r)}=\left\{\boldsymbol{\lambda}^{(r,s)}\right\}_{s=1}^{N_r},\boldsymbol{\zeta}^{(r)}=\left\{\left\{\boldsymbol{\zeta}^{(r,s,i)}\right\}_{s=1}^{N_r}\right\}_{i=1}^{m},\boldsymbol{\mu}^{(r)}=(\mu_1^{(r)},\ldots,\mu_{N_r}^{(r)})^T.
$$

A similar approach is applied in the later parts of the paper. For each $s\in\mathcal{N}_r$ let us define

$$
\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})=\left\{\overline{\boldsymbol{\zeta}}^{(r,s,i)}: \quad \overline{\lambda}_i^{(r,s)}>0\right\},
$$

which is a set of worst-case scenarios for $\boldsymbol{\zeta}$ determining that the optimal value for (5) cannot be better than $\overline{z}^{(r)}$. Since the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ are defined with respect to the given optimal dual solution, they are all finite.

The following theorem states that at least one of the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ for which $|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})|>1$ must be divided as a result of splitting $\mathcal{Z}_{r,s}$ in order for the optimal value $\overline{z}^{(r')}$ of the problem after the subsequent splitting rounds to be better than $\overline{z}^{(r)}$.

Theorem 1 *Assume that problem (5) satisfies Slater's condition, $\overline{\mathbf{x}}^{(r)}$ is the optimal primal solution, and $\overline{\boldsymbol{\lambda}}^{(r)}$ $\overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ is the optimal dual solution. Assume that at a splitting round $r' > r$ there exists a sequence of distinct numbers $\{j_1, j_2, ..., j_{N_r}\} \subset \mathcal{N}_{r'}$ such that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',j_s}$ for each $1 \leq s \leq N_r$, that is, each set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ remains not divided, staying a part of some uncertainty subset. Then, it holds that the optimal value $\overline{z}^{(r')}$ after the r'-th splitting round is equal to $\overline{z}^{(r)}$.* □

Proof. We construct a lower bound for the problem after the r'-th round with value $\overline{z}^{(r)}$ by choosing proper $\boldsymbol{\lambda}^{(r',s)}, \boldsymbol{\mu}^{(r')}, \boldsymbol{\zeta}^{(r',s,i)}$. Without loss of generality we assume that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',s}$ for all $s \in N_r$. We take the dual problem of the problem after the r'-th splitting round in the form (7). We assign the following values:

$$\lambda_i^{(r',s)} = \begin{cases} \overline{\lambda}_i^{(r,s)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases}$$
$$\mu_s^{(r')} = \begin{cases} \overline{\mu}_s^{(r)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases}$$
$$\boldsymbol{\zeta}^{(r',s,i)} = \begin{cases} \bar{\boldsymbol{\zeta}}^{(r,s,i)} & \text{if } s \leq N_r, \quad \overline{\lambda}_i^{(r,s)} > 0 \\ \text{any } \boldsymbol{\zeta}^{(r',s,i)} \in \mathcal{Z}_{r',s} & \text{otherwise.} \end{cases}$$

Such variables are dual feasible and give an objective value to the dual equal to $\overline{z}^{(r)}$. Since the dual objective value provides a lower bound on the primal problem after the r'-th round, the theorem follows. □

The above result provides an important insight. If there exist sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ with more than one element each, then at least one of such sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ should be divided in the splitting process. Otherwise, by Theorem 1, one can construct a lower bound showing that the resulting objective value cannot improve. On the other hand, if no such $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ exists, then splitting should stop since, by Theorem 1, the optimal value cannot improve.

Corollary 1 *If for optimal $\overline{\boldsymbol{\lambda}}^{(r,s)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ it holds that:*

$$\left|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})\right| \leq 1, \quad \forall s \in \mathcal{N}_r,$$

then $\overline{z}^{(r)} = \overline{z}_{adj}$, where $\overline{z}_{adj}$ is the optimal value of (2).

Proof. A lower-bound program with a scenario set $\overline{\mathcal{Z}} = \cup_{s \in \mathcal{N}_r} \overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ has an optimal value at most $\overline{z}_{\text{adj}}$. By duality arguments similar to Theorem 1, the optimal value of such a lower bound problem must be equal to $\overline{z}^{(r)}$. This, combined with the fact that $\overline{z}^{(r)} \geq \overline{z}_{\text{adj}}$ gives $\overline{z}^{(r)} = \overline{z}_{\text{adj}}$. □

Theorem 1 does not tell us which of the sets $\mathcal{Z}_{r,s}$ have to be split - it says only that at least one of $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ has to be split for which $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ contains more than one element. Moreover, if there exists more than one dual optimal $\overline{\boldsymbol{\lambda}}^{(r,s)}$, each of them may imply different sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ to be divided. In other words, conducting a 'proper' (in the sense of Theorem 1) splitting round with respect to sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, implied by the given $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ could, in the general case, not be 'proper' with respect to sets $\overline{\mathcal{Z}}_{r,s}(\widehat{\boldsymbol{\lambda}}^{(r)})$ implied by another dual optimal $\widehat{\boldsymbol{\lambda}}^{(r)}, \hat{\boldsymbol{\zeta}}^{(r)}$. However, such a situation did not occur in any of the numerical experiments conducted in this paper.

In the following section we consider the question how to find the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ to be divided.

2.3.2. Finding the sets of scenarios to be divided In this section we propose concrete methods of identifying the sets of scenarios to be divided. Such sets should be 'similar' to the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ in the sense that they should consist of scenarios $\boldsymbol{\zeta}$ that are a part of the optimal solution to the dual problem (7). If this condition is satisfied, such sets are expected to result in splitting decisions leading to improvements in the objective function value, in line with Theorem 1.

Active constraints. The first method of constructing scenario sets to be divided relies on the fact that for a given optimal solution $\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2^{(r)}$ to (5), a $\overline{\lambda}_i^{(r,s)} > 0$ corresponds to an active primal constraint. That means, for each $s \in \mathcal{N}_r$ we can define the set:

$$\Phi_{r,s}\left(\overline{\mathbf{x}}^{(r)}\right) = \left\{\boldsymbol{\zeta}: \quad \exists_i : \mathbf{a}_{1,i}^T(\boldsymbol{\zeta})\overline{\mathbf{x}}_1 + \mathbf{a}_{2,i}^T(\boldsymbol{\zeta})\overline{\mathbf{x}}_2^{(r,s)} = b_i\right\}.$$

Though some $\Phi_{r,s}\left(\overline{\mathbf{x}}^{(r)}\right)$ may contain infinitely many elements, one can approximate it by finding a single scenario for each constraint, solving the following problem for each s, i:

$$\begin{aligned} \min_{\boldsymbol{\zeta}} \; & b_i - \mathbf{a}_{1,i}^T(\boldsymbol{\zeta})\overline{\mathbf{x}}_1 + \mathbf{a}_{2,i}^T(\boldsymbol{\zeta})\overline{\mathbf{x}}_2^{(r,s)} \\ & \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}. \end{aligned} \tag{8}$$

If for given s, i the optimal value of (8) is 0, we add the optimal $\boldsymbol{\zeta}$ to the set $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$. In the general case, such a set may include $\boldsymbol{\zeta}$'s for which there exists no $\overline{\lambda}_i^{(r,s)} > 0$ being a part of optimal dual solution.

Using the KKT vector of the robust problem. The active constraints approach may result in having an unnecessarily large number of critical scenarios found. Therefore, there is a need for a way to obtain the values of $\overline{\boldsymbol{\lambda}}^{(r)}$ to choose only the scenarios $\boldsymbol{\zeta}^{(r,s,i)}$ for which it holds that $\overline{\lambda}_i^{(r,s)} > 0$. This requires us to solve the problem (7) by solving its convex reformulation.

Here, we choose to achieve this by removing the nonconvexity of problem (7), which requires an additional assumption that each $\mathcal{Z}_{r,s}$ is representable by a finite set of convex constraints:

$$\mathcal{Z}_{r,s} = \{\boldsymbol{\zeta}: \quad h_{r,s,j}(\boldsymbol{\zeta}) \leq 0, \quad j = 1, ..., I_{r,s}\}, \quad \forall s \in \mathcal{N}_r, \tag{9}$$

where each $h_{r,s,j}(.)$ is a closed convex function. Note that this representation allows for the use of hyperplanes to split, as affine functions are also convex. For an overview of sets representable in this way we refer to Ben-Tal et al. (2014), mentioning here only that such formulation entails also conic sets. With such a set definition, by results of Gorissen et al. (2014), we can transform (7) to an equivalent convex problem by substituting $\lambda_i^{(r,s)}\boldsymbol{\zeta}^{(r,s,i)} = \boldsymbol{\xi}^{(r,s,i)}$. Combining this with the definition of the rows of matrices $\mathbf{A}_1, \mathbf{A}_2$, we obtain the following problem, equivalent to (7):

$$\begin{aligned}
\max_{\boldsymbol{\lambda}^{(r)},\boldsymbol{\mu}^{(r)},\boldsymbol{\xi}^{(r)}} \quad & -\sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}b_i \\
\text{s.t.} \quad & \sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\mathbf{P}_{1,i}\boldsymbol{\xi}^{(r,s,i)} + \sum_{s\in\mathcal{N}_r}\mu_s^{(r)}\mathbf{c}_1 = 0 \\
& \sum_{i=1}^{m}\mathbf{P}_{2,i}\boldsymbol{\xi}^{(r,s,i)} + \mu_s^{(r)}\mathbf{c}_2 = 0, \quad \forall s\in\mathcal{N}_r \\
& \sum_{s\in\mathcal{N}_r}\mu_s^{(r)} = 1 \\
& \boldsymbol{\lambda}^{(r,s)} \geq 0, s\in\mathcal{N}_r \\
& \boldsymbol{\mu}^{(r)} \geq 0 \\
& \lambda_i^{(r,s)}h_{s,j}\left(\frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}}\right) \leq 0, \quad \forall s\in\mathcal{N}_r, \quad i=1,\dots,m, \quad j=1,...,I_{r,s}.
\end{aligned} \tag{10}$$

Problem (10) is convex in the decision variables - it involves constraints that are either linear in the decision variables or that involve *perspective functions* of convex functions, see Boyd and Vandenberghe (2004). Optimal variables for (10), with substitution

$$\boldsymbol{\zeta}^{(r,s,i)} = \begin{cases} \frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}} \text{ for } \lambda_i^{(r,s)} > 0 \\ \boldsymbol{\zeta}^{(r,s,i)} \in \mathcal{Z}_{r,s} \text{ for } \overline{\lambda}_i^{(r,s)} = 0, \end{cases}$$

are optimal for (7). Hence, one may construct the sets of points to be split as:

$$\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) = \left\{ \frac{\overline{\boldsymbol{\xi}}^{(r,s,i)}}{\overline{\lambda}_i^{(r,s)}} : \quad \overline{\lambda}_i^{(r,s)} > 0 \right\}.$$

Thus, in order to obtain a set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, one needs the solution to the convex problem (10). It turns out that this solution can be obtained at no extra cost apart from solving (5) if we assume representation (9) and that the tractable robust counterpart of (5) satisfies Slater's condition - one can use then its optimal KKT vector.

The tractable robust counterpart of (5), constructed using the methodology of Ben-Tal et al. (2014), is:

$$\begin{aligned}
\min_{z^{(r)},\mathbf{x}_1,\mathbf{x}_2^{(r,s)},\mathbf{v}^{(s,i,j)},u_j^{(s,i)}} \quad & z^{(r)} \\
\text{s.t.} \quad & \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2^{(r,s)} \leq z^{(r)}, \quad s\in\mathcal{N}_r \\
& \textstyle\sum_{j=1}^{I_{r,s}} u_j^{(s,i)}h_{s,i,j}^*\left(\frac{\mathbf{v}^{(s,i,j)}}{u_j^{(s,i)}}\right) \leq b_i, \quad \forall s\in\mathcal{N}_r, \forall 1\leq i\leq m \\
& \textstyle\sum_{j=1}^{I_{r,s}} \mathbf{v}^{s,i,j} = \mathbf{P}_{1,i}^T\mathbf{x}_1 + \mathbf{P}_{2,i}^T\mathbf{x}_2^{(r,s)}, \quad \forall s\in\mathcal{N}_r, \forall 1\leq i\leq m.
\end{aligned} \tag{11}$$

Let us denote the Lagrange multipliers of the three subsequent constraint types by $\mu_s^{(r)}, \lambda_i^{(r,s)}, \boldsymbol{\xi}^{(r,s,i)}$, respectively. Now we can formulate the theorem stating that the KKT vector of the optimal solution to (11) gives the optimal solution to (10).

