informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 73, No. 6, November–December 2025, pp. 3345–3358
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Technical Note—Multistage Robust Mixed-Integer Programming

Krzysztof Postek,[a] Ward Romeijnders,[b] Wolfram Wiesemann[c,*]

[a] Delft University of Technology, 2628 CD Delft, Netherlands; [b] University of Groningen, 9712 CP Groningen, Netherlands; [c] Imperial College London, London SW7 2AZ, United Kingdom

*Corresponding author

Contact: k.s.postek@tudelft.nl, https://orcid.org/0000-0002-4028-3725 (KP); w.romeijnders@rug.nl, https://orcid.org/0000-0002-6918-5635 (WR); ww@imperial.ac.uk, https://orcid.org/0000-0003-3076-1591 (WW)

Received: September 26, 2023
Revised: September 11, 2024; November 29, 2024
Accepted: December 5, 2024
Published Online in Articles in Advance: January 30, 2025

Area of Review: Optimization

https://doi.org/10.1287/opre.2023.0520

Abstract. Multistage robust optimization, in which decisions are taken sequentially as new information becomes available about uncertain problem parameters, is a very versatile yet computationally challenging paradigm for decision making under uncertainty. In this technical note, we propose a new model and solution approach for multistage robust mixed-integer programs, which may contain both continuous and discrete decisions at any time stage. Our model builds upon the finite adaptability scheme developed for two-stage robust optimization problems, and it allows us to decompose the multistage problem into a large number of much simpler two-stage problems. We discuss how these two-stage problems can be solved both exactly and approximately, and we report numerical results for route planning and location-transportation problems.

Funding: The research of Ward Romeijnders has been supported by the Netherlands Organisation for Scientific Research [Grant VI.Vidi.211.066]. W. Wiesemann gratefully acknowledges funding from the EPSRC [Grants EP/R045518/1 and EP/W003317/1]. The research of K. Postek has been partly supported by the Netherlands Organisation for Scientific Research [Grant VI.Veni.191E.035].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2023.0520.

Keywords: robust optimization • multistage problems • mixed-integer optimization

1. Introduction

Real-life decisions almost inevitably need to be taken under considerable uncertainty about key problem parameters such as future customer demands, raw material prices, exchange rates, equipment outages, and traffic conditions. Among the many paradigms that have been developed to safeguard decisions against uncertainty, such as simulation-based optimization, stochastic programming, and Markov decision processes, the relatively young field of robust optimization stands out because of its promise to scale to the large problem sizes commonly encountered in application areas. This is achieved by replacing probabilistic descriptions of the uncertainty with a set-based characterization under which the uncertain problem parameters can attain any value from a prespecified *uncertainty set*, and a decision is sought that performs best in view of the worst anticipated parameter realizations (Ben-Tal et al. 2009, Bertsimas et al. 2011a, Bertsimas and den Hertog 2022).

Whereas robust optimization has initially been developed for single-stage problems where all decisions are taken here-and-now, subsequent research has studied multistage formulations where the uncertain problem parameters are revealed gradually over time and future recourse decisions can depend on the parameters that have already been observed. In contrast to single-stage robust optimization problems that can often be solved in polynomial time, multistage robust optimization problems are NP-hard even if only two stages are considered, all decisions are continuous, and the objective function, as well as all constraints, are linear (Guslitser 2002). Two-stage robust optimization problems with continuous recourse have been solved exactly via Benders' decomposition, column-and-constraint generation, Fourier-Motzkin elimination, copositive programming, or iteratively lifting the uncertainty set; none of these techniques scale well beyond two stages. Multistage robust optimization problems with more than two stages are typically solved approximately by restricting the recourse decisions to decision rules or via an adaptation of stochastic dual dynamic programming. We refer to Ben-Tal et al. (2009), Delage and Iancu (2015), Yanıkoğlu et al. (2019), and Bertsimas and den Hertog (2022) for surveys of the literature.

All of the aforementioned approaches have in common that they require the recourse decisions to be continuous. The ubiquitous presence of integer recourse decisions in practical applications, such as whether to place an order, build a facility, or change the operation of a plant, has sparked significant interest in the development of solution schemes for multistage robust mixed-integer problems. However, the presence of integer recourse decisions prohibits a straightforward application of the aforementioned decision rules, and it precludes the use of strong convex duality results upon which most of the tractable reformulations from the continuous robust optimization literature rely. Two-stage robust optimization problems with mixed-integer recourse decisions have been solved exactly by semi-infinite programming techniques (Zhao and Zeng 2012) and (hybrid) Benders' decomposition (Hashemi Doulabi et al. 2021) as well as approximately by K-adaptability schemes (Bertsimas and Caramanis 2010, Hanasusanto et al. 2015, Subramanyam et al. 2020) that restrict the second-stage decisions to one out of K candidate solutions that are optimized over in the first stage. We are not aware of any successful attempts to generalize either of these techniques to multistage robust mixed-integer problems with more than two time stages, however.

In this technical note, we propose to approximate multistage robust mixed-integer programs by a finite adaptability formulation. Our formulation selects, in each time stage, the best mixed-integer state decision from a finite set of preselected candidate decisions. Continuous control decisions that only affect the feasibility and optimality within a stage, on the other hand, can be selected optimally from predefined polyhedral regions. We show that in contrast to the original multistage robust mixed-integer program, the finite adaptability approximation admits an equivalent nested formulation that can be solved via backward recursion. This allows us to reduce the monolithic multistage problem to a number of much simpler two-stage problems, many of which can be solved in parallel. We show how the arising two-stage problems can be solved exactly or approximately. We also discuss various heuristic strategies to select sets of candidate state decisions for each stage, and we show how to deterministically bound the suboptimality incurred by the current choice of candidate state decision sets. Our contributions can be summarized as follows:

(i) We conservatively approximate multistage robust mixed-integer programs via a finite adaptability approximation that admits an equivalent decomposition into two-stage subproblems. To our best knowledge, this is the first approach that decomposes multistage robust mixed-integer programs into smaller subproblems that can be solved independently.

(ii) The subproblems of our decomposition scheme constitute two-stage robust optimization problems with a particularly benign structure. We show how these problems can be solved exactly through mixed-integer programming reformulations, as well as approximately by extending recent results from the literature on two-stage robust optimization.

(iii) We demonstrate the promise of our framework in the context of two numerical experiments: a route planning problem involving a graph with 100 nodes and 40 time stages as well as a location-transportation problem involving up to 20 facilities, 40 customer sites, and 10 time stages. Our source codes, as well as all data sets, are released open-source to facilitate comparison with alternative approaches as well as reuse in applications.[1]

Our solution approach builds upon the multistage robust mixed-integer programming literature, which can be broadly categorized into two streams: (i) generalizations of the decision rule schemes developed for continuous problems, and (ii) uncertainty set-partitioning schemes. In the following, we summarize the most prominent approaches of both streams, and we subsequently explain how our proposed method differs from the literature.

The first decision rule architecture for multistage robust mixed-integer programs has been proposed by Bertsimas and Caramanis (2007). The authors model the recourse decisions as affine functions of features formed from the observed parameters. To ensure integrality of the discrete recourse decisions, the parameter realizations are rounded up to the closest integers, and the intercepts and slopes of the corresponding affine functions are restricted to integer numbers. The resulting semi-infinite mixed-integer linear program (MILP) is solved by constraint sampling, which results in probabilistic feasibility and optimality guarantees. Bertsimas and Georghiou (2015), on the other hand, model the continuous recourse decisions as piecewise affine functions of the observed parameters, whereas the discrete decisions (which are assumed to be binary) attain the value of one precisely when certain piecewise affine functions of the parameters are nonnegative. The problem can be formulated as a semi-infinite MILP that is solved by an iterative procedure that alternates between determining the optimal decision rules for a fixed set of parameter realizations and identifying new worst-case parameter realizations for the updated decision rules. For computational reasons, the algorithm is typically terminated prematurely, which implies that the solution may violate the constraints for some scenarios, and the true worst-case costs may exceed the worst-case costs estimated by the procedure. For the same problem class, Bertsimas and Georghiou (2018) model the binary recourse decisions as affine functions of nonanticipative binary features formed from the observed parameters. Binarity of the recourse decisions is achieved by requiring the intercepts and slopes of these affine functions to be integer and by restricting the image of the decision rules to be binary. Using convex duality arguments, the authors obtain a finite-dimensional MILP whose size grows exponentially in general but remains polynomial if the uncertainty set is a hyperrectangle and each feature mapping involves a single parameter. The approach has been extended to endogenous (decision-dependent) uncertainty by Feng et al. (2021). Bertsimas et al. (2011b) show that for multistage robust mixed-integer programs in which only the cost vector and the constraint right-hand sides are uncertain, the cost of a finitely adaptable solution is bounded by a constant times the optimal cost of a fully adaptable solution. More recently, Daryalal et al. (2023) propose a primal-dual approach for multistage robust mixed-integer programming. The authors develop a conservative primal approximation that replaces the state decisions (which link consecutive time stages) with decision rules, whereas the control decisions (which only affect the feasibility and objective function value within a stage) are solved optimally. For integer recourse decisions, the authors propose a new class of piecewise constant decision rules that enumerate the admissible values of each integer decision and select the value based on the realization of a nonanticipate affine decision rule. The resulting two-stage robust mixed-integer program is solved via column-and-constraint generation or K-adaptability schemes. The authors assess the suboptimality of their primal solution through the computation of a dual bound that applies an affine decision rule approximation to a strong Lagrange dual of the original problem and that solves the resulting two-stage stochastic program using bilinear programming or cutting plane techniques.

The first uncertainty set-partitioning scheme for multistage robust mixed-integer programs has been proposed by Vayanos et al. (2011), who prepartition the uncertainty set in each time stage into hyperrectangles and select affine decision rules (for the continuous recourse decisions) as well as constant decision rules (for the integer recourse decisions) over each hyperrectangle. By invoking convex duality arguments, the authors solve the corresponding conservative approximation exactly as a finite-dimensional MILP. The authors also discuss how their approach generalizes to the presence of endogenous (decision-dependent) uncertainty. Subsequent extensions of the uncertainty set-partitioning paradigm have continued to rely on affine and constant decision rules for the continuous and discrete recourse decisions, respectively, but they seek to refine the partitions adaptively in view of the incumbent solutions. Postek and den Hertog (2016) alternate between determining the best decision rules for a fixed partition of the uncertainty set and identifying subsets of the partition where multiple parameter realizations result in binding constraints, thus indicating the need to further subdivide these subsets to obtain better decisions. In the multistage case, the stagewise partitions form a tree structure that is similar to the scenario trees in stochastic programming. This partitioning scheme is very effective when the adaptive problem at hand has a small number of time stages, but it exhibits an exponential growth if many time stages are involved. Romeijnders and Postek (2021) refine the splitting technique of Postek and den Hertog (2016) in the presence of integer recourse decisions, where subsets of the partition may need to be split even if they do not generate binding constraints. The authors show that critical parameter realizations can be identified from the LP relaxations in the branch-and-bound tree that determines the decision rules. Bertsimas and Dunning (2015) propose an alternative scheme where the uncertainty set in each time stage is partitioned by a Voronoi diagram that is constructed from the binding scenarios, thus eliminating the need to choose splits manually. Because their approach splits each subset of the partition in every iteration, however, the size of the partition grows quickly in the number of algorithm iterations, even for two-stage problems.

In contrast to the aforementioned decision rule architectures and uncertainty set-partitioning schemes, Goerigk and Hartisch (2021) model multistage robust pure integer programs with pure integer uncertainties as quantified integer linear programs, which constitute two-person zero-sum games between an existential player (the "minimizer") and a universal player (the "maximizer"). The resulting problems can be solved with an open-source quantified integer programming solver.

