A nonparametric algorithm for optimal stopping based on robust optimization

Bradley Sturt
Department of Information and Decision Sciences
University of Illinois at Chicago, bsturt@uic.edu

Optimal stopping is a class of stochastic dynamic optimization problems with applications in finance and operations management. In this paper, we introduce a new method for solving stochastic optimal stopping problems with known probability distributions. First, we use simulation to construct a robust optimization problem that approximates the stochastic optimal stopping problem to any arbitrary accuracy. Second, we characterize the structure of optimal policies for the robust optimization problem, which turn out to be simple and finite-dimensional. Harnessing this characterization, we develop exact and approximation algorithms for solving the robust optimization problem, which in turn yield policies for the stochastic optimal stopping problem. Numerical experiments show that this combination of robust optimization and simulation can find policies that match, and in some cases significantly outperform, those from state-of-the-art algorithms on low-dimensional, non-Markovian optimal stopping problems from options pricing.

Key words: stochastic dynamic optimization; robust optimization; optimal stopping; options pricing.

1. Introduction

Consider the following class of stochastic dynamic optimization problems. A sequence of random states are incrementally revealed to a decision maker. After observing the state in each period, the decision maker chooses whether to continue to the next period or stop and receive a reward that depends on the current state. The problem is to find a control policy, called a stopping rule, for selecting when to stop the process to maximize the expected reward.

Such optimal stopping problems are widely-studied and arise in a variety of domains like finance, promotion planning (Feng and Gallego 1995), and organ transplantation (David and Yechiali 1985). In particular, optimal stopping has considerable importance to industry for the pricing of financial derivatives. With a record trading volume that exceeded seven billion contracts in 2020, equity options are among the most widely-traded type of financial derivative (Reuters 2021), and financial firms depend on solving optimal stopping problems to determine accurate prices for American-style options, the most common type of equity option.

In this paper, we study a general class of optimal stopping problems in which the sequence of random states is driven by a non-Markovian probability distribution. We recall that a sequence of random states is non-Markovian if the state in the next time period (e.g., a stock’s price tomorrow) has a probability distribution which depends both on the state in the current time period (e.g., the stock’s price today) as well as the states in the past periods (e.g., the stock’s price yesterday). This class of optimal stopping problems has witnessed a surge of interest as financial firms increasingly use non-Markovian probability distributions to accurately model the volatility patterns of stocks (Gatheral et al. 2018, Leão et al. 2019, Becker et al. 2019, Bezerra et al. 2020, Goudenège et al. 2021).
2020, Bayer et al. 2020). Optimal stopping problems with non-Markovian probability distributions also occur when using popular dimensionality-reduction techniques for pricing high-dimensional basket options (Bayer et al. 2019, p. 372) and pricing options when the probability distribution of underlying assets is accessed via a black-box simulator constructed from historical data (Ciocan and Mišić 2020, §5.5).

Despite their importance in practice, non-Markovian optimal stopping (NMOS) problems are “not easy to solve” (Leão et al. 2019, p. 982). The difficulty of these problems arises because the optimal decision in each period may depend on the entire history of the state process. In principle, an NMOS problem can be transformed into an equivalent Markovian optimal stopping problem by converting the original sequence of random states \(x_1, \ldots, x_T \in \mathcal{X}\) into a Markovian stochastic process \(X_1, \ldots, X_T \in \mathcal{X}_T\), where the new state \(X_t := (x_1, \ldots, x_t, 0, \ldots, 0) \in \mathcal{X}_T\) in each period \(t\) includes the entire state history of the original process. Unfortunately, the enlarged state space \(\mathcal{X}_T\) will be high-dimensional when the optimal stopping problem has many periods, and the difficulty of solving a Markovian optimal stopping problem explodes in the dimensionality of the state space.

To contend with the ‘curse-of-dimensionality’ that arises in NMOS problems, a natural approximation technique is to search only for stopping rules that are Markovian. Rather than depending on the entire history of the original sequence, a Markovian stopping rule makes a decision in each period \(t\) based only on the current original state \(x_t\) and knowledge that the sequence of random states was not stopped in any of the previous time periods. In general, the best Markovian stopping rule for an NMOS problem is not guaranteed to be an optimal stopping rule for the NMOS problem. However, recent numerical evidence demonstrates that Markovian stopping rules can lead to highly accurate approximations of optimal stopping rules in NMOS problems from options pricing; see Goudenège et al. (2020, §9.2), Ciocan and Mišić (2020, §5.5), Bayer et al. (2020, §3.5).