Theorem 2 *Suppose that problem (11) satisfies Slater's condition. Then, the components of the optimal KKT vector of (11) yield the optimal solution to (10).* □

Proof. The Lagrangian for problem (11) is:

$$\begin{aligned} L\left(z^{(r)}, \mathbf{x}^{(r)}, \mathbf{v}^{(s)}, \mathbf{u}^{(s)}, \boldsymbol{\lambda}^{(r)}, \boldsymbol{\mu}^{(r)}, \boldsymbol{\xi}^{(r)}\right) = {} & z^{(r)} + \sum_s \mu_s^{(r)} \left\{ \mathbf{c}_1^T \mathbf{x}_1 + \mathbf{c}_2^T \mathbf{x}_2^{(r,s)} - z^{(r)} \right\} + \\ & + \sum_{s,i} \lambda_i^{(r,s)} \left(\sum_j u_j^{(s,i)} h^*_{s,i,j} \left(\frac{\mathbf{v}^{s,i,j}}{u_j^{(s,i)}} \right) - b_i \right) + \\ & - \sum_{s,i} \left(\boldsymbol{\xi}^{(r,s,i)}\right)^T \left(\sum_j \mathbf{v}^{s,i,j} - \mathbf{P}_{1,i}^T \mathbf{x}_1 - \mathbf{P}_{2,i}^T \mathbf{x}_2^{(r,s)} \right) \end{aligned}$$

We now show that the Lagrange multipliers correspond to the decision variables with the corresponding names in problem (10), by deriving the Lagrange dual problem:

$$\begin{aligned} & \max_{\boldsymbol{\lambda}^{(r)} \geq 0, \boldsymbol{\mu}^{(r)} \geq 0, \boldsymbol{\xi}^{(r)}} \min_{\substack{z^{(r)}, \mathbf{x}^{(r)}, \\ \mathbf{v}^{(s)}, \mathbf{u}^{(s)}}} L\left(z^{(r)}, \mathbf{x}^{(r)}, \mathbf{v}^{(s)}, \mathbf{u}^{(s)}, \boldsymbol{\lambda}^{(r)}, \boldsymbol{\mu}^{(r)}, \boldsymbol{\xi}^{(r)}\right) = \\ & = \max_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\mu} \geq 0, \boldsymbol{\xi}} \left\{ \min_{z^{(r)}} \left(1 - \sum_s \mu_s^{(r)} \right) z^{(r)} + \min_{\mathbf{x}_1} \left(\sum_s \mu_s^{(r)} \mathbf{c}_1 + \sum_{s,i} \mathbf{P}_{1,i} \boldsymbol{\xi}^{(r,s,i)} \right)^T \mathbf{x}_1 \right. \\ & \qquad + \min_{\mathbf{x}_2^{(r,s)}} \sum_s \left(\mu_s^{(r)} \mathbf{c}_2 + \sum_i \mathbf{P}_{2,i} \boldsymbol{\xi}^{(r,s,i)} \right)^T \mathbf{x}_2^{(r,s)} \\ & \qquad \left. + \sum_{s,i,j} \min_{\mathbf{v}^{(s,i,j)}, u_j^{(s,i)}} \left\{ \lambda_i^{(r,s)} u_j^{(s,i)} h^*_{s,i,j} \left(\frac{\mathbf{v}^{s,i,j}}{u_j^{(s,i)}} \right) - \left(\boldsymbol{\xi}^{(r,s,i)} \right)^T \mathbf{v}^{s,i,j} \right\} \right\} \\ & = \max_{\boldsymbol{\lambda} \geq 0, \boldsymbol{\mu} \geq 0, \boldsymbol{\xi}} \left\{ -\sum_s \mu_s^{(r)} b_i \,\middle|\, 1 - \sum_s \mu_s^{(r)} = 0, \quad \sum_s \mu_s^{(r)} \mathbf{c}_1 + \sum_{s,i} \mathbf{P}_{1,i} \boldsymbol{\xi}^{(r,s,i)} = 0, \right. \\ & \qquad \left. \mu_s^{(r)} \mathbf{c}_2 + \sum_i \mathbf{P}_{2,i} \boldsymbol{\xi}^{(r,s,i)} = 0, \quad \forall s, \quad \lambda_i^{(r,s)} h_{s,i,j} \left(\frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}} \right) \leq 0 \quad \forall s,i,j \right\} \end{aligned}$$

Hence, one arrives at the problem equivalent to (10) and the theorem follows. □

In fact, Theorem 2 turns out to be a special case of a the result of Beck and Ben-Tal (2009). Due to Theorem 2, we know that the optimal solution to (10), and thus to (7), can be obtained at no extra computational effort since most of the solvers produce the KKT vector as a part of output.

As already noted in Section 2.3.1, the collections of sets $\left\{ \overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \right\}_{s=1}^{N_r}$ and $\left\{ \overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)}) \right\}_{s=1}^{N_r}$ may only be one of many possible collections of sets, of which at least one is to be divided. This is because different combinations of sets may correspond to different values of the optimal primal and dual variable values. Hence, there is no guarantee that even by dividing all the sets $\left\{ \overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \right\}_{s=1}^{N_r}$ or $\left\{ \overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)}) \right\}_{s=1}^{N_r}$, one separates 'all the $\boldsymbol{\zeta}$ scenarios that ought to be separated'. However, the approaches presented in this section are computationally tractable and may give a good practical performance, as shown in the numerical examples of Section 6.

3. Multiperiod problems

3.1. Description

In this section we extend the basic two-period methodology to the case with more than two periods, which requires a more extensive notation. The uncertain parameter and the decision vector are:

$$\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_{T-1} \end{bmatrix} \in \mathbb{R}^{L_1} \times ... \times \mathbb{R}^{L_{T-1}}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix} \in \mathbb{R}^{d_1} \times .... \times \mathbb{R}^{d_T}.$$

Value of the component $\boldsymbol{\zeta}_t$ is revealed at time t. The decision $\mathbf{x}_t$ is implemented at time t, after the value of $\boldsymbol{\zeta}_{t-1}$ is known but before $\boldsymbol{\zeta}_t$ is known. We introduce a special notation for the time-dependent parts of the vectors. The symbol $\mathbf{x}_{s:t}$, where $s \leq t$ shall denote the part of the vector $\mathbf{x}$ corresponding to periods s through t. We also define $L = \sum\limits_{t=1}^{T-1} L_i$ and $d = \sum\limits_{t=1}^{T} d_t$.

The considered robust multi-period problem is:

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \ & \mathbf{A}(\boldsymbol{\zeta})\mathbf{x} \leq \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{aligned} \tag{12}$$

where the matrix $\mathbf{A} : \mathbb{R}^L \to \mathbb{R}^{m \times d}$ is linear and its i-th row is denoted by $\mathbf{a}_i^T$. In the multi-period case we also split the set $\mathcal{Z}$ into a collection of sets $\mathcal{Z}_{r,s}$ where $\cup_{s \in \mathcal{N}_r} \mathcal{Z}_{r,s} = \mathcal{Z}$ for each r. By $\text{Proj}_t(\mathcal{Z}_{r,s})$ we denote the projection of the set $\mathcal{Z}_{r,s}$ onto the space corresponding to the uncertain parameters from the first t periods:

$$\text{Proj}_t(\mathcal{Z}_{r,s}) = \{\boldsymbol{\xi} : \quad \exists \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \quad \boldsymbol{\xi} = \boldsymbol{\zeta}_{1:t}\}.$$

Contrary to the two-period case, every subset $\mathcal{Z}_{r,s}$ shall correspond to a vector $\mathbf{x}^{(r,s)} \in \mathbb{R}^d$, i.e., a vector including decisions for all the periods.

In the two-period case, the time 1 decision was common for all the variants of decision variables. In the multi-period notation this condition would be written as $\mathbf{x}_1^{(r,s)} = \mathbf{x}_1^{(r,s+1)}$ for $1 \leq s \leq N_r - 1$. In the two-period case each of the uncertainty subsets $\mathcal{Z}_{r,s}$ corresponded to a separate variant $\mathbf{x}_2^{(r,s)}$, and given a $\boldsymbol{\zeta}$, any of them could be chosen if only it held at time 2 that $\boldsymbol{\zeta} \in \mathcal{Z}_{r,s}$. In this way, it was guaranteed that

$$\forall \boldsymbol{\zeta} \in \mathcal{Z} \quad \exists \mathbf{x}_2^{(r,s)} : \quad \mathbf{A}_1(\boldsymbol{\zeta})\mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta})\mathbf{x}_2^{(r,s)} \leq \mathbf{b}.$$

In the multi-period case the main obstacle is the need to establish *nonanticipativity constraints*, see Shapiro et al. (2014) for a discussion of nonanticipativity in the context of Stochastic Programming. Nonanticipativity requires that decisions made at time t can be based only on information available at that time.

In our context, we have that information about subsequent components of $\boldsymbol{\zeta}$ is revealed period after period, whereas at the same time decisions need to be implemented. In general up to time T one may not know to which $\mathcal{Z}_{r,s}$ the vector $\boldsymbol{\zeta}$ will *surely* belong to.

For instance, suppose that at time 1 the decision $\overline{\mathbf{x}}_1$ is implemented. At time 2, knowing only the value $\boldsymbol{\zeta}_1$ there may be many potential sets $\mathcal{Z}_{r,s}$ to which $\boldsymbol{\zeta}$ may belong and for which $\overline{\mathbf{x}}_2 = \mathbf{x}_2^{(r,s)}$ - all the $\mathcal{Z}_{r,s}$ for which $\boldsymbol{\zeta}_1 \in \mathrm{Proj}_1(\mathcal{Z}_{r,s})$. Suppose that a decision $\overline{\mathbf{x}}_2 = \mathbf{x}_2^{(r,s)}$ is chosen at time 2, for some s. Then, at time 3 there must exist a set $\mathcal{Z}_{r,s}$ such that $\boldsymbol{\zeta}_{1:2} \in \mathrm{Proj}_2(\mathcal{Z}_{r,s})$ and for which $\overline{\mathbf{x}}_{1:2} = \mathbf{x}_{1:2}^{(r,s)}$, so that its decision for time 3 can be implemented.

In general, at each time period $2 < t \leq T$ there must exist a set $\mathcal{Z}_{r,s}$ such that the vector $\boldsymbol{\zeta}_{1:t-1} \in \mathrm{Proj}_{t-1}(\mathcal{Z}_{r,s})$, and for which it holds that $\overline{\mathbf{x}}_{1:t-1} = \mathbf{x}_{1:t-1}^{(r,s)}$, where $\overline{\mathbf{x}}_{1:t-1}$ stands for the decisions already implemented. We propose an iterative splitting strategy ensuring that this postulate is satisfied.

In this strategy, the early-period decisions corresponding to various sets $\mathcal{Z}_{r,s}$ are identical, as long as it is not possible to distinguish to which of them the vector $\boldsymbol{\zeta}$ will belong. Our strategy facilitates simple determination of these equality constraints between various decisions and is based on the following notion.

Definition 1. A hyperplane defined by a normal vector $\mathbf{g} \in \mathbb{R}^L$ and intercept term $h \in \mathbb{R}$ is a time t splitting hyperplane (called later t-SH) if:

$$t = \min\left\{u : \mathbf{g}^T\boldsymbol{\zeta} = h \quad \Leftrightarrow \quad \mathbf{g}_{1:u}^T\boldsymbol{\zeta}_{1:u} = h, \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^L\right\}.$$

In other words, such a hyperplane is determined by a linear inequality that depends on $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_t$, but not on $\boldsymbol{\zeta}_{t+1}, \ldots, \boldsymbol{\zeta}_T$. We shall refer to a hyperplane by the pair $(\mathbf{g}, h)$.

We illustrate with an example how the first splitting can be done and how the corresponding equality structure between decision vectors $\mathbf{x}^{(r,s)}$ is determined.

Example 3. Consider a multi-period problem where $T = 3$, with a rectangular uncertainty set containing one dimensional ζ_1 and ζ_2, as depicted in Figure 3. We split the uncertainty set $\mathcal{Z}$ with a 1-SH $(\mathbf{g}, h)$. Then, two subsets result:

$$\mathcal{Z}_{1,1} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \mathbf{g}^T\boldsymbol{\zeta} \leq h\right\} \quad \text{and} \quad \mathcal{Z}_{1,2} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \mathbf{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

Now, there are two decision vectors $\mathbf{x}^{(1,1)}, \mathbf{x}^{(1,2)} \in \mathbb{R}^d$. Their time 1 decisions should be identical since they are implemented before the value of ζ_1 is known, allowing to determine whether $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{1,2}$. Thus, we add a constraint $\mathbf{x}_1^{(1,1)} = \mathbf{x}_1^{(1,2)}$. This splitting is illustrated in Figure 3.

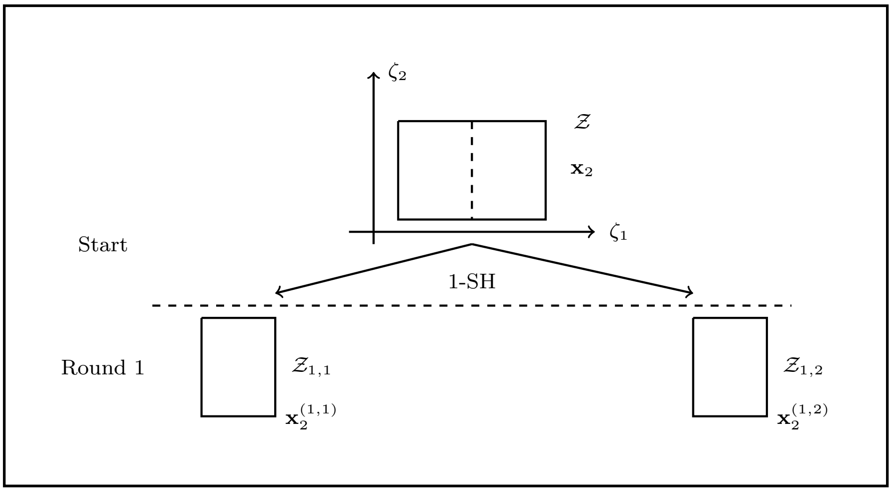

Figure 3 A multi-period problem after a single splitting with a time-1 splitting hyperplane. Only information about ζ_1 is needed to determine if $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$ (Round 1).