The method developed in this technical note complements the existing solution schemes for multistage robust mixed-integer programs. The uncertainty set-partitioning approaches from the literature are essentially free of hyperparameters and thus do not require any a priori knowledge about the problem. On the flip side, they construct set-based analogues of scenario trees that exhibit exponential growth in the number of time stages as well as, typically, the number of uncertain problem parameters per stage. They are thus ideally suited for smaller problems where the decision maker has little a priori insight into the structure of well-performing solutions. Most of the decision rule architectures, on the other hand, require the functional form of the recourse decisions to be selected up front and thus rely on domain knowledge of the decision maker. In contrast to the uncertainty set partitioning approaches, however, decision rule architectures tend to scale polynomially in the number of decision variables and time stages, which renders them promising for larger problems. Similar to the decision rule architectures, our method requires a priori knowledge about the problem to select suitable candidate decisions for each time stage. On the other hand, because our approach decomposes the overall problem into smaller two-stage subproblems, our method scales particularly well in the number of time stages. Moreover, and in contrast to all of the existing approaches, our method is ideally suited for parallelization because many of the subproblems can be solved in parallel. This is attractive in view of the recent growth in cloud computing services, which enable users to rent vast amounts of parallel computing resources at an hourly billing. On the other hand, because of its reliance on preselected candidate decisions, our algorithm is unlikely to perform well when a problem contains large numbers of continuous and/or general integer state variables (as opposed to binary ones); we expect existing approaches to be better suited for such problems.

The remainder of this technical note proceeds as follows. Section 2 formulates the multistage robust mixed-integer program of interest, it presents our finite adaptability approximation, and it elucidates how this approximation admits a decomposition into two-stage subproblems. Section 3 discusses exact and approximate solution approaches for the two-stage subproblems. Section 4 presents a progressive bound to estimate the suboptimality of our approximation, and it develops heuristic strategies to select candidate state decision sets for each stage. We conclude with numerical experiments in Section 5. Further details, including all proofs, are relegated to the Online Appendix.

Notation. For a vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)$ constructed from T subvectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$, we denote by $\boldsymbol{\xi}_t$ the t-th subvector, whereas $\boldsymbol{\xi}^t = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$ stacks all subvectors up to and including $\boldsymbol{\xi}_t$. The vectors $\mathbf{e}$ and $\mathbf{e}_i$ refer to the all-ones and the i-th basis vectors, respectively; their dimension will be clear from the context. Finally, we denote by $\mathbf{1}[\cdot]$ the indicator function that attains the value of one if the logical expression $\cdot$ is satisfied, and zero otherwise.

2. Problem Formulation

We are interested in multistage robust mixed-integer optimization problems of the form

$$
\begin{aligned}
\text{minimize} \quad & \max_{\boldsymbol{\xi}\in\Xi} \sum_{t=1}^{T} \boldsymbol{q}_t(\boldsymbol{\xi}_t)^\top \boldsymbol{x}_t(\boldsymbol{\xi}^t) + \boldsymbol{r}_t{}^\top \boldsymbol{y}_t(\boldsymbol{\xi}^t) \\
\text{subject to} \quad & \boldsymbol{T}_t(\boldsymbol{\xi}_t)\,\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \boldsymbol{W}_t(\boldsymbol{\xi}_t)\,\boldsymbol{x}_t(\boldsymbol{\xi}^t) + \boldsymbol{V}_t\,\boldsymbol{y}_t(\boldsymbol{\xi}^t) \\
& \qquad \geqslant \boldsymbol{h}_t(\boldsymbol{\xi}_t) \qquad \forall \boldsymbol{\xi}\in\Xi,\ \ \forall t = 1,\ldots,T \\
& \boldsymbol{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t \text{ and } \boldsymbol{y}_t(\boldsymbol{\xi}^t) \in \mathbb{R}^{n_t^2} \text{ for all } \boldsymbol{\xi}\in\Xi \\
& \qquad \text{and } t = 1,\ldots,T,
\end{aligned}
\tag{1}
$$

where the uncertain problem parameters $\boldsymbol{\xi}_t$ are revealed at the beginning of each time stage t, $t = 1, \ldots, T$, and the decisions $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ are taken afterward under the full knowledge of $\boldsymbol{\xi}^t = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$. A solution to Problem (1) is immunized against all parameter realizations $\boldsymbol{\xi} = \boldsymbol{\xi}^T$ in the uncertainty set Ξ. The cost vectors $\boldsymbol{q}_t$, the technology matrices $\boldsymbol{T}_t$, the recourse matrices $\boldsymbol{W}_t$, as well as the right-hand-side vectors $\boldsymbol{h}_t$, may all depend affinely on $\boldsymbol{\xi}_t$. We stipulate that $\boldsymbol{T}_1(\boldsymbol{\xi}_0) \equiv \mathbf{0}$ and $\boldsymbol{x}_0(\boldsymbol{\xi}^0) \equiv \mathbf{0}$; that is, the first-stage constraints only involve the first-stage decisions $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$. We assume that $\boldsymbol{\xi}_1$ is deterministic, and hence, $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$ are here-and-now decisions. We can interpret $\boldsymbol{x}_t$ as *state variables* because they couple constraints of consecutive stages, whereas $\boldsymbol{y}_t$ are *control variables* that only affect the costs and feasibility within stage t. Each set $\mathcal{X}_t \subseteq \mathbb{R}^{n_t^1}$ can be nonconvex and comprise both continuous and discrete decision variables. In contrast, the linear constraints of Problem (1) fully characterize the set of admissible $\boldsymbol{y}_t$; in particular, all control decisions are assumed to be continuous. To ease the exposition, we assume that the set of admissible $\boldsymbol{y}_t$ is bounded; this avoids tedious but otherwise straightforward case distinctions where infinite cost savings in one stage need to be weighed against infeasibility (that is, infinite costs) in another stage. We do *not* assume that Problem (1) has a relatively complete recourse; that is, there may be partial solutions $(\boldsymbol{x}_\tau, \boldsymbol{y}_\tau)_{\tau=1}^{t}$ satisfying the constraints up to time stage t that cannot be extended to complete solutions $(\boldsymbol{x}_\tau, \boldsymbol{y}_\tau)_{\tau=1}^{T}$ satisfying all constraints up to the final time stage T. The presence of potentially nonconvex stagewise feasible regions $\mathcal{X}_t$ renders Problem (1) strongly NP-hard, even in the absence of uncertainty. Even if the stagewise feasible regions $\mathcal{X}_t$ are polyhedra, however, Problem (1) remains strongly NP-hard because of the presence of adaptive decisions; see Guslitser (2002, theorem 3.5) and Subramanyam et al. (2020, proposition A.3).

The uncertainty set Ξ comprises discrete *interstage uncertainties* $\boldsymbol{\phi}_t \in \mathbb{R}^{k_t^1}$ that can be coupled over consecutive time stages as well as continuous *intrastage uncertainties* $\boldsymbol{\psi}_t \in \mathbb{R}^{k_t^2}$ that are stagewise rectangular apart from their potential dependence on $\boldsymbol{\phi}_t$. More precisely, Ξ emerges from stagewise uncertainty sets Ξ_t as follows. The parameters of the first stage are deterministic and are thus the only element of the singleton set $\Xi_1 = \{\boldsymbol{\xi}_1 = (\boldsymbol{\phi}_1, \boldsymbol{\psi}_1)\}$. For $t = 2, \ldots, T$, the stagewise uncertainty sets satisfy

$$
\begin{aligned}
\Xi_t(\boldsymbol{\phi}_{t-1}) = \{\boldsymbol{\xi}_t = (\boldsymbol{\phi}_t, \boldsymbol{\psi}_t) : \boldsymbol{\phi}_t \in \Phi_t(\boldsymbol{\phi}_{t-1}),& \\
\boldsymbol{U}_t(\boldsymbol{\phi}_t)\,\boldsymbol{\psi}_t \leqslant \boldsymbol{b}_t(\boldsymbol{\phi}_t)\}&,
\end{aligned}
$$

where $\boldsymbol{U}_t : \mathbb{R}^{k_t^1} \to \mathbb{R}^{l_t \times k_t^2}$ and $\boldsymbol{b}_t : \mathbb{R}^{k_t^1} \to \mathbb{R}^{l_t}$ can be arbitrary functions of $\boldsymbol{\phi}_t$. To keep the notation consistent, we stipulate that $\boldsymbol{\phi}_0 \equiv \mathbf{0}, \Phi_1(\boldsymbol{\phi}_0) \equiv \Phi_1 = \{\boldsymbol{\phi}_1\}$ and $\Xi_1(\boldsymbol{\phi}_0) \equiv \Xi_1$. We assume that the sets $\Phi_t(\boldsymbol{\phi}_{t-1})$ are finite and their unions $\Phi_t = \cup_{\boldsymbol{\phi}_{t-1}\in\Phi_{t-1}} \Phi_t(\boldsymbol{\phi}_{t-1})$, $t = 2, \ldots, T$, are "not too large" in a sense that we will discuss in more detail later on. We can then define the overall uncertainty set Ξ from the stagewise uncertainty sets Ξ_t as

$$
\Xi = \{\boldsymbol{\xi} = (\underbrace{(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)}_{=\boldsymbol{\xi}_t})_{t=1}^{T} : \boldsymbol{\xi}_1 \in \Xi_1,\ \boldsymbol{\xi}_t \in \Xi_t(\boldsymbol{\phi}_{t-1})\ \forall t = 2, \ldots, T\}.
\tag{2}
$$

Note that Ξ may fail to be stagewise rectangular; that is, $\Xi \subsetneq \times_{t=1}^{T} \{\boldsymbol{\xi}_t : \boldsymbol{\xi} \in \Xi\}$. To avoid technical but otherwise straightforward case distinctions, we assume in the following that Ξ is bounded.

The Uncertainty Set (2) is quite versatile. It allows us, for example, to model *lattice regime uncertainty sets*, under which the discrete regime ϕ_t in stage t is any element of $\Phi_t(\phi_{t-1})$, where $\Phi_t(\phi_{t-1})$ describes the children of the previous regime node ϕ_{t-1} in a lattice structure (cf. Figure 1(a)), as well as the *Markovian regime chain uncertainty set* under which the regime ϕ_t is any of the neighbors $\Phi_t(\phi_{t-1})$ of the previous regime node ϕ_{t-1} in a graph (cf. Figure 1(b)). In both cases, the discrete regimes can be complemented by continuous intraregime uncertainties. Online Appendix A reviews other classes of uncertainty sets that can be modeled as special cases of (2).

