K-Adaptability in Two-Stage Robust Binary Programming

Grani A. Hanasusanto
Department of Computing, Imperial College London, United Kingdom, g.hanasusanto11@imperial.ac.uk
Daniel Kuhn
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland, daniel.kuhn@epfl.ch
Wolfram Wiesemann
Imperial College Business School, Imperial College London, United Kingdom, ww@imperial.ac.uk

Over the last two decades, robust optimization has emerged as a computationally attractive approach to formulate and solve single-stage decision problems affected by uncertainty. More recently, robust optimization has been successfully applied to multi-stage problems with continuous recourse. This paper takes a step towards extending the robust optimization methodology to problems with integer recourse, which have largely resisted solution so far. To this end, we approximate two-stage robust binary programs by their corresponding K-adaptability problems, in which the decision maker pre-commits to K second-stage policies here-and-now and implements the best of these policies once the uncertain parameters are observed. We study the approximation quality and the computational complexity of the K-adaptability problem, and we propose two mixed-integer linear programming reformulations that can be solved with off-the-shelf software. We demonstrate the effectiveness of our reformulations for stylized instances of supply chain design, route planning and capital budgeting problems.

Key words: robust optimization, integer programming, two-stage problems

1. Introduction

Robust optimization offers a rigorous and efficient methodology to formulate and solve decision problems affected by uncertainty. In order to overcome the curse of dimensionality that plagues traditional approaches, robust optimization replaces probability distributions with uncertainty sets as the fundamental descriptors of uncertainty. The basic robust optimization problem can be formulated as follows.

$$\begin{array}{lll} \text{minimize} & \max_{\boldsymbol{\xi} \in \Xi} f(\boldsymbol{x}, \boldsymbol{\xi}) & \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} & \\ & \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\xi}) \leq \mathbf{0} & \forall \boldsymbol{\xi} \in \Xi \end{array} \tag{1}$$

Here, $\boldsymbol{x}$ and $\boldsymbol{\xi}$ represent the decision variables and uncertain problem parameters, respectively. Problem (1) determines a decision that minimizes the worst-case objective value over the uncertainty set Ξ, subject to satisfying the constraints for all parameter realizations $\boldsymbol{\xi} \in \Xi$. The set Ξ is usually of infinite cardinality, which renders problem (1) intractable in its general form. It turns out, however, that for an astounding variety of sets Ξ and functions f and $\boldsymbol{g}$, problem (1) can be reformulated as a finite-dimensional optimization problem that can be solved in polynomial time. We refer the reader to Ben-Tal et al. (2009) and Bertsimas et al. (2011a) for a detailed account of the state-of-the-art in robust optimization.

In recent years, the basic model (1) has been extended to multi-stage formulations where a sequence of parameter vectors is observed over time, and recourse decisions are taken whenever the value of a parameter vector becomes known. A recourse decision taken at some time t must therefore be modeled as a function of all the parameters observed up to time t. Multi-stage formulations faithfully model the dynamic nature of decision-making processes in practice, and they are essential to mitigate the conservatism of problem (1). Although multi-stage robust optimization problems are computationally intractable in general (Ben-Tal et al. 2004), approximation schemes based on linear and nonlinear decision rules can efficiently provide feasible decisions that are sometimes optimal (Anderson and Moore 1990, Gounaris et al. 2013, Iancu et al. 2013) and often surprisingly close to the optimal solution (Ben-Tal et al. 2004, Chen and Zhang 2009, Goh and Sim 2010, Kuhn et al. 2011, Georghiou et al. 2014). While these approximation schemes can accommodate for discrete here-and-now decisions, they cannot account for discrete recourse decisions.

Discrete recourse decisions have a long history in the related field of stochastic integer programming, where the uncertain problem parameters $\boldsymbol{\xi}$ are modeled as a random vector that is governed by a known probability distribution. Two-stage stochastic integer programs are often solved with the integer L-shaped method, which iteratively adds feasibility and optimality cuts to a relaxed formulation of the first-stage problem. Since the expected recourse function of a two-stage stochastic integer program is in general nonconvex and discontinuous, the introduced cuts are nonconvex themselves. Different classes of cuts can be generated from continuous relaxations of the second-stage problem, evaluations of the expected recourse function at fixed first-stage decisions, as well as cutting plane and branch-and-bound schemes, see Laporte and Louveaux (1993), Carøe and Tind (1998) and Sen and Sherali (2006). Alternatively, two-stage stochastic integer programs can be solved by scenario decomposition schemes, which dualize the nonanticipativity requirement of the first-stage decisions. The resulting problems constitute nonsmooth convex optimization problems where the objective function is evaluated by solving several mixed-integer single-scenario problems. The problem can be solved heuristically with nondifferentiable optimization techniques and subsequent rounding of the first-stage decisions, or it can be solved exactly through branch & bound algorithms or successive elimination of candidate solutions, see Carøe and Schultz (1999), Alonso-Ayuso et al. (2003) and Ahmed (2013). Other solution schemes for stochastic integer programs include stochastic branch & bound algorithms (Norkin et al. 1998), enumeration schemes using Gröbner basis methods from computational algebra (Schultz et al. 1998) and approaches that construct the convex envelope of the expected recourse function (Klein Haneveld et al. 1995, 1996). For a detailed review of the stochastic integer programming literature, we refer to van der Vlerk (1996–2007), Louveaux and Schultz (2003), Schultz (2003) and Romeijnders et al. (2014).

Approximation algorithms for two-stage stochastic binary programs appear to have first been developed by Dye et al. (2003) for a variant of the service-provisioning problem. The authors propose a linear programming (LP) based rounding scheme which affords a constant worst-case performance ratio. Ravi and Sinha (2004) adapt the approximation algorithms for various deterministic binary problems to their two-stage stochastic counterparts and obtain constant-factor approximation guarantees for the stochastic problem variants. Immorlica et al. (2004) study several classes of two-stage stochastic covering problems where the second-stage costs are multiples of the first-stage costs. The resulting problems possess a threshold property which allows to characterize whether a particular decision should be taken before or after the uncertain problem parameters are revealed. Shmoys and Swamy (2006) develop approximation algorithms for two-stage stochastic integer programs that access the probability distribution indirectly through a sampling oracle. The algorithms first solve an LP-relaxation of the two-stage stochastic integer program and then use an approximation scheme for the deterministic problem variant to round the fractional solutions. Swamy (2011) extends this result to risk-averse two-stage stochastic binary programs by replacing the expectation with a value-at-risk operator. Multi-stage extensions of the approximation algorithms are proposed by Gupta et al. (2005), Srinivasan (2007) and Swamy and Shmoys (2012).

The literature on robust optimization problems with discrete recourse decisions, on the other hand, is relatively sparse. Bertsimas and Goyal (2010) study the adaptability gap in two-stage robust mixed-integer linear programs (MILPs), which they define as the difference in objective values between the optimal solution and the best static solution (*i.e.*, where all decisions are taken here-and-now). The authors show that for certain classes of symmetric and nonnegative uncertainty sets, the adaptability gap is bounded by a factor of two if only the constraint right-hand sides are uncertain, whereas the gap increases to a factor of four if both the objective coefficients and the constraint right-hand sides are uncertain. The results have been generalized to asymmetric uncertainty sets by Bertsimas et al. (2011b).

Vayanos et al. (2011) develop a conservative approximation for multi-stage robust MILPs. The authors partition the uncertainty set Ξ into hyperrectangles and restrict the continuous and binary recourse decisions to affine and constant functions of $\boldsymbol{\xi}$ over each hyperrectangle, respectively. The resulting conservative approximation can be reformulated as a MILP, see also Gorissen et al. (2013).

Bertsimas and Caramanis (2007) present an approximation scheme for multi-stage robust MILPs where the constraints are satisfied with high probability. The authors restrict the recourse decisions to weighted linear combinations of basis functions of $\boldsymbol{\xi}$, and they obtain a finite-dimensional optimization problem through constraint sampling. In order to satisfy the integrality constraints, the authors suggest to restrict the weights and the images of the basis functions to integers.

Bertsimas and Georghiou (2013) propose an iterative approach to solve multi-stage robust MILPs with fixed recourse. The authors restrict the continuous recourse decisions to continuous and piecewise affine functions of $\boldsymbol{\xi}$. Binary recourse decisions take the value 1 for a specific realization of $\boldsymbol{\xi}$ whenever an associated continuous and piecewise affine function of $\boldsymbol{\xi}$ is nonpositive. The authors solve the problem with a cutting plane algorithm akin to semi-infinite programming schemes.

While this paper was under review, Bertsimas and Georghiou (2014) combined the ideas of Bertsimas and Caramanis (2007) and Georghiou et al. (2014) to develop piecewise constant binary decision rules for multi-stage robust MILPs with random recourse. While their generic problem reformulation scales exponentially in the number of uncertain problem parameters, the authors propose a polynomial-size MILP reformulation for the special case where the recourse decisions are restricted to linear combinations of translated Heaviside step functions.

In the related literature, there has been significant recent progress on approximation algorithms for specific classes of two-stage robust combinatorial problems. Dhamdhere et al. (2005) propose an approximation algorithm for several classes of two-stage robust covering problems where the uncertainty set contains a finite number of discrete scenarios. The algorithm provides constant-factor approximations for the Steiner tree, vertex cover and facility location problems as well as $\mathcal{O}(\log n)$- and $\mathcal{O}(\log n \log^2 n)$-approximations for the min-cut and min multi-cut problem, respectively. Feige et al. (2007) study two-stage robust covering problems where the uncertainty set comprises all subsets of a universe set with cardinality less than or equal to a prespecified constant. The authors propose an LP-based rounding scheme that affords an $\mathcal{O}(\log n \log m)$-approximation for the robust set covering problem, where n is the cardinality of the universe set and m is the number of cover sets. Gupta et al. (2010) improve this bound to $\mathcal{O}(\log n + \log m)$ by designing a guess-and-prune method that first guesses the worst-case second-stage costs and then identifies the set of costly elements that needs to be covered here-and-now. A similar guess-and-prune method has been proposed by Golovin et al. (2014) to provide a 2-approximation for the min-cut problem and a 3.39-approximation to the shortest path problem, respectively. Khandekar et al.

(2013) observe that the LP-based rounding scheme of Feige et al. (2007) does not seem to yield good approximation guarantees for certain classes of two-stage robust network design problems and propose constant-factor approximation algorithms for such problems. Multi-stage extensions of the approximation algorithms are proposed by Gupta et al. (2013).

There is also a rich literature on persistency in distributionally robust optimization, see Li et al. (2014). Here, $\boldsymbol{\xi}$ is modeled as a random vector that is governed by a probability distribution which is only partially known. The goal is to determine the expected optimal value of a mixed-integer problem depending on $\boldsymbol{\xi}$, as well as the probability that a particular binary variable attains the value 1 at optimality, under the most favorable probability distribution that is consistent with the available information.

In this paper, we study generic two-stage robust binary programs of the form

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\boldsymbol{\xi}\in\Xi}\left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\boldsymbol{y}\in\mathcal{Y}}\left\{\boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y} : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \le \boldsymbol{H}\boldsymbol{\xi}\right\}\right] \\ \text{subject to} & \boldsymbol{x}\in\mathcal{X}, \end{array} \tag{$\mathcal{P}$}$$

where $\mathcal{X}\subseteq\mathbb{R}^N_+$ and $\mathcal{Y}\subseteq\{0,1\}^M$ are bounded polyhedral sets,[1] $\Xi\subseteq\mathbb{R}^Q$, $\boldsymbol{C}\in\mathbb{R}^{Q\times N}$, $\boldsymbol{Q}\in\mathbb{R}^{Q\times M}$, $\boldsymbol{T}\in\mathbb{R}^{L\times N}$, $\boldsymbol{W}\in\mathbb{R}^{L\times M}$ and $\boldsymbol{H}\in\mathbb{R}^{L\times Q}$. The decisions $\boldsymbol{x}$ are here-and-now (or first-stage) decisions that are taken before the realization of the uncertain parameters $\boldsymbol{\xi}\in\Xi$ is known, whereas the wait-and-see (or second-stage) decisions $\boldsymbol{y}$ can adapt to the realization of $\boldsymbol{\xi}$. While the here-and-now decision $\boldsymbol{x}$ may contain continuous and binary components, we require all components of the wait-and-see decision $\boldsymbol{y}$ to be binary, that is, we study problems with pure binary recourse. The uncertainty set Ξ is described by a nonempty bounded polyhedron of the form $\Xi=\{\boldsymbol{\xi}\in\mathbb{R}^Q : \boldsymbol{A}\boldsymbol{\xi}\le\boldsymbol{b}\}$, where $\boldsymbol{A}\in\mathbb{R}^{R\times Q}$ and $\boldsymbol{b}\in\mathbb{R}^R$. Note that both the objective function and the constraint right-hand sides are *linear* in $\boldsymbol{\xi}$. We can account for *affine* dependencies on $\boldsymbol{\xi}$ by introducing an auxiliary parameter $\boldsymbol{\xi}_{Q+1}$ and augmenting the uncertainty set Ξ with the constraint $\boldsymbol{\xi}_{Q+1}=1$. As we will elaborate later on, the methods presented in this paper also extend to affine dependencies of the recourse matrix $\boldsymbol{W}$ and the technology matrix $\boldsymbol{T}$ on $\boldsymbol{\xi}$.

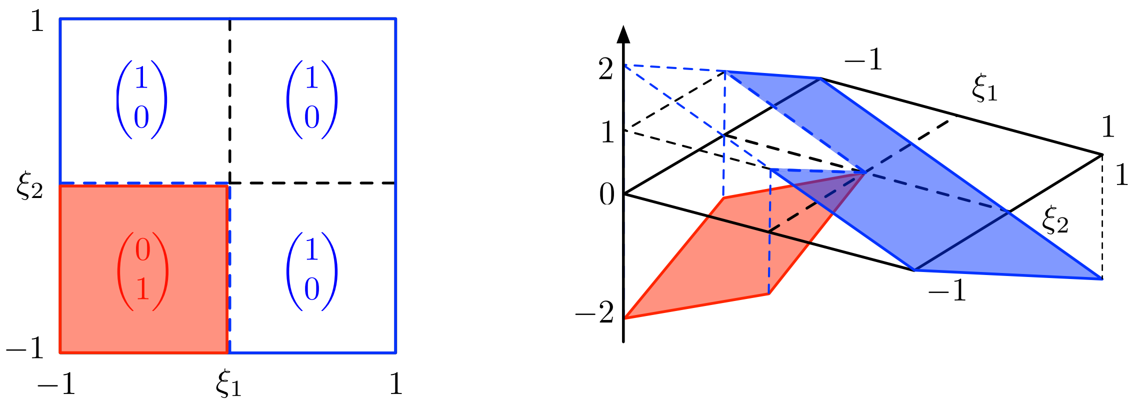

Figure 1 Optimal wait-and-see decisions $\boldsymbol{y}$ (left) and the associated objective function (right) for problem (2). The decision $\boldsymbol{y} = (1,0)^\top$ is optimal whenever $\boldsymbol{\xi}_1 > 0$ or $\boldsymbol{\xi}_2 > 0$, as well as for $\boldsymbol{\xi} = \mathbf{0}$, whereas the decision $\boldsymbol{y} = (0,1)^\top$ is optimal for $\boldsymbol{\xi} \leq \mathbf{0}$. The objective function is discontinuous on the set $\{\boldsymbol{\xi} \in [-1,0]^2 : \boldsymbol{\xi}_1\boldsymbol{\xi}_2 = 0, \boldsymbol{\xi} \neq \mathbf{0}\}$.

Problem $\mathcal{P}$ has diverse applications ranging from operations management (*e.g.*, facility location, layout planning, vehicle routing as well as production and project scheduling) to investment planning (*e.g.*, project selection) and game theory (*e.g.*, network fortification games). Unfortunately, the problem poses severe theoretical and computational challenges.