The problem to be solved after the first splitting round is analogous to the two-period case, with an equality constraint added:

$$\begin{aligned}\min_{z^{(1)},\mathbf{x}^{(1,s)}} \quad & z^{(1)} \\ \text{s.t.} \quad & \mathbf{c}^T\mathbf{x}^{(1,s)} \le z^{(1)}, \quad s=1,2 \\ & \mathbf{A}(\boldsymbol{\zeta})\,\mathbf{x}^{(s)} \le \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,s}, \quad s=1,2 \\ & \mathbf{x}_1^{(1,1)} = \mathbf{x}_1^{(1,2)}.\end{aligned}$$

The splitting process may be continued and multiple types of t-SHs are possible. To state our methodology formally, we define a parameter $t_{max}(\mathcal{Z}_{r,s})$ for each set $\mathcal{Z}_{r,s}$. If the set $\mathcal{Z}_{r,s}$ is a result of subsequent splits with various t-SH's, the number $t_{max}(\mathcal{Z}_{r,s})$ denotes the largest t of them. By convention, for the set $\mathcal{Z}$ it shall hold that $t_{max}(\mathcal{Z}) = 0$. The following rule defines how the subsequent sets can be split and what the values of the parameter t_{max} for each of the resulting sets are.

Rule 1. A set $\mathcal{Z}_{r,s}$ can be split only with a t-SH such that $t \ge t_{max}(\mathcal{Z}_{r,s})$. For the resulting two sets $\mathcal{Z}_{r+1,s'}, \mathcal{Z}_{r+1,s''}$ we define $t_{max}(\mathcal{Z}_{r+1,s'}) = t_{max}(\mathcal{Z}_{r+1,s''}) = t$. If the set is not split and in the next round it becomes the set $\mathcal{Z}_{r+1,s'}$ then $t_{max}(\mathcal{Z}_{r+1,s'}) = t_{max}(\mathcal{Z}_{r,s})$. □

The next rule defines the equality constraints for the problem after the $(r+1)$-th splitting round, based on the problem after the r-th splitting round.

Rule 2. Let a set $\mathcal{Z}_{r,s}$ be split with a t-SH into sets $\mathcal{Z}_{r+1,s'}, \mathcal{Z}_{r+1,s''}$. Then the constraint $\mathbf{x}_{1:t}^{(r+1,s')} = \mathbf{x}_{1:t}^{(r+1,s'')}$ is added to the problem after the $(r+1)$-th splitting round.

Assume the problem after splitting round r includes sets $\mathcal{Z}^{r,s}$ and $\mathcal{Z}^{r,u}$ with a constraint $\mathbf{x}_{1:k_s}^{(r,s)} = \mathbf{x}_{1:k_s}^{(r,u)}$, and the sets $\mathcal{Z}^{r,s}, \mathcal{Z}^{r,u}$ are split into $\mathcal{Z}^{r+1,s'}, \mathcal{Z}^{r+1,s''}$ and $\mathcal{Z}^{r+1,u'}, \mathcal{Z}^{r+1,u''}$, respectively. Then, the constraint $\mathbf{x}_{1:k_s}^{(r+1,s'')} = \mathbf{x}_{1:k_s}^{(r+1,u')}$ is added to the problem after the $(r+1)$-th splitting round. □

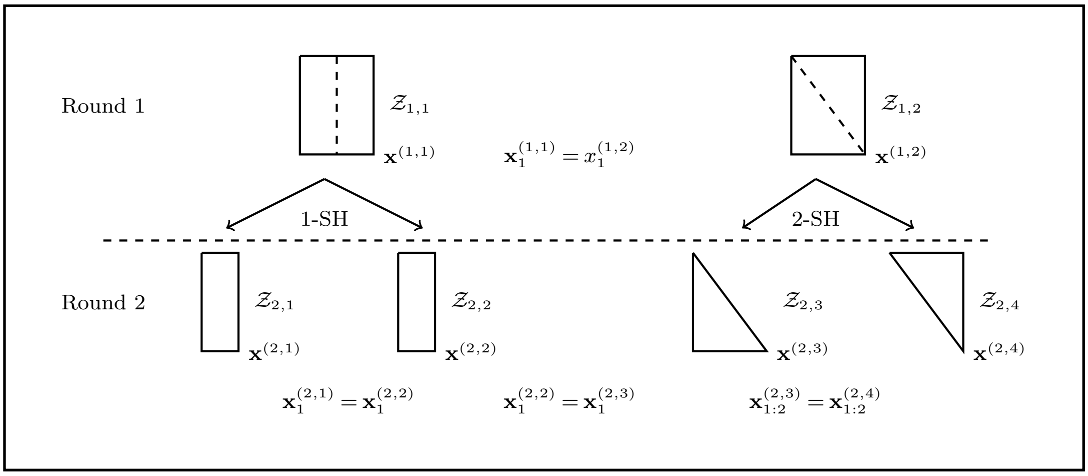

Figure 4 Example of second splitting round for the multi-period case. Only information about ζ_1 is needed to determine whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,1}$ or $\mathcal{Z}_{2,2}$ (Round 2). However, information about both ζ_1 and ζ_2 is needed to distinguish whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,3}$ or $\mathcal{Z}_{2,4}$ (Round 2).

The first part of Rule 2 ensures that the decision vectors $\mathbf{x}^{(r+1,s')}, \mathbf{x}^{(r+1,s'')}$ can differ only from time period $t+1$ on, since only then one can distinguish between the sets $\mathcal{Z}_{r,s'}, \mathcal{Z}_{r,s''}$. The second part of Rule 2 ensures that the dependence structure between decision vectors from stage r is not 'lost' after the splitting. Rule 2 as a whole ensures that $\mathbf{x}_{1:k_s}^{(r+1,s')} = \mathbf{x}_{1:k_s}^{(r+1,s'')} = \mathbf{x}_{1:k_s}^{(r+1,u')} = \mathbf{x}_{1:k_s}^{(r+1,u'')}$.

Rules 1 and 2 are not the only possible implementation of the splitting technique that respects the nonanticipativity restriction. However, their application in the current form does not require the decision maker to compare the sets $\mathcal{Z}_{r,s}$ for establishing the equality constraints between their corresponding decision vectors.

We illustrate the application of Rules 1 and 2 with a continuation of our example.

Example 4. By Rule 1 we have $t_{max}(\mathcal{Z}_{1,1}) = t_{max}(\mathcal{Z}_{1,2}) = 1$. Thus, each of the sets $\mathcal{Z}_{1,1}, \mathcal{Z}_{1,2}$ can be split with a t-SH where $t \geq 1$. We split the set $\mathcal{Z}_{1,1}$ with a 1-SH and the set $\mathcal{Z}_{1,2}$ with a 2-SH. The scheme of the second splitting round is given in Figure 4.

We obtain 4 uncertainty sets $\mathcal{Z}_{2,s}$ and 4 decision vectors $x^{(2,s)}$. The lower part of Figure 4 includes three equality constraints. The first constraint $\mathbf{x}_1^{(2,1)} = \mathbf{x}_1^{(2,2)}$ and the third constraint $\mathbf{x}_{1:2}^{(2,3)} = \mathbf{x}_{1:2}^{(2,4)}$ follow from the first part of Rule 2, whereas the second equality constraint $\mathbf{x}_1^{(2,2)} = \mathbf{x}_1^{(2,3)}$ is determined by the second part of Rule 2. The equality constraints imply that $\mathbf{x}_1^{(2,1)} = \mathbf{x}_1^{(2,2)} = \mathbf{x}_1^{(2,3)} = \mathbf{x}_1^{(2,4)}$.

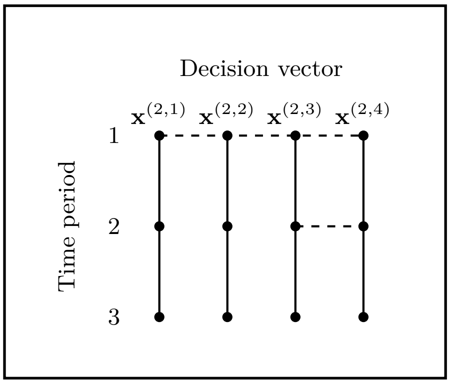

Figure 5 Time structure of the decision variants after the second splitting. Dashed horizontal lines denote the nonanticipativity (equality) constraints between decisions. The figure is motivated by Figure 3.2 in Chapter 3.1.4 of Shapiro et al. (2014).

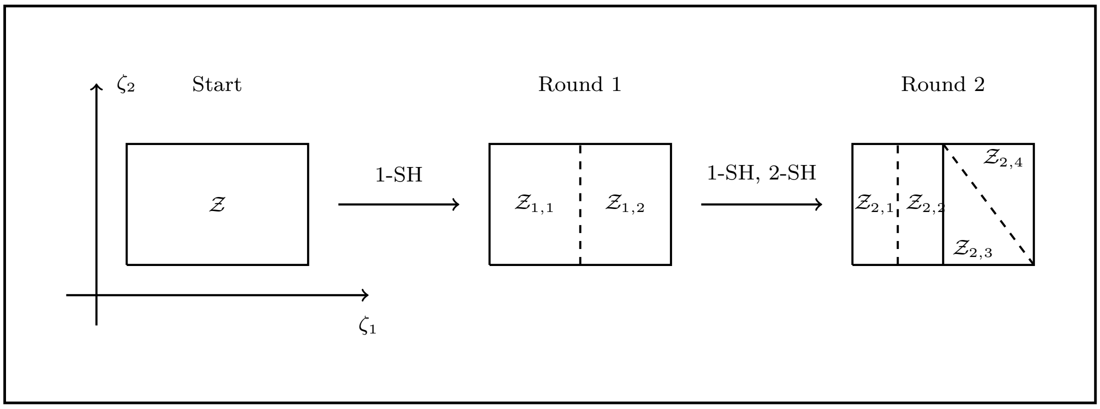

Figure 6 Example of the first two splitting rounds for the multi-period case. Only information about ζ_1 is needed to determine if a point $\boldsymbol{\zeta}$ belongs to (i) $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$ (Round 1), (ii) $\mathcal{Z}_{2,1}$ or $\mathcal{Z}_{2,2}$ (Round 2). However, information about both ζ_1 and ζ_2 is needed to distinguish whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,3}$ or $\mathcal{Z}_{2,4}$ (Round 2).

The problem after the second splitting round is:

$$\begin{aligned}
\min_{z^{(2)},\mathbf{x}^{(2,s)}} \quad & z^{(2)} \\
\text{s.t.} \quad & \mathbf{c}^T\mathbf{x}^{(2,s)} \leq z^{(2)}, \quad s=1,\dots,4 \\
& \mathbf{A}(\boldsymbol{\zeta})\,\mathbf{x}^{(2,s)} \leq \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{2,s}, \quad s=1,\dots,4 \\
& \mathbf{x}_1^{(2,1)} = \mathbf{x}_1^{(2,2)} \\
& \mathbf{x}_1^{(2,2)} = \mathbf{x}_1^{(2,3)} \\
& \mathbf{x}_{1:2}^{(2,3)} = \mathbf{x}_{1:2}^{(2,4)}.
\end{aligned}$$

The time structure of decisions for subsequent time periods is illustrated in Figure 5. Also, Figure 6 shows the evolution of the uncertainty set relations with the subsequent splits.

At time 1 there is only one possibility for the first decision. Then, at time 2 the value of ζ_1 is known and one can determine if $\boldsymbol{\zeta}$ is within the set $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$, or both.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1}$, further verification is needed to determine whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,1}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,2}$, to choose the correct variant of decisions for time 2 and later.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,2}$, the time 2 decision $\mathbf{x}_2^{(2,3)} = \mathbf{x}_2^{(2,4)}$ is implemented. Later, the value of $\boldsymbol{\zeta}_2$ is revealed and based on it, one determines whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,3}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,4}$. In the first case, decision $\mathbf{x}_3^{(2,3)}$ is implemented. In the second case, decision $\mathbf{x}_3^{(2,4)}$ is implemented.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \cap \mathcal{Z}_{1,2}$ (thus $\boldsymbol{\zeta}$ belongs to the tangent segment of the two sets, see Figure 6), then at time 2 one can implement either $\mathbf{x}_2^{(2,2)}$ or $\mathbf{x}_2^{(2,3)} = \mathbf{x}_2^{(2,4)}$. It is best to choose the decision for which the entire decision vector $\mathbf{x}^{(r,s)}$ gives the best worst-case objective value.

If one chooses $\mathbf{x}_2^{(2,3)} = \mathbf{x}_2^{(2,4)}$, then at time 2 it is known whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,3}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,4}$ (or both), and the sequence of decisions for later periods is chosen. If one chooses $\mathbf{x}_2^{(2,2)}$ then $\mathbf{x}_3^{(2,2)}$ is implemented. An analogous procedure holds for other possibilities.