Instead of solving the Multistage Robust Mixed-Integer Problem (1) directly, we propose to investigate the following *finite adaptability approximation*,

$$\begin{aligned}
&\text{minimize} && \max_{\boldsymbol{\xi}\in\Xi} \sum_{t=1}^{T} \boldsymbol{q}_t(\boldsymbol{\xi}_t)^\top \boldsymbol{x}_t(\boldsymbol{\xi}^t) + \boldsymbol{r}_t{}^\top \boldsymbol{y}_t(\boldsymbol{\xi}^t)\\
&\text{subject to} && \boldsymbol{T}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1}) + \boldsymbol{W}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_t(\boldsymbol{\xi}^t)\\
&&& \quad + \boldsymbol{V}_t\boldsymbol{y}_t(\boldsymbol{\xi}^t) \geqslant \boldsymbol{h}_t(\boldsymbol{\xi}_t) \quad \forall \boldsymbol{\xi}\in\Xi,\ \forall t = 1,\dots,T\\
&&& \boldsymbol{x}_t(\boldsymbol{\xi}^t) \in \mathcal{X}_t^{\mathrm{C}} \text{ and } \boldsymbol{y}_t(\boldsymbol{\xi}^t) \in \mathbb{R}^{n_t^2}\\
&&& \qquad \text{for all } \boldsymbol{\xi}\in\Xi \text{ and } t = 1,\dots,T,
\end{aligned} \tag{3}$$

where $\mathcal{X}_t^{\mathrm{C}} = \{\boldsymbol{x}_t^1,\dots,\boldsymbol{x}_t^{p_t}\} \subseteq \mathcal{X}_t$ is a finite set of preselected candidate decisions for stage t. The Finite Adaptability Approximation (3) differs from the Multistage Robust Mixed-Integer Problem (1) in that it selects optimal state decisions $\boldsymbol{x}_t$ from the restricted finite sets $\mathcal{X}_t^{\mathrm{C}}$ as opposed to the original sets $\mathcal{X}_t$. This conservative approximation turns out to be crucial for the development of our nested problem formulation and its associated solution method below. Note that in contrast to the K-adaptability schemes for two-stage robust problems (Bertsimas and Caramanis 2010, Hanasusanto et al. 2015, Subramanyam et al. 2020), the candidate decision sets $\mathcal{X}_t^{\mathrm{C}}$ in Problem (3) are fixed. Similar to the choice of decision rule architectures discussed in the introduction, the choice of suitable candidate decision sets $\mathcal{X}_t^{\mathrm{C}}$ requires some a priori insight into the structure of well-performing solutions to the Multistage Robust Mixed-Integer Problem (1). We will discuss in Section 4 different heuristic approaches for choosing candidate decision sets $\mathcal{X}_t^{\mathrm{C}}$ if such a priori knowledge is not available. We emphasize that Problem (3) selects the best decisions from each set $\mathcal{X}_t^{\mathrm{C}}$ adaptively based on the realization of $\boldsymbol{\xi}^t$ and that the problem therefore remains challenging, as it generalizes a known NP-hard problem (Subramanyam et al. 2020, proposition B.3).

The Finite Adaptability Approximation (3) gives rise to an equivalent nested problem formulation that can be solved by a backward recursion. Our nested formulation is defined by the *stage-t worst-case cost to-go problem*

$$\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\phi}_{t-1}) = \max_{\boldsymbol{\xi}_t\in\Xi_t(\boldsymbol{\phi}_{t-1})} Q_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\xi}_t), \tag{4a}$$

$t = 1,\dots,T$, where the *stage-t nominal cost to-go problem* satisfies

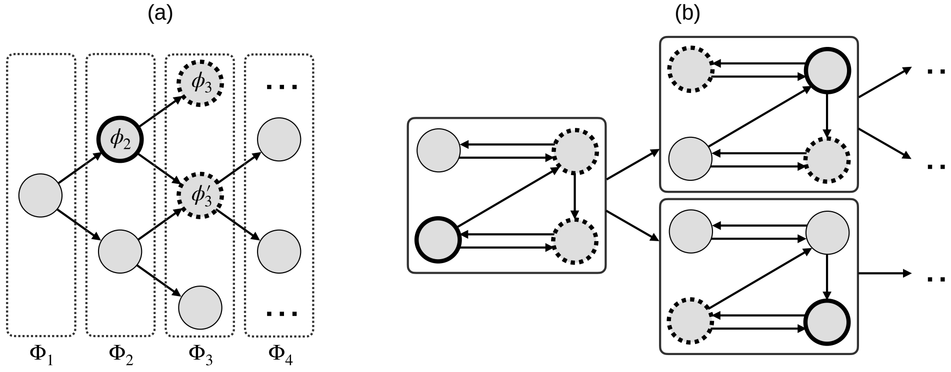
$$Q_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\xi}_t) = \left[\begin{array}{ll}\text{minimize} & \boldsymbol{q}_t(\boldsymbol{\xi}_t)^\top\boldsymbol{x}_t + \boldsymbol{r}_t{}^\top\boldsymbol{y}_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t;\boldsymbol{\phi}_t)\\ \text{subject to} & \boldsymbol{T}_t(\boldsymbol{\xi}_t)\hat{\boldsymbol{x}}_{t-1} + \boldsymbol{W}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_t + \boldsymbol{V}_t\boldsymbol{y}_t \geqslant \boldsymbol{h}_t(\boldsymbol{\xi}_t)\\ & \boldsymbol{x}_t\in\mathcal{X}_t^{\mathrm{C}},\ \boldsymbol{y}_t\in\mathbb{R}^{n_t^2}\end{array}\right]. \tag{4b}$$

In Problem (4b), we stipulate that $Q_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\xi}_t) = +\infty$ if the minimization is infeasible, and we set the boundary condition $\mathcal{Q}_{T+1}(\boldsymbol{x}_T;\boldsymbol{\phi}_T) \equiv 0$ for all $\boldsymbol{x}_T \in \mathcal{X}_T^{\mathrm{C}}$ and $\boldsymbol{\phi}_T \in \Phi_T$. For $t = 1$, we also set $\hat{\boldsymbol{x}}_0 \equiv \mathbf{0}$ and abbreviate $\mathcal{Q}_1(\hat{\boldsymbol{x}}_0,\boldsymbol{\phi}_0)$ by $\mathcal{Q}_1$.

Proposition 1. *The Finite Adaptability Approximation* (3) *and the stage-1 worst-case cost to-go problem* $\mathcal{Q}_1$ *share the same optimal value. In particular,* (3) *is infeasible if and only if* $\mathcal{Q}_1 = \infty$.

The equivalence of multistage problems and their nested formulations is well understood in the stochastic

Figure 1. Lattice Regime and Markovian Regime Chain Uncertainty Sets



Notes. (a) Lattice regime. (b) Markovian regime chain. In both cases, the regimes are represented as nodes. In (a), we have $\Phi_3(\phi_2) = \{\phi_3, \phi_3'\}$. In (b), the sets $\Phi_{t+1}(\phi_t)$ corresponding to the nodes ϕ_t highlighted by bold strokes are the nodes with dotted strokes.

programming community when the worst-case approach is replaced with the expected value (Rockafellar and Wets 1998, theorem 14.60). Shapiro (2017) studies the more general case where the expected value is replaced with a generic risk measure. Whereas this result includes our worst-case approach, the nested formulation of Shapiro (2017) does not account for any structure in the support of the random vectors, and his stagewise cost to-go problems therefore depend on the entire parameter sequence $\boldsymbol{\xi}^T$. In contrast, the stage-t worst-case cost to-go problems in our derivations only depend on the realization $\boldsymbol{\phi}_{t-1}$ of the preceding interstage uncertainty. This parsimonious dependence of the worst-case cost to-go problems on the past uncertainty realizations lies at the heart of our approach; it is facilitated by the design of our generalized rectangularity of Ξ, and it is key for the tractability of our approach.

In the following, we first discuss how the stage-t worst-case cost to-go problems $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\phi}_{t-1})$ can be solved by a backward recursion (Algorithm 1). Here, the key step repeatedly solves two-stage subproblems $\mathcal{Q}_t$, given the (previously computed) values of the worst-case cost to-go problems $\mathcal{Q}_{t+1}$. Section 3 is devoted to the exact and approximate solution of these subproblems.

Algorithm 1 (Evaluation of the Stage-t Worst-Case Cost To-Go Problems $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\phi}_{t-1})$)

1. **Initialization.** Set $t = T$ as well as $\mathcal{Q}_{T+1}(\boldsymbol{x}_T;\boldsymbol{\phi}_T) = 0$ for all $\boldsymbol{x}_T \in \mathcal{X}_T^{\mathrm{C}}$ and $\boldsymbol{\phi}_T \in \Phi_T$.
2. **Iteration.** Compute $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1};\boldsymbol{\phi}_{t-1})$ for all $\hat{\boldsymbol{x}}_{t-1} \in \mathcal{X}_{t-1}^{\mathrm{C}}$ and $\boldsymbol{\phi}_{t-1} \in \Phi_{t-1}$.
3. **Termination.** If $t > 1$, update $t \leftarrow t-1$ and repeat Step 2. Otherwise, terminate.

Note that the subproblems $\mathcal{Q}_t$ occurring in the same time stage t are independent of one another and can thus be solved in parallel, leading to a theoretical maximum speedup of $(1/T)\cdot\sum_{t=1}^{T}|\mathcal{X}_t^{\mathrm{C}}|\cdot|\Phi_{t-1}|$ over the sequential solution of all subproblems. Once Algorithm 1 has been executed, the optimal first-stage decisions $(\boldsymbol{x}_1^\star,\boldsymbol{y}_1^\star)$ can be obtained by solving a single instance of the nominal problem Q_1. In principle, Algorithm 1 is sufficient for solving the Finite Adaptability Approximation (3) in a receding-horizon fashion, whereupon observing $\boldsymbol{\xi}_t$, we construct and solve a new $(T-t+1)$-stage problem. However, because Algorithm 1 solves *all* worst-case cost to-go problems $\mathcal{Q}_t$, they do not need to be recomputed at later time stages. The following algorithm exploits this property.

Algorithm 2 (Optimal Policy for the Finite Adaptability Approximation (3))

1. **Initialization.** Solve all stage-t worst-case cost to-go problems via Algorithm 1.
2. **First stage.** Implement $(\boldsymbol{x}_1^\star,\boldsymbol{y}_1^\star)$ as determined by the problem $\mathcal{Q}_1$. Set $t = 2$.
3. **Iteration.** Upon observation of the uncertain parameters $\boldsymbol{\xi}_t$, solve the stage-t nominal cost to-go problem $Q_t(\boldsymbol{x}_{t-1}^\star;\boldsymbol{\xi}_t)$ and implement the corresponding stage-t decisions $(\boldsymbol{x}_t^\star,\boldsymbol{y}_t^\star)$.
4. **Termination.** If $t < T$, update $t \leftarrow t+1$ and repeat Step 3. Otherwise, terminate.

We next show that Algorithm 2 *implicitly* defines an optimal policy $\{(\boldsymbol{x}_t^\star,\boldsymbol{y}_t^\star)\}_{t=1}^T$ for the Finite Adaptability Approximation (3) in response to a stagewise revealed scenario $\boldsymbol{\xi} \in \Xi$.

Proposition 2. *Assume that Problem (3) is feasible. Then the policy $\{(\boldsymbol{x}_t^\star,\boldsymbol{y}_t^\star)\}_{t=1}^T$ that is implicitly defined by Algorithm 2 in response to the parameter realizations $\boldsymbol{\xi} \in \Xi$ is optimal in (3).*

Remark 1 (Problem Extensions). Problem (1) allows for the possibility that the state decision $\boldsymbol{x}_t$ is taken before $\boldsymbol{\xi}_t$ is observed. In this case, we expand the state decision $\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1})$ to $\boldsymbol{x}'_{t-1}(\boldsymbol{\xi}^{t-1}) = (\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1}), \boldsymbol{z}_{t-1}(\boldsymbol{\xi}^{t-1}))$, and we add the constraint $\boldsymbol{x}_t(\boldsymbol{\xi}^t) = \boldsymbol{z}_{t-1}(\boldsymbol{\xi}^{t-1})$ for all $\boldsymbol{\xi} \in \Xi$ and $t = 1,\ldots,T$, with the understanding that $\boldsymbol{z}_0(\boldsymbol{\xi}^0) \equiv \boldsymbol{0}$. Similarly, the stage-$t$ constraints can depend on the entire history of decisions $\boldsymbol{x}_1(\boldsymbol{\xi}^1),\ldots,\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1})$, as opposed to just the decision $\boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1})$ of the previous time stage, if we expand all $\boldsymbol{x}_t(\boldsymbol{\xi}^t)$ to $\boldsymbol{x}'_t(\boldsymbol{\xi}^t) = (\boldsymbol{x}_t(\boldsymbol{\xi}^t), \boldsymbol{z}_t(\boldsymbol{\xi}^t))$ and add the constraints $\boldsymbol{z}_t(\boldsymbol{\xi}^t) = \boldsymbol{x}'_{t-1}(\boldsymbol{\xi}^{t-1})$ for all $\boldsymbol{\xi} \in \Xi$ and $t = 1,\ldots,T$, again with the understanding that $\boldsymbol{z}_0(\boldsymbol{\xi}^0) \equiv \boldsymbol{0}$.