Example 1 *Consider the following instance of problem $\mathcal{P}$ where $N = 0$, that is, no first stage decision is taken.*

$$\begin{aligned} \text{maximize} \quad & \min_{\boldsymbol{y} \in \{0,1\}^2} \left\{ (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2)(\boldsymbol{y}_2 - \boldsymbol{y}_1) : \boldsymbol{y}_1 + \boldsymbol{y}_2 = 1, \right. \\ & \qquad\qquad\qquad \left. \boldsymbol{y}_1 \geq \boldsymbol{\xi}_1,\ \boldsymbol{y}_1 \geq \boldsymbol{\xi}_2 \right\} \\ \text{subject to} \quad & \boldsymbol{\xi} \in \mathbb{R}^2 \\ & -1 \leq \boldsymbol{\xi}_q \leq 1,\ q = 1,2 \end{aligned} \tag{2}$$

Figure 1 illustrates the optimal wait-and-see decision $\boldsymbol{y}$ as a function of $\boldsymbol{\xi}$, as well as the associated objective function. The figure shows that the parameter region where a particular wait-and-see decision is optimal can be non-closed and nonconvex. Moreover, the objective function can be both nonconcave and discontinuous in $\boldsymbol{\xi}$. As a result, the optimal value of the problem may not be attained. In fact, the supremum of problem (2) *is* 1*, and it is attained in the limit by the parameter* *sequences* $\boldsymbol{\xi}_n = (1/n, -1)^\top$ *and* $\boldsymbol{\xi}_n = (-1, 1/n)^\top$, $n \in \mathbb{N}$. *If we were to solve a static problem in which the decision* $\boldsymbol{y}$ *is taken before the realization of* $\boldsymbol{\xi}$ *is known, then* $\boldsymbol{y} = (1,0)^\top$ *would be the only feasible decision, resulting in an objective value of* 2.

Instead of solving problem $\mathcal{P}$ directly, we propose to solve its associated K-adaptability problem,

$$\begin{array}{ll} \text{minimize} & \max_{\boldsymbol{\xi}\in\Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\boldsymbol{x} + \min_{k\in\mathcal{K}} \left\{ \boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{y}^k \,:\, \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{H}\boldsymbol{\xi} \right\} \right] \\ \text{subject to} & \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}^k \in \mathcal{Y},\ k\in\mathcal{K}, \end{array} \tag{$\mathcal{P}_K$}$$

where $\mathcal{K} = \{1, \ldots, K\}$. In this problem, we determine K non-adjustable second-stage policies $\boldsymbol{y}^k$ here-and-now, that is, before the value of $\boldsymbol{\xi}$ is known. Once the value of $\boldsymbol{\xi}$ is observed, the best policy among the feasible ones is implemented. If all policies are infeasible for some $\boldsymbol{\xi} \in \Xi$, then we interpret the maximum and minimum in $\mathcal{P}_K$ as supremum and infimum, that is, the K-adaptability problem evaluates to $+\infty$. Problem $\mathcal{P}_K$ is a conservative approximation of the two-stage robust binary program $\mathcal{P}$, and both problems have the same optimal value if $K = |\mathcal{Y}| < \infty$. The hope is that in practice, a much smaller number of second-stage policies $\boldsymbol{y}^k$ suffices to closely approximate the optimal solution to $\mathcal{P}$. We note that problem $\mathcal{P}_K$ may be of interest in its own right. In fact, in applications where two-stage problems are solved repeatedly (*e.g.*, vehicle routing and production scheduling), complete flexibility in the second stage may be too expensive as it leads to a high process variability. Instead, a small number of alternative plans may be sought, the best of which is implemented once further information about the uncertain problem parameters becomes known. A similar point can be made for emergency response planning, where a small number of contingency plans may be preferred to the impromptu solution of a second-stage optimization problem.

The K-adaptability problem $\mathcal{P}_K$ has been first proposed by Bertsimas and Caramanis (2010). In that paper, the authors relate the difference between the optimal values of problems $\mathcal{P}$ and $\mathcal{P}_K$ to correlations between the uncertain coefficients in the second-stage constraints, and they provide necessary conditions for the K-adaptability problem to improve upon the static formulation where all decision are taken here-and-now. The authors also develop a reformulation of the K-adaptability problem as a finite-dimensional bilinear program for the case where $K = 2$.

This paper aims to provide new insights into the theoretical and computational properties of the K-adaptability problem. To this end, we characterize the problem's complexity in terms of the number of second-stage policies K needed to recover the original two-stage robust binary program $\mathcal{P}$, as well as the effort required to evaluate the objective function in $\mathcal{P}_K$ for a fixed here-and-now decision. It turns out that in both cases, the watershed of computational tractability is marked by the presence of the parameters $\boldsymbol{\xi}$ in the constraints. We also derive explicit MILP reformulations for the K-adaptability problem with objective and constraint uncertainty. To our best knowledge, we present the first approximation scheme for two-stage robust binary programs that scales to practically relevant problems and that avoids the intricacies of sampling methods.

The contributions of this paper can be summarized as follows.

1. We analyze the approximation quality of problem $\mathcal{P}_K$. We find that two-stage robust binary programs can be solved exactly by solving the corresponding K-adaptability problems with $K = \min\{M, Q\} + 1$ policies if the parameters $\boldsymbol{\xi}$ only enter the objective function of the problem. If, on the other hand, the constraints depend on $\boldsymbol{\xi}$, then $K = |\mathcal{Y}|$ policies may be required for an exact solution of the two-stage robust binary program.

2. We study the computational complexity of problem $\mathcal{P}_K$. We find that evaluating the objective function in $\mathcal{P}_K$ is tractable if the parameters $\boldsymbol{\xi}$ only enter the objective function of the problem, whereas the evaluation may become strongly $\mathcal{NP}$-hard if the constraints depend on $\boldsymbol{\xi}$.

3. We derive MILP reformulations for problem $\mathcal{P}_K$ with objective and constraint uncertainty. Our results generalize the MILP formulation for the 2-adaptability problem that has been proposed by Bertsimas and Caramanis (2010). In comparison to the latter formulation, our model extends to any number of policies K, and the model size grows polynomially in the description of Ξ.

We should also point out two objectives that the paper *fails* to achieve. First and foremost, our formulations do not readily extend to continuous recourse decisions. We believe that such decisions could be incorporated via decision rules of the type presented by Ben-Tal et al. (2004), Chen and Zhang (2009), Goh and Sim (2010), Kuhn et al. (2011) and Georghiou et al. (2014). For the sake of brevity, however, we leave this extension as future work. Secondly, we have not been able to extend our formulations to multi-stage models without incurring an exponential growth in problem size. We thus regard this paper as a first step towards the solution of two-stage and multi-stage robust integer programs, and we hope that it will spur further research into this challenging problem class.

The remainder of the paper is structured as follows. We study the K-adaptability problem with objective and constraint uncertainty in Sections 2 and 3, respectively. In both sections, we investigate the approximation quality and the computational complexity of the respective problem, and we provide a MILP reformulation of the problem. We present the results of numerical experiments in Section 4, and we conclude in Section 5. All proofs are relegated to the Electronic Companion.

Notation We use bold lower-case and upper-case letters for vectors and matrices, while scalars are printed in regular font. Random objects are notationally distinguished from deterministic quantities by a tilde sign. We denote by $\mathbf{e}_k$ the kth canonical basis vector, while $\mathbf{e}$ denotes the vector whose components are all ones. By a slight abuse of notation, we sometimes use the maximum and minimum operators even when attainment of the optimum is not guaranteed; in such cases, the operators should be understood as suprema and infima, respectively. Finally, for a logical expression $\mathcal{E}$, we define the indicator function $\mathbb{I}[\mathcal{E}]$ as $\mathbb{I}[\mathcal{E}] = 1$ if $\mathcal{E}$ is true and 0 otherwise.

2. The K-Adaptability Problem with Objective Uncertainty

In this section, we assume that the parameters $\boldsymbol{\xi}$ only enter the objective function of the two-stage robust binary program $\mathcal{P}$:

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\xi}\in\Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\boldsymbol{x} + \min_{\boldsymbol{y}\in\mathcal{Y}} \left\{\boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{y} : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \le \boldsymbol{h}\right\}\right] \\ &\text{subject to} && \boldsymbol{x}\in\mathcal{X}, \end{aligned} \tag{$\mathcal{PO}$}$$

where $\boldsymbol{h}\in\mathbb{R}^L$. We study the K-adaptability problem associated with $\mathcal{PO}$:

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\xi}\in\Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\boldsymbol{x} + \min_{k\in\mathcal{K}} \left\{\boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{y}^k : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \le \boldsymbol{h}\right\}\right] \\ &\text{subject to} && \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}^k\in\mathcal{Y},\ k\in\mathcal{K} \end{aligned} \tag{3}$$

For several reasons, problem (3) constitutes a natural starting point for our investigation. Firstly, the MILP reformulation of (3) requires similar techniques as the MILP reformulation of the generic K-adaptability problem $\mathcal{P}_K$ while avoiding some of the technicalities that arise in the latter case. Secondly, problem (3) enjoys superior approximation and complexity properties that are lost once we study the generic problem. Finally and perhaps most importantly, problem (3) arises naturally in a number of application domains, such as traveling salesman and vehicle routing problems with uncertain travel times, network expansion problems with uncertain costs, facility location problems with uncertain future customer demands and layout planning problems with uncertain future production quantities.

Remark 1 (Progressive Approximation of $\mathcal{PO}$) *Since the second-stage constraints in problem $\mathcal{PO}$ do not depend on $\boldsymbol{\xi}$, we can derive a lower (progressive) bound on the optimal value of $\mathcal{PO}$ by disregarding the integrality requirement for $\boldsymbol{y}$, applying the classical min-max theorem to exchange the order of the maximization problem over $\boldsymbol{\xi} \in \Xi$ and the minimization problem over $\boldsymbol{y} \in \mathcal{Y}$ and subsequently dualizing the maximization problem. This progressive bound is tight whenever the second-stage constraints are totally unimodular in $\boldsymbol{y}$ for every $\boldsymbol{x} \in \mathcal{X}$, in which case the integrality requirement for $\boldsymbol{y}$ is superfluous. In our numerical experiments, we will use this progressive approximation of $\mathcal{PO}$ to obtain upper (conservative) bounds on the loss of optimality when we approximate problem $\mathcal{PO}$ by problem* (3)*.*

We can simplify problem (3) by moving the second-stage constraints to the first stage.

Observation 1 *Problem* (3) *is equivalent to*

$$\begin{array}{ll} \text{minimize} & \max\limits_{\boldsymbol{\xi} \in \Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C} \boldsymbol{x} + \min\limits_{k \in \mathcal{K}} \boldsymbol{\xi}^\top \boldsymbol{Q} \boldsymbol{y}^k \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{y}^k \in \mathcal{Y}, k \in \mathcal{K} \\ & \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{h} \qquad \forall k \in \mathcal{K}. \end{array} \tag{$\mathcal{PO}_K$}$$

Remark 2 *A solution* $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}}) \in \mathcal{X} \times \mathcal{Y}^K$ *is feasible in problem* $\mathcal{PO}_K$ *if it satisfies the second-stage constraints* $\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{h}$ *for* all $k \in \mathcal{K}$*, whereas it is feasible in problem* (3) *if it satisfies the second-stage constraints for* some $k \in \mathcal{K}$*.*

The parameter realizations $\boldsymbol{\xi}$ for which the k-th policy $\boldsymbol{y}^k$ is optimal are

$$\left\{ \boldsymbol{\xi} \in \Xi : \boldsymbol{\xi}^\top \boldsymbol{Q}\, \boldsymbol{y}^k \leq \boldsymbol{\xi}^\top \boldsymbol{Q}\, \boldsymbol{y}^{k'} \;\; \forall k' \in \mathcal{K} \right\}. \tag{4}$$

In contrast to Example 1, where the parameters $\boldsymbol{\xi}$ enter the constraints of the problem, the set (4) constitutes a closed and convex polyhedron. Moreover, one can readily show that the objective function of $\mathcal{PO}_K$ is continuous in $\boldsymbol{x}$ and $\boldsymbol{y}^k$, $k \in \mathcal{K}$. Since the feasible region of $\mathcal{PO}_K$ is compact, $\mathcal{PO}_K$ thus attains its minimum whenever the problem is feasible. Note that the objective function in $\mathcal{PO}_K$ is convex in $\boldsymbol{x}$, but it typically fails to be convex in $\boldsymbol{y}^k$. One readily verifies that the objective function in $\mathcal{PO}_K$ can be evaluated through a polynomial-time solvable linear program (LP) if we replace the inner minimization with an epigraph formulation.[2] For later reference, we make this statement explicit.

Observation 2 *For a fixed decision* $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$*, the objective function in* $\mathcal{PO}_K$ *can be evaluated in polynomial time.*

By construction, the K-adaptability problem $\mathcal{PO}_K$ constitutes a conservative approximation to the two-stage robust binary program $\mathcal{PO}$, that is, the optimal value of $\mathcal{PO}_K$ bounds the optimal value of $\mathcal{PO}$ from above. It is then natural to ask how much adaptability is required so that both problems have identical optimal values.

Theorem 1 *The* K*-adaptability problem* $\mathcal{PO}_K$ *has the same optimal value as the two-stage robust binary program* $\mathcal{PO}$ *if we choose* $K \geq \min\{\dim \mathcal{Y}, \operatorname{rk} \boldsymbol{Q}\} + 1$ *policies, where* $\dim \mathcal{Y}$ *denotes the affine dimension of* $\mathcal{Y}$ *and* $\operatorname{rk} \boldsymbol{Q}$ *the row rank of the matrix* $\boldsymbol{Q}$*.*

Remark 3 *By construction, we have* $\dim \mathcal{Y} \leq M$ *and* $\operatorname{rk} \boldsymbol{Q} \leq Q$*. Also, we can assume that* $\operatorname{rk} \boldsymbol{Q} \leq \dim \Xi + 1$*. Otherwise, one can show that there is a matrix* $\boldsymbol{Q}'$ *such that* $\operatorname{rk} \boldsymbol{Q}' \leq \dim \Xi + 1$ *and the optimal value and the optimal solutions to* $\mathcal{PO}_K$ *do not change if we replace* $\boldsymbol{Q}$ *with* $\boldsymbol{Q}'$*.*

Theorem 1 raises hope that the K-adaptability problem $\mathcal{PO}_K$ serves as a good approximation of the two-stage robust binary program $\mathcal{PO}$ even if the number of second-stage policies K is small. This distinguishes two-stage robust binary programs from their stochastic counterparts.

Example 2 *Consider the following two-stage stochastic program without a first-stage decision:*

$$\mathbb{E}_{\mathbb{P}}\left[\min_{\boldsymbol{y}\in\{0,1\}^Q} \tilde{\boldsymbol{\xi}}^\top \boldsymbol{y}\right], \qquad \text{where } \mathbb{P}\left[\tilde{\boldsymbol{\xi}}=\boldsymbol{\xi}\right] = 2^{-Q}\cdot \mathbb{I}\left[\boldsymbol{\xi}\in\{-1,1\}^Q\right].$$

Here, the expectation is taken with respect to the uniform distribution on the vertices of the $-1/+1$ hypercube in $\mathbb{R}^Q$. For a given parameter realization $\boldsymbol{\xi}\in\{-1,1\}^Q$, $\boldsymbol{y}=\sum_{q=1}^Q \mathbb{I}[\boldsymbol{\xi}_q=-1]\,\mathbf{e}_q$ is the unique optimal second-stage decision. We thus conclude that the corresponding stochastic K-adaptability problem only attains the same optimal value if $K=2^Q$, that is, if all policies $\boldsymbol{y}\in\{0,1\}^Q$ are considered.

2.1. Reformulation as a Mixed-Integer Linear Program

In the previous section we have shown that the evaluation of the objective function in $\mathcal{PO}_K$ can be formulated as an LP if the decisions $\boldsymbol{x}$ and $\boldsymbol{y}^k$, $k\in\mathcal{K}$, are fixed. If we dualize this LP and linearize the ensuing bilinear terms, then we obtain an equivalent MILP reformulation of the problem $\mathcal{PO}_K$.