In general, the problem after the r-th splitting round has N_r subsets $\mathcal{Z}_{r,s}$ and decision vectors $x^{(r,s)}$. Its formulation is:

$$\begin{aligned} \min_{z^{(r)},\mathbf{x}^{(r,s)}} \quad & z^{(r)} \\ & \mathbf{c}^T \mathbf{x}^{(r,s)} \le z^{(r)}, \quad s \in \mathcal{N}_r \\ & \mathbf{A}(\boldsymbol{\zeta})\, \mathbf{x}^{(r,s)} \le \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \quad s \in \mathcal{N}_r \\ & \mathbf{x}_{1:k_s}^{(r,s)} = \mathbf{x}_{1:k_s}^{(r,s+1)}, \quad s \in \mathcal{N}_r \setminus \{N_r\}, \end{aligned} \tag{13}$$

where k_s is the number of the first time period decisions that are required to be identical for decision vectors $\mathbf{x}^{(r,s)}$ and $\mathbf{x}^{(r,s+1)}$. When Rules 1 and 2 are applied in the course of splitting, a complete set of numbers k_s is obtained from Rule 2 and at most $N_r - 1$ such constraints are needed. This corresponds to the sets $\{\mathcal{Z}_{r,s}\}_{s=1}^{N_r}$ being ordered in a line and having equality constraints only between the adjacent sets, see Figure 4, where after the second splitting round equality constraints are required only between $\mathbf{x}^{(2,1)}$ and $\mathbf{x}^{(2,2)}$, $\mathbf{x}^{(2,2)}$ and $\mathbf{x}^{(2,3)}$, and between $\mathbf{x}^{(2,3)}$ and $\mathbf{x}^{(2,4)}$.

3.2. Lower bounds

Similar to the two-period case, one can obtain lower bounds for the adjustable robust solution. The lower bound problem differs from the two-period case since the uncertain parameter may have a multi-period equality structure of the components that can be exploited.

Let $\overline{\mathcal{Z}} = \left\{\boldsymbol{\zeta}^{(1)}, \dots, \boldsymbol{\zeta}^{(|\overline{\mathcal{Z}}|)}\right\} \subset \mathcal{Z}$ be a finite set of scenarios for the uncertain parameter. Then, the optimal solution to

$$\begin{aligned} \min_{w,\mathbf{x}_2^{(i)},i=1,\dots,|\overline{\mathcal{Z}}|} \quad & w \\ \text{s.t. } & \mathbf{c}^T \mathbf{x}_2^{(i)} \le w, \quad i = 1, \dots, \left|\overline{\mathcal{Z}}\right| \\ & \mathbf{A}\left(\boldsymbol{\zeta}^{(i)}\right) \mathbf{x}^{(i)} \le \mathbf{b}, \quad i = 1, \dots, \left|\overline{\mathcal{Z}}\right| \\ & \mathbf{x}_{1:t}^{(i)} = \mathbf{x}_{1:t}^{(j)} \quad \forall_{i,j,t} : \boldsymbol{\zeta}_{1:t}^{(i)} = \boldsymbol{\zeta}_{1:t}^{(j)} \end{aligned} \tag{14}$$

is a lower bound for problem (13).

In the multi-period case it is required that for each decision vectors $\mathbf{x}^{(i)}, \mathbf{x}^{(j)}$ whose corresponding uncertain scenarios are identical up to time t the corresponding decisions must be the same up to time t as well (nonanticipativity restriction). This is needed since up to time t one cannot distinguish between $\boldsymbol{\zeta}^{(i)}$ and $\boldsymbol{\zeta}^{(j)}$ and the decisions made should be the same. The equality structure between the decision vectors $\mathbf{x}^{(i)}$ can be obtained efficiently (using at most $\left|\overline{\mathcal{Z}}\right| - 1$ vector equalities) if uncertain parameter is one-dimensional in each time period - one achieves it by sorting the set $\overline{\mathcal{Z}}$ lexicographically.

3.3. How to split

3.3.1. General theorem We assume that (13) satisfies Slater's condition. By the result of Ben-Tal and Beck (2009) the dual of (13) is equivalent to:

$$
\begin{aligned}
\max\; & - \sum_{s \in \mathcal{N}_r} \sum_{i=1}^{m} \lambda_i^{(r,s)} b_i \\
\text{s.t.}\; & \sum_{i=1}^{m} \lambda_i^{(r,s)} \mathbf{a}_i\left(\boldsymbol{\zeta}^{(r,s,i)}\right) + \mu_s^{(r)} \mathbf{c} + \begin{bmatrix} \nu_s^{(r)} \\ 0 \end{bmatrix} - \begin{bmatrix} \nu_{s-1}^{(r)} \\ 0 \end{bmatrix} = 0, \quad \forall 1 < s < N_r \\
& \sum_{i=1}^{m} \lambda_i^{(r,1)} \mathbf{a}_i\left(\boldsymbol{\zeta}^{(r,1,i)}\right) + \mu_1^{(r)} c + \begin{bmatrix} \nu_1^{(r)} \\ 0 \end{bmatrix} = 0 \\
& \sum_{i=1}^{m} \lambda_i^{(r,N_r)} \mathbf{a}_i\left(\boldsymbol{\zeta}^{(r,N_r,i)}\right) + \mu_{N_r}^{(r)} c - \begin{bmatrix} \nu_{r,N_r-1}^{(r)} \\ 0 \end{bmatrix} = 0 \\
& \sum_{s \in \mathcal{N}_r} \mu_s^{(r)} = 1 \\
& \boldsymbol{\lambda}^{(r)}, \boldsymbol{\mu}^{(r)} \geq 0 \\
& \boldsymbol{\zeta}^{(r,s,i)} \in \mathcal{Z}_{r,s}, \quad \forall s \in \mathcal{N}_r, \quad \forall 1 \leq i \leq m.
\end{aligned}
\tag{15}
$$

Because of Slater's condition, strong duality holds and for an optimal primal solution $\overline{\mathbf{x}}^{(r)}$ with objective value $\overline{z}^{(r)}$ there exist $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\nu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ such that the optimal value of (15) is attained and is equal to $\overline{z}^{(r)}$. For each subset $\mathcal{Z}_{r,s}$ we define:

$$
\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) = \left\{ \overline{\boldsymbol{\zeta}}^{(r,s,i)} \in \mathcal{Z}_{r,s} : \quad \overline{\boldsymbol{\lambda}}_{s,i}^{(r)} > 0 \right\}.
$$

Then, the following result holds, stating that at least one of the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, for which $\left|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})\right| > 1$, should be split.

Theorem 3 *Assume that problem (13) satisfies Slater's condition, $\overline{\mathbf{x}}^{(r)}$ is the the optimal primal solution, and that $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\nu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ are the optimal dual variables. Assume that at any splitting round $r' > r$ there exists a sequence of distinct numbers $\{j_1, j_2, ..., j_{N_r}\} \subset \mathcal{N}_{r'}$ such that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',j_s}$ and for each $1 \leq s \leq N_r$ it holds that $\mathcal{Z}_{r',j_s}$ results from splitting the set $\mathcal{Z}_{r,s}$. Then, the optimal value $\overline{z}^{(r')}$ is the same as $\overline{z}^{(r)}$, that is, $\overline{z}^{(r')} = \overline{z}^{(r)}$.* □

Proof. We construct a lower bound for the problem after the r'-th round with value $\overline{z}^{(r)}$. Without loss of generality we assume that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',s}$ for all $1 \leq s \leq N_r$. By Rules 1 and 2, the problem after the r'-th splitting round implies equality constraints $\mathbf{x}_{1:k_s}^{(r',s)} = \mathbf{x}_{1:k_s}^{(r',s+1)}$, where $1 \leq s \leq N_r - 1$. Take the dual (15) of the problem after the r'-th splitting round. We assign the following values for $\boldsymbol{\lambda}^{(r')}, \boldsymbol{\mu}^{(r')}$:

$$\begin{aligned}
\lambda_i^{(r',s)} &= \begin{cases} \overline{\lambda}_i^{(r,s)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases} \\
\mu_s^{(r')} &= \begin{cases} \overline{\mu}_s^{(r)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases} \\
\nu_s^{(r')} &= \begin{cases} \overline{\nu}_s^{(r)} & \text{for } 1 \leq s \leq N_r - 1 \\ 0 & \text{otherwise} \end{cases} \\
\boldsymbol{\zeta}^{(r',s,i)} &= \begin{cases} \bar{\boldsymbol{\zeta}}^{(r,s,i)} & \text{if } 1 \leq s \leq N_r, \quad \overline{\lambda}_i^{(r,s)} > 0 \\ \text{any } \boldsymbol{\zeta}^{(r',s,i)} \in \mathcal{Z}_{r',s,i} & \text{otherwise.} \end{cases}
\end{aligned}$$

These values are dual feasible and give an objective value to the dual problem equal to $\overline{z}^{(r)}$. Since the dual objective value provides a lower bound for the primal problem, the objective function value for the problem after the r'-th round cannot be better than $\overline{z}^{(r)}$. $\square$

Similar to the two-period case, one can prove that if each of the sets $\overline{\mathcal{Z}}_{r,s}$ has at most one element, then the splitting process may stop since the optimal objective value cannot be better than $\overline{z}^{(r)}$.

3.3.2. Finding the sets of scenarios to be divided For the multi-period case, the same observations hold that have been made in the case of the two-period problem. That is, one may construct sets $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$ by searching for the scenarios $\boldsymbol{\zeta}$ corresponding to active primal constraints, or sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ by using the optimal KKT variables of the tractable counterpart of (13). The latter approach is preferred for its inclusion only of the critical scenarios in the meaning of Theorem 3.

4. Problems with integer variables

4.1. Methodology

A particularly difficult application field for adjustable robust decision rules is when some of the decision variables are integer. Our methodology can be particularly useful since the decisions are fixed numbers for each of the uncertainty subset $\mathcal{Z}_{r,s}$. A general multiperiod robust adjustable problem with integer and continuous variables can be solved through splitting in the same fashion as in Section 2 and 3.

Suppose, taking the notation of Section 3, that the indices of components of the vector x to be integer belong to a set $\mathcal{I}$. Then, the mixed-integer version of problem (13) has only an additional integer condition:

$$\begin{array}{rl} \min\limits_{z^{(r)},\mathbf{x}^{(r,s)}} & z^{(r)} \\ & \mathbf{c}^T\mathbf{x}^{(r,s)} \leq z^{(r)}, \quad s \in \mathcal{N}_r \\ & \mathbf{A}\left(\zeta\right)\mathbf{x}^{(r,s)} \leq \mathbf{b}, \quad \forall \zeta \in \mathcal{Z}_{r,s}, \quad s \in \mathcal{N}_r \\ & \mathbf{x}^{(r,s)}_{1:k_s} = \mathbf{x}^{(r,s+1)}_{1:k_s}, \quad i \in \mathcal{N}_r \setminus \{N_r\} \\ & \mathbf{x}^{(r,s)}_i \in \mathbb{Z}, \quad \forall s \in \mathcal{N}_r, \forall i \in \mathcal{I}. \end{array} \tag{16}$$

To obtain lower bounds, we propose the analogues of the strategies given in Sections 2.2 and 3.2, with the integer condition.

4.2. Finding the sets of scenarios to be divided

For mixed integer optimization the available duality tools are substantially weaker than for problems with continuous variables. One can utilize the subadditive duality theorems to derive results 'similar' to the ones from Section 2.3 and 3.3, but they are not applicable in practice. Two approaches that seem intuitively correct are: (1) separating scenarios responsible for constraints that are 'almost active' for the optimal solution $\overline{\mathbf{x}}^{(r)}$, (2) separating scenarios found on the basis of the LP relaxation of problem (16). We now discuss these two approaches.

Almost active constraints. In the continuous case, the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$ were found by identifying $\boldsymbol{\zeta}$'s generating active constraints for the optimal primal solution. One can also apply this approach in the mixed-integer case, with a correction due to the fact that in mixed-integer problems the notion of 'active constraints' loses its proper meaning - in general case the worst-case value of a left-hand side is not a continuous function of the decision variable $\mathbf{x}$. For that reason, it may happen that:

$$\sup_{\boldsymbol{\zeta} \in \mathcal{Z}_{r,s}} \mathbf{a}_i(\boldsymbol{\zeta})^T \mathbf{x}^{(r,s)} < b_i,$$

even for constraints that are critical - being elements of a set of constraints prohibiting the optimal objective value of (16) from being better than $\overline{z}^{(r)}$. However, for each $s \in \mathcal{N}_r$ one can define an approximate set $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$ of $\boldsymbol{\zeta}$'s corresponding to 'almost active' constraints. To find such $\boldsymbol{\zeta}$'s, for a precision level $\epsilon > 0$ and $s \in \mathcal{N}_r, 1 \leq i \leq m$ one solves the following problem:

$$\begin{array}{l} \min\limits_{\zeta} \; b_i - \mathbf{a}_i\left(\boldsymbol{\zeta}\right)^T \overline{\mathbf{x}}^{(r,s)} - \epsilon \\ \text{s.t. } \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}. \end{array} \tag{17}$$

If the result is a nonpositive optimal value, then one can add the optimal solution ζ to the set $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$. However, this strategy may be subject to scaling problems since ϵ may imply a different degree of 'almost activeness' for different constraints. One may try to mitigate this issue by normalizing the coefficients of each constraint before solving problem (17).

KKT vector of the LP relaxation. Another approach for problems with integer variables, less sensitive to scaling issues, is to determine the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ corresponding to the LP relaxation of problem (16). This approach is expected to perform well in problems where the optimal mixed integer solution is close to the optimal solution of the LP relaxation.

4.3. Problems with constraint-wise uncertainty

Some optimization problems involve constraint-wise uncertainty, that is, $\boldsymbol{\zeta}$ can be split into disjoint blocks in such a way that the data of each uncertain constraint depends on a separate block of $\boldsymbol{\zeta}$, and the uncertainty set $\mathcal{Z}$ is a direct product of uncertainty sets corresponding to the constraints (see Ben-Tal et al. (2004)). A special case are problems where uncertainty is present only in the objective function. Though in most applications this is not the case, this issue deserves a separate treatment. From Ben-Tal et al. (2004) we know that for problems with continuous decisions and constraint-wise uncertainty the optimal value obtained with adjustable decisions is equal to the one obtained with the static robust solution. However, in problems with integer decisions, adjustability may still yield an improvement in the objective function.