Remark 2 (Alternative Formulations). Whereas in theory, a nested formulation akin to (4) could be formulated for a finite adaptability approximation that selects the candidate decision sets $\mathcal{X}_t^{\mathrm{C}}$ as part of the optimization, or even more generally for the Multistage Robust Mixed-Integer Problem (1), the corresponding backward recursion would have to compute the worst-case cost to-go for all possible decisions $\hat{\boldsymbol{x}}_{t-1} \in \mathcal{X}_{t-1}$, as opposed to the candidate decisions $\hat{\boldsymbol{x}}_{t-1} \in \mathcal{X}_{t-1}^{\mathrm{C}}$ only in order to quantify the future worst-case cost to-go in each two-stage subproblem. This unfavorable scaling would limit the applicability of the resulting solution scheme to very small problem instances.

3. Solution of the Two-Stage Subproblems $\mathcal{Q}_t$

The key step in Algorithm 1 for our Finite Adaptability Approximation (3) is the solution of the two-stage subproblems $\mathcal{Q}_t$. The presence of discrete wait-and-see decisions $\boldsymbol{x}_t$ in these problems renders most of the existing exact solution approaches for two-stage robust optimization problems (cf. Section 1) inapplicable. Indeed, Fourier-Motzkin elimination and iterative uncertainty set lifting schemes fundamentally require all second-stage decisions to be continuous, whereas Benders' decomposition and column-and-constraint generation depend on strong convex duality of the second stage to identify worst-case uncertainty realizations for fixed first-stage decisions. We are only aware of two exact solution approaches for two-stage robust optimization problems with mixed-integer recourse. The first one is the nested column-and-constraint generation scheme of Zhao and Zeng (2012). Applied to our setting, however, this method would require the two-stage subproblems $\mathcal{Q}_t$ to be solved by extreme point uncertainty realizations $\boldsymbol{\xi}_t \in \text{ext}\,\Xi_t(\boldsymbol{\phi}_{t-1})$, which is not guaranteed to be the case in our context. Indeed, one can readily construct instances of $\mathcal{Q}_t$ without control variables and either objective uncertainty or right-hand side uncertainty whose worst-case uncertainty realizations are not extreme points, and the two-stage subproblems in our two applications of Section 5 are not optimized by extreme point worst-case parameter realizations in general. The second exact solution approach for two-stage robust optimization problems with mixed-integer recourse is a hybrid Benders' decomposition because of Hashemi Doulabi et al. (2021). This algorithm, however, requires the uncertainty to only affect the constraint right-hand sides, the uncertainty set to be discrete and finitely enumerable, and the recourse matrix to be block diagonal. Our problem $\mathcal{Q}_t$ is also reminiscent of the K-adaptability problems studied by Bertsimas and Caramanis (2010), Hanasusanto et al. (2015), and Subramanyam et al. (2020). In contrast to those problems, however, the candidate decision sets $\mathcal{X}_t^{\mathrm{C}}$ in $\mathcal{Q}_t$ have finite cardinality. This allows us to design exact solution schemes despite the presence of continuous second-stage decisions $\boldsymbol{y}_t$.

In the following, we derive exact reformulations (Section 3.1) as well as conservative approximations (Section 3.2) of the two-stage subproblems $\mathcal{Q}_t$.

3.1. Exact Solution

To solve the two-stage subproblems $\mathcal{Q}_t$ exactly, we derive an equivalent mixed-integer bilinear programming formulation of $\mathcal{Q}_t$. To this end, we from now on denote by m_t the number of second-stage constraints $\boldsymbol{T}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_{t-1} + \boldsymbol{W}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_t + \boldsymbol{V}_t\boldsymbol{y}_t \geqslant \boldsymbol{h}_t(\boldsymbol{\xi}_t)$ in stage t. Online Appendix B then identifies and discusses four special cases of this formulation that lead to LP or MILP reformulations of polynomial size: (i) the constraints of Q_t do not depend on the uncertainty realization $\boldsymbol{\xi}_t$, (ii) the control variables $\boldsymbol{y}_t$ are absent in Q_t, (iii) the recourse matrix $\boldsymbol{V}_t$ in Q_t is invertible, and (iv) the recourse matrix $\boldsymbol{V}_t$ in Q_t has a block-diagonal structure.

Proposition 3. *Fix $\hat{\boldsymbol{x}}_{t-1} \in \mathcal{X}_{t-1}^{\mathrm{C}}$ and $\boldsymbol{\phi}_{t-1} \in \Phi_{t-1}$. If the stage-t nominal cost to-go problem $Q_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\xi}_t)$ is feasible for some $\boldsymbol{\xi}_t \in \Xi_t(\boldsymbol{\phi}_{t-1})$, then the two-stage subproblem $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\phi}_{t-1})$ has the equivalent reformulation*

$$\begin{aligned}
&\text{maximize} && \theta \\
&\text{subject to} && \theta \leqslant \boldsymbol{q}_t(\boldsymbol{\xi}_t)^\top \boldsymbol{x}_t^i + [\boldsymbol{h}_t(\boldsymbol{\xi}_t) - \boldsymbol{T}_t(\boldsymbol{\xi}_t)\hat{\boldsymbol{x}}_{t-1} - \boldsymbol{W}_t(\boldsymbol{\xi}_t)\boldsymbol{x}_t^i]^\top \\
&&& \quad \boldsymbol{\lambda}_t^i + \mathcal{Q}_{t+1}(\boldsymbol{x}_t^i; \boldsymbol{\phi}_t) && \forall i = 1, \ldots, p_t \\
&&& \boldsymbol{V}_t^\top \boldsymbol{\lambda}_t^i = \boldsymbol{r}_t && \forall i = 1, \ldots, p_t \\
&&& \theta \in \mathbb{R},\ \boldsymbol{\xi}_t \in \Xi_t(\boldsymbol{\phi}_{t-1}),\ \boldsymbol{\lambda}_t^i \in \mathbb{R}_+^{m_t},\ i = 1, \ldots, p_t.
\end{aligned} \tag{5}$$

3.2. Conservative Approximation

In cases where the two-stage subproblems $\mathcal{Q}_t$ cannot be solved exactly, we propose to use conservative approximations $\overline{\mathcal{Q}}_t$ that restrict the adaptivity of the state or control variables. The following observation justifies the use of stagewise conservative approximations $\overline{\mathcal{Q}}_t$ in a multistage setting.

Observation 1 (Conservative Approximations: Propagation of Upper Bounds).

(i) If $\mathcal{Q}_{t+1}$ in each nominal stage-t cost to-go problem Q_t is replaced with an upper bound $\overline{\mathcal{Q}}_{t+1}(\boldsymbol{x}_t; \boldsymbol{\phi}_t) \geqslant \mathcal{Q}_{t+1}(\boldsymbol{x}_t; \boldsymbol{\phi}_t)$, $t = 1, \ldots, T$, $\boldsymbol{x}_t \in \mathcal{X}_t^{\mathrm{C}}$ and $\boldsymbol{\phi}_t \in \Phi_t$, then the resulting approximations $\overline{\mathcal{Q}}_t$ of $\mathcal{Q}_t$ are conservative: $\overline{\mathcal{Q}}_t(\boldsymbol{x}_{t-1}; \boldsymbol{\phi}_{t-1}) \geqslant \mathcal{Q}_t(\boldsymbol{x}_{t-1}; \boldsymbol{\phi}_{t-1})$ for all t, $\boldsymbol{x}_{t-1}$ and $\boldsymbol{\phi}_{t-1}$.

(ii) If $\overline{\mathcal{Q}}_1 < \infty$, then the policy $\{(\boldsymbol{x}_t^\star, \boldsymbol{y}_t^\star)\}_{t=1}^T$ that is implicitly defined by Algorithm 2 using the upper bounds $\overline{\mathcal{Q}}_t$, $t = 1, \ldots, T$, is feasible in (3) and attains worst-case costs of at most $\overline{\mathcal{Q}}_1$.

Observation 1 shows that $\overline{\mathcal{Q}}_{t+1}(\boldsymbol{x}_t^\star; \boldsymbol{\phi}_t)$ remains an upper bound on the worst-case cost to-go when decision $\boldsymbol{x}_t^\star \in \mathcal{X}_t^{\mathrm{C}}$ is taken by Algorithm 2, even when the worst-case cost to-go functions in later time stages are also replaced by conservative approximations.

In the remainder of this section, we discuss two different conservative bounds based on approximations of the state and control variables. Because both approximations distinguish between the interstage uncertainties $\boldsymbol{\phi}_t$ and the intrastage uncertainties $\boldsymbol{\psi}_t$ in time stage t, we define $\Psi_t(\boldsymbol{\phi}_t) = \{\boldsymbol{\psi}_t : (\boldsymbol{\phi}_t, \boldsymbol{\psi}_t) \in \Xi_t(\boldsymbol{\phi}_{t-1})\}$ as the set of conditional intrastage uncertainty realizations.

Approximation 1 (State Approximation). *The state variables $\boldsymbol{x}_t$ in each two-stage subproblem $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\phi}_t)$ no longer adapt to the exact values of the intrastage uncertainties $\boldsymbol{\psi}_t$, but only to their set memberships $\{\mathbf{1}[\boldsymbol{\psi}_t \in \Psi_{tj}(\boldsymbol{\phi}_t)]\}_{j=1}^l$ across a preselected polyhedral partition $\{\Psi_{tj}(\boldsymbol{\phi}_t)\}_{j=1}^l$ of the conditional intrastage uncertainty realizations $\Psi_t(\boldsymbol{\phi}_t)$.*

Note that under Approximation 1, the state variables $\boldsymbol{x}_t$ remain fully adaptive in the interstage uncertainties $\boldsymbol{\phi}_t$. Approximation 1 then adopts a piecewise constant decision rule approximation in $\boldsymbol{\psi}_t$ for the state variables $\boldsymbol{x}_t$. The number l of subsets in the partition $\{\Psi_{tj}(\boldsymbol{\phi}_t)\}_{j=1}^l$ of $\Psi_t(\boldsymbol{\phi}_t)$ can vary with the time stage t as well as the values of $\boldsymbol{\phi}_{t-1}$, $\boldsymbol{\phi}_t$ and $\hat{\boldsymbol{x}}_{t-1}$; for ease of exposition, however, we notationally suppress this potential dependence.

Observation 2. Denote by $\eta^\star(\boldsymbol{\phi}_t, j, \boldsymbol{x}_t)$, $\boldsymbol{\phi}_t \in \Phi_t(\boldsymbol{\phi}_{t-1})$, $j = 1, \ldots, l$ and $\boldsymbol{x}_t \in \mathcal{X}_t^{\mathrm{C}}$, the optimal value of the following worst-case optimization problem with continuous recourse.

$$\begin{array}{ll} \text{maximize} & \left[\begin{array}{ll} \text{minimize} & \boldsymbol{q}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)^\top \boldsymbol{x}_t + \boldsymbol{r}_t{}^\top \boldsymbol{y}_t + \overline{\mathcal{Q}}_{t+1}(\boldsymbol{x}_t; \boldsymbol{\phi}_t) \\ \text{subject to} & \boldsymbol{T}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)\hat{\boldsymbol{x}}_{t-1} + \boldsymbol{W}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)\boldsymbol{x}_t \\ & \quad + \boldsymbol{V}_t \boldsymbol{y}_t \geqslant \boldsymbol{h}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t) \\ & \boldsymbol{y}_t \in \mathbb{R}^{n_t^2} \end{array}\right] \\ \text{subject to} & \boldsymbol{\psi}_t \in \Psi_{tj}(\boldsymbol{\phi}_t). \end{array}$$

Under Approximation 1, the optimal value of the two-stage subproblem $\overline{\mathcal{Q}}_t$ coincides with

$$\max\{\min\{\eta^\star(\boldsymbol{\phi}_t, j, \boldsymbol{x}_t) : \boldsymbol{x}_t \in \mathcal{X}_t^{\mathrm{C}}\} : \boldsymbol{\phi}_t \in \Phi_t(\boldsymbol{\phi}_{t-1}) \text{ and } j = 1, \ldots, l\}.$$

The $|\Phi_t(\boldsymbol{\phi}_{t-1})| \cdot l \cdot p_t$ worst-case optimization problems of Observation 2 can be solved in parallel. Together with our earlier observation of the parallel solution of the subproblems in each time stage, there are thus two levels of parallelization. We can solve $|\Phi_{t-1}| \cdot |\Phi_t| \cdot l \cdot p_{t-1} \cdot p_t$ worst-case optimization problems in parallel in each time stage t. In contrast to the two-stage subproblems $\mathcal{Q}_t$, the worst-case optimization problems in Observation 2 contain no first-stage decisions, discrete uncertainties, or discrete recourse decisions, and they can be solved using column-and-constraint generation (Zeng and Zhao 2013) or iterative liftings of the uncertainty set (Georghiou et al. 2020).