Theorem 2 *Problem $\mathcal{PO}_K$ is equivalent to the following MILP.*

$$\begin{aligned}
\text{minimize}\quad & \boldsymbol{b}^\top\boldsymbol{\alpha} \\
\text{subject to}\quad & \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}^k\in\mathcal{Y},\ k\in\mathcal{K}, \\
& \boldsymbol{z}^k\in\mathbb{R}^M_+,\ k\in\mathcal{K},\ \boldsymbol{\alpha}\in\mathbb{R}^R_+,\ \boldsymbol{\beta}\in\mathbb{R}^K_+ \\
& \boldsymbol{A}^\top\boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{Q}\boldsymbol{z}^k,\ \mathbf{e}^\top\boldsymbol{\beta}=1 \\
& \left.\begin{array}{l}
\boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{h} \\
\boldsymbol{z}^k\le\boldsymbol{y}^k,\ \boldsymbol{z}^k\le\boldsymbol{\beta}_k\mathbf{e} \\
\boldsymbol{z}^k\ge(\boldsymbol{\beta}_k-1)\mathbf{e}+\boldsymbol{y}^k
\end{array}\right\}\ \forall k\in\mathcal{K}.
\end{aligned} \tag{5}$$

We stress that the size of the MILP in the statement of Theorem 2 grows polynomially in the size of the input data for the K-adaptability problem $\mathcal{PO}_K$.

Remark 4 (Integer Recourse Decisions) *The restriction to binary recourse decisions in $\mathcal{PO}$ is motivated by the proof of Theorem 2, which employs an exact linearization of the bilinear terms that emerge from dualizing the objective function of the K-adaptability problem $\mathcal{PO}_K$. This is in contrast to the stochastic integer programming literature, where binary recourse decisions are sometimes required to approximate the convex hull of the second-stage problem. Sherali and Fraticelli (2002), for example, assume binarity to use lift-and-project cuts, while Sen and Higle (2005) require binarity to guarantee the finite convergence of their cutting plane scheme.*

3. The K-Adaptability Problem with Constraint Uncertainty

We now study the generic K-adaptability problem $\mathcal{P}_K$, where the parameter vector $\boldsymbol{\xi}$ enters both the objective function and the constraint right-hand sides. Throughout this section, we assume that $\mathcal{X} \subseteq \{0,1\}^N$. We remark that the reformulations presented in this section extend to affine dependencies of the recourse matrix $\boldsymbol{W}$ the technology matrix $\boldsymbol{T}$ on $\boldsymbol{\xi}$. We elaborate on these generalizations in Remark 8 at the end of this section.

Contrary to problem $\mathcal{PO}_K$, where $\boldsymbol{\xi}$ only enters the objective function, the generic K-adaptability problem $\mathcal{P}_K$ is less well-behaved. Indeed, we have seen in Example 1 that the parameter realizations $\boldsymbol{\xi}$ for which a particular policy $\boldsymbol{y}^k$ is optimal can form a non-closed and nonconvex region. Moreover, the objective function in $\mathcal{P}_K$ may be nonconvex and discontinuous. Also, it is considerably more challenging to obtain a good progressive approximation of the problem.

Remark 5 (Progressive Approximation of $\mathcal{P}$) *In contrast to problem $\mathcal{PO}$, it is difficult to derive a good lower (progressive) bound on the optimal value of $\mathcal{P}$ by disregarding the integrality of the second-stage decisions. In fact, we cannot employ the classical min-max theorem to exchange the order of the maximization over $\boldsymbol{\xi} \in \Xi$ and the minimization over $\boldsymbol{y} \in \mathcal{Y}$ due to the coupling of $\boldsymbol{y}$ and $\boldsymbol{\xi}$ in the constraints. Alternatively, we could dualize the minimization over $\boldsymbol{y} \in \mathcal{Y}$ and subsequently dualize the maximization over $\boldsymbol{\xi} \in \Xi$. The second dualization would involve a nonconvex problem, however, and the resulting duality gap turns out to be large in our experiments. Instead, we can obtain a lower bound on the optimal value of $\mathcal{P}$ by discretizing the uncertainty set Ξ into a finite* *set of scenarios and solving the resulting scenario-based two-stage robust optimization problem as a MILP (Hadjiyiannis et al. 2011, Bertsimas and Georghiou 2013). In Section 4, we will use this progressive approximation of $\mathcal{P}$ to obtain upper (conservative) bounds on the loss of optimality when we approximate problem $\mathcal{P}$ by problem $\mathcal{P}_K$.*

Observation 2 states that the objective function in $\mathcal{PO}_K$ can be evaluated in polynomial time. We now show that this is no longer the case for the generic K-adaptability problem $\mathcal{P}_K$.

Theorem 3 *Evaluating the objective function in problem $\mathcal{P}_K$*

(i) can be done in polynomial time up to any accuracy if K is fixed, and

(ii) is strongly $\mathcal{NP}$-hard otherwise.

Theorem 1 has shown that if $\boldsymbol{\xi}$ only enters the objective function of the two-stage robust binary program $\mathcal{P}$, then the associated K-adaptability problem with $K = \min\{\dim \mathcal{Y}, \operatorname{rk} \boldsymbol{Q}\} + 1$ policies attains the same optimal value. In contrast, we now show that every feasible policy $\boldsymbol{y} \in \mathcal{Y}$ may be required in the K-adaptability problem if $\boldsymbol{\xi}$ enters the constraints of $\mathcal{P}$.

Theorem 4 *The K-adaptability problem $\mathcal{P}_K$ may attain a strictly higher optimal value than the two-stage robust binary program $\mathcal{P}$ for any number of policies $K < |\mathcal{Y}|$.*

3.1. Reformulation as a Mixed-Integer Linear Program

In this section, we derive a MILP formulation for the generic K-adaptability problem $\mathcal{P}_K$. We have seen in Example 1 that the presence of $\boldsymbol{\xi}$ in the constraints implies that problem $\mathcal{P}_K$ may not attain its optimal value. As such, it is not surprising that our MILP formulation is not exact. Instead, we derive an ϵ-approximation that converges to problem $\mathcal{P}_K$ in a meaningful way as the approximation parameter ϵ approaches 0.

We will derive our MILP formulation in several steps. We first reformulate the generic K-adaptability problem $\mathcal{P}_K$ as a variant of the problem $\mathcal{PO}_K$ where the uncertainty set Ξ is parameterized by a vector $\boldsymbol{\ell}$ that encodes which second-stage policies $\boldsymbol{y}^k$ are feasible. Since the resulting uncertainty sets $\Xi(\boldsymbol{\ell})$ fail to be closed, we replace them with closed inner approximations $\Xi_\epsilon(\boldsymbol{\ell})$ that are parameterized by $\epsilon > 0$. We then show that both the objective functions and the optimal values of the approximate optimization problems converge to the objective function and the optimal value of the exact problem as ϵ approaches 0. Finally, we provide a MILP formulation for the approximate optimization problems.

We begin with a reformulation of the problem $\mathcal{P}_K$ that shifts the second-stage constraints $\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{H}\boldsymbol{\xi}$ from the objective function to the definition of the uncertainty set. To this end, we replace Ξ with a family of uncertainty sets parameterized by a vector $\boldsymbol{\ell}$.

Proposition 1 *The K-adaptability problem $\mathcal{P}_K$ is equivalent to*

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\ell} \in \mathcal{L}} \max_{\boldsymbol{\xi} \in \Xi(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\boldsymbol{x} + \min_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k = 0}} \boldsymbol{\xi}^\top \boldsymbol{Q}\boldsymbol{y}^k \right] \\ &\text{subject to} && \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y}^k \in \mathcal{Y},\ k \in \mathcal{K}, \end{aligned} \tag{6}$$

where $\mathcal{L} = \{0, \ldots, L\}^K$, L is the number of second-stage constraints in the objective function of $\mathcal{P}_K$, and the uncertainty sets $\Xi(\boldsymbol{\ell})$, $\boldsymbol{\ell} \in \mathcal{L}$, are defined as

$$\Xi(\boldsymbol{\ell}) = \left\{ \boldsymbol{\xi} \in \Xi : \begin{array}{ll} \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{H}\boldsymbol{\xi} & \forall k \in \mathcal{K} : \boldsymbol{\ell}_k = 0 \\ [\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} > [\boldsymbol{H}\boldsymbol{\xi}]_{\boldsymbol{\ell}_k} & \forall k \in \mathcal{K} : \boldsymbol{\ell}_k \neq 0 \end{array} \right\}.$$

For ease of exposition, we notationally suppress the dependence of $\Xi(\boldsymbol{\ell})$ on $\boldsymbol{x}$ and $\{\boldsymbol{y}^k\}_{k \in \mathcal{K}}$.

Remark 6 *The components of $\boldsymbol{\ell} \in \mathcal{L}$ encode which second-stage policies are feasible for the parameter realizations $\boldsymbol{\xi} \in \Xi(\boldsymbol{\ell})$. In particular, policy $\boldsymbol{y}^k$ is feasible if $\boldsymbol{\ell}_k = 0$, whereas it violates the $\boldsymbol{\ell}_k$-th second-stage constraint in the objective function of problem $\mathcal{P}_K$ if $\boldsymbol{\ell}_k \neq 0$. Although a policy can violate multiple constraints, it is sufficient to record one of those violations.*

Problem (6) *resembles problem $\mathcal{PO}_K$, the K-adaptability problem with objective uncertainty. In contrast to $\mathcal{PO}_K$, however,* (6) *involves multiple uncertainty sets, and the shapes of those uncertainty sets depend on the decisions $\boldsymbol{x}$ and $\boldsymbol{y}^k$. Robust optimization problems with decision-dependent uncertainty set have been explored by Spacey et al. (2012) and Vayanos et al. (2011).*

Example 3 *The following instance of problem* (6) *corresponds to the 2-adaptability problem associated with problem* (2) *in Example 1:*

$$\begin{aligned} \text{minimize} \quad & \max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \min_{\substack{k\in\mathcal{K}: \\ \boldsymbol{\ell}_k=0}} \left[(\boldsymbol{\xi}_1+\boldsymbol{\xi}_2)(\boldsymbol{y}_2^k-\boldsymbol{y}_1^k)\right] \\ \text{subject to} \quad & \boldsymbol{y}^1,\boldsymbol{y}^2\in\{0,1\}^2 \\ & \boldsymbol{y}_1^k+\boldsymbol{y}_2^k=1 \qquad\qquad \forall k=1,2, \end{aligned}$$

where $\mathcal{L}=\{0,1,2\}^2$ *and*

$$\Xi(\boldsymbol{\ell})=\left\{\boldsymbol{\xi}\in[-1,1]^2 : \begin{array}{lll} \boldsymbol{y}_1^k\geq\boldsymbol{\xi}_1 \text{ if } \boldsymbol{\ell}_k=0, & \boldsymbol{y}_1^k<\boldsymbol{\xi}_1 \text{ if } \boldsymbol{\ell}_k=1, & k=1,2, \\ \boldsymbol{y}_1^k\geq\boldsymbol{\xi}_2 \text{ if } \boldsymbol{\ell}_k=0, & \boldsymbol{y}_1^k<\boldsymbol{\xi}_2 \text{ if } \boldsymbol{\ell}_k=2, & k=1,2 \end{array}\right\}.$$

For a fixed decision $(\boldsymbol{x},\{\boldsymbol{y}^k\}_{k\in\mathcal{K}})\in\mathcal{X}\times\mathcal{Y}^K$, the objective function in problem (6) evaluates to $+\infty$ if and only if there is $\boldsymbol{\ell}\in\mathcal{L}$, $\boldsymbol{\ell}>\boldsymbol{0}$, such that $\Xi(\boldsymbol{\ell})\neq\emptyset$. In fact, for any $\boldsymbol{\ell}>\boldsymbol{0}$ and $\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})$, the minimization over $k\in\mathcal{K}$ evaluates to $+\infty$. Conversely, the minimization over $k\in\mathcal{K}$ attains finite values for all $\boldsymbol{\ell}\in\mathcal{L}$, $\boldsymbol{\ell}\not>\boldsymbol{0}$, which implies that the objective function in (6) attains a finite value whenever $\Xi(\boldsymbol{\ell})=\emptyset$ for all $\boldsymbol{\ell}>\boldsymbol{0}$. Note also that by construction of $\Xi(\boldsymbol{\ell})$, we have $\Xi(\boldsymbol{\ell})\neq\emptyset$ for $\boldsymbol{\ell}>\boldsymbol{0}$ if and only if there are parameter realizations $\boldsymbol{\xi}\in\Xi$ for which the decision $(\boldsymbol{x},\{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ violates the second-stage constraints in problem $\mathcal{P}_K$. Thus, the objective function in (6) attains a finite value for the decision $(\boldsymbol{x},\{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ if and only if $(\boldsymbol{x},\{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ is feasible in the problem $\mathcal{P}_K$. In the following, we will use the index sets $\partial\mathcal{L}=\{\boldsymbol{\ell}\in\mathcal{L}:\boldsymbol{\ell}\not>\boldsymbol{0}\}$ and $\mathcal{L}_+=\{\boldsymbol{\ell}\in\mathcal{L}:\boldsymbol{\ell}>\boldsymbol{0}\}$ to refer to the uncertainty sets for which the decision $(\boldsymbol{x},\{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ satisfies or violates the second-stage constraints in problem (6), respectively.

Note that problem (6) involves an exponential number $(L+1)^K$ of uncertainty sets. As we will see shortly, the size of our approximate MILP reformulation for (6) also grows exponentially in K. This is not surprising as Theorem 3 states that the objective function of $\mathcal{P}_K$ (and hence, of problem (6)) can be evaluated in polynomial time if and only if the number of policies K is fixed.

The sets $\Xi(\boldsymbol{\ell})$ in problem (6) are not closed in general. The following example shows that we cannot naïvely replace the strict inequalities in $\Xi(\boldsymbol{\ell})$ with weak inequalities.

Example 4 *The instance of problem* (6) *presented in Example 3 has an optimal value of* 1*, which is attained by the solution* $\boldsymbol{y}^1 = (1,0)^\top$ *and* $\boldsymbol{y}^2 = (0,1)^\top$*. Since the objective value of* (6) *is finite for this solution, we conclude that* $\Xi(\boldsymbol{\ell}) = \emptyset$ *for all* $\boldsymbol{\ell} \in \mathcal{L}_+$*. For example, we easily verify that*

$$\Xi(1,1) = \left\{\boldsymbol{\xi} \in [-1,1]^2 : \boldsymbol{\xi}_1 > \boldsymbol{y}_1^1,\ \boldsymbol{\xi}_1 > \boldsymbol{y}_1^2\right\} = \emptyset.$$

If we were to replace the strict inequalities in $\Xi(\boldsymbol{\ell})$*,* $\boldsymbol{\ell} \in \mathcal{L}$*, with weak inequalities, then we readily verify that* $\Xi(1,1)$ *would become nonempty. For* $\boldsymbol{\ell} = (1,1)$*, the minimization over* $k \in \mathcal{K}$ *in problem* (6) *would then be taken over the empty set and thus evaluate to* $+\infty$*, whereas the maximization over* $\boldsymbol{\xi} \in \Xi(1,1)$ *would be taken over a nonempty set. Thus, the objective function in* (6) *would evaluate to* $+\infty$*, that is, the solution* $(\boldsymbol{y}^1, \boldsymbol{y}^2)$ *is infeasible in the variant of problem* (6) *where the strict inequalities in* $\Xi(\boldsymbol{\ell})$*,* $\boldsymbol{\ell} \in \mathcal{L}$*, are replaced with weak inequalities.*

In the following, we will employ closed inner approximations $\Xi_\epsilon(\boldsymbol{\ell})$ of the sets $\Xi(\boldsymbol{\ell})$ that are parameterized by a scalar $\epsilon > 0$:

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\boldsymbol{\ell} \in \mathcal{L}} \max_{\boldsymbol{\xi} \in \Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\, \boldsymbol{x} + \min_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k = 0}} \boldsymbol{\xi}^\top \boldsymbol{Q}\, \boldsymbol{y}^k\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y}^k \in \mathcal{Y},\ k \in \mathcal{K}, \end{array} \tag{6_ϵ}$$

where the approximate uncertainty sets $\Xi_\epsilon(\boldsymbol{\ell})$ are defined as

$$\Xi_\epsilon(\boldsymbol{\ell}) = \left\{ \boldsymbol{\xi} \in \Xi : \begin{array}{ll} \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{H}\boldsymbol{\xi} & \forall k \in \mathcal{K} : \boldsymbol{\ell}_k = 0 \\ [\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} \geq [\boldsymbol{H}\boldsymbol{\xi}]_{\boldsymbol{\ell}_k} + \epsilon & \forall k \in \mathcal{K} : \boldsymbol{\ell}_k \neq 0 \end{array} \right\}.$$

Note that by construction, the approximate uncertainty sets $\Xi_\epsilon(\boldsymbol{\ell})$ are closed. Lemma 1 in the Electronic Companion proves the intuitive fact that the sets $\Xi_\epsilon(\boldsymbol{\ell})$, $\boldsymbol{\ell} \in \mathcal{L}$, converge to the sets $\Xi(\boldsymbol{\ell})$ as ϵ approaches 0. We now show that this convergence of uncertainty sets carries over to a convergence of the objective functions and the optimal values of the approximate problems (6_ϵ).