As far as we are aware, only two papers until now have suggested methods which are theoretically capable of finding the best Markovian stopping rules to NMOS problems. Belomestny (2011a) analyzes simulation-based methods which optimize directly over spaces of Markovian stopping rules and suggests a nonparametric space of Markovian stopping rules based on \(\epsilon\)-nets; however, he does not propose any concrete algorithms for optimizing over this nonparametric space of Markovian stopping rules. Ciocan and Mišić (2020) propose optimizing over Markovian stopping rules which are restricted to decision trees with fixed depth. Due to the computational intractability of optimizing over all decision trees of fixed depth, the authors develop greedy heuristics which are shown to limit the range of attainable decision trees, and overcoming this limitation of the heuristics “is not obvious, especially in light of the structure of the optimal stopping problem that is leveraged to efficiently optimize split points in our construction” (Ciocan and Mišić 2020, p. 22).

We take a different approach to the aforementioned literature, and in doing so make our contributions to optimal stopping, by drawing on the traditionally unrelated field of robust optimization. Over the past two decades, robust optimization has emerged as a leading tool in operations research for dynamic decision-making when uncertainty is driven by unknown or ambiguous probability distributions (Ben-Tal et al. 2009, Delage and Iancu 2015). In this paper, we show that robust optimization can be combined with simulation to develop algorithms for finding Markovian stopping rules to NMOS problems with known probability distributions. Compared to Belomestny (2011a) and Ciocan and Mišić (2020), our method for optimal stopping does not restrict the space
of Markovian stopping rules to any parametric class, and we develop concrete algorithms which are guaranteed to yield \( \epsilon \)-optimal Markovian stopping rules for general classes of NMOS problems.

At a high level, our method is comprised of the following main steps. We first use Monte-Carlo simulation to generate sample paths of the sequence of random states. From those sample paths, we construct a robust optimization problem that serves as a proxy for the NMOS problem. As the number of sample paths grows to infinity, the optimal objective value and optimal Markovian stopping rules of the robust optimization problem are shown to converge to those of the NMOS problem. Because the joint probability distribution of the sequence of random states is known, we can simulate any desired number of sample paths when constructing the robust optimization problem. By solving the robust optimization problem, we thus obtain stopping rules for the NMOS problem.

Once the robust optimization problem is constructed from the simulated sample paths, we are faced with the necessary task of solving the robust optimization problem. A key challenge is that the robust optimization problem requires optimizing over the infinite-dimensional space of all Markovian stopping rules. Infinite-dimensional optimization problems are generally intractable from a computational perspective, which raises key questions: can optimal or provably near-optimal Markovian stopping rules for our robust optimization problem be found efficiently? More broadly, do optimal Markovian stopping rules for our robust optimization problem even exist? We answer these questions by making the following technical developments:

- We characterize the structure of optimal Markovian stopping rules for the robust optimization problem, which turn out to be simple and finite-dimensional. In particular, we show that the optimal Markovian stopping rules for the robust optimization problem can be represented exactly using \( N \) integers, where \( N \) is the number of simulated sample paths.
- Harnessing the above characterization, we develop an exact reformulation of the robust optimization problem as a zero-one bilinear program over totally unimodular constraints.
- By exploiting the structure of the exact reformulation, we design a tractable (polynomial-time) heuristic algorithm for approximating the robust optimization problem.

Along the way, we also establish theoretical guarantees:

- We prove that the optimality gap of the heuristic approximation algorithm is equal to zero for any optimal stopping problem with two periods.
- For optimal stopping problems with three or more periods, we show that finding the optimal Markovian stopping rules for the robust optimization problem is strongly NP-hard.
- We develop a general upper bound on the optimality gap of our heuristic approximation algorithm with respect to an intermediary problem.

To the best of our knowledge, this is the first work to combine robust optimization and simulation to design algorithms for a class of stochastic dynamic optimization problems with known probability distributions. Robust optimization has been studied extensively in the context of dynamic decision-making under uncertainty since Ben-Tal et al. (2004), and Delage and Iancu (2015) provide an excellent introduction to dynamic robust optimization problems. A key technical result of our paper, Theorem 4, builds on a rich tradition of characterizing the structure of optimal policies to dynamic robust optimization problems, exemplified by Bertsimas et al. (2010), Iancu et al. (2013), Mamani
et al. (2017), Sun and Van Mieghem (2019), and Delage and Iancu (2015, §4). Differing from prior work, our robust optimization problem provides an approximation of a stochastic dynamic optimization problem to any arbitrary accuracy; consequently, our characterization of optimal policies for a robust optimization problem yields a new class of $\epsilon$-optimal policies for a stochastic dynamic optimization problem.