Instead of approximating the state variables, we can also approximate the control variables.

Approximation 2 (Control Approximation). *The control variables $\boldsymbol{y}_t$ in each two-stage subproblem $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\phi}_t)$ no longer adapt to the exact values of the intrastage uncertainties $\boldsymbol{\psi}_t$; instead, they are optimally chosen from an optimally selected set of affine decision rules $\{\boldsymbol{y}_{tj} : \Psi_t(\boldsymbol{\phi}_t) \to \mathbb{R}^{n_t^2}\}_{j=1}^l$.*

Under Approximation 2, the control variables $\boldsymbol{y}_t$ remain fully adaptive in the interstage uncertainties $\boldsymbol{\phi}_t$. Instead of selecting $\boldsymbol{y}_t$ optimally in response to the observation of $\boldsymbol{\psi}_t$, Approximation 2 selects the l best affine decision rules $\{\boldsymbol{y}_{tj} : \Psi_t(\boldsymbol{\phi}_t) \to \mathbb{R}^{n_t^2}\}_{j=1}^l$ here-and-now (that is, upon observation of $\boldsymbol{\phi}_t$ but before observing $\boldsymbol{\psi}_t$) and subsequently computes $\boldsymbol{y}_t$ from the best of these l decision rules wait-and-see (that is, upon observation of $\boldsymbol{\psi}_t$). Similar to the number l in Approximation 1, the number l of decision rules in Approximation 2 can vary with the values of t, $\boldsymbol{\phi}_{t-1}$, $\boldsymbol{\phi}_t$, and $\hat{\boldsymbol{x}}_{t-1}$.

Observation 3. Denote by $\eta^\star(\boldsymbol{\phi}_t)$, $\boldsymbol{\phi}_t \in \Phi_t(\boldsymbol{\phi}_{t-1})$, the optimal value of the following (p_t, l)-adaptable two-stage robust MILP that optimizes over affine decision rules $\{\boldsymbol{y}_{tj}\}_{j=1}^l$ here-and-now and over combinations $(k, j) \in \{1, \ldots, p_t\} \times \{1, \ldots, l\}$ wait-and-see.

$$\begin{array}{ll} \text{minimize} & \left(\max_{\boldsymbol{\psi}_t \in \Psi_t(\boldsymbol{\phi}_t)} \left[\begin{array}{ll} \text{minimize} & \boldsymbol{q}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)^\top \boldsymbol{x}_t^k + \boldsymbol{r}_t{}^\top \boldsymbol{y}_{tj}(\boldsymbol{\psi}_t) \\ & \quad + \overline{\mathcal{Q}}_{t+1}(\boldsymbol{x}_t^k; \boldsymbol{\phi}_t) \\ \text{subject to} & \boldsymbol{T}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)\hat{\boldsymbol{x}}_{t-1} + \boldsymbol{W}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)\boldsymbol{x}_t^k \\ & \quad + \boldsymbol{V}_t \boldsymbol{y}_{tj}(\boldsymbol{\psi}_t) \geqslant \boldsymbol{h}_t(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t) \\ & (k, j) \in \{1, \ldots, p_t\} \times \{1, \ldots, l\} \end{array}\right]\right) \\ \text{subject to} & \boldsymbol{y}_{tj} : \Xi_t(\boldsymbol{\phi}_{t-1}) \to \mathbb{R}^{n_t^2}, j = 1, \ldots, l. \end{array}$$

Under Approximation 2, the optimal value of the two-stage subproblem $\overline{\mathcal{Q}}_t$ coincides with

$$\max\{\eta^\star(\boldsymbol{\phi}_t) : \boldsymbol{\phi}_t \in \Phi_t(\boldsymbol{\phi}_{t-1})\}.$$

The $|\Phi_t(\boldsymbol{\phi}_{t-1})|$ (p_t, l)-adaptable two-stage robust MILPs of Observation 3 can be solved in parallel. Together with our earlier observation of the parallel solution of the subproblems in each time stage, we can thus solve $|\Phi_{t-1}| \cdot |\Phi_t| \cdot p_{t-1}$ (p_t, l)-adaptable two-stage robust optimization problems in parallel in each time stage t. In contrast to the two-stage subproblems $\mathcal{Q}_t$, the (p_t, l)-adaptable two-stage robust MILPs in Observation 3 contain neither discrete uncertainties nor continuous recourse decisions, and they are thus amenable to standard solution schemes for K-adaptable two-stage robust optimization problems (Hanasusanto et al. 2015, Subramanyam et al. 2020).

4. Lower Bounds and Selection of Candidate State Decision Sets

So far, we assumed that the sets $\mathcal{X}_t^{\mathrm{C}}$ of candidate state decisions in the Finite Adaptability Approximation (3) are prespecified and fixed. This is reasonable if domain expertise can be leveraged to design promising candidate state decisions, but it may place an undue burden on the decision maker when such information is not available. To address such situations, we develop, in this section, a greedy heuristic that iteratively expands the sets $\mathcal{X}_t^{\mathrm{C}}$ based on optimal solutions to lower bounds of the Multistage Robust Mixed-Integer Optimization Problem (1). The lower bounds are useful in their own right to estimate the suboptimality of a solution to (3), regardless of whether the sets $\mathcal{X}_t^{\mathrm{C}}$ are being expanded or kept fixed.

Our expansion heuristic is motivated by a necessary and a sufficient condition for the improvement of the optimal value of the Finite Adaptability Approximation (3). To ease the exposition, we denote by $\mathcal{X}^{\mathrm{C}} = \{\mathcal{X}_1^{\mathrm{C}}, \ldots, \mathcal{X}_T^{\mathrm{C}}\}$ the collection of all candidate state decision sets, and for a fixed choice of $\mathcal{X}^{\mathrm{C}}$ and Ξ, we let $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ denote both the Formulation (3) and its optimal value.

Proposition 4 (Necessary and Sufficient Conditions for Improvement). *Consider two collections of candidate state decision sets $\mathcal{X}^{\mathrm{C}} = \{\mathcal{X}_t^{\mathrm{C}}\}_{t=1}^T$ and $\hat{\mathcal{X}}^{\mathrm{C}} = \{\hat{\mathcal{X}}_t^{\mathrm{C}}\}_{t=1}^T$ with $\hat{\mathcal{X}}_\tau^{\mathrm{C}} = \mathcal{X}_\tau^{\mathrm{C}} \cup \{\boldsymbol{x}'_\tau\}$ for some $\tau = 1, \ldots, T$ and $\boldsymbol{x}'_\tau \in \mathcal{X}_t$, and $\hat{\mathcal{X}}_t^{\mathrm{C}} =$*

$\mathcal{X}_t^{\mathrm{C}}$ for all $t \neq \tau$. A necessary condition for $P(\hat{\mathcal{X}}^{\mathrm{C}}, \Xi) < P(\mathcal{X}^{\mathrm{C}}, \Xi)$ is the existence of $\boldsymbol{\phi}_{\tau-1} \in \Phi_{\tau-1}$, $\hat{\boldsymbol{x}}_{\tau-1} \in \mathcal{X}_{\tau-1}^{\mathrm{C}}$, $\boldsymbol{\xi}_\tau \in \Xi_\tau(\boldsymbol{\phi}_{\tau-1})$ and $\boldsymbol{y}'_\tau \in \mathbb{R}^{n_\tau^2}$ such that

(1) $\boldsymbol{T}_\tau(\boldsymbol{\xi}_\tau)\hat{\boldsymbol{x}}_{\tau-1} + \boldsymbol{W}_\tau(\boldsymbol{\xi}_\tau)\boldsymbol{x}'_\tau + \boldsymbol{V}_\tau \boldsymbol{y}'_\tau \geqslant \boldsymbol{h}_\tau(\boldsymbol{\xi}_\tau)$, *and*

(2) $\boldsymbol{q}_\tau(\boldsymbol{\xi}_\tau)^\top \boldsymbol{x}'_\tau + \boldsymbol{r}_\tau{}^\top \boldsymbol{y}'_\tau + \mathcal{Q}_{\tau+1}(\boldsymbol{x}'_\tau; \boldsymbol{\phi}_\tau) < Q_\tau(\hat{\boldsymbol{x}}_{\tau-1}; \boldsymbol{\xi}_\tau)$.

A sufficient condition for $P(\hat{\mathcal{X}}^{\mathrm{C}}, \Xi) < P(\mathcal{X}^{\mathrm{C}}, \Xi)$ is the existence of an optimal solution $(\boldsymbol{x}^\star, \boldsymbol{y}^\star)(\boldsymbol{\xi})$ to $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ such that for all corresponding worst-case uncertainties $\boldsymbol{\xi}^\star \in \Xi$, conditions (1) and (2) hold for $\hat{\boldsymbol{x}}_{\tau-1} = \boldsymbol{x}^\star_{\tau-1}([\boldsymbol{\xi}^\star]^{\tau-1})$ and $\boldsymbol{\xi}_\tau = \boldsymbol{\xi}^\star_\tau$.

In Proposition 4, $\boldsymbol{\xi}^\star \in \Xi$ is a worst-case uncertainty if $\sum_{t=1}^T \boldsymbol{q}_t(\boldsymbol{\xi}_t^\star)^\top \boldsymbol{x}_t^\star([\boldsymbol{\xi}^\star]^t) + \boldsymbol{r}_t{}^\top \boldsymbol{y}_t^\star([\boldsymbol{\xi}^\star]^t) = P(\mathcal{X}^{\mathrm{C}}, \Xi)$. The conditions (1) and (2) correspond to the feasibility and objective improvement of $(\hat{\boldsymbol{x}}_t, \hat{\boldsymbol{y}}_t)$ in the stage-t nominal cost to-go problem Q_t, respectively. The necessary condition in Proposition 4 is not sufficient because the candidate state solution $\boldsymbol{x}'_\tau$ may only improve nominal stage-t problems $Q_t(\boldsymbol{x}_{\tau-1}; \boldsymbol{\xi}_\tau)$ that do not involve worst-stage uncertainties $\boldsymbol{\xi}_\tau$ or optimal decisions $\boldsymbol{x}_{\tau-1}$ in $P(\mathcal{X}^{\mathrm{C}}, \Xi)$. Likewise, the sufficient condition in Proposition 4 is not necessary because it is possible that by adding $\boldsymbol{x}'_\tau$ to $\mathcal{X}_\tau^{\mathrm{C}}$, the worst-case cost to-go $\mathcal{Q}_t$ and nominal cost to-go Q_t decrease in early stages $t = 1, \ldots, \tau - 1$ for previously suboptimal decisions, thus leading to an improvement without the sufficient condition being satisfied.

Motivated by Proposition 4, we limit our attention to worst-case uncertainties $\boldsymbol{\xi}^\star$ when selecting new candidate state decisions to add to $\mathcal{X}^{\mathrm{C}}$. This idea is formalized in the following algorithm.