Proposition 2 *Denote by* dom (6) *and* dom (6_ϵ) *the effective domains of problems* (6) *and* (6_ϵ)*, respectively, that is, the sets of decisions* $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k \in \mathcal{K}}) \in \mathcal{X} \times \mathcal{Y}^K$ *for which the objective values in the respective problems are finite (*i.e., *do not evaluate to* $+\infty$*). We then have:*

(i) $\mathrm{dom}\,(6_\epsilon) = \mathrm{dom}\,(6)$ *for sufficiently small* $\epsilon > 0$*, and*

(ii) over their effective domains, the objective functions in (6_ϵ) *converge uniformly to the objective function in* (6) *as* ϵ *approaches* 0*.*

Note that the statements *(i)* and *(ii)* in Proposition 2 immediately imply that the optimal values of the problems (6_ϵ) converge to the optimal value of problem (6) as ϵ approaches 0.

We now provide a MILP reformulation for the approximate optimization problem (6_ϵ).

Theorem 5 *The approximate problem* (6_ϵ) *is equivalent to the mixed-integer bilinear program*

$$
\begin{array}{ll}
\text{minimize} & \tau \\
\text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y}^k \in \mathcal{Y}, k \in \mathcal{K},\ \tau \in \mathbb{R} \\
& \left.\begin{array}{l}
\boldsymbol{\lambda}(\boldsymbol{\ell}) \in \Delta_K(\boldsymbol{\ell}),\ \boldsymbol{\alpha}(\boldsymbol{\ell}) \in \mathbb{R}^R_+,\ \boldsymbol{\beta}^k(\boldsymbol{\ell}) \in \mathbb{R}^L_+, k \in \mathcal{K},\ \boldsymbol{\gamma}(\boldsymbol{\ell}) \in \mathbb{R}^K_+ \\
\boldsymbol{b}^\top \boldsymbol{\alpha}(\boldsymbol{\ell}) - \displaystyle\sum_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k = 0}} (\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k)^\top \boldsymbol{\beta}^k(\boldsymbol{\ell}) + \displaystyle\sum_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k \neq 0}} \left([\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} - \epsilon \right) \gamma_k(\boldsymbol{\ell}) \leq \tau \\
\boldsymbol{A}^\top \boldsymbol{\alpha}(\boldsymbol{\ell}) - \displaystyle\sum_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k = 0}} \boldsymbol{H}^\top \boldsymbol{\beta}^k(\boldsymbol{\ell}) + \displaystyle\sum_{\substack{k \in \mathcal{K}: \\ \boldsymbol{\ell}_k \neq 0}} \boldsymbol{H}_{\boldsymbol{\ell}_k} \gamma_k(\boldsymbol{\ell}) = \boldsymbol{C}\boldsymbol{x} + \displaystyle\sum_{k \in \mathcal{K}} \boldsymbol{\lambda}_k(\boldsymbol{\ell}) \boldsymbol{Q}\boldsymbol{y}^k
\end{array}\right\} \forall \boldsymbol{\ell} \in \partial\mathcal{L}, \\
& \left.\begin{array}{l}
\boldsymbol{\alpha}(\boldsymbol{\ell}) \in \mathbb{R}^R_+,\ \boldsymbol{\gamma}(\boldsymbol{\ell}) \in \mathbb{R}^K_+ \\
\boldsymbol{b}^\top \boldsymbol{\alpha}(\boldsymbol{\ell}) + \displaystyle\sum_{k \in \mathcal{K}} \left([\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} - \epsilon \right) \gamma_k(\boldsymbol{\ell}) \leq -1 \\
\boldsymbol{A}^\top \boldsymbol{\alpha}(\boldsymbol{\ell}) + \displaystyle\sum_{k \in \mathcal{K}} \boldsymbol{H}_{\boldsymbol{\ell}_k} \gamma_k(\boldsymbol{\ell}) = \boldsymbol{0}
\end{array}\right\} \forall \boldsymbol{\ell} \in \mathcal{L}_+,
\end{array}
\tag{7}
$$

where $\boldsymbol{H}_{\boldsymbol{\ell}_k}$ *denotes the* $\boldsymbol{\ell}_k$*-th row of matrix* $\boldsymbol{H}$ *as a column vector.*

Remark 7 *Since all bilinear terms in problem* (7) *constitute products of one continuous and one binary variable, the problem can be reformulated as a MILP using standard Big-M techniques (Hillier 2009). For the sake of brevity, we omit the resulting MILP formulation.*

Remark 8 *Theorem 5 can be generalized to instances of the* K*-adaptability problem* $\mathcal{P}_K$ *where the technology matrix* $\boldsymbol{T}$ *or the recourse matrix* $\boldsymbol{W}$ *depend on* $\boldsymbol{\xi}$ *since we can absorb such dependencies in the right-hand side matrix* $\boldsymbol{H}$*. Assume, for example, that* $\boldsymbol{T}(\boldsymbol{\xi}) = \sum_{q=1}^{Q} \boldsymbol{T}^q \boldsymbol{\xi}_q$ *and* $\boldsymbol{W}(\boldsymbol{\xi}) =$

$\sum_{q=1}^{Q} \boldsymbol{W}^q \boldsymbol{\xi}_q$ for $\boldsymbol{T}^q \in \mathbb{R}^{L\times N}$ and $\boldsymbol{W}^q \in \mathbb{R}^{L\times M}$, $q = 1,\ldots,Q$. Using analogous arguments as in the proof of Theorem 5, one can derive a reformulation similar to (7) where $\boldsymbol{T} = \boldsymbol{0}$, $\boldsymbol{W} = \boldsymbol{0}$ and the expression $(\boldsymbol{T}^1\boldsymbol{x} \cdots \boldsymbol{T}^Q\boldsymbol{x}) + (\boldsymbol{W}^1\boldsymbol{y} \cdots \boldsymbol{W}^Q\boldsymbol{y})$ is subtracted from the matrix $\boldsymbol{H}$. Assuming that the here-and-now decisions $\boldsymbol{x}$ are binary, this reformulation can be linearized using Big-M techniques.

We emphasize that both the number of variables and the number of constraints in problem (7) scale with $|\mathcal{L}| = (L+1)^K$, that is, the reformulation (7) of the generic K-adaptability problem $\mathcal{P}_K$ scales exponentially in the number of policies K. For a fixed decision $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}}) \in \mathcal{X} \times \mathcal{Y}^K$, problem (7) reduces to an LP whose optimal value is identical to the objective function value of $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ in problem (6_ϵ). Moreover, an inspection of the proofs of Proposition 2 and Lemma 1 in the Electronic Companion reveals that for any given accuracy $\kappa > 0$, we can choose the approximation parameter ϵ in problem (6_ϵ) so that the objective functions in (6) and (6_ϵ) differ by at most κ and the bit length of ϵ is polynomial in the size of the input data for $\mathcal{P}_K$ and the bit length of κ^{-1}. We thus obtain the following result, which we have already anticipated in Theorem 3.

Corollary 1 *The objective function in $\mathcal{P}_K$ can be evaluated in polynomial time up to any accuracy if the number of policies K is fixed.*

4. Numerical Experiments

To gain a better understanding of the trade-offs between adaptability, approximation quality and computational effort, we apply the methods of the previous sections to stylized formulations of supply chain design route planning and capital budgeting problems. The supply chain design and route planning problems can be modeled as instances of the problem $\mathcal{PO}_K$, whereas the capital budgeting problem is an instance of the generic K-adaptability problem $\mathcal{P}_K$. We also compare our methods with the binary decision rules proposed by Bertsimas and Georghiou (2014). All optimization problems in this section are solved using the YALMIP modeling language by Löfberg (2004) and the Gurobi Optimizer 5.6 (Gurobi Optimization 2014). Unless stated otherwise, we use the Gurobi default settings and a time limit of 7,200 seconds.

4.1. Supply Chain Design

We consider a capacity expansion problem where a company seeks to build F factories at candidate sites $s \in \mathcal{S} = \{1, \ldots, S\}$, $S \geq F$, to serve customers $c \in \mathcal{C} = \{1, \ldots, C\}$ with uncertain demands $\boldsymbol{\xi}_c \in \mathbb{R}_+$ at minimum transportation costs. Each customer must be served by a single factory, and each factory can serve up to B customers. The transportation costs for serving customer $c \in \mathcal{C}$ from site $s \in \mathcal{S}$ amount to $\boldsymbol{c}_{sc}\boldsymbol{\xi}_c$, where $\boldsymbol{c}_{sc} \in \mathbb{R}_+$ can be interpreted as the per-unit transportation costs. We assume that the customer demands are only known to reside in the uncertainty set

$$\Xi = \left\{ \boldsymbol{\xi} \in [0, 100]^C : \boldsymbol{\xi} \leq \overline{\boldsymbol{\xi}},\ \mathbf{e}^\top \boldsymbol{\xi} = 100 \right\},$$

where $\overline{\boldsymbol{\xi}} \in \mathbb{R}^C_+$ denotes the vector of maximally anticipated customer demands. The uncertainty set expresses the view that the cumulative customer demands are known, but their breakdown by customer is uncertain. It is only known that the demand of customer $c \in \mathcal{C}$ is bounded above by $\overline{\boldsymbol{\xi}}_c$. Note that the size of the demand bounds $\overline{\boldsymbol{\xi}}$ determines the degree of uncertainty.

The problem can be formulated as the following instance of problem $\mathcal{PO}$:

$$\begin{aligned}
\text{minimize} \quad & \max_{\boldsymbol{\xi} \in \Xi} \min_{\boldsymbol{y} \in \{0,1\}^{S \times C}} \left\{ \sum_{(s,c) \in \mathcal{S} \times \mathcal{C}} \boldsymbol{c}_{sc} \boldsymbol{\xi}_c \boldsymbol{y}_{sc} : \boldsymbol{y}_{sc} \leq \boldsymbol{x}_s \ \forall (s,c) \in \mathcal{S} \times \mathcal{C}, \right. \\
& \qquad\qquad \left. \sum_{s \in \mathcal{S}} \boldsymbol{y}_{sc} = 1 \ \forall c \in \mathcal{C},\ \sum_{c \in \mathcal{C}} \boldsymbol{y}_{sc} \leq B \ \forall s \in \mathcal{S} \right\} \\
\text{subject to} \quad & \boldsymbol{x} \in \{0,1\}^S,\ \mathbf{e}^\top \boldsymbol{x} = F
\end{aligned}$$

Note that in this problem, the second-stage constraints are totally unimodular in $\boldsymbol{y}$ for every $\boldsymbol{x} \in \{0,1\}^S$, which implies that the progressive approximation of problem $\mathcal{PO}$ presented in Remark 1 is actually tight. This choice of problem is intentional as it will allow us to numerically assess the suboptimality of K-adaptable solutions for small values of K, where Theorem 1 is not applicable.

For our numerical experiments, we generate random test instances with $N \in \{10, 15, \ldots, 40\}$ candidate sites and customers. For each instance, we select N points (x_n, y_n) uniformly at random from the interval $[0, 10]^2$, where (x_n, y_n) represents the location of both the n-th candidate site and the n-th customer. We identify the per-unit transportation cost from site s to customer c with the

		Number of locations N						
	K	10	15	20	25	30	35	40
$\overline{\boldsymbol{\xi}} = 50\mathbf{e}$	2	100%/1s/0%	100%/3m:05s/0%	22%/1h:24m:31s/4.13%	0%/-/5.70%	0%/-/7.24%	0%/-/7.06%	0%/-/8.71%
	3	100%/5s/0%	89%/32m:43s/1.14%	0%/-/2.29%	0%/-/3.10%	0%/-/3.84%	0%/-/4.22%	0%/-/5.31%
	4	100%/16s/0%	6%/1h:06m:16s/0.46%	0%/-/1.61%	0%/-/1.61%	0%/-/2.18%	0%/-/2.85%	0%/-/3.62%
$\overline{\boldsymbol{\xi}} = 100\mathbf{e}$	2	100%/<1s/0%	100%/36s/0%	43%/1h:05m:26s/4.22%	1%/54m:56s/6.18%	0%/-/7.22%	0%/-/7.46%	0%/-/8.69%
	3	100%/5s/0%	83%/25m:35s/1.90%	0%/-/2.57%	0%/-/3.01%	0%/-/3.97%	0%/-/4.48%	0%/-/5.08%
	4	100%/17s/0%	12%/50m:07s/0.50%	0%/-/1.47%	0%/-/1.36%	0%/-/2.24%	0%/-/2.68%	0%/-/3.36%

Table 1 Summary of the results for the supply chain design problem. Each entry in the table documents the percentage of instances solved within the time limit, the average solution time for the instances solved within the time limit and the average optimality gap for the instances not solved to optimality. All results are averaged over 100 instances.

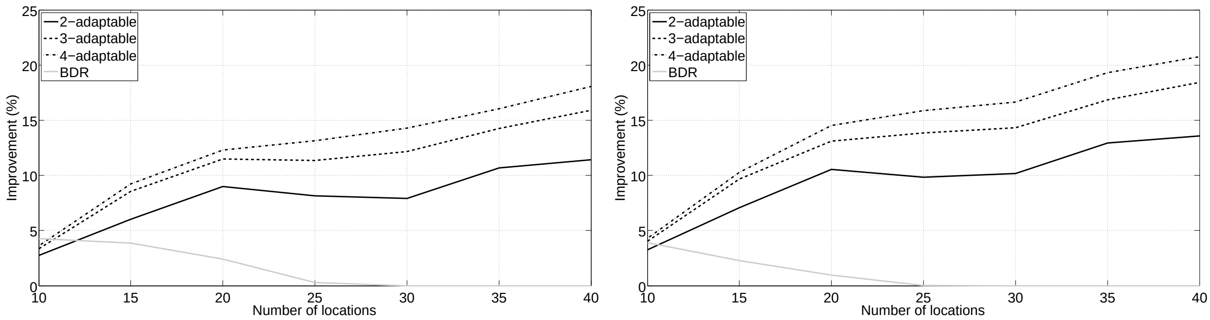

Figure 2 Improvement of the best 2-, 3- and 4-adaptable solutions and the binary decision rules of Bertsimas and Georghiou (2014) determined within the set time limit over the static solutions for the supply chain design problem with $\overline{\boldsymbol{\xi}} = 50\mathbf{e}$ (left) and $\overline{\boldsymbol{\xi}} = 100\mathbf{e}$ (right). The figures show the improvements for problems with $N = 10, 15, \ldots, 40$ locations as averages over 100 instances. For each instance size, we show the maximum improvement of the binary decision rules with up to 7 basis functions per component of $\boldsymbol{\xi}$.

Euclidean distance between the respective locations, that is, $\boldsymbol{c}_{sc} = \|(x_s, y_s) - (x_c, y_c)\|_2$. We assume that $F = N/5$ factories are to be built, each factory serves $B = N/F$ customers, and we consider homogeneous upper demand bounds $\overline{\boldsymbol{\xi}} \in \{50\mathbf{e}, 100\mathbf{e}\}$. The resulting instances of the K-adaptability problem (5) have $\mathcal{O}(KN^2)$ binary variables. To facilitate the solution of these large-scale MILPs, we first solve auxiliary problems where we fix $\boldsymbol{\beta} = K^{-1}\mathbf{e}$ in (5). We then feed the optimal solutions to these problems as initial feasible solutions to the unrestricted MILP (5) where $\boldsymbol{\beta}$ is a vector of decision variables.