Our method for optimal stopping builds upon the work of Bertsimas et al. (2020), which used robust optimization to approximate a class of stochastic dynamic linear optimization problems with unknown probability distributions. In particular, the present paper expands on the aforementioned work in three significant ways: (i) we provide the first characterization of optimal policies for a dynamic robust optimization problem that averages over multiple uncertainty sets; (ii) we develop the first exact algorithms for solving these dynamic robust optimization problems; (iii) we show that the dynamic robust optimization model developed in Bertsimas et al. (2020) and Sturt (2020) can be combined with simulation to design algorithms for solving a class of stochastic dynamic optimization problems with known probability distributions.

We conclude with numerical experiments that demonstrate the value of our robust optimization-based algorithms in several settings. First, we consider a simple one-dimensional non-Markovian optimal stopping problem with fifty periods, and we compare the robust optimization algorithm to existing methods based on approximate dynamic programming (Longstaff and Schwartz 2001) and parametric stopping rules (Ciocan and Mišić 2020). The experiments show that our method can find stopping rules that significantly outperform those found by the other techniques, while maintaining a comparable computational cost. In particular, the experiments reveal that our method can strictly outperform alternative algorithms for finding Markovian stopping rules to NMOS problems that are based on backwards recursion. Second, we consider a widely-studied and important problem of pricing high-dimensional Bermudan barrier options with over fifty periods. Across several variants of this problem, we demonstrate that our combination of robust optimization and simulation can find stopping rules that match, and in some cases significantly outperform, those from state-of-the-art algorithms.

The rest of our paper has the following organization. §1.1 provides a review of methods for solving optimal stopping problems with Markovian and non-Markovian probability distributions. §2 formalizes the problem setting and introduces our robust optimization-based method. §3 characterizes the structure of optimal policies for the robust optimization problem. §4 develops tractable algorithms and computational complexity results for the robust optimization problem. §5 illustrates the performance of our algorithms in numerical experiments. With the exception of the characterization of the structure of optimal policies for the robust optimization problem (Theorem 4), all technical proofs can be found in the appendices.

### 1.1. Other Related Literature

Many methods based on approximate dynamic programming (ADP) have been developed for optimal stopping problems with high-dimensional Markovian stochastic processes. The most popular ADP methods for these optimal stopping problems are based on Monte-Carlo simulation and regression, which originate with Carriere (1996), Longstaff and Schwartz (2001) and Tsitsiklis and
Van Roy (2001). Given sample paths of the entire stochastic process, these methods use backwards recursion and regression to obtain approximations of the value function, and exercise policies are then obtained by proceeding greedily with respect to the approximate value functions. The efficacy of regression-based methods hinges on selecting a parametrization of basis functions for the value function that strikes a balance between approximation quality and sample complexity. Nonparametric choices for the basis functions, e.g., Laguerre polynomials, are discussed in the aforementioned works and subsequently analyzed in works such as Clément et al. (2002), Glasserman and Yu (2004), Egloff (2005), Belomestny (2011b) and Zanger (2020). In §5, we provide numerical comparisons of our proposed algorithms to ADP techniques in the context of NMOS problems.

A variety of other nonparametric methods have been developed for solving optimal stopping problems with Markovian probability distributions, such as quantization-based approximations of value functions (Bally and Pages 2003) and scenario tree discretizations of the sequence of random states (Broadie and Glasserman 1997). Recent works have also considered using deep learning to learn the continuation function, including Becker et al. (2019) and Goudené et al. (2020). Methods to compute upper bounds on optimal stopping problems grew in interest due to the independent works of Haugh and Kogan (2004) and Rogers (2002), and duality-based algorithms to obtain upper bounds which combine simulation and suboptimal stopping rules were first proposed by Andersen and Broadie (2004). Other works which harness dual representations to solve optimal stopping problems include Desai et al. (2012), Belomestny (2013), and Goldberg and Chen (2018), among many others.