Algorithm 3 (Iterative Expansion of the Candidate State Decision Sets)

1. **Initialization.** Initialize the collection of candidate state decision sets $\mathcal{X}^{\mathrm{C}}$.
2. **Conservative bound.** Solve the current finite adaptability problem $P(\mathcal{X}^{\mathrm{C}}, \Xi)$.
3. **Progressive bound.** Identify a set $\hat{\Xi} \subseteq \Xi$ of worst-case uncertainties and solve $P(\mathcal{X}, \hat{\Xi})$, where $\mathcal{X} = \{\mathcal{X}_t\}_{t=1}^T$. Identify the candidate state decision(s) to add to $\mathcal{X}^{\mathrm{C}}$.
4. **Termination.** Stop if $\mathcal{X}^{\mathrm{C}}$ has not been updated. Otherwise, go back to Step 2.

Algorithm 3 produces a sequence of upper and lower bounds to the Multistage Robust Mixed-Integer Optimization Problem (1) by solving finite adaptability problems $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ with updated candidate state decision sets $\mathcal{X}^{\mathrm{C}}$ and relaxations of (1) that only involve subsets $\hat{\Xi} \subseteq \Xi$ of the uncertain parameter realizations, respectively. Algorithm 3 constitutes a template that allows for several ways of (i) initializing and updating the collection $\mathcal{X}^{\mathrm{C}}$, (ii) constructing reduced uncertainty sets $\hat{\Xi}$, and (iii) solving the upper and lower bounds $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ and $P(\mathcal{X}, \hat{\Xi})$.

We initialize the collection $\mathcal{X}^{\mathrm{C}}$ of candidate state decision sets by solving a static version of the Multistage Robust Mixed-Integer Optimization Problem (1) and identifying $\mathcal{X}_t^{\mathrm{C}}$ with the decision $\boldsymbol{x}_t^\star \in \mathcal{X}_t$ taken by the optimal solution to the static problem. Likewise, we update $\mathcal{X}^{\mathrm{C}}$ by adding to each set $\mathcal{X}_t^{\mathrm{C}}$ the decision $\boldsymbol{x}_t^\star \in \mathcal{X}_t$ taken by an optimal solution to the lower bound $P(\mathcal{X}, \hat{\Xi})$. This approach is heuristic as it uses the sufficient (but not necessary) condition of Proposition 4.

We consider two alternative constructions of the reduced uncertainty set $\hat{\Xi}$. In the first, we identify $\hat{\Xi}$ with all worst-case uncertainties $\boldsymbol{\xi}^\star$ determined during the solution of the upper bound problem $P(\mathcal{X}^{\mathrm{C}}, \Xi)$, which results in a set $\hat{\Xi}$ of finite cardinality. In the second approach, we set $\hat{\Xi} = \{\boldsymbol{\xi} = (\boldsymbol{\phi}^\star, \boldsymbol{\psi}) \in \Xi : \boldsymbol{\phi}^\star \in \Phi^\star\}$, where $\Phi^\star$ denotes the set of worst-case interstage uncertainties in problem $P(\mathcal{X}^{\mathrm{C}}, \Xi)$. In other words, the second approach considers all uncertainty realizations that emerge from combinations of the worst-case interstage uncertainties with any admissible intrastage uncertainties. If necessary, the reduced uncertainty set $\hat{\Xi}$ resulting from either approach can be restricted further by selecting subsets of the worst-case parameter realizations.

We solve the upper bound problems $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ using Algorithm 1. As part of this algorithm, we solve the emerging two-stage subproblems $\mathcal{Q}_t$ either exactly (using the techniques of Section 3.1 and Online Appendix B) or approximately (using the methods of Section 3.2). The lower bound problems $P(\mathcal{X}, \hat{\Xi})$, on the other hand, can be solved as large-scale MILPs as long as the reduced uncertainty set $\hat{\Xi}$ is finite (i.e., it takes the form of a scenario tree or a scenario fan). If $\hat{\Xi}$ has infinite cardinality, on the other hand, then we solve a single-stage robust optimization problem in which the decisions $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ only adapt to the interstage uncertainties $\boldsymbol{\phi}_t$, but not to the intrastage uncertainties $\boldsymbol{\psi}_t$. Note that this approach no longer provides a lower bound on the Multistage Robust Mixed-Integer Optimization Problem (1), but it remains a valid heuristic to choose candidate state decisions to add to $\mathcal{X}^{\mathrm{C}}$.

5. Numerical Experiments

We compare our iterative solution scheme for the Finite Adaptability Approximation (3) with different benchmark approaches for solving the Multistage Robust Mixed-Integer Problem (1) in two case studies. The first case study concerns a route planning problem in which the stagewise feasible regions $\mathcal{X}_t$ are sufficiently benign for our approach to solve the problem exactly within seconds. We compare the worst-case routes of our method with (i) the static robust route, a fixed here-and-now route that minimizes the worst-case route length before any arc lengths are observed; (ii) the worst-case crystal ball route over all possible uncertainty realizations, which benefits from observing all arc lengths ahead of time; and (iii) a variant of our adaptive robust route where the decisions $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ need to be taken *before* the uncertainties $\boldsymbol{\xi}_t$ are observed. The second case study concerns a transportation-location planning problem where the stagewise feasible regions $\mathcal{X}_t$ grow exponentially in the problem description; here, we use our iterative state variable selection procedure from Section 4 in combination with the piecewise constant state decision approximation from Section 3.2 to approximately solve the problem to a high accuracy. We show that our method outperforms two benchmark models. The first model restricts the state decisions $\boldsymbol{x}_t$ to adapt only to the interstage uncertainties ϕ_t, but not to the intrastage uncertainties $\boldsymbol{\psi}_t$, whereas the second model additionally replaces the uncertain intrastage demand realizations $\boldsymbol{d}_t(\boldsymbol{\xi}_t)$ with their nominal values. All experiments are run on Intel Xeon 2.66-GHz cluster processors with 16 GB memory and four cores.

5.1. Route Planning

We consider a dynamic robust route planning problem on a directed graph $G = (N, A)$ with nodes $N = \{1, \ldots, n\}$ and arcs $A \subseteq V \times V$. The goal is to determine an adaptive shortest path from the start node 1 to the terminal node n that minimizes the worst-case route length when the arc lengths are uncertain and nonstationary over time. The problem can be formulated as the following instance of our Multistage Robust Mixed-Integer Optimization Problem (1):

$$
\begin{aligned}
\text{minimize} \quad & \max_{\boldsymbol{\xi} \in \Xi} \sum_{t=1}^{T} \sum_{(i,j) \in A} y_{tij}(\boldsymbol{\xi}^t) \\
\text{subject to} \quad & y_{tij}(\boldsymbol{\xi}^t) \geqslant r_{tij}(\boldsymbol{\xi}_t)(x_{tj}(\boldsymbol{\xi}^t) + x_{t-1,i}(\boldsymbol{\xi}^{t-1}) - 1) \\
& \qquad\qquad\qquad \forall \boldsymbol{\xi} \in \Xi,\ \forall (i,j) \in A, \\
& \qquad\qquad\qquad \forall t = 1, \ldots, T \\
& x_{tj}(\boldsymbol{\xi}^t) + x_{t-1,i}(\boldsymbol{\xi}^{t-1}) \leqslant 1 \quad \forall \boldsymbol{\xi} \in \Xi,\ \forall (i,j) \notin A, \\
& \qquad\qquad\qquad \forall t = 1, \ldots, T \\
& \boldsymbol{x}_T(\boldsymbol{\xi}^T) = \mathbf{e}_n \qquad\qquad \forall \boldsymbol{\xi} \in \Xi \\
& \boldsymbol{x}_t(\boldsymbol{\xi}^t) \in \{\mathbf{e}_i : i = 1, \ldots, n\},\ \boldsymbol{y}_t(\boldsymbol{\xi}^t) \geqslant \mathbf{0} \\
& \qquad\qquad\qquad \forall \boldsymbol{\xi} \in \Xi,\ \forall t = 1, \ldots, T.
\end{aligned}
$$

Here, $r_{tij}(\boldsymbol{\xi}_t)$ represents the uncertain length of arc $(i,j) \in A$ in time stage $t = 1, \ldots, T$. The state variable $\boldsymbol{x}_t$ records the node entered at time stage t, with $\boldsymbol{x}_0(\boldsymbol{\xi}^0) \equiv \mathbf{e}_1$, and the components of the auxiliary control variable $\boldsymbol{y}_t$ evaluate (at optimality) to $y_{tij}(\boldsymbol{\xi}^t) = r_{tij}(\boldsymbol{\xi}_t)$ if arc (i,j) is traversed in time stage t and $y_{tij}(\boldsymbol{\xi}^t) = 0$ otherwise. The objective function evaluates the worst-case cumulative length of the arcs traversed over time. The first constraint ensures that $y_{tij}(\boldsymbol{\xi}^t) \geqslant r_{tij}(\boldsymbol{\xi}_t)$ precisely when we traverse from node i to node j in time stage t, the second constraint ensures that this is possible only when $(i,j) \in A$, and the third constraint ensures that we reach node n by time T. Our formulation assumes that the arc lengths are observed prior to selecting which arc to traverse next. This is commonly the case in practice, where modern navigation systems utilize real-time traffic information that is very accurate at any given moment, whereas traffic conditions in the near future can be unpredictable because of rapid changes in traffic flow.

For our numerical experiments, we generate random graphs with nodes $N = \{1, \ldots, 100\}$ that are located uniformly at random on the "N"-shaped subset $[0, 10]^2 \setminus ([2, 4] \times [0, 8] \cup [6, 8] \times [2, 10])$ of the two-dimensional square $[0, 10]^2$; the exceptions are the start node 1 and the terminal node $n = 100$, which are located at the bottom left corner $(0, 0)$ and top right corner $(10, 10)$, respectively. To construct the arc set, we start with $A = N \times N$ (including all self-loops) and remove all arcs that cross the boundaries of our N-shaped set. We subsequently remove arcs in order of decreasing arc lengths until either $|A| = 6n$, or the removal of any further arcs would make the graph disconnected. This eliminates trivial problem instances in which the shortest path contains few arcs. Note that the inclusion of self-loops allows the decision maker to reside at the current node in any time stage; in particular, the decision maker can arrive and reside at the terminal node n prior to time stage T. To model the uncertain arc lengths $r_{tij}(\boldsymbol{\xi}_t)$, we construct a budget uncertainty set of the form

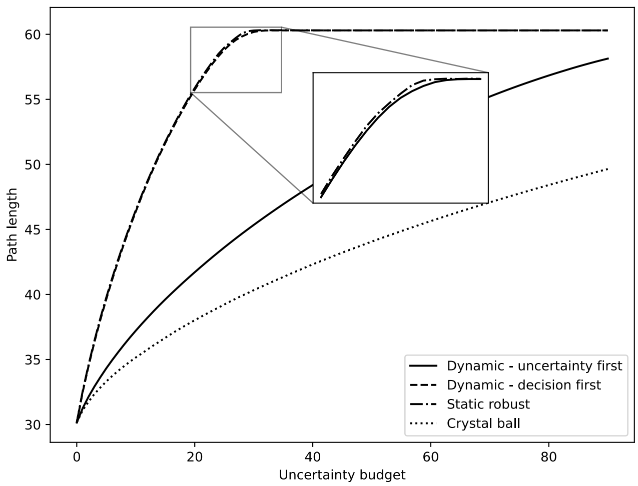
$$
\begin{aligned}
\Xi_t(\phi_{t-1}) := \{\boldsymbol{\xi}_t = (\phi_t, \boldsymbol{\psi}_t) : \phi_t \in \Phi_t(\phi_{t-1}),\ \boldsymbol{\psi}_t \in [0, 1]^{|A|}, \\
\mathbf{e}^\top \boldsymbol{\psi}_t \leqslant \phi_{t-1} - \phi_t,\ \psi_{tii} = 0 \quad \forall i \in N\}.
\end{aligned}
$$

Here, the intrastage uncertainties $\boldsymbol{\psi}_t$ determine the arc length excesses via $r_{tij}(\boldsymbol{\xi}_t) := (1 + \psi_{tij}) r^0_{ij}$, where the nominal length r^0_{ij} of arc $(i,j) \in A$ is set to the Euclidean distance between nodes i and j, and the interstage uncertainty budget ϕ_t evolves according to the set $\Phi_t(\phi_{t-1}) := \{\phi_t \in \mathbb{N}_0 : \phi_t \leqslant \phi_{t-1}\}$ for some initial budget ϕ_0 (selected below). We set the time horizon to $T = 40$.