Table 1 shows that only the small problem instances with 10–20 candidate sites and customers can be solved to optimality within the set time limit. Moreover, Gurobi reports large optimality gaps for those instances that cannot be solved within the time limit (not shown in the table).

Suspecting that the actual optimality gaps may be much smaller, we compare the objective values of the best feasible solutions found by Gurobi with the optimal values of the respective instances of the two-stage robust binary program $\mathcal{PO}$, see Remark 1. Table 1 reveals that these gaps are all below 10%. We thus conclude that for the considered class of supply chain design problems, *(i)* a small degree of adaptability is sufficient for problem $\mathcal{PO}_K$ to provide a close approximation to problem $\mathcal{PO}$, and *(ii)* even though Gurobi cannot certify optimality within the set time limit, it reliably produces near-optimal solutions to problem (5).

Figure 2 visualizes the improvement of the 2-, 3- and 4-adaptable solutions and the binary decision rules of Bertsimas and Georghiou (2014) over the static solutions where all decisions are taken here-and-now. The solution times for the binary decision rule problems range between 5m:53s for the smallest instances and more than the set time limit of two hours for the larger instances. The figure reveals that the additional flexibility of the K-adaptability formulations leads to a significant improvement over the static solutions, and that this improvement increases with the number of locations N and the size of the uncertainty set. While the binary decision rules are competitive for smaller instances, the improvement over the static solutions decreases with the problem size. This seems to be caused by an unfavorable scaling behavior of the associated MILP reformulations. In fact, while the smaller instances can be solved to optimality for binary decision rules with up to 7 basis functions per component of $\boldsymbol{\xi}$, the larger instances cannot be solved within the time limit even for decision rules with one basis function per uncertain problem parameter.

4.2. Route Planning

We consider a shortest path problem that is defined on a directed, arc-weighted graph $G=(V,A,\boldsymbol{w})$ with nodes $V=\{1,\ldots,N\}$, arcs $A\subseteq V\times V$ and weights $\boldsymbol{w}_{ij}(\boldsymbol{\xi})\in\mathbb{R}_+$, $(i,j)\in A$. We assume that the arc weights are functions of an uncertain parameter vector $\boldsymbol{\xi}$ that is only known to reside in an uncertainty set Ξ. The goal is to determine K paths from a start node $s\in V$ to a terminal node $t\in V$, $s\neq t$, here-and-now (that is, before observing $\boldsymbol{\xi}$) such that the worst-case length of the shortest among the K paths is minimized. This problem can be formulated as the following instance of problem $\mathcal{PO}_K$.

$$\begin{array}{ll} \text{minimize} & \max_{\boldsymbol{\xi}\in\Xi} \min_{k\in\mathcal{K}} \sum_{(i,j)\in A} \boldsymbol{w}_{ij}(\boldsymbol{\xi})\, \boldsymbol{y}_{ij}^k \\ \text{subject to} & \boldsymbol{y}_{ij}^k \in \{0,1\}\,,\ (i,j)\in A \text{ and } k\in\mathcal{K} \\ & \sum_{(j,l)\in A} \boldsymbol{y}_{jl}^k \geq \sum_{(i,j)\in A} \boldsymbol{y}_{ij}^k + \mathbb{I}[j=s] - \mathbb{I}[j=t] \qquad \forall j\in V,\ \forall k\in\mathcal{K} \end{array}$$

In this problem, the second-stage constraints are totally unimodular in $\boldsymbol{y}$, which implies that the progressive approximation of problem $\mathcal{PO}$ presented in Remark 1 is tight. It is unlikely, however, that the resulting formulation is of any practical interest since the route planning problem does not involve any first-stage decisions. In contrast, the K-adaptability problem has important applications in emergency preparedness planning, where the goal could be to determine K different routes for transporting relief supplies or evacuating citizens in the event of a hypothetical disaster.

For our numerical experiments, we generate random graphs with $N \in \{20, 25, \ldots, 50\}$ nodes. In each problem instance, the nodes correspond to N points (x_n, y_n) that are chosen uniformly at random from the interval $[0, 10]^2$, $n = 1, \ldots, N$. We choose the pair of nodes with the largest Euclidean distance as the designated start and terminal nodes. As for the arc set A, we begin with a fully connected graph and remove 70% of the arcs $(i, j) \in A$ in order of decreasing Euclidean distance $\|(x_i, y_i) - (x_j, y_j)\|_2$, that is, starting with the longest arcs. This eliminates trivial instances where the shortest path contains very few arcs. We choose a budget uncertainty set of the form

$$\Xi = \left\{ (\boldsymbol{\xi}_{ij})_{(i,j)\in A} : \boldsymbol{\xi}_{ij} \in [0,1]\ \forall (i,j)\in A,\ \sum_{(i,j)\in A} \boldsymbol{\xi}_{ij} \leq B \right\},$$

and we set the arc weights to $\boldsymbol{w}_{ij}(\boldsymbol{\xi}) = (1+\boldsymbol{\xi}_{ij}/2)\, \|(x_i, y_i) - (x_j, y_j)\|_2$. Thus, the travel time between each pair of adjacent nodes varies between 100% and 150% of the Euclidean distance between the nodes, and at most B arcs attain their maximum travel times. In our experiments, we consider the uncertainty budgets $B \in \{3, 6\}$.

The two-stage robust binary formulation of the route planning problem involves $\mathcal{O}(N^2)$ uncertain problem parameters, and the corresponding instances of the K-adaptability problem (5) contain

		Number of nodes N						
	K	20	25	30	35	40	45	50
$B=3$	2	100%/8s/0%	99%/2m:48s/3.68%	69%/18m:51s/5.69%	17%/38m:55s/5.70%	6%/49m:09s/5.96%	0%/-/6.48%	0%/-/6.75%
	3	97%/7m:43s/1.60%	31%/21m:13s/1.14%	6%/26m:03s/1.71%	0%/-/2.23%	0%/-/2.59%	0%/-/3.14%	0%/-/3.44%
	4	51%/18m:23s/0.23%	6%/47m:31s/0.30%	0%/-/0.66%	0%/-/0.96%	0%/-/1.28%	0%/-/1.62%	0%/-/1.89%
$B=6$	2	100%/7s/0%	99%/3m:17s/8.00%	67%/22m:52s/10.72%	16%/46m:59s/10.76%	5%/48m:08s/11.29%	0%/-/11.79%	0%/-/12.31%
	3	97%/7m:08s/4.06%	38%/29m:31s/3.52%	6%/25m:37s/4.54%	0%/-/5.58%	0%/-/6.19%	0%/-/6.92%	0%/-/7.55%
	4	55%/13m:15s/0.91%	7%/53m:28s/1.23%	0%/-/2.09%	0%/-/2.88%	0%/-/3.50%	0%/-/4.11%	0%/-/4.74%

Table 2 Summary of the results for the route planning problem. The entries have the same interpretation as in Table 1.

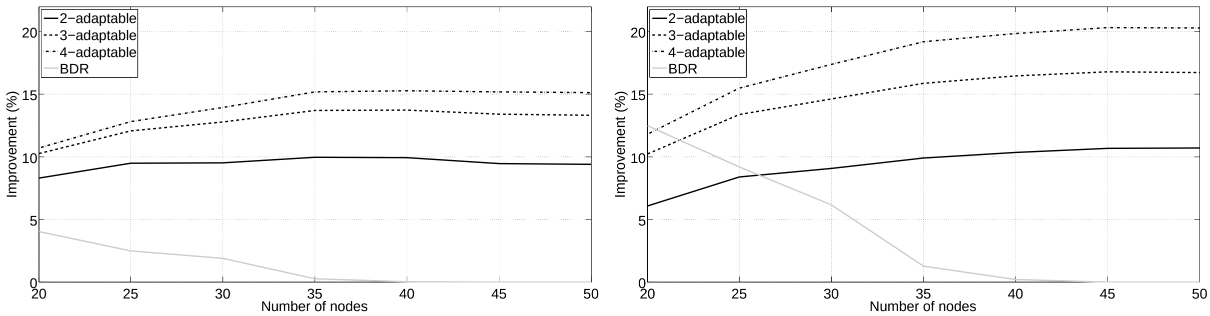

Figure 3 Improvement of the best 2-, 3- and 4-adaptable solutions and the binary decision rules of Bertsimas and Georghiou (2014) determined within the set time limit over the static solutions for the route planning problem with $B=3$ (left) and $B=6$ (right). The figures show the improvements for problems with $N=20,25,\ldots,50$ nodes as averages over 100 instances. For each instance size, we show the maximum improvement of the binary decision rules with up to 7 basis functions per component of $\boldsymbol{\xi}$.

$\mathcal{O}(KN^2)$ binary variables. As such, it is not surprising that the route planning problem is difficult to solve. This is reflected in Table 2, which shows that most of the problem instances cannot be solved to optimality within the set time limit. However, we observe that for most of the instances which could not be solved to optimality, the optimality gap (relative to problem $\mathcal{PO}$) is below 10%. Moreover, the optimality gaps decrease as the number of policies K increases. This suggests that Gurobi finds near-optimal solutions for those instances, and that the gap for small K is owed to the gap between the optimal values of the problems $\mathcal{PO}_K$ and $\mathcal{PO}$. This suspicion is strengthened by an investigation of the solver reports, which show that the terminal solution is typically found within 50 seconds.

Figure 3 visualizes the improvement of the 2-, 3- and 4-adaptable solutions and the binary decision rules of Bertsimas and Georghiou (2014) over the static solutions. The solution times for

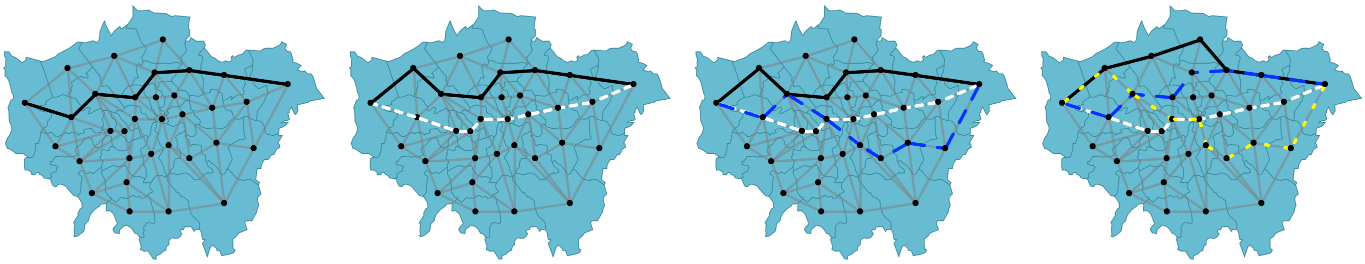

Figure 4 Optimal 2-, 3- and 4-adaptable routes from Hillingdon to Havering, two boroughs of London, UK.

the binary decision rule problems range between 18m:53s for the smallest instances and more than the set time limit of two hours for the larger instances. The figure shows that as a function of the number of nodes N, the improvement of the K-adaptability formulations initially increases but subsequently levels off and eventually decreases. In fact, since the uncertainty budget B does not grow with N, one can show that the outperformance of the fully adaptable solutions to problem $\mathcal{PO}$ over the static solutions goes to zero as N approaches infinity. The figure also shows that the improvement increases with B, which specifies the size of the uncertainty set. The scaling behavior of the binary decision rules implies that they are primarily competitive for smaller instance sizes.

We close with an illustration of the solutions to the K-adaptable route planning problem. Imagine that a decision maker aims to travel from Hillingdon to Havering, two boroughs in London, UK. We can formulate this problem as an instance of our route planning problem. To this end, we identify the nominal arc weights with the travel times between neighboring boroughs as reported by Google Maps during a rush hour period.[3] Figure 4 visualizes the optimal static, 2-, 3-, and 4-adaptable solutions for $B = 5$. The static solution has a worst-case travel time of 1h:21m:00s, whereas the 2-, 3-, and 4-adaptable routes have travel times of 1h:12m:04s, 1h:11m:09s, and 1h:10m:56s in the worst case, respectively. The solution to the respective instance of problem $\mathcal{PO}$ has a worst-case travel time of 1h:10m:50s, which confirms that the 4-adaptable solution is nearly optimal. We emphasize that the optimal 4-adaptable solution does not contain the three routes from the optimal 3-adaptable solution. In fact, one can show that any 4-adaptable solution which contains the three routes from the 3-adaptable solution is strictly inferior to the optimal 4-adaptable solution.

4.3. Capital Budgeting

We consider an investment planning problem where 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 $\boldsymbol{c}_i(\boldsymbol{\xi})$ and uncertain profits $\boldsymbol{r}_i(\boldsymbol{\xi})$, which are modeled as functions of an uncertain vector $\boldsymbol{\xi}$ of risk factors. The company can invest in a project before or after observing the risk factors $\boldsymbol{\xi}$. Early investments enjoy a first-mover advantage (*e.g.* in the form of technological leadership, switching costs or the preemption of scarce resources), whereas a postponed investment in project i incurs the same costs $\boldsymbol{c}_i(\boldsymbol{\xi})$ but generates only a fraction $\theta \in [0, 1)$ of the profits $\boldsymbol{r}_i(\boldsymbol{\xi})$.

Assuming that the risk factors $\boldsymbol{\xi}$ are only known to reside in an uncertainty set Ξ, the problem can be formulated as the following instance of the generic K-adaptability problem $\mathcal{P}_K$:

$$\begin{array}{ll} \text{maximize} & \displaystyle\min_{\boldsymbol{\xi} \in \Xi} \max_{\boldsymbol{y} \in \{0,1\}^N} \left\{ \boldsymbol{r}(\boldsymbol{\xi})^\top (\boldsymbol{x} + \theta \boldsymbol{y}) : \boldsymbol{c}(\boldsymbol{\xi})^\top (\boldsymbol{x} + \boldsymbol{y}) \le B, \ \boldsymbol{x} + \boldsymbol{y} \le \mathbf{e} \right\} \\ \text{subject to} & \boldsymbol{x} \in \{0,1\}^N \end{array}$$

In this formulation, the decisions $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ attain the value 1 if and only if an early or late investment in project i is undertaken, respectively. Note that in the model, the projects' costs $\boldsymbol{c}(\boldsymbol{\xi})$ are already deducted from the projects' profits $\boldsymbol{r}(\boldsymbol{\xi})$ in the objective function. In contrast to the previous examples, both the objective function and the constraints of the capital budgeting problem are affected by uncertainty.

For our numerical experiments, we generate random problem instances with $N \in \{5, 10, \ldots, 30\}$ projects. In all instances, we use 4 risk factors that reside in the hyperrectangle $\Xi = [-1, 1]^4$. We model the project costs and profits as affine functions of these factors:

$$\boldsymbol{c}_i(\boldsymbol{\xi}) = (1 + \boldsymbol{\Phi}_i^\top \boldsymbol{\xi}/2) \boldsymbol{c}_i^0 \qquad \text{and} \qquad \boldsymbol{r}_i(\boldsymbol{\xi}) = (1 + \boldsymbol{\Psi}_i^\top \boldsymbol{\xi}/2) \boldsymbol{r}_i^0$$

Here, $\boldsymbol{c}_i^0$ and $\boldsymbol{r}_i^0$ are the nominal costs and profits of project i, respectively, whereas $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_i$ represent the i-th rows of the factor loading matrices $\boldsymbol{\Phi}, \boldsymbol{\Psi} \in \mathbb{R}^{N \times 4}$ as column vectors. We choose the nominal costs $\boldsymbol{c}^0$ uniformly at random from the interval $[0, 10]^N$, and we set the nominal profits to $\boldsymbol{r}^0 = \boldsymbol{c}^0/5$. The components in each row of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are chosen uniformly at random from the

	Number of projects N					
K	5	10	15	20	25	30
2	100%/<1s/0%	100%/4s/0%	100%/8m:32s/0%	2%/1h:27m:12s/32.40%	0%/-/59.36%	0%/-/60.20%
3	100%/1s/0%	74%/20m:10s/9.89%	0%/-/35.00%	0%/-/36.66%	0%/-/36.37%	0%/-/36.18%
4	100%/36s/0%	0%/-/26.49%	0%/-/26.61%	0%/-/26.61%	0%/-/26.35%	0%/-/26.16%

Table 3 Summary of the results for the capital budgeting problem. The entries have the same interpretation as in Table 1.

unit simplex in $\mathbb{R}^4$, which implies that the costs and profits of each project can deviate by up to 50% from their nominal values. We set the investment budget to $B = \mathbf{e}^\top \boldsymbol{c}^0/2$, and we assume that postponed investments only generate 80% of the profits, that is, $\theta = 0.8$. We set $\epsilon = 10^{-4}$.