Up to now, we have proposed splitting the sets $\mathcal{Z}_{r,s}$ by means of dividing a set $\overline{\mathcal{Z}}_{r,s}$ containing at least two critical scenarios belonging to $\mathcal{Z}_{r,s}$. However, in case of constraint-wise uncertainty it will hold that for each constraint there is only one worst-case scenario, corresponding to a different block of $\boldsymbol{\zeta}$. Thus, splitting the uncertainty sets in order to separate the worst-case scenarios belonging to *the same* uncertainty subset cannot be applied. In such a situation, one has to resort to ad-hoc methods of finding another critical scenario within $\mathcal{Z}_{r,s}$, which may depend on the properties of the problem at hand. We present such a heuristic approach in the route planning experiment of Section 6.3.

5. Heuristics

In this section we propose heuristics for choosing the hyperplanes to split sets $\mathcal{Z}_{r,s}$ (by splitting their corresponding sets $\overline{\mathcal{Z}}_{r,s}$) in the $(r+1)$-th splitting round, for constructing the lower bound scenario sets $\overline{\mathcal{Z}}$, and for deciding when to stop the splitting algorithm.

From now on we fix the optimal primal solution after the r-th splitting round $\overline{\mathbf{x}}^{(r)}$ and the sets $\overline{\mathcal{Z}}_{r,s}$, making no distinction between the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r)})$ obtained by using the optimal KKT vector of the problems' LP relaxations and the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ obtained by searching constraint-wise for scenarios that make the constraints (almost) active. We only consider splitting of sets $\mathcal{Z}_{r,s}$ for which $\left|\overline{\mathcal{Z}}_{r,s}\right| > 1$.

5.1. Choosing the t for the t-SHs

In multi-period problems one must determine the t for the t-SH, and this choice should balance two factors. Intuitively, the set $\mathcal{Z}_{r,s}$ should be split with a $t \geq t_{max}(\mathcal{Z}_{r,s})$ for which the components $\boldsymbol{\zeta}_t$ are most dispersed over $\boldsymbol{\zeta} \in \overline{\mathcal{Z}}_{r,s}$. On the other hand, choosing a high value of t in an early splitting round reduces the range of possible t-SHs in later rounds because of Rule 1.

We propose that each $\mathcal{Z}_{r,s}$ is split with a t-SH for which the components $\boldsymbol{\zeta}_t$ show biggest dispersion within the set $\overline{\mathcal{Z}}_{r,s}$ (measured, for example, with variance) and where $t_{max}(\mathcal{Z}_{r,s}) \leq t \leq t_{max}(\mathcal{Z}_{r,s}) + q$, with q being a predetermined number. If the dispersion equals 0 for all $t_{max}(\mathcal{Z}_{r,s}) \leq t \leq t_{max}(\mathcal{Z}_{r,s}) + q$ then we propose to choose the smallest $t \geq t_{max}(\mathcal{Z}_{r,s})$ such that the components $\boldsymbol{\zeta}_t$ show a nonzero dispersion within $\overline{\mathcal{Z}}_{r,s}$.

5.2. Splitting hyperplane heuristics

In this subsection we provide propositions for constructing the splitting hyperplanes.

Heuristic 1. The idea of this heuristic is to determine the two most distant scenarios in $\overline{\mathcal{Z}}_{r,s}$ and to choose a hyperplane that separates them strongly.

Find the $\boldsymbol{\zeta}^{(a)}, \boldsymbol{\zeta}^{(b)} \in \overline{\mathcal{Z}}_{r,s}$ maximizing $\left\| \boldsymbol{\zeta}^{(i)}_{1:t} - \boldsymbol{\zeta}^{(j)}_{1:t} \right\|_2$ over $\boldsymbol{\zeta}^{(i)}, \boldsymbol{\zeta}^{(j)} \in \overline{\mathcal{Z}}_{r,s}$. Then, split the set $\mathcal{Z}_{r,s}$ with a t-SH defined by:

$$\mathbf{g}_j = \begin{cases} \zeta^{(a)}_j - \zeta^{(b)}_j & \text{if } j \leq t \\ 0 & \text{otherwise} \end{cases}, \quad h = \frac{\mathbf{g}^T \left(\boldsymbol{\zeta}^{(a)} + \boldsymbol{\zeta}^{(b)} \right)}{2}.$$

If (8) or (17) is used to find critical binding scenarios, then these problems could have multiple binding scenarios. Then, the separation of optimal facets may yield better results than of a single $\boldsymbol{\zeta}$ found to be optimal for (8), (17). Then, the heuristic would separate the two most distant facets with, for example, their bisector hyperplane.

Heuristic 2. The idea of this heuristic is to divide the set $\overline{\mathcal{Z}}_{r,s}$ into two sets whose cardinalities differ by as little as possible.

Choose an arbitrary normal vector g for the t-SH. Then, determine the intercept term h such that the term $\left| \left| \overline{\mathcal{Z}}^-_{r,s} \right| - \left| \overline{\mathcal{Z}}^+_{r,s} \right| \right|$ is minimized, with

$$\overline{\mathcal{Z}}^-_{r,s} = \overline{\mathcal{Z}}_{r,s} \cap \left\{ \boldsymbol{\zeta} : \mathbf{g}^T \boldsymbol{\zeta} \leq h \right\}, \quad \overline{\mathcal{Z}}^+_{r,s} = \overline{\mathcal{Z}}_{r,s} \cap \left\{ \boldsymbol{\zeta} : \mathbf{g}^T \boldsymbol{\zeta} \geq h \right\}.$$

The best h can be found using binary search.

HEURISTIC 3. The idea of this heuristic is to split the set $\mathcal{Z}_{r,s}$ with a hyperplane, and to manipulate the late period decisions while keeping the early-period decisions fixed, in such a way that the maximum worst-case 'objective function' for the two sets is minimized. We describe it for the multi-period case.

Choose an arbitrary normal vector g for the t-SH. For a given intercept h define the two sets:

$$\mathcal{Z}_{r+1,s}^{h-} = \mathcal{Z}_{r,s} \cap \left\{ \boldsymbol{\zeta} : \mathbf{g}^T \boldsymbol{\zeta} \leq h \right\}, \quad \mathcal{Z}_{r+1,s}^{h+} = \mathcal{Z}_{r,s} \cap \left\{ \boldsymbol{\zeta} : \mathbf{g}^T \boldsymbol{\zeta} \geq h \right\}.$$

For a fixed $\mathbf{g}$ we define the following function (note that the formulation only includes the constraints related to the given s):

$$\begin{aligned} \tau(h) = \min_{\mathbf{x}^{(r,s')}, \mathbf{x}^{(r,s'')}, w} \quad & w \\ \text{s.t. } & \mathbf{c}_1^T \mathbf{x}_1 + \mathbf{c}_2^T \mathbf{x}_2^{(r,s')} \leq w \\ & \mathbf{c}_1^T \mathbf{x}_1 + \mathbf{c}_2^T \mathbf{x}_2^{(r,s'')} \leq w \\ & \mathbf{A}_1(\boldsymbol{\zeta}) \mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta}) \mathbf{x}_2^{(r,s')} \leq \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r+1,s}^{h-} \\ & \mathbf{A}_1(\boldsymbol{\zeta}) \mathbf{x}_1 + \mathbf{A}_2(\boldsymbol{\zeta}) \mathbf{x}_2^{(r,s'')} \leq \mathbf{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r+1,s}^{h+} \\ & \mathbf{x}_{1:t_{max}(\mathcal{Z}_{r,s})}^{(r,s')} = \mathbf{x}_{1:t_{max}(\mathcal{Z}_{r,s})}^{(r,s'')} = \overline{\mathbf{x}}_{1:t_{max}(\mathcal{Z}_{r,s})}^{(r,s)}. \end{aligned} \tag{18}$$

Equality constraints ensure that the decision variables related by equality constraints to other decision vectors stay with the same values (not to lose the feasibility of the decision vectors for sets $\mathcal{Z}_{r,p}$, where $p \neq s$). The aim is to minimize $\tau(h)$ over the domain of h for which both $\mathcal{Z}_{r+1,s}^{h-}$ and $\mathcal{Z}_{r+1,s}^{h+}$ are nonempty. Function $\tau(h)$ is quasiconvex in h, which has been noted in a different setting in Bertsimas et al. (2010).

5.3. Constructing the lower bound scenario sets

The key premise is that the size of the set $\overline{\mathcal{Z}}^{(r)}$ (the lower bound scenario set after the r-th splitting round) should be kept limited since each additional scenario increases the size of the lower bound problem. Hence, it is important that the limited number of scenarios covers set $\mathcal{Z}$ well.

Summing the scenario sets. One approach is to use $\overline{\mathcal{Z}}^{(r)} = \cup_{s \in \mathcal{N}_r} \overline{\mathcal{Z}}_{r,s}$ after each splitting round, since $\overline{\mathcal{Z}}_{r,s}$ approximates the set of the scenarios that are part of the current dual optimal solution, yielding a bound on the optimal value of the objective function.

To reduce the size of $\overline{\mathcal{Z}}^{(r)}$, we propose that $\overline{\mathcal{Z}}$ contains at most k elements of each $\overline{\mathcal{Z}}_{r,s}$, where k is a predetermined number. This approach implies that the lower bound sequence $\left\{ \overline{w}^{(r)} \right\}$, where $\overline{w}^{(r)}$ is the optimal value of the lower bound problem after the r-th splitting round, needs not be nondecreasing in r.

Incremental building of a scenario set. To ensure a nondecreasing lower bound sequence, one can construct the sets incrementally, starting with $\overline{\mathcal{Z}}^{(1)}$ after the first splitting round and enlarging it with new scenarios after each splitting round. We describe a possible variant of this idea for the multi-period case.

Assume that problem (14) has been solved after the r-th splitting round, the lower-bounding scenario set is $\overline{\mathcal{Z}}^{(r)}$, the optimal value of the lower-bounding problem is $\overline{w}^{(r)}$, and $\overline{\mathbf{x}}^{(i)}$, $i = 1, \ldots, |\overline{\mathcal{Z}}^{(r)}|$, are the decision vectors from the lower bound problem after the r-th splitting round. Suppose that after the $(r+1)$-th splitting round there is a candidate scenario $\boldsymbol{\zeta}' \in \overline{\mathcal{Z}}_{r+1,s}$ for being added to the lower-bound scenario set $\overline{\mathcal{Z}}^{(r+1)}$. Then, scenario $\boldsymbol{\zeta}'$ is added if (1) there is no $1 \leq i \leq |\overline{\mathcal{Z}}^{(r)}|$ such that $A(\boldsymbol{\zeta}')\left(\overline{\mathbf{x}}^{(i)}\right) \leq b$, (2) there exists no $\mathbf{x}^{(\boldsymbol{\zeta}')}$ such that the optimal value to the problem:

$$
\begin{aligned}
\max_{\kappa, \mathbf{x}^{(\boldsymbol{\zeta}')}} \quad & \kappa \\
\text{s.t.} \quad & \mathbf{c}^T \mathbf{x}^{(\boldsymbol{\zeta}')} \leq \overline{w}^{(r)} - \kappa \\
& A(\boldsymbol{\zeta}')\overline{\mathbf{x}}^{(i)} \leq b, \quad \forall i \\
& \mathbf{x}^{(\boldsymbol{\zeta}')}_{1:t} = \mathbf{x}^{(i)}_{1:t} \quad \forall 1 \leq i \leq |\overline{\mathcal{Z}}^{(r)}|, \quad \forall t : \boldsymbol{\zeta}'_{1:t} = \boldsymbol{\zeta}^{(i)}_{1:t},
\end{aligned}
$$

is nonnegative. Condition (1) excludes the case when there exists already $\boldsymbol{\zeta}^{(i)} \in \overline{\mathcal{Z}}^{(r)}$ whose corresponding decision vector $\mathbf{x}^{(i)}$ is robust to $\boldsymbol{\zeta}'$. Condition (2) excludes the case when it is possible to construct a decision vector for $\boldsymbol{\zeta}'$ satisfying the nonanticipativity constraints in relation to decision vectors corresponding to $\boldsymbol{\zeta} \in \overline{\mathcal{Z}}^{(r)}$, and yielding an objective value $\mathbf{c}^T \mathbf{x}^{(\boldsymbol{\zeta}')} \leq \overline{w}^{(r)}$. Such a scenario brings no value as it is known that a lower bound obtained using $\boldsymbol{\zeta}'$ in addition to $\overline{\mathcal{Z}}^{(r)}$ would be at most equal to the lower bound obtained using only $\overline{\mathcal{Z}}^{(r)}$.

Simple heuristic. We propose also an approach that combines approximately the properties of the two propositions above and is fast at the same time. The idea is to build up the lower-bounding set iteratively and add from each $\overline{\mathcal{Z}}_{r,s}$ the k scenarios whose sum of distances from the elements of $\overline{\mathcal{Z}}^{(r-1)}$ is largest. The distance between two vectors is measured by the 2-norm.

5.4. Stopping the algorithm

As the splitting continues, the computational workload related to solving the split problem grows because of the number of variables and uncertainty subsets. We propose three stopping rules for the splitting method: (1) when the objective value is closer to the lower bound than a predetermined threshold level, (2) when the limit of total computational time is reached, (3) when the maximum number of splitting rounds is reached.