The static problem, where the uncertain arc lengths are stationary and observed after choosing the shortest path, can be reduced to the solution of multiple deterministic shortest path problems (Bertsimas and Sim 2003). In our finite adaptability approximation of the dynamic problem, we employ full adaptivity; that is, we set $\mathcal{X}^{\mathrm{C}}_t = \mathcal{X}_t$ for all t; this is possible because $|\mathcal{X}_t| = n$ in this case study. We solve the resulting problem with Algorithm 1 (cf. Section 2), where the subproblems $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\phi}_{t-1})$ are solved exactly using Corollary 3 and Remarks 3 and 4 from Online Appendix B. In fact, each subproblem $\mathcal{Q}_t(\hat{\boldsymbol{x}}_{t-1}; \boldsymbol{\phi}_{t-1})$ of our finite adaptability approximation can be solved to any fixed precision ϵ in time $\mathcal{O}(\log \epsilon^{-1})$ via a binary search. All instances of the static, as well as the adaptive route planning formulation, were solved within 10 seconds each, which is why we omit detailed runtime comparisons in this section.

Figure 2 compares the average worst-case route lengths over 50 randomly generated instances and initial uncertainty budgets $\phi_0 \in \{0, \ldots, T\}$ of four alternative methods: (i) the static robust route ("static robust")

Figure 2. Worst-Case Route Lengths of the Static, the Crystal Ball, and Our Two Dynamic Solutions Across 50 Randomly Generated Instances and Varying Uncertainty Budgets ϕ_0



that solves a single-stage robust optimization problem to determine a fixed here-and-now route to minimize the worst-case route length before any arc lengths are observed; (ii) the worst-case crystal ball route ("crystal ball") over all possible uncertainty realizations, which benefits from observing all arc lengths ahead of time; (iii) our adaptive robust route ("dynamic—uncertainty first"); and (iv) a variant of our adaptive robust route where the decisions (x_t, y_t) need to be taken *before* the uncertainties $\boldsymbol{\xi}_t$ are observed ("dynamic—decision first"; cf. Remark 1). As expected, the figure confirms that the worst-case route length of our adaptive solution is bounded from above and below by the static and the crystal ball solution, respectively. Moreover, we can see that the static solution "saturates" quickly even for relatively small uncertainty budgets, in which case the adversarial nature maximally perturbs all arc lengths along the worst-case optimal path. This illustrates the value of adaptivity in the context of robust route planning. Interestingly, the problem variant "dynamic—decision first" performs very similarly to the static robust policy in our experiments. This highlights the importance of selecting an appropriate temporal structure for our model that reflects the timing of information discovery in practice.

To further study the differences between static and adaptive routes, we run a simulation study where, in each time stage t, the adversarial nature randomly selects each arc $(i,j) \in A$ emanating from the decision maker's present location $x_{t-1,i}(\boldsymbol{\xi}^{t-1}) = 1$ with probability p, and the lengths of the selected arc are maximally increased (subject to the remaining uncertainty budget). Figure 3 compares the static robust route (in blue) with different adaptive robust routes (in red) that respond in a worst-case optimal way to randomly drawn arc length sequences $\boldsymbol{r}_1(\boldsymbol{\xi}_1), \ldots, \boldsymbol{r}_T(\boldsymbol{\xi}_T)$ for the initial uncertainty budgets $\phi_0 \in \{0, 15, 30, 45\}$. In the figure, the optimal static route changes as the initial budget increases from zero to 15, but it subsequently returns to its original form when the budget increases further. Indeed, an initial budget of 30 is sufficient to maximally perturb all arcs on the worst-case optimal path, in which case the static robust problem becomes equivalent to a nominal shortest-path problem. For an initial uncertainty budget of zero, the simulated paths of our adaptive route coincide with the optimal static route. For all other uncertainty budgets, however, our adaptive route reacts to the revealed arc lengths; this is illustrated by the red arcs, whose thickness is proportional to the traversal frequency across 1,000 simulated sequences of arc lengths. It is worth noting that the initial arc selected by our adaptive route often differs from the one chosen by the optimal static route. This shows that future adaptivity needs to be accounted for by the decision taken here-and-now, and it demonstrates that a receding-horizon implementation of a static problem may result in suboptimal solutions.

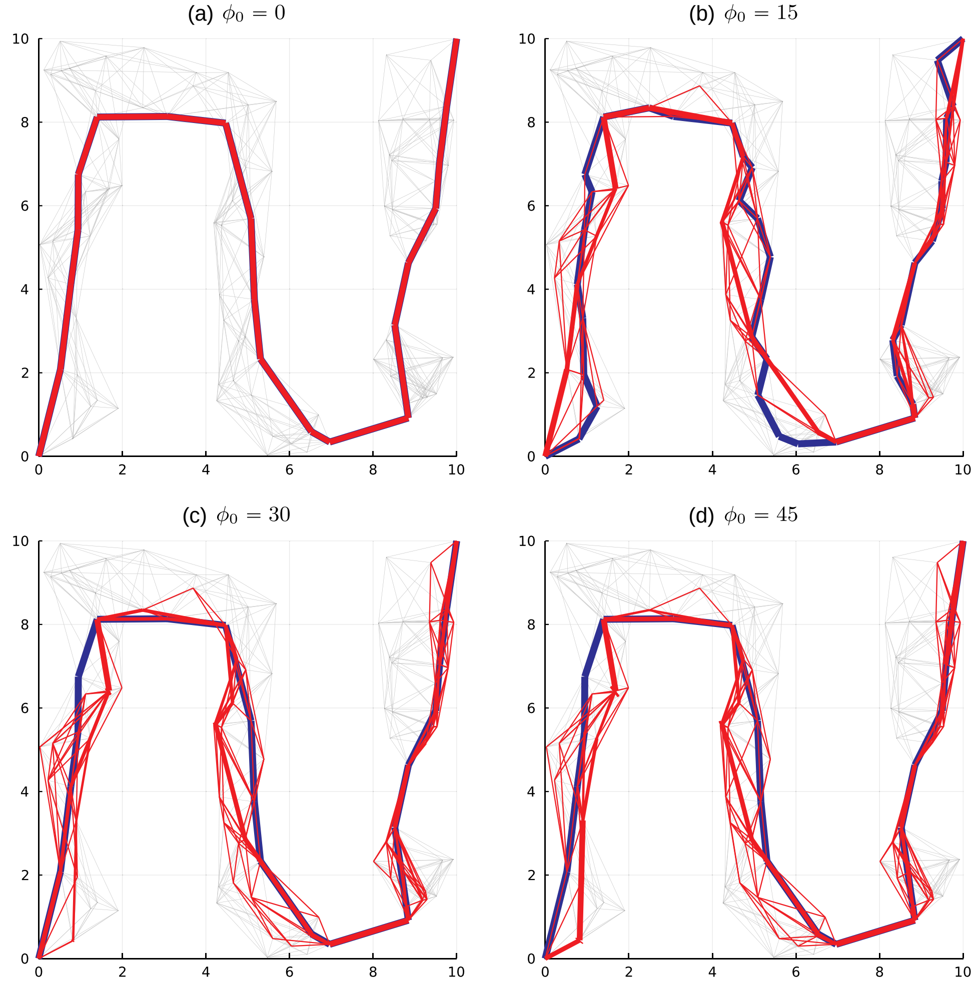
5.2. Location-Transportation Planning

We study a dynamic location-transportation problem where facilities can be built at locations $i = 1, \ldots, m$ to serve uncertain customer demands at locations $j = 1, \ldots, n$ over a time span of T stages. The problem can be formulated as the following instance of our Multistage Robust Mixed-Integer Optimization Problem (1):

$$
\begin{aligned}
\text{minimize} \quad & \max_{\boldsymbol{\xi} \in \Xi} \sum_{t=1}^{T} \boldsymbol{p}_t^\top \boldsymbol{y}_t(\boldsymbol{\xi}^t) - \boldsymbol{c}_t^\top \boldsymbol{z}_t(\boldsymbol{\xi}^t) - \boldsymbol{b}_t^\top \boldsymbol{x}_t(\boldsymbol{\xi}^t) \\
\text{subject to} \quad & \sum_{i=1}^{m} y_{tij}(\boldsymbol{\xi}^t) + z_{tj}(\boldsymbol{\xi}^t) = d_{tj}(\boldsymbol{\xi}_t) \\
& \qquad \forall \boldsymbol{\xi} \in \Xi, \ \forall j = 1, \ldots, n, \ \forall t = 1, \ldots, T \\
& \sum_{j=1}^{n} y_{tij}(\boldsymbol{\xi}^t) \leqslant C_i \cdot x_{ti}(\boldsymbol{\xi}^t) \quad \forall \boldsymbol{\xi} \in \Xi, \ \forall i = 1, \ldots, m, \ \forall t = 1, \ldots, T \\
& \boldsymbol{x}_t(\boldsymbol{\xi}^t) \geqslant \boldsymbol{x}_{t-1}(\boldsymbol{\xi}^{t-1}) \quad \forall \boldsymbol{\xi} \in \Xi, \ \forall t = 1, \ldots, T \\
& \boldsymbol{x}_t(\boldsymbol{\xi}^t) \in \{0,1\}^m, \ \boldsymbol{y}_t(\boldsymbol{\xi}^t) \in \mathbb{R}_+^{mn}, \ \boldsymbol{z}_t(\boldsymbol{\xi}^t) \in \mathbb{R}_+^n \\
& \qquad \forall \boldsymbol{\xi} \in \Xi, \ \forall t = 1, \ldots, T.
\end{aligned}
$$

Here, p_{tij} and c_{tj} denote the profit and penalty for serving or disregarding one unit of demand at location j from location i in time stage t, respectively; b_{ti} represents the cost of operating a facility at location i in time stage t; d_{tj} are the uncertain customer demands at location j in time stage t; and C_i is the capacity of the candidate facility at location i. The state variables x_{ti} record whether a facility exists at location i in time stage t, whereas the control variables y_{tij} and z_{tj} record the demand at location j that

Figure 3. (Color online) Comparison of the Optimal Static Route with 1,000 Simulated Paths of the Optimal Adaptive Route on Randomly Generated Instances with Four Different Initial Uncertainty Budgets



Note. The optimal static route is in blue, and the optimal adaptive route is in red; thickness corresponds to traversal frequency.

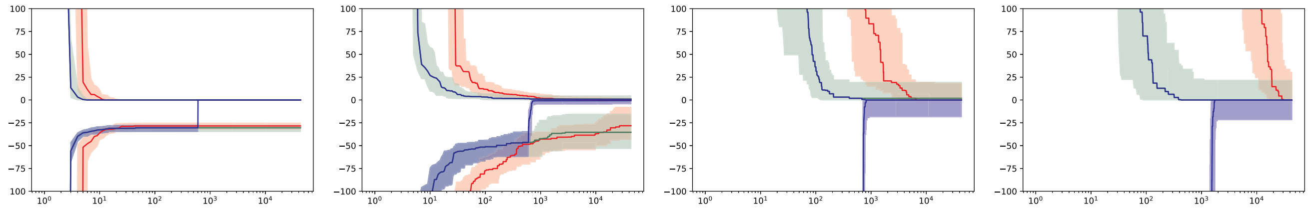
is served from location i or disregarded in time stage t, respectively. The objective function computes the cumulative worst-case difference of sales profits and penalties as well as facility operating costs. The first constraint stipulates that all customer demands are either served or disregarded, the second constraint ensures that the facility capacities are obeyed, and the third constraint prohibits the closure of previously opened facilities.