The K-adaptability formulation (7) corresponding to the capital budgeting problem has $\mathcal{O}(KN)$ binary variables and $\mathcal{O}(2^K KN)$ continuous variables. Table 3 shows that only the small instances with 2 policies and up to 15 projects can be solved to optimality within the set time limit. To assess the optimality gaps of those instances that cannot be solved within the time limit, we employ the progressive bound for problem $\mathcal{P}$ presented in Remark 5. As in our route planning experiment, the optimality gaps tend to get smaller with larger values of K, which indicates that the optimality gaps are primarily owed to the gap between the optimal values of the problem $\mathcal{P}_K$ and the discretized version of problem $\mathcal{P}$, as opposed to the suboptimality of the solution to $\mathcal{P}_K$ determined by Gurobi.

Figure 5 shows the improvement of the 2-, 3- and 4-adaptable solutions and the binary decision rules proposed by Bertsimas and Georghiou (2014) over the static solutions. The figure reveals that the improvement of the K-adaptability formulations increases with the number of projects N, but that it saturates as N increases. This is owed to the fact that for the considered class of capital budgeting problems, the outperformance of the fully adaptable solutions to problem $\mathcal{PO}$ over the static solutions is bounded by a constant. Contrary to the previous examples, the binary decision rules of Bertsimas and Georghiou (2014) compare favourably with the K-adaptable solutions. This is due to the fact that *(i)* the uncertainty set is rectangular, which implies that the reformulation for the binary decision rules is exact, and *(ii)* the number of uncertain problem parameters does not increase with the problem size, which allows us to optimize over binary decision rules with up to 5 basis functions per uncertain problem parameter even for larger problem instances.

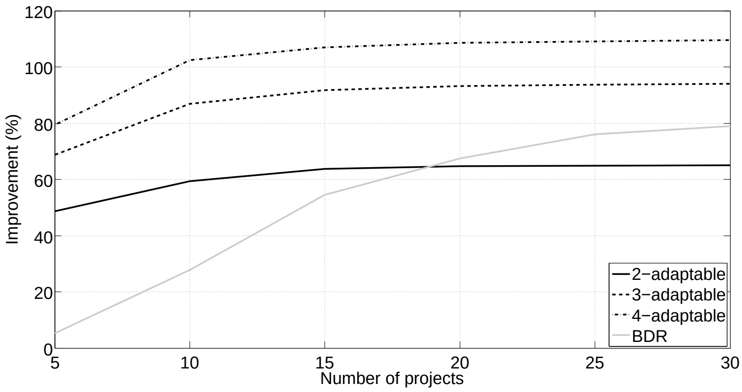

Figure 5 Improvement of the best 2-, 3- and 4-adaptable solutions and the binary decision rules of Bertsimas and Georghiou (2014) determined within the set time limit over the static solutions for the capital budgeting problem. The figure shows the improvements for problems with $N = 5, 10, \ldots, 30$ projects as averages over 100 instances. For each instance size, we show the maximum improvement of the binary decision rules with up to 7 basis functions per component of $\boldsymbol{\xi}$.

5. Conclusion

In our opinion, robust optimization has succeeded as a methodology due to its adherence to three fundamental guiding principles. First and foremost, the literature aims to propose robust optimization problems that are of *similar complexity* as their deterministic counterparts. For two-stage and multi-stage problems, this typically implies that one has to resort to approximate problem formulations (*e.g.*, using affine or piecewise affine decision rules). This leads us to the second guiding principle, which stipulates that any approximation undertaken should be of *conservative nature*, that is, it should not introduce any spurious solutions that violate the constraints of the original problem. Finally, it should be possible to quantify the *degree of suboptimality* and refine the approximation scheme if the optimality gap is judged to be unacceptable.

The approach proposed in this paper is aligned with all three principles. We provide a reformulation for the problem $\mathcal{PO}_K$ that is of comparable complexity as the associated deterministic integer program, and the same holds true for our reformulation of the problem $\mathcal{P}_K$ if we fix the number of policies K. With the exception of the ϵ-approximation in Section 3 (which is unlikely to be of practical concern), the proposed K-adaptability problems are conservative approximations of the two-stage robust binary programs $\mathcal{PO}$ and $\mathcal{P}$. The Remarks 1 and 5 outline how the suboptimality of these approximations can be measured, and Theorems 1 and 4 show that in both cases, the suboptimality can be reduced to zero by increasing the number of considered policies K.

Endnotes

1. Boundedness of $\mathcal{X}$ and $\mathcal{Y}$ is assumed for ease of exposition only. Our findings extend to unbounded sets, but our proofs would require further case distinctions.

2. Here and in the following, 'polynomial time' is understood relative to the length of the input data for the problem.

3. Google Maps: `https://maps.google.co.uk`.

References

Ahmed, S. 2013. A scenario decomposition algorithm for 0-1 stochastic programs. *Operations Research Letters* **41**(6) 565–569.

Alonso-Ayuso, A., L. F. Escudero, M. T. Ortuño. 2003. BFC, a branch-and-fix coordination algorithmic framework for solving some types of stochastic pure and mixed 0-1 programs. *European Journal of Operational Research* **151**(3) 503–519.

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

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

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

Bertsimas, D., D. B. Brown, C. Caramanis. 2011a. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, D., C. Caramanis. 2007. Adaptability via sampling. *Proceedings of the 46th IEEE Conference on Decision and Control*. 4717–4722.

Bertsimas, D., C. Caramanis. 2010. Finite adaptibility in multistage linear optimization. *IEEE Transactions on Automatic Control* **55**(12) 2751–2766.

Bertsimas, D., A. Georghiou. 2013. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. Available on Optimization Online.

Bertsimas, D., A. Georghiou. 2014. Binary decision rules for multistage adaptive mixed-integer optimization. Available on Optimization Online.

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

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

Carøe, C., R. Schultz. 1999. Dual decomposition in stochastic integer programming. *Operations Research Letters* **24**(1–2) 37–45.

Carøe, C. C., J. Tind. 1998. L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming* **83**(1–3) 451–464.

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

Dhamdhere, K., V. Goyal, R. Ravi, M. Singh. 2005. How to pay, come what may: Approximation algorithms for demand-robust covering problems. *46th Annual IEEE Symposium on Foundations of Computer Science*. 367–376.

Dye, S., L. Stougie, A. Tomasgard. 2003. The stochastic single resource service-provision problem. *Naval Research Logistics* **50**(8) 869–887.

Feige, U., K. Jain, M. Mahdian, V. Mirrokni. 2007. Robust combinatorial optimization with exponential scenarios. M. Fischetti, D. P. Williamson, eds., *Integer Programming and Combinatorial Optimization*. Springer, 439–453.

Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.

Georghiou, A., W. Wiesemann, D. Kuhn. 2014. Generalized decision rule approximations for stochastic programming via liftings. *Forthcoming in Mathematical Programming A*.

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

Golovin, D., V. Goyal, V. Polishchuk, R. Ravi, M. Sysikaski. 2014. Improved approximations for two-stage min-cut and shortest path problems under uncertainty. *Mathematical Programming A* 1–28.

Gorissen, B. L., I. Yanikoglu, D. den Hertog. 2013. Hints for practical robust optimizations. Available on SSRN.

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

Gupta, A., V. Nagarajan, R. Ravi. 2010. Thresholded covering algorithms for robust and max-min optimization. S. Abramsky, C. Gavoille, C. Kirchner, F. Meyer auf der Heide, P. G. Spirakis, eds., *Automata, Languages and Programming*. Springer, 262–274.

Gupta, A., V. Nagarajan, V. V. Vazirani. 2013. Thrifty algorithms for multistage robust optimization. M. Goemans, J. Correa, eds., *Integer Programming and Combinatorial Optimization*. Springer, 217–228.

Gupta, A., M. Pál, R. Ravi, A. Sinha. 2005. What about Wednesday? Approximation algorithms for multi-stage stochastic optimization. C. Chekuri, K. Jansen, J. D. P. Rolim, L. Trevisan, eds., *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*. Springer, 86–98.

Gurobi Optimization, Inc. 2014. Gurobi optimizer reference manual. URL `http://www.gurobi.com`.

Hadjiyiannis, M. J., P. Goulart, D. Kuhn. 2011. A scenario approach for measuring the suboptimality of linear decision rules in two-stage robust optimization. *Proceedings of the 50th IEEE Conference on Decision and Control*. 7386–7391.

Hillier, F. S. 2009. *Introduction to Operations Research*. McGraw-Hill.

Iancu, D. A., M. Sharma, M. Sviridenko. 2013. Supermodularity and affine policies in dynamic robust optimization. *Operations Research* **61**(4) 941–956.

Immorlica, N., D. Karger, M. Minkoff, V. S. Mirrokni. 2004. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. 691–700.

Khandekar, R., G. Kortsarz, V. Mirrokni, M. R. Salavatipour. 2013. Two-stage robust network design with exponential scenarios. *Algorithmica* **65**(2) 391–408.

Klein Haneveld, W. K., L. Stougie, M. H. van der Vlerk. 1995. On the convex hull of the simple integer recourse objective function. *Annals of Operations Research* **56**(1) 209–224.

Klein Haneveld, W. K., L. Stougie, M. H. van der Vlerk. 1996. An algorithm for the construction of convex hulls in simple integer recourse programming. *Annals of Operations Research* **64**(1) 67–81.

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

Laporte, G., F. V. Louveaux. 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* **13**(3) 133–142.

Li, X., K. Natarajan, C.-P. Teo, Z. Zheng. 2014. Distributionally robust mixed integer linear programs: Persistency model with applications. *European Journal of Operational Research* **233**(3) 459–473.

Löfberg, J. 2004. YALMIP: A toolbox for modeling and optimization in MATLAB. *IEEE International Symposium on Computer Aided Control Systems Design*. 284–289.

Louveaux, F. V., R. Schultz. 2003. Stochastic integer programming. A. Ruszczyński, A. Shapiro, eds., *Stochastic Programming*, *Handbooks in Operations Research and Management Science*, vol. 10, chap. 4. Elsevier, 213–266.

Norkin, V. I., Y. M. Ermoliev, A. Ruszczyński. 1998. On optimal allocation of indivisibles under uncertainty. *Operations Research* **46**(3) 381–395.

Ravi, R., A. Sinha. 2004. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. D. Bienstock, G. Nemhauser, eds., *Integer Programming and Combinatorial Optimization*. Springer, 101–115.

Romeijnders, W., L. Stougie, M. H. van der Vlerk. 2014. Approximation in two-stage stochastic integer programming. *Surveys in Operations Research and Management Science* **19**(1) 17–33.

Schultz, R. 2003. Stochastic programming with integer variables. *Mathematical Programming* **97**(1–2) 285–309.

Schultz, R., L. Stougie, M. H. van der Vlerk. 1998. Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis. *Mathematical Programming* **83**(1–3) 229–252.

Sen, S., J. L. Higle. 2005. The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming A* **104**(1) 1–20.

Sen, S., H. D. Sherali. 2006. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming A* **106**(2) 203–223.

Sherali, H. D., B. M. P. Fraticelli. 2002. A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization* **22**(1–4) 319–342.

Shmoys, D. B., C. Swamy. 2006. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *Journal of the ACM* **53**(6) 978–1012.

Spacey, S. A., W. Wiesemann, D. Kuhn, W. Luk. 2012. Robust software partitioning with multiple instantiation. *INFORMS Journal on Computing* **24**(3) 500–515.

Srinivasan, A. 2007. Approximation algorithms for stochastic and risk-averse optimization. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete algorithms*. 1305–1313.

Swamy, C. 2011. Risk-averse stochastic optimization: Probabilistically-constrained models and algorithms for black-box distributions. *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. 1627–1646.

Swamy, C., D. Shmoys. 2012. Sampling-based approximation algorithms for multistage stochastic optimization. *SIAM Journal on Computing* **41**(4) 975–1004.

van der Vlerk, M. H. 1996–2007. Stochastic integer programming bibliography. `http://www.eco.rug.nl/mally/spbib.html`.

Vayanos, P., D. Kuhn, B. Rustem. 2011. Decision rules for information discovery in multi-stage stochastic programming. *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*. 7368–7373.

E-Companion: Proofs

Proof of Observation 1. If there is no $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \leq \boldsymbol{h}$, then problem (3) and problem $\mathcal{PO}_K$ are both infeasible and the statement holds true. In the remainder of the proof, we thus assume that there is $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \leq \boldsymbol{h}$.

Any feasible solution $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k \in \mathcal{K}})$ to problem $\mathcal{PO}_K$ is also feasible in (3), and it has the same objective value in both problems since all second-stage policies $\boldsymbol{y}^k$ satisfy the second-stage constraints in (3). Conversely, take any feasible solution $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k \in \mathcal{K}})$ to problem (3). Since Ξ is nonempty and the problem is feasible, at least one policy $\boldsymbol{y}^{k^\star}$ must satisfy the second-stage constraints. By replacing any policy $\boldsymbol{y}^k$ that violates the second-stage constraints with $\boldsymbol{y}^{k^\star}$ if necessary, we obtain a feasible solution to problem $\mathcal{PO}_K$ that attains the same objective value. $\square$

Proof of Theorem 1. Assume that $\dim \mathcal{Y} \leq \operatorname{rk} \boldsymbol{Q}$ and consider problem $\mathcal{PO}_K$ with $K = |\mathcal{Y}|$ policies. Note that $|\mathcal{Y}| < \infty$ because $\mathcal{Y} \subseteq \{0,1\}^M$. For this choice of K, problem $\mathcal{PO}_K$ has the same optimal objective value as $\mathcal{PO}$. Indeed, one readily verifies that any feasible solution $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k \in \mathcal{K}})$ to problem $\mathcal{PO}_K$ can be transformed into a feasible solution to problem $\mathcal{PO}$ that attains the same objective value and vice versa. We can reformulate the objective function of $\mathcal{PO}_K$ as

$$\max_{\boldsymbol{\xi} \in \Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\, \boldsymbol{x} + \min_{\boldsymbol{\lambda} \in \Delta_K} \sum_{k \in \mathcal{K}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\, \boldsymbol{y}^k \right],$$

where $\Delta_K = \{\boldsymbol{\lambda} \in \mathbb{R}^K_+ : \mathbf{e}^\top \boldsymbol{\lambda} = 1\}$ denotes the unit simplex in $\mathbb{R}^K$. Since the terms in the objective function are linear in $\boldsymbol{\xi}$ and linear in $\boldsymbol{\lambda}$, we can employ the classical min-max theorem to exchange the order of the maximum and minimum operators. Problem $\mathcal{PO}_K$ is thus equivalent to

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\boldsymbol{\xi} \in \Xi} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\, \boldsymbol{x} + \sum_{k \in \mathcal{K}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\, \boldsymbol{y}^k \right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X},\; \boldsymbol{y}^k \in \mathcal{Y},\, k \in \mathcal{K},\; \boldsymbol{\lambda} \in \Delta_K \\ & \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \leq \boldsymbol{h} \qquad\qquad \forall k \in \mathcal{K}. \end{array} \tag{EC.1}$$

We claim that problem (EC.1) has the same optimal value as

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\xi}\in\Xi}\left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \sum_{k\in\mathcal{D}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] \\ &\text{subject to} && \boldsymbol{x}\in\mathcal{X},\ \ \boldsymbol{y}^k\in\mathcal{Y},\, k\in\mathcal{D},\ \ \boldsymbol{\lambda}\in\Delta_{D+1} \\ &&& \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{h} \qquad\qquad \forall k\in\mathcal{D}, \end{aligned} \tag{EC.2}$$

where $\mathcal{D}=\{1,\ldots,D+1\}$, $D=\dim\mathcal{Y}$ and Δ_{D+1} is the unit simplex in $\mathbb{R}^{D+1}$. Note that the optimal value of (EC.2) provides an upper bound on the optimal value of (EC.1) since $D+1=\dim\mathcal{Y}+1\le|\mathcal{Y}|=K$. We now show that both problems indeed attain the same optimal value. To this end, fix an optimal solution $(\boldsymbol{x}^\star,\{\boldsymbol{y}^{\star,k}\}_{k\in\mathcal{K}},\boldsymbol{\lambda}^\star)$ to problem (EC.1) and define $\boldsymbol{y}'\in\mathbb{R}^M$ as $\boldsymbol{y}'=\sum_{k\in\mathcal{K}}\boldsymbol{\lambda}^\star_k\cdot\boldsymbol{y}^{\star,k}$. By construction, $\boldsymbol{y}'$ is contained in the convex hull of the K policies $\boldsymbol{y}^{\star,k}\in\mathbb{R}^M$, $k\in\mathcal{K}$. From Carathéodory's theorem we then conclude that $\boldsymbol{y}'$ is indeed contained in the convex hull of $D+1$ policies $\boldsymbol{y}^{\star,k_1},\ldots,\boldsymbol{y}^{\star,k_{D+1}}$, $k_1,\ldots,k_{D+1}\in\mathcal{K}$ (not necessarily all different). Choose $\boldsymbol{\lambda}'\in\Delta_{D+1}$ such that $\boldsymbol{y}'=\sum_{d\in\mathcal{D}}\boldsymbol{\lambda}'_d\cdot\boldsymbol{y}^{\star,k_d}$. We readily verify that $(\boldsymbol{x}^\star,\{\boldsymbol{y}^{\star,k_d}\}_{d\in\mathcal{D}},\boldsymbol{\lambda}')$ is a feasible solution to problem (EC.2) that attains the same objective value as $(\boldsymbol{x}^\star,\{\boldsymbol{y}^{\star,k}\}_{k\in\mathcal{K}},\boldsymbol{\lambda}^\star)$ in problem (EC.1). We thus conclude that both problems attain the same optimal value.