6. Numerical experiments

6.1. Capital budgeting

The first numerical experiment involves no fixed recourse and is the capital budgeting problem taken from Hanasusanto et al. (2014). In this problem, a company can allocate an investment budget of B to a subset of projects $i \in \{1, \ldots, N\}$. Each project i has uncertain costs $c_i(\zeta)$ and uncertain profits $r_i(\zeta)$, modelled as affine functions of an uncertain vector ζ of risk factors. The company can invest in a project before or after observing the risk factors ζ. A postponed investment in project i incurs the same costs $c_i(\zeta)$, but yields only a fraction $\theta \in [0, 1)$ of the profits $r_i(\zeta)$.

The problem of maximizing the worst-case return can be formulated as:

$$\begin{aligned} \max \ & R \\ \text{s.t.} \ & R \leq \mathbf{r}(\zeta)^T(\mathbf{x} + \theta \mathbf{y}), \quad \forall \zeta \in \mathcal{Z} \\ & \mathbf{c}(\zeta)^T(\mathbf{x} + \mathbf{y}) \leq B, \quad \forall \zeta \in \mathcal{Z} \\ & \mathbf{x} + \mathbf{y} \leq 1 \\ & \mathbf{x}, \mathbf{y} \in \{0, 1\}^N, \end{aligned}$$

where the decisions x_i and y_i attain value 1 if and only if an early or late investment in project i is undertaken, respectively. The uncertainty set is $\mathcal{Z} = [-1, 1]^F$, where F is the number of risk factors.

We adopt the same random data setting as Hanasusanto et al. (2014). In all instances we use $F = 4$. The project costs and profits are modelled as:

$$c_i(\zeta) = (1 + \mathbf{\Phi}_i^T \boldsymbol{\zeta}/2)c_i^0, \qquad r_i(\zeta) = (1 + \mathbf{\Psi}_i^T \boldsymbol{\zeta}/2)r_i^0, \quad i = 1, \ldots, N.$$

Parameters c_i^0 and r_i^0 are the nominal costs and profits of project i, whereas Φ_i and Ψ_i represent the i-th rows of the factor loading matrices $\mathbf{\Phi}, \mathbf{\Psi} \in \mathbb{R}^{N \times 4}$ as column vectors. The nominal costs c^0 are sampled uniformly from $[0, 10]^N$, and the nominal profits are set to $r^0 = c^0/5$. The components in each row of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are sampled uniformly from the unit simplex in $\mathbb{R}^4$. The investment budget is set to $B = 1^T c^0/2$, and we set $\theta = 0.8$. Table 1 gives the results of Hanasusanto et al. (2014), who apply a K-adaptability approach and sample 100 instances for each combination of N and K (the number of time-2 decision variants) and try to solve it to optimality within a time limit of 2h per instance.

We sample 50 instances for each N and conduct 8 splitting rounds for $N = 5, 10$, 6 for $N = 15, 20$ and 4 for $N = 25, 30$ (for smaller problems one can allow more splitting rounds to obtain better objectives and still operate within reasonable time limits). To split the uncertainty sets we use the worst-case scenarios coming from the optimal KKT vector of the LP relaxation of the robust MILP problems (see Section 2.3.2). In each splitting round we split all subsets $\mathcal{Z}_{r,s}$ for which $|\overline{\mathcal{Z}}_{r,s}| > 1$. The splitting hyperplanes are constructed using Heuristic 1 (see Section 5.2). The upper bound scenario sets are

	$K=2$			$K=3$			$K=4$		
N	(1) (%)	(2) s	(3) (%)	(1) (%)	(2) s	(3) (%)	(1) (%)	(2) s	(3) (%)
5	100	<1	48.67	100	1	68.71	100	36	79.50
10	100	4	59.34	74	1210	86.91	0	-	102.48
15	100	512	63.69	0	-	91.75	0	-	106.93
20	2	5232	64.78	0	-	93.20	0	-	108.61
25	0	-	64.85	0	-	93.72	0	-	109.10
30	0	-	64.98	0	-	94.08	0	-	109.42

Table 1 Results of Hanasusanto et al. (2014) for the capital budgeting problem. K is the number of time-2 decision variants allowed and N is the number of projects. The columns are (1) - percentage of instances solved to optimality within 2h, (2) - average solution time of the instances solved within 2h, (3) - average objective improvements (including the suboptimal solutions from Gurobi for the instances not solved within 2h.

constructed according to the ‘simple heuristic’ (see Section 5.3) with $k=2$. The after-splitting robust MILP problems are solved with Gurobi precision set to 0.5%. All problems were formulated using CVX package and solved with Gurobi solver on an Intel Core 2.66GHz computer.

Apart from the worst-case results, for each instance we conduct a simulation study by sampling from $[-1,1]^4$ uniformly 500 scenarios of the risk factors’ values and computing the objective function values obtained using the static robust solution and our splitting-based adjustable solution.

Table 2 gives the results of our methodology. All the instances have been solved fast, with the largest average time equal to 26.81s. We remark here that, typically for problems with binary variables, the distribution of the solution times is heavy-tailed, and whereas most of the instances are solved within 2-3s, some instances take much more time and influence the average times in this way. Our methodology performs worse on the small instances, which the ‘more exact’ method of Hanasusanto et al. (2014) can solve efficiently in short time. For larger instances our improvements in the objective value are close to the best values of Hanasusanto et al. (2014) for larger instances $N=20,25,30$ - ours being $107.81, 105.33, 106.88\%$ versus their $108.61, 109.10, 109.42\%$, respectively.

We also compare the running time performance of our method to the results of Hanasusanto et al. (2014) though we should mention that the main objective of Hanasusanto et al. (2014) was to find the best solution using a fixed number of time 2 policies. For larger instances ($N \geq 15$) the results of Hanasusanto et al. (2014) are based on suboptimal solutions from Gurobi obtained after 2 hours of computation per instance (see Table 1), whereas our method uses on average less than 27s per instance, with most of the mean times being less than 7s. Upon request, we obtained the Gurobi output of Hanasusanto et al. (2014). It reveals that in majority of instances studied by them, the objective value obtained by the solver after 60s is within 5% of the end objective value obtained after the time limit of 7200s, given in Table 1.

The right part of Table 2 gives the average-case improvements obtained using the adjustable decisions. The improvements are significantly smaller than the worst-case improvements, stabilizing around the level of 25% for larger N.

Table 2 Our results for the capital budgeting problem. 'Splitting rounds' denotes the number of splitting rounds conducted. 'Initial gap' is the optimality gap for the static robust solution and the lower bound obtained after the first splitting round. 'Final gap' is the optimality gap computed with the objective value and lower bound after the last splitting round. 'Average case improvement' denotes the increase of the average-case objective value obtained with the adjustable decisions, relative to the one yielded by the static solution. The relative optimality gaps are computed as $\frac{(\text{UB}-\text{LB})}{0.5(\text{UB}+\text{LB})} * 100\%$, where LB is the objective function value and UB is the upper bound value.

Splitting rounds	N	Obj improvement (%)	Initial gap (%)	Final gap (%)	Average case improvement (%)	Mean time (s)
8	5	57.89	106.09	39.00	12.11	5.40
	10	93.81	100.15	27.68	20.36	26.81
6	15	102.63	100.00	24.29	23.13	4.72
	20	107.81	100.00	22.24	24.79	5.43
4	25	105.33	100.00	23.48	24.30	3.96
	30	106.88	100.00	22.93	24.80	6.54

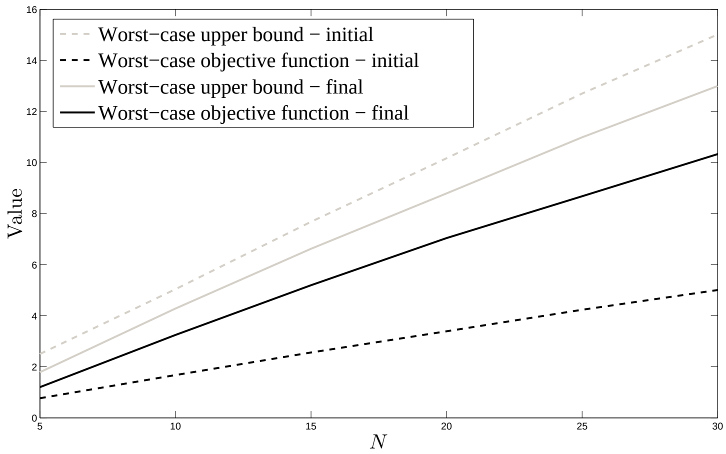

Figure 7 Capital budgeting problem. Plots of initial and final upper bound on the worst-case objective function values and the initial and final worst-case objective function values (average over all problem instances for each N).

Figure 7 shows the average (over problem instances for given N) improvements of the worst-case objective functions and the upper bounds for all N. One can see that the relative gap between the upper bound values and the worst-case objective values decreases significantly with the number N of projects.

We summarize now the results of the first numerical example. Hanasusanto et al. (2014) give good worst-case objective value improvements with a small number of time-2 decision variants (at most 4) after a longer computation time, whereas our splitting method gives such improvements after a short computation time, but with more time-2 decision variants. For example, 9 splitting rounds typically result in a division of the uncertainty set $\mathcal{Z}$ into more than 10 parts, each with a corresponding time-2 decision variant. Thus, our methodology is preferred when it is the computation time, and not the number of decision variants, that is to be kept low.

6.2. Lot sizing

As the second numerical experiment we consider a multi-stage lot sizing problem taken from Bertsimas and Georghiou (2015). The problem entails a single product, T time periods, and the following parameters:

- ζ_t, where $t = 1, \ldots, T$, is the uncertain demand in period t
- l_t, where $t = 2, \ldots, T$, is the lowest possible demand in period t
- u_t, where $t = 2, \ldots, T$, is the highest possible demand in period t
- c_{y_n}, where $n = 1, \ldots, N$, is the per product unit of buying a fixed quantity q_n
- c_x is the ordering cost per product unit for purchases that are delivered in the subsequent period
- c_h is the holding cost per product unit
- $\overline{x}_{\text{tot},t}$, where $t = 2, \ldots, T$, is the cumulative orders limit up to time period t.

The variables are:

- I_t, where $t = 1, \ldots, T$, is the level of available inventory after period t
- x_t, where $t = 1, \ldots, T-1$ is the product amount ordered in period t, after $\zeta_1, \ldots, \zeta_t$ is known, and delivered in period $t+1$, at unit price c_x
- y_{nt}, where $n = 1, \ldots, N$, $t = 2, \ldots, T$, is a binary decision made after $\zeta_1, \ldots, \zeta_t$ is known, whether to buy a fixed quantity q_n of the product in time period t, delivered in the same time period.

The difference between the ordering decisions x_t and y_{nt} is thus that x_t stands for continuous ordering decisions that result in products being delivered with a delay of one time period, and y_{nt} stands for a fixed-size product amount delivered immediately.

The problem is to minimize the worst-case combined ordering and holding costs (referred later to as the ‘total cost’), subject to cumulative ordering constraints:

$$
\begin{aligned}
\min\ & z \\
\text{s.t.}\ & \sum_{t=2}^{T} \left(c_x x_{t-1}(\zeta_{1:t-1}) + c_h I_t(\zeta_{1:t}) + \textstyle\sum_{n=1}^{N} c_{y_n} q_n y_{nt}(\zeta_{1:t}) \right) \leq z, \quad \forall \zeta \in \mathcal{Z} \\
& \left.\begin{array}{l}
I_t(\zeta_{1:t}) = I_{t-1}(\zeta_{1:t-1}) + x_{t-1}(\zeta_{1:t-1}) + \sum_{n=1}^{N} q_n y_{nt}(\zeta_{1:t}) - \zeta_t \\
0 \leq x_{t-1}(\zeta_{1:t-1}) \\
0 \leq I_t(\zeta_{1:t}) \\
\sum\limits_{j=1}^{t-1} x_j(\zeta_{1:j}) \leq \overline{x}_{\text{tot},t}
\end{array}\right\} \begin{array}{c} \forall t = 2, \ldots, T, \\ \forall \zeta \in \mathcal{Z} \end{array} \\
& y_{nt}(\zeta_{1:t}) \in \{0, 1\}, \quad \forall n, t \\
& x_t(\zeta_{1:t}) \geq 0, \quad \forall t
\end{aligned}
\tag{19}
$$

where

$$
\mathcal{Z} = \{\zeta : \zeta_1 = 1, \quad l_t \leq \zeta_t \leq u_t, \quad t = 2, \ldots, T\}.
$$

Table 3 Results of Bertsimas and Georghiou (2015) for the lot sizing problem. The relative optimality gaps are computed as $\frac{(\text{UB}-\text{LB})}{0.5(\text{UB}+\text{LB})} * 100\%$, where UB is the objective function value and LB is the lower bound value. 'Nonadaptive gap' denotes the relative optimality gap computed for the solution where the integer decisions are static and the linear decision rules are implemented for the continuous decision variables. '$\mathcal{PB}_t(1)$ gap' denotes relative optimality gap computed for the solution obtained using the binary adjustability technique used by the authors and where the linear decision rules are implemented for the continuous decision variables. 1% and 5% at the top of the Table are two variants of solver precision used when solving the MILP problems.