For our numerical experiments, we generate random problem instances with $m \in \{5, 10, 15, 20\}$ facility locations and $n = 2m$ customer locations that are located uniformly at random on the two-dimensional square $[0, 10]^2$. We set the per-unit sales profits p_{tij} to 80 minus d times the Euclidean distance between the locations of facility i and customer j, where the per-unit transportation costs are $d \in \{1, 5, 10\}$. The per-unit penalty $c_{tj} \in \{10, 25, 50\}$ for unserved demand is the same across all customer locations j and time stages t. We set the facility capacities to $C_i = 90$ for all locations i, whereas the operating costs are set to $b_{ti} = 80 \cdot \gamma \cdot C_i$, where $\gamma \in \{0.6, 0.7, 0.8\}$ is the facility cost factor. With this choice, the per-period operating cost b_{ti} of facility i is amortized by its sales revenues if it utilizes $100\% \cdot \gamma$ of its production capacity C_i to serve customer demands. We employ a lattice regime chain uncertainty set for the customer demands (cf. Section 2), where the demand regime ϕ_t satisfies $\phi_1 = 100$ and $\phi_t \in \Phi_t(\phi_{t-1}) = \{\phi_{t-1}, \phi_{t-1} + \Delta\phi\}$ for $t = 2, \ldots, T$, where the demand increments $\Delta\phi$ are selected from the set $\{10, 20, 30\}$. The stagewise uncertainty sets are

$$\Xi_t(\phi_{t-1}) = \Big\{ \boldsymbol{\xi}_t = (\phi_t, \boldsymbol{\psi}_t) \in \Phi_t(\phi_{t-1}) \times \mathbb{R}^4_+ : \mathbf{e}^\top \boldsymbol{d}_t(\boldsymbol{\xi}_t) = \phi_t,\ 0.8 \cdot \frac{\phi_t}{4}\mathbf{e} \leqslant \boldsymbol{\psi}_t \leqslant 1.2 \cdot \frac{\phi_t}{4}\mathbf{e} \Big\},$$

where the customer demands satisfy $d_{tj}(\boldsymbol{\xi}_t) = \sum_{f=1}^{4}$

Figure 4. (Color online) Average and 25%–75% Quartile Ranges of the Upper and Lower Bounds Generated from our Approach, the Intrastage Uncertainty Model, and the Nominal Intrastage Model



Notes. Averages are shown by bold lines, 25%–75% quartile ranges are shaded, upper and lower bounds are in blue, the intrastage uncertainty model is in red, and the nominal intrastage model is shown in green. Shown are the relative optimality gaps (ordinates) as functions of the runtime in seconds (abscissae) for $m = 5, 10, 15, 20$ (from left to right), averaged over all other $3^4 = 81$ parameter choices. Some of the lower bounds are absent because their optimality gaps lie outside the displayed region.

$\left[\frac{1}{\mathrm{dist}(j,f)} / \sum_{k=1}^{n} \frac{1}{\mathrm{dist}(k,f)}\right] \psi_{tf}$. Here, $\boldsymbol{\psi}_t \in \mathbb{R}^4_+$ is a vector of risk factors, and $\mathrm{dist}(j,f)$ is the Euclidean distance between the customer location j and the f-th corner of the square $[0, 10]^2$. The uncertainty set reflects the idea that the overall demand in time stage t is known to be ϕ_t, but its geographic breakdown is governed by uncertain risk factors that are only known to reside in a hypercube. In summary, our problem instances are characterized by the five parameters m, d, c_{tj}, γ, and $\Delta\phi$. We generate 50 random problem instances for each of these $4 \times 3^4 = 324$ parameter combinations and set the time horizon to $T = 10$.

Because the stagewise feasible regions $\mathcal{X}_t$ of the state decisions scale exponentially in the number of facility locations m, we solve the finite adaptability approximation of the dynamic transportation-location problem with our iterative state variable selection procedure outlined in Algorithm 3 (cf. Section 4). In each iteration, we solve the finite adaptability problem $P(\mathcal{X}^{\mathrm{C}}, \Xi)$ using Approximation 1 with piecewise constant state decisions based on a single partition (cf. Section 3.2), and we identify the set $\hat{\Xi}$ of worst-case uncertainties with the two worst scenarios in this problem. We compare our method with two benchmark models. The first model ("intrastage uncertainty") restricts the state decisions $\boldsymbol{x}_t$ to adapt only to the interstage uncertainties ϕ_t, but not to the intrastage uncertainties $\boldsymbol{\psi}_t$. The resulting problem can be reformulated as a large MILP that combines a scenario tree for the demand regimes ϕ_t with robust constraints for the intrastage demand realizations $\boldsymbol{d}_t(\boldsymbol{\xi}_t)$. The second model ("nominal intrastage") additionally replaces the uncertain intrastage demand realizations $\boldsymbol{d}_t(\boldsymbol{\xi}_t)$ with their nominal values $d^0_{tj} = \frac{1}{4}\sum_{f=1}^{4}\left[\frac{1}{\mathrm{dist}(j,f)} / \sum_{k=1}^{n} \frac{1}{\mathrm{dist}(k,f)}\right] \cdot \phi_t$. The resulting model is smaller and thus easier to solve, but it constitutes a weaker approximation of the original problem. In our experiments, we use the incumbent solution of nominal intrastage that has been determined after 10 minutes runtime to warm start our iterative state variable selection procedure. We set the runtime limit of each approach to 12 hours per problem instance.

Figure 4 presents numerical results for different parameter settings. Note that each of the three methods provides both upper (conservative) and lower (progressive) bounds on the optimal value of the dynamic location-transportation problem. In particular, upper bounds can be obtained via Algorithm 1, whereas the lower bounds result from computing the crystal ball solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ over a reduced uncertainty set $\hat{\Xi}$ that is extracted from the upper bound problem. Our results show that the nominal intrastage model can obtain feasible solutions quickly, but the lower bounds tend to provide poor estimates of the optimal value. The intrastage uncertainty model, on the other hand, can result in good upper and lower bounds, but its runtime grows excessively as a function of the problem size. Our iterative approach, finally, provides tighter bounds than both benchmark methods under almost all parameter settings. The kinks in our lower bounds for $m = 5$ and $m = 10$ in Figure 4 are because of the transition from the incumbent solution to the nominal intrastage model after 10 minutes to the lower bounds produced by Algorithm 3. We note that whereas the final gap between our lower and upper bounds is typically small; in most problem instances, the bounds do not collapse. We present further details of our numerical results in Online Appendix C.

Acknowledgments

The authors would like to sincerely thank the area editor and the associate editor as well as the two referees for their constructive and insightful feedback that has led to substantial improvements of the manuscript.

Endnote

[1] See http://www.doc.ic.ac.uk/~wwiesema/multi-stage-ro.zip.

References

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

Bertsimas D, Caramanis C (2007) Adaptability via sampling. *46th IEEE Conf. Decision Control* (IEEE, Piscataway, NJ), 4717–4722.

Bertsimas D, Caramanis C (2010) Finite adaptability in multistage linear optimization. *IEEE Trans. Automatic Control* 55(12):2751–2766.

Bertsimas D, den Hertog D (2022) *Robust and Adaptive Optimization* (Dynamic Ideas, Charlestown, MA).

Bertsimas D, Dunning I (2015) Multistage robust mixed-integer optimization with adaptive partitions. *Oper. Res.* 64(4):980–998.

Bertsimas D, Georghiou A (2015) Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Oper. Res.* 63(3):610–627.

Bertsimas D, Georghiou A (2018) Binary decision rules for multistage adaptive mixed-integer optimization. *Math. Programming* 167(2):395–433.

Bertsimas D, Sim M (2003) Robust discrete optimization and network flows. *Math. Programming* 98(1–3):49–71.

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

Bertsimas D, Goyal V, Sun XA (2011b) A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Math. Oper. Res.* 36(1):24–54.

Daryalal M, Arslan AN, Bodur M (2023) Two-stage and Lagrangian dual decision rules for multistage adaptive robust optimization. Preprint, submitted May 10, https://arxiv.org/abs/2305.06190.

Delage E, Iancu DA (2015) Robust multistage decision making. *INFORMS Tutorials in Operations Research* (INFORMS, Catonsville, MD), 20–46.

Feng W, Feng Y, Zhang Q (2021) Multistage robust mixed-integer optimization under endogenous uncertainty. *Eur. J. Oper. Res.* 294(2):460–475.

Georghiou A, Tsoukalas A, Wiesemann W (2020) A primal-dual lifting scheme for two-stage robust optimization. *Oper. Res.* 68(2):572–590.

Goerigk M, Hartisch M (2021) Multistage robust discrete optimization via quantified integer programming. *Comput. Oper. Res.* 135:105434.

Guslitser E (2002) Uncertainty-immunized solutions in linear programming. Master's thesis, Technion - Israel Institute of Technology, Haifa, Israel.

Hanasusanto GA, Kuhn D, Wiesemann W (2015) *K*-adaptability in two-stage robust binary programming. *Oper. Res.* 63(4):877–891.

Hashemi Doulabi H, Jaillet P, Pesant G, Rousseau L-M (2021) Exploiting the structure of two-stage robust optimization models with exponential scenarios. *INFORMS J. Comput.* 33(1):143–162.

Postek K, den Hertog D (2016) Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS J. Comput.* 28(3):553–574.

Rockafellar RT, Wets RJ-B (1998) *Variational Analysis*, Grundlehren der mathematischen Wissenschaften, vol. 317 (Springer, Berlin, Heidelberg).

Romeijnders W, Postek K (2021) Piecewise constant decision rules via branch-and-bound based scenario detection for integer adjustable robust optimization. *INFORMS J. Comput.* 33(1): 390–400.

Shapiro A (2017) Interchangeability principle and dynamic equations in risk averse stochastic programming. *Oper. Res. Lett.* 45(4):377–381.

Subramanyam A, Gounaris CE, Wiesemann W (2020) *K*-adaptability in two-stage mixed-integer robust optimization. *Math. Programming Comput.* 12(2):193–224.

Vayanos P, Kuhn D, Rustem B (2011) Decision rules for information discovery in multi-stage stochastic programming. *Proc. 50th IEEE Conf. Decision Control* (IEEE, Piscataway, NJ), 7368–7373.

Yanıkoğlu I, Gorissen BL, den Hertog D (2019) A survey of adjustable robust optimization. *Eur. J. Oper. Res.* 277(3):799–813.

Zeng B, Zhao L (2013) Solving two-stage robust optimization problems using a column-and-constraint generation method. *Oper. Res. Lett.* 41(5):457–561.

Zhao L, Zeng B (2012) An exact algorithm for two-stage robust optimization with mixed integer recourse problems. Technical report, University of South Florida, Tampa, FL.

Krzysztof Postek is a senior data scientist at the Boston Consulting Group, specializing in optimization. His research focuses on optimization under uncertainty, robust optimization, and their applications in engineering and operations management. He has received funding from prestigious grants, including the NWO Veni and Research Talent grants.

Ward Romeijnders is professor of optimization under uncertainty at the Department of Operations at the University of Groningen. His research focuses on decision making under uncertainty, often in the presence of integer decisions. Currently, he is the secretary of the Stochastic Programming Society and associate editor of *Mathematical Methods of Operations Research*.

Wolfram Wiesemann is professor of analytics and operations as well as the head of the Analytics, Marketing & Operations Department at Imperial College Business School. His research interests evolve around decision making under uncertainty, with applications to supply chain management, healthcare, and energy. He is an elected member of the boards of the Mathematical Optimization Society and the Stochastic Programming Society, and he currently serves as editor-in-chief of *Operations Research Letters* as well as a department coeditor for *Management Science*.