Employing the classical min-max theorem, we can reformulate problem (EC.2) as

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\xi}\in\Xi}\left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{k\in\mathcal{D}}\, \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] \\ &\text{subject to} && \boldsymbol{x}\in\mathcal{X},\ \ \boldsymbol{y}^k\in\mathcal{Y},\, k\in\mathcal{D} \\ &&& \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{h} \qquad\qquad \forall k\in\mathcal{D}, \end{aligned}$$

which we readily identify as the $(D+1)$-adaptability problem. We have thus shown the statement of the theorem for $\dim\mathcal{Y}\le\operatorname{rk}\boldsymbol{Q}$. We can proceed analogously for the case $\operatorname{rk}\boldsymbol{Q}<\dim\mathcal{Y}$ if we define $\boldsymbol{y}'\in\mathbb{R}^M$ as $\boldsymbol{y}'=\sum_{k\in\mathcal{K}}\boldsymbol{\lambda}^\star_k\cdot\boldsymbol{Q}\boldsymbol{y}^{\star,k}$. This concludes the proof. □

Proof of Theorem 2. For fixed $\boldsymbol{x}$ and $\boldsymbol{y}^k$, $k \in \mathcal{K}$, we can employ an epigraph formulation to express the evaluation of the objective function in $\mathcal{PO}_K$ as

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \tau \\ \text{subject to} \quad & \boldsymbol{\xi} \in \mathbb{R}^Q,\ \tau \in \mathbb{R} \\ & \boldsymbol{A}\boldsymbol{\xi} \le \boldsymbol{b} \\ & \tau \le \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k \qquad \forall k \in \mathcal{K}. \end{aligned}$$

Strong LP duality implies that this LP has the same optimal value as its dual problem,

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{b}^\top \boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathbb{R}^R_+,\ \boldsymbol{\beta} \in \mathbb{R}^K_+ \\ & \boldsymbol{A}^\top \boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{x} + \sum_{k \in \mathcal{K}} \beta_k \boldsymbol{Q}\boldsymbol{y}^k \\ & \mathbf{e}^\top \boldsymbol{\beta} = 1. \end{aligned}$$

The result now follows if we replace the bilinear terms $\boldsymbol{\beta}_k \boldsymbol{y}^k$ with auxiliary variables $\boldsymbol{z}^k \in \mathbb{R}^M_+$, $k \in \mathcal{K}$, subject to the constraints that

$$\boldsymbol{z}^k = \boldsymbol{\beta}_k \boldsymbol{y}^k \quad \Longleftrightarrow \quad \boldsymbol{z}^k \le \boldsymbol{y}^k,\ \boldsymbol{z}^k \le \boldsymbol{\beta}_k \mathbf{e},\ \boldsymbol{z}^k \ge (\boldsymbol{\beta}_k - 1)\mathbf{e} + \boldsymbol{y}^k.$$

This reformulation exploits the fact that $\mathbf{0} \le \boldsymbol{\beta}$, $\boldsymbol{y}^k \le \mathbf{e}$ and that $\boldsymbol{y}^k$ is binary. □

Proof of Theorem 3. For ease of exposition, we deferred the proof of the first statement to Corollary 1 in the main text. In the following, we focus on the proof of the second statement.

We recall that the strongly $\mathcal{NP}$-hard 0/1 Integer Programming (IP) feasibility problem is defined as follows (Garey and Johnson 1979).

0/1 INTEGER PROGRAMMING FEASIBILITY.

Instance. Given are $\boldsymbol{A} \in \mathbb{R}^{R \times Q}$ and $\boldsymbol{b} \in \mathbb{R}^R$.

Question. Is there a vector $\boldsymbol{\xi} \in \{0,1\}^Q$ such that $\boldsymbol{A}\boldsymbol{\xi} \le \boldsymbol{b}$?

We aim to reduce the IP feasibility problem to evaluating the objective function of an instance of problem $\mathcal{P}_K$. To this end, fix an instance $(\boldsymbol{A}, \boldsymbol{b})$ of the IP feasibility problem, and let $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^Q : \boldsymbol{A\xi} \leq \boldsymbol{b},\, \boldsymbol{0} \leq \boldsymbol{\xi} \leq \mathbf{e}\}$ denote the LP relaxation of the IP problem's feasible region. Set $\boldsymbol{y}^k = \mathbf{e}_k$, $k = 1, \ldots, 2Q$, where $\mathbf{e}_k$ denotes the k-th unit basis vector in $\mathbb{R}^{2Q}$. We claim that the answer to the IP feasibility problem is affirmative if and only if

$$\max_{\boldsymbol{\xi} \in \Xi} \min_{k \in \mathcal{K}} \left\{ (\boldsymbol{\xi}^\top,\, \mathbf{e}^\top - \boldsymbol{\xi}^\top)\, \boldsymbol{y}^k \,:\, \boldsymbol{y}^k_q \leq 2\boldsymbol{\xi}_q,\; \boldsymbol{y}^k_{Q+q} \leq 2(1 - \boldsymbol{\xi}_q),\, q = 1, \ldots, Q \right\} \tag{EC.3}$$

is greater than or equal to 1. Note that (EC.3) can be interpreted as the objective function of an instance of problem $\mathcal{P}_K$ if we set $N = 0$, $M = K = 2Q$ and $\mathcal{Y} = \{\boldsymbol{y} \in \{0,1\}^{2Q} : \mathbf{e}^\top \boldsymbol{y} = 1\}$.

To prove our claim, assume first that the answer to the IP feasibility problem is affirmative, that is, there is $\boldsymbol{\xi}^\star \in \Xi$ such that $\boldsymbol{\xi}^\star \in \{0,1\}^Q$. The policy $\boldsymbol{y}^q$, $q = 1, \ldots, Q$, satisfies the constraints in (EC.3) for $\boldsymbol{\xi} = \boldsymbol{\xi}^\star$ if and only if $\boldsymbol{\xi}^\star_q = 1$, in which case the objective function within the minimization evaluates to $(\boldsymbol{\xi}^{\star\top},\, \mathbf{e}^\top - \boldsymbol{\xi}^{\star\top})\, \mathbf{e}_q = \boldsymbol{\xi}^\star_q = 1$. Likewise, the policy $\boldsymbol{y}^{Q+q}$, $q = 1, \ldots, Q$, satisfies the constraints in (EC.3) for $\boldsymbol{\xi} = \boldsymbol{\xi}^\star$ if and only if $\boldsymbol{\xi}^\star_q = 0$, in which case the objective function within the minimization evaluates to $(\boldsymbol{\xi}^{\star\top},\, \mathbf{e}^\top - \boldsymbol{\xi}^{\star\top})\, \mathbf{e}_{Q+q} = 1 - \boldsymbol{\xi}^\star_q = 1$. We thus conclude that the minimum in (EC.3) evaluates to 1 for $\boldsymbol{\xi}^\star$, which implies that the value of the overall expression (EC.3) is greater than or equal to 1.

Assume now that the value of the expression in (EC.3) is greater than or equal to 1. Fix any maximizer $\boldsymbol{\xi}^\star$ for this expression and consider any component $\boldsymbol{\xi}^\star_q$ of $\boldsymbol{\xi}^\star$, $q = 1, \ldots, Q$. If $\boldsymbol{\xi}^\star_q \geq 1/2$, then $\boldsymbol{y}^q = \mathbf{e}_q$ is a feasible second-stage policy, and we conclude that $(\boldsymbol{\xi}^{\star\top},\, \mathbf{e}^\top - \boldsymbol{\xi}^{\star\top})\, \mathbf{e}_q = \boldsymbol{\xi}^\star_q \geq 1$. If $\boldsymbol{\xi}^\star_q \leq 1/2$, on the other hand, then $\boldsymbol{y}^{Q+q} = \mathbf{e}_{Q+q}$ is a feasible second-stage policy, and we conclude that $(\boldsymbol{\xi}^{\star\top},\, \mathbf{e}^\top - \boldsymbol{\xi}^{\star\top})\, \mathbf{e}_{Q+q} = 1 - \boldsymbol{\xi}^\star_q \geq 1$, that is, $\boldsymbol{\xi}^\star_q \leq 0$. Since $\boldsymbol{\xi}^\star_q \in [0,1]$ by construction of Ξ, we thus conclude that $\boldsymbol{\xi}^\star_q \in \{0,1\}$. Since q was chosen arbitrarily, the answer to the IP feasibility problem is affirmative. This concludes the proof. □

Proof of Theorem 4. Consider the following instance of problem $\mathcal{P}$ where no first-stage decision is taken.

$$\begin{aligned} &\text{maximize} && \min_{\boldsymbol{y}\in\{0,1\}^Q} \{0 : \boldsymbol{y}_q - \boldsymbol{\xi}_q \le 1/2,\ \boldsymbol{\xi}_q - \boldsymbol{y}_q \le 1/2,\ q=1,\ldots,Q\} \\ &\text{subject to} && \boldsymbol{\xi}\in\mathbb{R}^Q \\ &&& 0\le \boldsymbol{\xi}_q \le 1,\ q=1,\ldots,Q \end{aligned} \tag{EC.4}$$

The minimization in the objective function of (EC.4) has a nonempty feasible region. In fact, for any $\boldsymbol{\xi}\in[0,1]^Q$, the second-stage policy $\boldsymbol{y}(\boldsymbol{\xi})$ defined through $[\boldsymbol{y}(\boldsymbol{\xi})]_q = \mathbb{I}(\boldsymbol{\xi}_q \ge 1/2),\ q=1,\ldots,Q$, satisfies the second-stage constraints in the objective function of (EC.4). We thus conclude that the optimal value of problem (EC.4) is 0.

We claim that the optimal value of the K-adaptability problem associated with (EC.4),

$$\begin{aligned} &\text{minimize} && \max_{\boldsymbol{\xi}\in[0,1]^Q} \min_{k\in\mathcal{K}} \left\{0 : \boldsymbol{y}^k_q - \boldsymbol{\xi}_q \le 1/2,\ \boldsymbol{\xi}_q - \boldsymbol{y}^k_q \le 1/2, q=1,\ldots,Q\right\} \\ &\text{subject to} && \boldsymbol{y}^k \in \{0,1\}^Q,\ k\in\mathcal{K}, \end{aligned} \tag{EC.5}$$

is unbounded whenever $K < |\mathcal{Y}| = 2^Q$. In fact, assume that a solution $\{\boldsymbol{y}^k\}_{k\in\mathcal{K}}$ to (EC.5) does not contain the policy $\boldsymbol{y}^\star \in \{0,1\}^Q$. Then the objective function in (EC.5) is unbounded since there is no policy $\boldsymbol{y}^k$, $k\in\mathcal{K}$, that satisfies the second-stage constraints in the objective function of (EC.5) for the parameter realization $\boldsymbol{\xi} = \boldsymbol{y}^\star$. We thus conclude that the problems (EC.4) and (EC.5) only attain the same optimal values if $K = |\mathcal{Y}|$. □

Proof of Proposition 1. We first show that $\{\Xi(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\mathcal{L}}$ is a cover of Ξ, that is, $\bigcup_{\boldsymbol{\ell}\in\mathcal{L}} \Xi(\boldsymbol{\ell}) = \Xi$. To this end, fix any $\boldsymbol{\xi}\in\Xi$ and choose $\boldsymbol{\ell}\in\mathcal{L}$ such that

$$\boldsymbol{\ell}_k = \begin{cases} 0 & \text{if } \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \le \boldsymbol{H}\boldsymbol{\xi}, \\ \min\left\{l\in\{1,\ldots,L\} : [\boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k]_l > [\boldsymbol{H}\boldsymbol{\xi}]_l\right\} & \text{otherwise} \end{cases} \qquad \forall k\in\mathcal{K}.$$

By construction, we have $\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})$. Conversely, the definition of $\Xi(\boldsymbol{\ell})$ implies that $\Xi(\boldsymbol{\ell})\subseteq\Xi$ for all $\boldsymbol{\ell}\in\mathcal{L}$. We thus conclude that $\{\Xi(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\mathcal{L}}$ is indeed a cover of Ξ. Note that $\{\Xi(\boldsymbol{\ell})\}_{\boldsymbol{\ell}\in\mathcal{L}}$ is in general not a partition of Ξ since the sets $\Xi(\boldsymbol{\ell})$ overlap whenever a policy $\boldsymbol{y}^k$ violates multiple constraints.

Our findings imply that the objective function of problem $\mathcal{P}_K$ satisfies

$$\begin{aligned}
&\max_{\boldsymbol{\xi}\in\Xi}\left[\boldsymbol{\xi}^\top\boldsymbol{C}\boldsymbol{x}+\min_{k\in\mathcal{K}}\left\{\boldsymbol{\xi}^\top\boldsymbol{Q}\boldsymbol{y}^k : \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{H}\boldsymbol{\xi}\right\}\right]\\
=&\max_{\boldsymbol{\xi}\in\bigcup_{\boldsymbol{\ell}\in\mathcal{L}}\Xi(\boldsymbol{\ell})}\left[\boldsymbol{\xi}^\top\boldsymbol{C}\boldsymbol{x}+\min_{k\in\mathcal{K}}\left\{\boldsymbol{\xi}^\top\boldsymbol{Q}\boldsymbol{y}^k : \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{H}\boldsymbol{\xi}\right\}\right]\\
=&\max_{\boldsymbol{\ell}\in\mathcal{L}}\max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})}\left[\boldsymbol{\xi}^\top\boldsymbol{C}\boldsymbol{x}+\min_{k\in\mathcal{K}}\left\{\boldsymbol{\xi}^\top\boldsymbol{Q}\boldsymbol{y}^k : \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{H}\boldsymbol{\xi}\right\}\right].
\end{aligned}$$

We now note that by definition of the set $\Xi(\boldsymbol{\ell})$, the indices $k\in\mathcal{K}$ for which the associated second-stage policies $\boldsymbol{y}^k$ satisfy the constraints $\boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k\le\boldsymbol{H}\boldsymbol{\xi}$ are precisely those indices $k\in\mathcal{K}$ for which $\boldsymbol{\ell}_k=0$. This concludes the proof. □

The proof of Proposition 2 requires the following auxiliary result, which we prove first.