		1% optimality			5% optimality		
N	T	$\mathcal{PB}_t(1)$ gap (%)	Nonadaptive gap (%)	Mean time (s)	$\mathcal{PB}_t(1)$ gap (%)	Nonadaptive gap (%)	Mean time (s)
2	2	0	17.6	0.1	0.6	17.6	0.4
	4	24.2	68.6	50.6	27.3	68.6	45.5
	6	37.4	62.0	4833.8	38.9	62.1	956.8
	8	37.9	84.4	27531.1	38.0	84.4	19573.1
	10	39.7	89.9	35716.6	42.0	89.9	31464.1
3	2	0	27.6	0.1	1.2	27.6	0.1
	4	17.2	73.3	3381.8	23.9	73.3	781.6
	6	34.5	66.2	9181.0	38.4	66.1	3298.1
	8	37.6	83.4	28742.7	38.1	83.7	21885.5
	10	-	89.7	-	41.1	90.7	39141.5

The above problem is transformed by eliminating the variables I_t for $t = 2, \dots, T$. The adjustable variables are x_t, allowed to depend on $\zeta_{1:t}$ for $t = 1, \dots, T-1$ and y_{nt}, allowed to depend on $\zeta_{1:t}$ for $t = 2, \dots, T$.

Problem parameters are sampled as in Bertsimas and Georghiou (2015). Ordering costs are chosen from $c_x \in [0;5]$ and $c_{y_n} \in [0;10]$, separately for all $n = 1, \dots, N$, such that $c_x < c_{y_n}$. In this way, the per-item costs of the fixed-size lots of products is always higher than of the product amounts ordered in continuous decisions, and hence, the only advantages of fixed-size lots are in their immediate delivery.

Holding costs are elements of $c_h \in [0;10]$ with the fixed ordering quantities set to $q_n = 100/N$ for all $n = 1, \dots, N$. The cumulative ordering budget is set to $\overline{x}_{\text{tot},t} = \sum_{s=1}^{t-1} \bar{x}_s$ for $t = 2, \dots, T$, with $\bar{x}_t \in [0;100]$ and the lower and upper bounds for the demand are sampled uniformly as $l_t \in [0;25]$ and $u_t \in [75;100]$, $t = 2, \dots, T$. We assume that the initial inventory level I_1 equals zero. Table 3 gives the results obtained by Bertsimas and Georghiou (2015) using their methodology of piecewise linear decision rules for the decision variables.

We sample and solve 50 instances of the problem for $N = 2, 3$ and $T = 2, 4, \dots, 10$. Since $q_n = 100/N$ for all n and the splitting method facilitates the use of integer non-binary variables, we may substitute the N binary decision variables for each period by a single integer variable: $z_t(\zeta_{1:t}) = \sum_{n=1}^{N} y_{nt}(\zeta_{1:t})$ for all $t = 2, \dots, T$, such that $0 \le z_t(\zeta_{1:t}) \le N$ for all t. To see that this is possible, consider a fixed time period t and assume w.l.o.g. that $c_{y_1} \le \dots \le c_{y_N}$. We know that if at the optimal solution exactly z_t

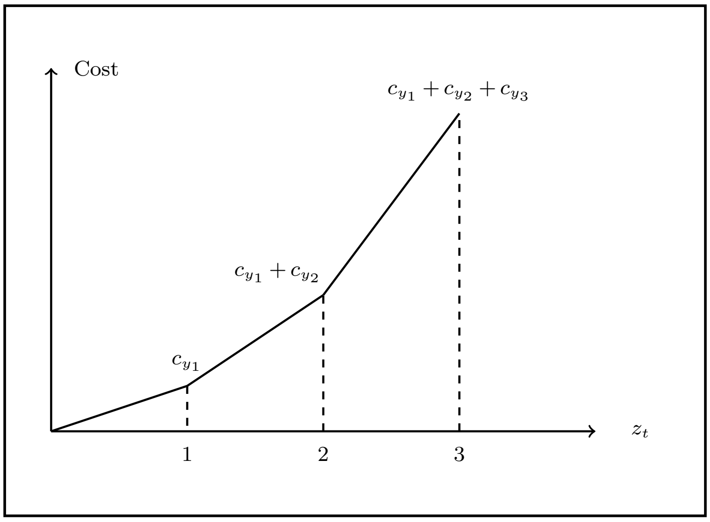

Figure 8 $N=3$. The ordering cost of the fixed-size lots at time t at the optimal solution is a convex, piecewise linear function of z_t.

of the variables $y_{1t},\ldots,y_{Nt}$ have value 1, those will be the variables corresponding to the z_t smallest c_{y_n}'s:

$$\begin{aligned}\min_{y_{nt}}\left\{\sum_{n=1}^{N} c_{y_n}q_n y_{nt}\,\middle|\,\sum_{n=1}^{N} y_{nt}=z_t\right\} &= \min_{y_{nt}}\left\{\frac{100}{N}\sum_{n=1}^{N} c_{y_n}y_{nt}\,\middle|\,\sum_{n=1}^{N} y_{nt}=z_t\right\}\\ &= \frac{100}{N}\sum_{n=1}^{z_t} c_{y_n}\\ &= \frac{100}{N}\max_{n\in\{1,\ldots,N\}}\left\{c_{y_n}(z_t-(n-1))+\sum_{k=1}^{n-1} c_{y_k}\right\}.\end{aligned} \tag{20}$$

The last equality follows from the fact that when $c_{y_1}\le\ldots\le c_{y_N}$, then the sum of z_t smallest c_{y_n} is a convex piecewise linear function of z_t, which can be reformulated as a maximum over N linear functions of z_t, see Figure 8. For that reason, the obtained formulation can substitute the component $\sum_{n=1}^{N} c_{y_n}q_n y_{nt}$ in the objective function of (19) without losing the problem's convexity.

Since problem (19) involves fixed recourse only, we study also the impact of using linear decision rules for the continuous variables $x_t(\zeta_{1:t})$. In such case we set $x_t(\zeta_{1:t})$ to be an affine function of $\zeta_1,\ldots,\zeta_t$:

$$x_t(\zeta_{1:t})=\alpha_{t,0}+\sum_{j=1}^{t}\alpha_{t,j}\zeta_j,\quad \forall t=1,\ldots,T-1,$$

where $\alpha_{t,j}$ are then treated as decision variables implemented in period t.

Each problem instance is solved in four ways: 1) applying static decisions to all variables 2) applying linear decision rules to the continuous variables and static decisions to the integer variables 3) applying only the splitting methodology to all variables 4) applying the splitting methodology to all variables, combined with linear decision rules for the continuous decisions (the parameters $\alpha_{t,j}$ can also differ after splitting of the uncertainty set).

Table 4 Our results for the lot sizing problem for $N=2$. LDR stands for the solution with linear decision rules for the continuous decision variables and static decisions for the integer variables, S stands for only our splitting methodology applied to all variables, S+LDR stands for a combination of set splitting with linear decision rules for the continuous variables. 'Objective improvement' is the decrease in the average worst-case objective value reduction, relative to the static robust solution. Optimality gaps are computed as in Table 3. 'Initial gap' is the optimality gap for the static robust solution and the lower bound obtained after the first splitting round. 'Final gap' is the optimality gap computed with the objective value and lower bound after the last splitting round. The asterisk indicates the fact that for $T=2,4$ the lower bound scenario sets include also all vertices of the uncertainty set $\mathcal{Z}$. All the static robust problems were solved in less than 2s.

	Objective improvement (%)			Initial gap (%)	Final gap (%)			Mean time (s)	
T	LDR	S	S+LDR		LDR	S	S+LDR	S	S+LDR
2	0	11.39	11.38	51.02∗	51.02∗	15.49∗	15.51∗	2.36	2.77
4	31.64	28.07	42.32	85.78∗	52.46∗	57.34∗	34.04∗	5.67	7.69
6	43.77	30.29	54.94	113.14	69.22	87.51	47.39	5.64	10.09
8	48.91	26.32	61.01	125.59	78.73	107.17	54.68	7.54	15.03
10	52.09	22.43	64.21	134.65	86.16	121.02	61.85	9.23	24.23

For each instance we conduct 4 splitting rounds. For splitting we use the worst-case scenario sets obtained using optimal KKT vectors from the robust counterpart of the LP relaxation of the problem (see Sections 2.3.2 and 3.3.2). In each splitting round we split all subsets $\mathcal{Z}_{r,s}$ for which $|\overline{\mathcal{Z}}_{r,s}|>1$. Time periods t for the t-SHs are chosen according to the biggest variance of uncertain demands from subsequent periods with $q=2$ (see Section 5.1). Splitting hyperplanes are constructed using Heuristic 1 (see Section 5.2). The scenario sets for the lower bound problems are constructed according to the 'simple heuristic' (see Section 5.3) with $k=2$. For $T=2,4$ the lower bound scenario sets include also all vertices of the uncertainty set $\mathcal{Z}$. The after-splitting robust MILP problems are solved with Gurobi precision (the relative duality gap when the solver stops) equal to 0.1%. All problems were formulated using CVX package and solved with Gurobi solver on an Intel Core 2.66GHz computer.

Tables 4 and 5 give our results for $N=2$ and $N=3$, respectively. All methodologies offer substantial improvements in the objective value compared to the static robust solution. Also, combination of our splitting methodology with linear decision rules (S+LDR) gives a strong combined effect - the objective value improves significantly more than using any of the methods S or LDR separately - by as much as 64.82% for $N=3, T=10$, compared to 53.21% for LDR and 21.42% for S. For $T=2$ the linear decision rules cannot bring any improvement because x_1 is a scalar. One can observe that for problems with larger T our methodology gives better objective improvements. Also, the relative optimality gaps decrease significantly in all cases, mostly due to improvements in the objective function. All problems have been solved fast, with the maximum mean time equal to 55.82s.

We compare now our results to those of Bertsimas and Georghiou (2015). The main difference between the methods lies in the fact that decision rules proposed by Bertsimas and Georghiou (2015) satisfy the problem's constraints with a high probability (99%), obtained using Hoeffding bounds, whereas

Table 5 Our results for the lot sizing problem for $N=3$. Terminology is the same as in Table 4.

	Objective improvement (%)			Initial gap (%)	Final gap (%)			Mean time (s)	
T	LDR	S	S+LDR		LDR	S	S+LDR	S	S+LDR
2	0	22.94	22.94	61.90∗	61.90∗	17.64∗	17.64∗	2.25	2.61
4	32.66	31.70	47.22	95.06∗	62.30∗	65.09∗	39.14∗	5.24	7.39
6	43.99	29.41	56.86	118.38	78.36	96.55	54.14	5.85	9.80
8	50.14	25.13	62.05	129.06	85.27	113.58	61.48	7.11	14.18
10	53.21	21.42	64.82	136.55	92.22	125.08	68.88	9.18	55.82

Table 6 Lot sizing problem. Average-case performance of the solutions obtained using the three methodologies in comparison to the static robust solution. 'Average-case improvement' is the average reduction of the total cost, relative to the total costs obtained with the static solution for the given demand scenario.

	Average-case improvement (%)					
	$N=2$			$N=3$		
T	LDR	S	S+LDR	LDR	S	S+LDR
2	0.00	18.55	18.55	0.00	14.59	14.13
4	21.87	22.91	31.90	24.51	26.51	37.65
6	30.80	23.24	41.02	33.69	22.72	45.81
8	35.23	20.00	48.05	40.94	18.83	51.56
10	39.67	16.95	51.68	43.79	15.78	55.07

our methodology ensures 100% robustness by design. Comparing the numbers from Tables 3 (column '$\mathcal{PB}_t(1)$ gap'), 4, and 5 (columns 'Final gap (%) - S+LDR'), one can see that our methodology performs worse in terms of the final optimality gap. For example, for $N=2, T=4$ our result is 39.16% compared to their 24.2% for $N=2, T=4$. This can be partly explained by the difference between the types of robustness, and also by different way of choosing the scenarios for the lower bounding problems. On the other hand, our method provides significantly faster computation times which, combined with full robustness, may be an appealing property. In particular, this is visible on larger instances, with our mean solution times being significantly lower, e.g., our 55.82s compared to 39141.5s for $N=3, T=10$.

In addition to the worst-case results, for each solved instance we conduct a simulation study. In each of them we sample uniformly 500 demand scenario realizations $l \leq d_{\text{realized}} \leq u$ and compute the average total costs incurred by each of the four solutions. Table 6 gives the results on average-case improvements relative to the static robust solution. The table shows that our method not only offers substantial improvements on the worst-case basis, but also in terms of the average-case total cost, in particular when combined with the linear decision rules for the continuous variables.

To sum up the results of this numerical example, the main benefits of our approach have been: 1) fast computation time even for large problems, corresponding to the number of splitting rounds (the more splitting rounds, the better the improvement in the objective, but also the longer computation time), 2) substantial improvements in the objective function value, 3) robustness to the entire uncertainty set after each splitting round.

6.3. Route planning

We consider another numerical example from Hanasusanto et al. (2014), the route planning problem, where the uncertainty occurs only in the objective function. On this example, we shall see that our methodology depends heavily on having multiple uncertain constraints that give rise to different worst-case scenarios for the uncertain parameter.

The problem at hand is a shortest path problem that is defined on a directed, arc-weighted graph $G = (V, A, \mathbf{w})$ with nodes $V = \{1, \ldots, N\}$, arcs $A \subseteq V \times V$, and weights $w_{ij}(\xi) \in \mathbb{R}_+$, $(i, j) \in A$. We assume that the arc weights $w_{i,j}$ are functions of an uncertain parameter vector $\boldsymbol{\zeta}$ that is only known to reside in an uncertainty set $\mathcal{Z}$. The goal is to determine the shortest worst-case path from a start node $b \in V$ to a terminal node $e \in V$, $b \neq e$, before the value of $\boldsymbol{\zeta}$ is known. Hanasusanto et al. (2014) consider the number of possible paths to be fixed and equal to K.

In our setting, we begin with the following robust problem, equivalent to having a single path. The binary variable x_{ij} is equal to 1 if arc (i, j) is a part of the path from b to e:

$$
\begin{aligned}
\min_{z,\mathbf{x}} \ & z \\
\text{s.t.} \ & \sum_{(i,j)\in A} w_{ij}(\boldsymbol{\zeta}) x_{ij} \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\
& x_{ij} \in \{0,1\}, \quad (i,j) \in A \\
& \sum_{(j,l)\in A} x_{jl} \geq \sum_{(i,j)\in A} x_{ij} + \mathbb{I}(j = b) - \mathbb{I}(j = e), \ \forall j = 1, \ldots, N,
\end{aligned}
\tag{21}
$$

where $\mathbb{I}(\cdot)$ is the indicator function. In our method the set $\mathcal{Z}$ is split into subsets, to each of which a separate route vector $\mathbf{x}^{(r,s)}$ shall correspond. That is, when $\boldsymbol{\zeta} \in \mathcal{Z}_{r,s}$, then the corresponding path vector $\mathbf{x}^{(r,s)}$ is chosen. Such a problem has the property that finding the optimal path for each $\mathcal{Z}_{r,s}$ can be solved as a separate optimization problem, solving thus N_r smaller problems instead of one large problem.

As visible in (21), there is only one uncertain constraint in this problem. For that reason, we use in this case the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r,s)})$ obtained by searching for the critical scenarios based on the primal solution $\mathbf{x}^{(r)}$. However, as there is only one uncertain constraint, solving problem (17) for this constraint results in only one critical scenario $\overline{\boldsymbol{\zeta}}$. However, we need at least two distinct scenarios to be divided with a splitting hyperplane.

We propose, for a given subset $\mathcal{Z}_{r,s}$ with the corresponding optimal solution vector $\mathbf{x}^{(r,s)}$, to choose the second member of $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r,s)})$ according to the following procedure:

1. Find an alternative route from b to e that uses at most $\lfloor \theta \mathbf{1}^T \overline{\mathbf{x}}^{(r,s)} \rfloor$ arcs from the path corresponding to $\overline{\mathbf{x}}^{(r,s)}$, where $0 \leq \theta \leq 1$ denotes the fraction of the arcs from the 'old' path allowed to use. Denote the new alternative vector by $\widetilde{\mathbf{x}}^{(r,s)}$.

Table 7 Results of Hanasusanto et al. (2014) for the route planning problem. K is the number of time-2 decisions, B denotes the size of the uncertainty set, and N is the number of nodes. The columns are (1) - percentage of instances solved to optimality within 2h (%), (2) - average solution time of the instances solved within 2h (in seconds), (3) - average objective improvements (%).

	$B=3$									$B=6$								
	$K=2$			$K=3$			$K=4$			$K=2$			$K=3$			$K=4$		
N	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
20	100	8	8.31	97	463	10.26	51	1103	10.70	100	7	6.07	97	428	10.23	55	795	11.79
25	99	168	9.49	31	1273	12.06	6	2851	12.81	99	197	8.39	38	1771	13.36	7	3208	15.48
30	69	1131	9.51	6	1563	12.79	0	-	13.94	67	1372	9.06	6	1537	14.61	0	-	17.38
35	17	2335	9.97	0	-	13.70	0	-	15.20	16	2819	9.90	0	-	15.86	0	-	19.20
40	6	2949	9.94	0	-	13.74	0	-	15.28	5	2888	10.36	0	-	16.47	0	-	19.84
45	0	-	9.46	0	-	13.39	0	-	15.19	0	-	10.67	0	-	16.78	0	-	20.32
50	0	-	9.38	0	-	13.31	0	-	15.14	0	-	10.71	0	-	16.73	0	-	20.30

2. Find the worst-case scenario $\widetilde{\boldsymbol{\zeta}}$ corresponding to $\widetilde{\mathbf{x}}^{(r,s)}$ and add it to $\overline{\mathcal{Z}}_{r,s}(\overline{\mathbf{x}}^{(r,s)})$. If no feasible $\widetilde{\mathbf{x}}^{(r,s)}$ exists, take $\widetilde{\boldsymbol{\zeta}}$ to be the scenario that *minimizes* the uncertain distance corresponding to path $\overline{\mathbf{x}}^{(r,s)}$ (in contrast to the worst-case scenarios, which *maximize* the uncertain distance corresponding to path $\overline{\mathbf{x}}^{(r,s)}$).

We consider three values for θ: 0, 0.5, and 0.9. and adopt the same data setting as Hanasusanto et al. (2014). Table 7 presents the improvement results obtained by Hanasusanto et al. (2014). We sample 40 problem instances and, for each instance, we allow 90s for the subsequent splitting rounds, with at most 10s for each optimization problem to solve. Afterwards, we allow a solution time of 60s for each problem. For splitting the uncertainty subsets we use Heuristic 1.

Table 8 presents the results on the improvement in the objective function value. One can see that the method of Hanasusanto et al. (2014) performs significantly better than our approach, with the difference growing with N and the number K of possible time-2 decisions they use. For example, whereas for problems with $B=3$ and $N=20$ we obtain improvement of 5.50% compared to their 8.31% for $K=2$ and 10.70% for $K=4$, for problems with $B=6$ and $N=50$ we get 6.70% and they obtain 20.30% for $K=4$. Additionally, one can see that our approach performs best for $\theta=0$, decreasing with larger values of θ.

The big difference between the performances of our methodology and the one of Hanasusanto et al. (2014) is most likely due to the fact that their methodology optimizes the fixed number of K of decisions. This is implicitly equivalent to optimizing the division of the uncertainty set into K regions corresponding to K possible decisions (each possible time 2 decision has its 'share' of the uncertainty set on which it is at least as good as the other decision). On the other hand, in our methodology the splits are chosen in a relatively simple manner, by means of heuristics, which in this particular case do not perform very well if it is even not known exactly which scenarios should be separated by the splitting hyperplane. The impact of the difference of allowed solution time - 7200s by Hanasusanto

Table 8 Our results for the route planning problem. Average solution time over all instances was equal to 184s.

	Improvement (%)					
	$\theta = 0$		$\theta = 0.5$		$\theta = 0.9$	
N	$B = 3$	$B = 6$	$B = 3$	$B = 6$	$B = 3$	$B = 6$
20	5.50	3.91	4.29	3.88	1.65	2.80
25	6.15	5.01	4.44	4.27	2.63	3.18
30	6.26	5.12	4.18	5.44	1.78	2.73
35	6.34	6.01	4.39	5.73	2.02	3.21
40	5.68	6.90	4.09	5.89	1.07	3.65
45	5.28	7.96	3.76	5.35	1.79	2.38
50	5.45	6.70	2.79	5.18	0.61	2.85

et al. (2014) and average time of 184s in our case - is not expected to be substantial as Hanasusanto et al. (2014) report that in their case the terminal solution was attained typically after 60s.

Remark 1. The results of the route planning experiment leads to a remark that in fact, the methodology of Hanasusanto et al. (2014) could be used to construct 'best' splitting hyperplanes in multi-period problems with only binary decision variables. In such a case, in case of splitting Heuristic 3 (see Section 5.2), one would no longer keep the normal vector $\mathbf{g}$ of the hyperplane fixed, but would optimize it jointly with h. This, however, would only be possible for problems where the methodology of Hanasusanto et al. (2014) is applicable, i.e., to problems with adjustable binary variables.

7. Conclusions

In this paper we have introduced the method of iterative splitting of the uncertainty set for multi-period robust mixed-integer linear optimization problems. We have provided theory on how to determine efficiently which scenarios of the uncertain parameter are more important to be separated than others and how to obtain lower bounds for the adjustable worst-case value. Based on these theoretical results, we have proposed several heuristics for each part of the method.

Our approach can be used to a variety of problems. In particular, this applies to problems with a non-fixed recourse and adjustable integer variables (also non-binary), where implementation of other decision rules may not be possible or may involve large computational effort. For adjustable continuous variables in the non-fixed recourse setting, our method bypasses the challenge of dealing with interactions of uncertain parameters, as would be the case with linear or polynomial decision rules.

For fixed recourse problems the splitting method can be combined with other decision rules, such as linear decision rules, allowing them to take different forms over different parts of the uncertainty set. The second numerical experiment reveals that such a combination gives a strong joint effect. Our iterative method guarantees robustness of the decisions to the entire uncertainty set after each of the splitting rounds. Thus, depending on time constraints, the decision maker can set how many splitting rounds to conduct, with each additional round costing additional effort but bringing potentially extra improvements in the objective value.

Numerical experiments conducted on problems from Bertsimas and Georghiou (2015) and Hanasusanto et al. (2014) have shown our methodology to perform well on problems involving non-constraint-wise uncertainty. In both cases was our method outperformed on small problem instances. However, as the problems grow, our methodology was giving comparable results after only a fraction of the computation time of other authors.

We give now potential directions for further research. First, more theoretical results can be obtained regarding the choice of best splits of the uncertainty sets, and in particular, the 'best' distribution of the splits in time. Secondly, it is important to obtain better lower bound values, possibly by combining our method with results of other authors, e.g., Kuhn et al. (2011). Last, it is interesting to investigate whether our method, combined with the results of Ben-Tal et al. (2014), can be used efficiently in multistage nonlinear robust optimization problems.

Acknowledgements We are grateful to Angelos Georghiou and Grani Hanasusanto for their explanations and sharing with us the exact results of their papers. We are also grateful to the two anonymous referees for their valuable comments on the content and exposition of the paper, and to Bram Gorissen for his critical reading of the paper.

References

Ben-Tal, A., Goryashko, A., Guslitzer, E., & Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, Vol. 99(2), pp. 351-376.

Beck, A., & Ben-Tal, A. (2009). Duality in robust optimization: primal worst equals dual best. *Operations Research Letters*, Vol. 37(1), pp. 1-6.

Ben-Tal, A., El Ghaoui, L. & Nemirovski, A. (2009). *Robust optimization*. (Princeton University Press)

Ben-Tal, A., Den Hertog, D., & Vial, J.-Ph. (2014). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, Vol. 149(1-2), pp. 265-299.

Bertsimas, D., & Caramanis, C. (2007). Adaptability via sampling. In *46th IEEE Conference on Decision and Control* (pp. 4717-4722). IEEE.

Bertsimas, D., & Caramanis, C. (2010). Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control*, Vol. 55(12), pp. 2751-2766.

Bertsimas, D., & Goyal, V. (2010). On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, Vol. 35(2), pp. 284-305.

Bertsimas, D., Iancu, D.A., & Parrilo, P.A. (2010). Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, Vol. 35(2), pp. 363-394.

Bertsimas, D., Goyal, V., & Sun, X.A. (2011a). A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, Vol. 36(1), pp. 24-54.

Bertsimas, D., Iancu, D.A., & Parrilo, P.A. (2011b). A hierarchy of near-optimal policies for multistage adaptive optimization. In *IEEE Transactions on Automatic Control*, Vol. 56(12), pp. 2809-2824.

Bertsimas, D., & Georghiou, A. (2014a). Binary decision rules for multistage adaptive mixed-integer optimization. Available at: `http://www.optimization-online.org/DB_FILE/2014/08/4510.pdf.`

Bertsimas, D., Goyal, V., & Lu, B. Y. (2014b). A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, Vol. 150(2), pp. 281-319.

Bertsimas, D., & Georghiou, A. (2015). Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, Vol 63(3), pp. 610-627.

Birge, J. R., & Wets, R. J-B. (1986) Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Mathematical Programming Study*, Vol. 27, pp. 54-102.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming.* Springer Science & Business Media.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press.

Chen, X., Sim, M., & Sun, P. (2007). A robust optimization perspective on stochastic programming. *Operations Research*, Vol. 55(6), pp. 1058-1071.

Chen, X., & Zhang, Y. (2009). Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, Vol. 57(6), pp. 1469-1482.

Frauendorfer, K., & Kall, P., (1988) A solution method for SLP recourse problems with arbitrary multivariate distributions - the independent case. *Problems in Control and Information Theory*, Vol 17, pp. 177-205.

Gorissen, B. L., Ben-Tal, A., Blanc, H., & Den Hertog, D. (2014). Technical note - Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, Vol. 62(3), pp. 672-679.

Goyal, V., & Lu, B. Y. (2014). On the adaptivity gap in two-stage robust linear optimization under uncertain constraints. Available at: `http://www.optimization-online.org/DB_FILE/2014/12/4706.pdf`.

Grant, M., & Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`.

Gurobi Optimization, Inc. (2014). Gurobi Optimizer Reference Manual, `http://www.gurobi.com`.

Hadjiyiannis, M.J., Goulart, P.J., & Kuhn, D. (2011). A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. *IEEE Conference on Decision and Control and European Control Conference*, Orlando, USA.

Hanasusanto, G.A., Kuhn, D., & Wiesemann, W. (2014). K-adaptability in two-stage robust binary programming. Available online at: `http://www.optimization-online.org/DB_FILE/2014/03/4294.pdf.`

Kuhn, D., Wiesemann, W., & Georghiou, A. (2011). Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, Vol. 130(1), pp. 177-209.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory (Vol. 16).* SIAM.

Vayanos, P., Kuhn, D., & Rustem, B. (2011). Decision rules for information discovery in multi-stage stochastic programming. In *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC),* (pp. 7368-7373).