Lemma 1 *Consider the uncertainty sets* $\Xi(\boldsymbol{\ell})$ *and* $\Xi_\epsilon(\boldsymbol{\ell})$ *defined in* (6) *and* (6_ϵ)*, respectively. Fix* $\boldsymbol{\ell}\in\mathcal{L}$ *and assume that* $\Xi(\boldsymbol{\ell})\neq\emptyset$*. Then the sets* $\Xi_\epsilon(\boldsymbol{\ell})$ *converge to* $\Xi(\boldsymbol{\ell})$ *in Hausdorff distance as* ϵ *approaches* 0*, that is, we have*

$$\lim_{\epsilon\downarrow 0} d^{\mathrm{H}}(\Xi_\epsilon(\boldsymbol{\ell}),\Xi(\boldsymbol{\ell}))=0,$$

where the Hausdorff distance between two sets X *and* Y *is defined as*

$$d^{\mathrm{H}}(X,Y)=\max\left\{\sup_{\boldsymbol{x}\in X}\inf_{\boldsymbol{y}\in Y}\|\boldsymbol{x}-\boldsymbol{y}\|,\ \sup_{\boldsymbol{y}\in Y}\inf_{\boldsymbol{x}\in X}\|\boldsymbol{x}-\boldsymbol{y}\|\right\}.$$

Proof. Note that $\Xi_\epsilon(\boldsymbol{\ell})=\Xi(\boldsymbol{\ell})$ for all $\epsilon>0$ if $\boldsymbol{\ell}=\boldsymbol{0}$. In the remainder, we thus assume that $\boldsymbol{\ell}\neq\boldsymbol{0}$. We first show that $\Xi_\epsilon(\boldsymbol{\ell})\neq\emptyset$ for sufficiently small $\epsilon>0$. To this end, select any $\overline{\boldsymbol{\xi}}\in\Xi(\boldsymbol{\ell})$ and define

$$\epsilon'=\min\left\{[\boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k}-[\boldsymbol{H}\overline{\boldsymbol{\xi}}]_{\boldsymbol{\ell}_k} : k\in\mathcal{K},\ \boldsymbol{\ell}_k\neq 0\right\}.$$

By construction, $\epsilon'>0$ and $\overline{\boldsymbol{\xi}}\in\Xi_\epsilon(\boldsymbol{\ell})$ for all $\epsilon\in(0,\epsilon']$, that is, $\Xi_\epsilon(\boldsymbol{\ell})\neq\emptyset$ for sufficiently small ϵ. From now on, we always assume that ϵ is chosen so that $\Xi_\epsilon(\boldsymbol{\ell})$ is nonempty.

By definition of $\Xi_\epsilon(\boldsymbol{\ell})$, we have that $\Xi_\epsilon(\boldsymbol{\ell})\subseteq\Xi(\boldsymbol{\ell})$. Hence, we only have to prove that

$$\forall\kappa>0\ \exists\overline{\epsilon}>0 :\ \sup_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})}\inf_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell})}\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|\le\kappa\ \forall\epsilon\in(0,\overline{\epsilon}].$$

Assume to the contrary that there is $\kappa > 0$ such that

$$\forall i \in \mathbb{N}\, \exists \epsilon_i \in (0, 1/i],\, \boldsymbol{\xi}^i \in \Xi(\boldsymbol{\ell}) \,:\, \left\|\boldsymbol{\xi}^i - \boldsymbol{\xi}'\right\| > \kappa \;\; \forall \boldsymbol{\xi}' \in \Xi_\epsilon(\boldsymbol{\ell}).$$

Since $\Xi(\boldsymbol{\ell}) \subseteq \Xi$ and Ξ is compact, we can apply the Bolzano-Weierstrass Theorem to conclude that the sequence $\boldsymbol{\xi}^i$ has an accumulation point $\boldsymbol{\xi}^\star \in \mathrm{cl}\,\Xi(\boldsymbol{\ell})$. However, by construction $\boldsymbol{\xi}^\star$ satisfies $\|\boldsymbol{\xi}^\star - \boldsymbol{\xi}'\| \geq \kappa$ for all $\boldsymbol{\xi}' \in \bigcup_{i \in \mathbb{N}} \Xi_{\epsilon_i}(\boldsymbol{\ell}) = \Xi(\ell)$, which contradicts the statement that $\boldsymbol{\xi}^\star \in \mathrm{cl}\,\Xi(\boldsymbol{\ell})$. We thus conclude that our assumption was wrong, that is, the assertion in the statement of the lemma indeed holds. □

Proof of Proposition 2. In view of the first statement, Lemma 1 implies that for sufficiently small $\epsilon > 0$, the approximate uncertainty sets $\Xi_\epsilon(\boldsymbol{\ell})$, $\boldsymbol{\ell} \in \mathcal{L}$, are nonempty if and only if the exact uncertainty sets $\Xi(\boldsymbol{\ell})$ are. Thus, we have $\mathrm{dom}\,(6_\epsilon) = \mathrm{dom}\,(6)$ for sufficiently small $\epsilon > 0$.

As for the second statement, fix any $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}}) \in \mathrm{dom}\,(6)$ and assume that $\epsilon > 0$ is small enough so that $\mathrm{dom}\,(6_\epsilon) = \mathrm{dom}\,(6)$. Denote by φ_ϵ and φ the objective values of $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}})$ in problems (6_ϵ) and (6), respectively. We then have that

$$\begin{aligned}
|\varphi_\epsilon - \varphi| &\overset{\text{(a)}}{=} \max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] - \max_{\boldsymbol{\ell}'\in\mathcal{L}} \max_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}'^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k'\in\mathcal{K}:\\ \boldsymbol{\ell}'_{k'}=0}} \boldsymbol{\xi}'^\top \boldsymbol{Q}\,\boldsymbol{y}^{k'}\right] \\
&= \max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \min_{\boldsymbol{\ell}'\in\mathcal{L}} \min_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell}')} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] - \left[\boldsymbol{\xi}'^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k'\in\mathcal{K}:\\ \boldsymbol{\ell}'_{k'}=0}} \boldsymbol{\xi}'^\top \boldsymbol{Q}\,\boldsymbol{y}^{k'}\right] \\
&\overset{\text{(b)}}{\leq} \max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \min_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] - \left[\boldsymbol{\xi}'^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\substack{k'\in\mathcal{K}:\\ \boldsymbol{\ell}_{k'}=0}} \boldsymbol{\xi}'^\top \boldsymbol{Q}\,\boldsymbol{y}^{k'}\right] \\
&\overset{\text{(c)}}{\leq} \max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \min_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell})} \max_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] - \left[\boldsymbol{\xi}'^\top \boldsymbol{C}\,\boldsymbol{x} + \boldsymbol{\xi}'^\top \boldsymbol{Q}\,\boldsymbol{y}^k\right] \\
&\overset{\text{(d)}}{\leq} \left(\max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi(\boldsymbol{\ell})} \min_{\boldsymbol{\xi}'\in\Xi_\epsilon(\boldsymbol{\ell})} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|\right) \cdot \left(\max_{k\in\mathcal{K}} \left\|\boldsymbol{C}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{y}^k\right\|\right),
\end{aligned}$$

where (a) holds because $\varphi \geq \varphi_\epsilon$ since $\Xi_\epsilon(\boldsymbol{\ell}) \subseteq \Xi(\boldsymbol{\ell})$ for all $\boldsymbol{\ell} \in \mathcal{L}$, (b) and (c) follow since we restrict the minimizations to $\boldsymbol{\ell}' = \boldsymbol{\ell}$ and to the minimizer in the second pair of parentheses, respectively, and (d) is due to the Cauchy-Schwarz inequality. Lemma 1 implies that the first product term in the last expression can be made arbitrarily small by choosing $\epsilon > 0$ appropriately. The second product term, on the other hand, is bounded over $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}}) \in \text{dom}\,(6)$ since $\mathcal{X}$ and $\mathcal{Y}$ are bounded. For sufficiently small $\epsilon > 0$, we can thus bound the difference $|\varphi_\epsilon - \varphi|$ uniformly over $(\boldsymbol{x}, \{\boldsymbol{y}^k\}_{k\in\mathcal{K}}) \in \text{dom}\,(6)$ by an arbitrarily small constant, which concludes the proof. $\square$

Proof of Theorem 5. The objective function of the approximate problem (6_ϵ) is identical to

$$\max_{\boldsymbol{\ell}\in\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\boldsymbol{\lambda}\in\Delta_K(\boldsymbol{\ell})} \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k \right],$$

where $\Delta_K(\boldsymbol{\ell}) = \{\boldsymbol{\lambda} \in \mathbb{R}^K_+ : \mathbf{e}^\top\boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda}_k = 0\ \forall k \in \mathcal{K} : \boldsymbol{\ell}_k \neq 0\}$. Note that $\Delta_K(\boldsymbol{\ell}) = \emptyset$ if and only if $\boldsymbol{\ell} > \mathbf{0}$. If $\Xi_\epsilon(\boldsymbol{\ell}) = \emptyset$ for all $\boldsymbol{\ell} \in \mathcal{L}_+$, then the problem is equivalent to

$$\max_{\boldsymbol{\ell}\in\partial\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \min_{\boldsymbol{\lambda}\in\Delta_K(\boldsymbol{\ell})} \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k \right],$$

and we can apply the classical min-max theorem (since $\Delta_K(\boldsymbol{\ell})$ is nonempty for all $\boldsymbol{\ell} \in \partial\mathcal{L}$) to obtain the equivalent reformulation

$$\max_{\boldsymbol{\ell}\in\partial\mathcal{L}} \min_{\boldsymbol{\lambda}\in\Delta_K(\boldsymbol{\ell})} \max_{\boldsymbol{\xi}\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k \right],$$

which in turn is equivalent to

$$\min_{\substack{\boldsymbol{\lambda}(\boldsymbol{\ell})\in\Delta_K(\boldsymbol{\ell}),\\ \boldsymbol{\ell}\in\partial\mathcal{L}}} \max_{\boldsymbol{\ell}\in\partial\mathcal{L}} \max_{\boldsymbol{\xi}\in\Xi_\epsilon(\boldsymbol{\ell})} \left[\boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k(\boldsymbol{\ell}) \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k \right].$$

If, on the other hand, $\Xi_\epsilon(\boldsymbol{\ell}) \neq \emptyset$ for some $\boldsymbol{\ell} \in \mathcal{L}_+$, then the objective function in (6_ϵ) evaluates to $+\infty$. Using an epigraph reformulation, we thus conclude that (6_ϵ) is equivalent to the problem

$$\begin{array}{lll}
\text{minimize} & \tau & \\
\text{subject to} & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y}^k \in \mathcal{Y},\, k \in \mathcal{K}, & \\
& \tau \in \mathbb{R},\ \boldsymbol{\lambda}(\boldsymbol{\ell}) \in \Delta_K(\boldsymbol{\ell}),\, \boldsymbol{\ell} \in \partial\mathcal{L} & \\
& \tau \geq \boldsymbol{\xi}^\top \boldsymbol{C}\,\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k(\boldsymbol{\ell}) \cdot \boldsymbol{\xi}^\top \boldsymbol{Q}\,\boldsymbol{y}^k & \forall \boldsymbol{\ell} \in \partial\mathcal{L},\, \forall \boldsymbol{\xi} \in \Xi_\epsilon(\boldsymbol{\ell}) \\
& \Xi_\epsilon(\boldsymbol{\ell}) = \emptyset & \forall \boldsymbol{\ell} \in \mathcal{L}_+.
\end{array} \tag{EC.6}$$

The semi-infinite constraint associated with $\boldsymbol{\ell} \in \partial \mathcal{L}$ is satisfied if and only if the optimal value of

$$\begin{aligned}
\text{maximize} \quad & \Big[\boldsymbol{C}\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k(\boldsymbol{\ell})\boldsymbol{Q}\boldsymbol{y}^k\Big]^\top \boldsymbol{\xi} \\
\text{subject to} \quad & \boldsymbol{\xi} \in \mathbb{R}^Q \\
& \boldsymbol{A}\boldsymbol{\xi} \le \boldsymbol{b} \\
& \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k \le \boldsymbol{H}\boldsymbol{\xi} \qquad && \forall k \in \mathcal{K} : \boldsymbol{\ell}_k = 0 \\
& [\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} \ge [\boldsymbol{H}\boldsymbol{\xi}]_{\boldsymbol{\ell}_k} + \epsilon \qquad && \forall k \in \mathcal{K} : \boldsymbol{\ell}_k \ne 0
\end{aligned}$$

does not exceed τ. Strong linear programming duality implies that this problem attains the same optimal value as its dual problem, which is given by

$$\begin{aligned}
\text{minimize} \quad & \boldsymbol{b}^\top \boldsymbol{\alpha} - \sum_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} (\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k)^\top \boldsymbol{\beta}^k + \sum_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k\ne 0}} \left([\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} - \epsilon\right)\gamma_k \\
\text{subject to} \quad & \boldsymbol{\alpha} \in \mathbb{R}^R_+,\ \boldsymbol{\beta}^k \in \mathbb{R}^L_+,\ k \in \mathcal{K},\ \boldsymbol{\gamma} \in \mathbb{R}^K_+ \\
& \boldsymbol{A}^\top \boldsymbol{\alpha} - \sum_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k=0}} \boldsymbol{H}^\top \boldsymbol{\beta}^k + \sum_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k\ne 0}} \boldsymbol{H}_{\boldsymbol{\ell}_k}\gamma_k = \boldsymbol{C}\boldsymbol{x} + \sum_{k\in\mathcal{K}} \boldsymbol{\lambda}_k(\boldsymbol{\ell})\boldsymbol{Q}\boldsymbol{y}^k.
\end{aligned}$$

Strong duality holds because the dual problem is always feasible. Indeed, one can show that the compactness of Ξ implies that $\{\boldsymbol{A}^\top\boldsymbol{\alpha} : \boldsymbol{\alpha} \ge \boldsymbol{0}\} = \mathbb{R}^Q$. Note that the first constraint set in problem (7) ensures that the optimal value of this dual problem does not exceed τ for all $\boldsymbol{\ell} \in \partial\mathcal{L}$.

The last constraint in (EC.6) is satisfied for $\boldsymbol{\ell} \in \mathcal{L}_+$ whenever the linear program

$$\begin{aligned}
\text{maximize} \quad & 0 \\
\text{subject to} \quad & \boldsymbol{\xi} \in \mathbb{R}^Q \\
& \boldsymbol{A}\boldsymbol{\xi} \le \boldsymbol{b} \\
& [\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} \ge [\boldsymbol{H}\boldsymbol{\xi}]_{\boldsymbol{\ell}_k} + \epsilon \qquad \forall k \in \mathcal{K}
\end{aligned}$$

is infeasible. Strong LP duality implies that this is the case whenever the dual problem,

$$\begin{aligned}
\text{minimize} \quad & \boldsymbol{b}^\top \boldsymbol{\alpha} + \sum_{\substack{k\in\mathcal{K}:\\ \boldsymbol{\ell}_k\ne 0}} \left([\boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y}^k]_{\boldsymbol{\ell}_k} - \epsilon\right)\gamma_k \\
\text{subject to} \quad & \boldsymbol{\alpha} \in \mathbb{R}^R_+,\ \boldsymbol{\gamma} \in \mathbb{R}^K_+ \\
& \boldsymbol{A}^\top \boldsymbol{\alpha} + \sum_{k\in\mathcal{K}} \boldsymbol{H}_{\boldsymbol{\ell}_k}\gamma_k = \boldsymbol{0},
\end{aligned}$$

is unbounded. Note that strong duality holds because $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\gamma} = \mathbf{0}$ are feasible in the dual problem. Since the feasible region of the dual problem constitutes a cone, the dual problem is unbounded if and only if it admits a feasible solution that achieves an objective value of -1 or less. This is precisely what the last constraint set in problem (7) requires. $\square$

Acknowledgments

The authors gratefully acknowledge the two anonymous reviewers whose comments led to substantial improvements of the paper, as well as Angelos Georghiou for providing the implementation of the binary decision rule algorithm of Bertsimas and Georghiou (2014). This research was supported by EPSRC under grant EP/I014640/1.