In the context of non-Markovian optimal stopping, methods have been developed which address settings that are different from ours in non-trivial ways. Leão et al. (2019) and Bezerra et al. (2020) develop discretization schemes for NMOS problems over continuous time and restrict the class of probability distributions to those based on the Brownian motion. In contrast to these works, our paper develops algorithms for finding Markovian stopping rules for discrete-time optimal stopping problems, and we do not require any parametric assumptions on the probability distributions of the underlying stochastic processes. NMOS problems can also be addressed by the scenario tree method of Broadie and Glasserman (1997) and the recursive-dual algorithm of Goldberg and Chen (2018), provided that one can perform Monte-Carlo simulation on the conditional probability distribution of the stochastic process in each time period. In contrast to these methods, the algorithms in this paper require only the ability to simulate sample paths of the entire stochastic process and are shown in numerical experiments to be practically tractable in low-dimensional NMOS problems with dozens of time periods. Within the optimal stopping literature, our method is most closely related to a class of simulation-based methods which optimize directly over spaces of deterministic stopping rules, as explored by García (2003), Andersen (1999), Belomestny (2011a), Gemmrich (2012), Ciocan and Mišić (2020), and Glasserman (2013, §8.2). We discuss connections between our method and this stream of literature in §2.3, and a discussion of the challenges of using dynamic programming to find the best Markovian stopping rules for NMOS problems can be found in Appendix A.
2. Robust Optimization for Stochastic Optimal Stopping

2.1. Problem Setting

We consider stochastic optimal stopping problems defined by the following components:

**States:** Let $x \equiv (x_1, \ldots, x_T)$ denote a sequence of random states, where the state $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$ in each period $t$ is a random vector of dimension $d$. For example, the state in each period may represent the prices of multiple assets at that point in time. The joint probability distribution of this stochastic process is assumed to be known and accessible through a simulator which generates independent sample paths of the entire stochastic process.

**Policies:** Let $\mu \equiv (\mu_1, \ldots, \mu_T)$ represent a collection of exercise policies, where the exercise policy in each period $t$ is a function of the form $\mu_t : \mathcal{X} \rightarrow \{\text{Stop, Continue}\}$. Speaking intuitively, each exercise policy is a partitioning of the state space into regions for stopping and continuing. From the exercise policies, the corresponding Markovian stopping rule $\tau_{\mu} : \mathcal{X}^T \rightarrow \{1, \ldots, T\} \cup \{\infty\}$ is a function that maps a realization of the stochastic process to stopping period:

$$\tau_{\mu}(x) \triangleq \min\{t \in \{1, \ldots, T\} : \mu_t(x_t) = \text{Stop}\}.$$

Throughout this paper, a minimization problem with no feasible solutions is defined equal to $\infty$.

**Rewards:** Let $g : \{1, \ldots, T\} \cup \{\infty\} \times \mathcal{X}^T \rightarrow \mathbb{R}_+$ be a known and deterministic function which maps a stopping period and a realization of the entire stochastic process to a reward. The assumption that the reward function is nonnegative is common in many applications of optimal stopping, and we assume throughout the paper that a stochastic process that is never stopped yields a reward of zero: $g(\infty, x) \equiv 0$. It follows from the definition of the reward function that the reward from stopping on any period $t$ may in general depend on the states of the stochastic process in previous or future time periods.

**Problem:** With the above notation and inputs, the goal of this paper is to solve stochastic optimal stopping problems of the form

$$\sup_{\mu} \mathbb{E}[g(\tau_{\mu}(x), x)],$$

where the optimization is taken over the space of all Markovian stopping rules. In the following sections, we introduce and analyze a new simulation-based method for solving this class of stochastic dynamic optimization problems.

2.2. The Robust Optimization Method

Our proposed method for solving stochastic optimal stopping problems of the form (OPT) consists of the following steps. We first simulate sample paths of the stochastic process $x \equiv (x_1, \ldots, x_T)$. Let $N$ denote the number of sample paths and the values of the sample paths be denoted by $x^i \equiv (x^i_1, \ldots, x^i_T)$ for each $i = 1, \ldots, N$. 
We assume that the sample paths are independent and identically distributed realizations of the entire (possibly non-Markovian) stochastic process. We next choose the following robustness parameter:

$$\epsilon \geq 0.$$ 

The purpose of the robustness parameter will become clear momentarily, and a discussion on how to choose the number of sample paths and the robustness parameter is deferred until §2.5. With these parameters, let the uncertainty set around sample path $i$ on period $t$ be defined as

$$\mathcal{U}_t^i \triangleq \{y_t \in \mathcal{X} : \|y_t - x_t^i\|_{\infty} \leq \epsilon\}.$$ 

For notational convenience, denote the uncertainty set around sample path $i$ across all periods by

$$\mathcal{U}^i \triangleq \mathcal{U}_1^i \times \cdots \times \mathcal{U}_T^i.$$ 

Hence, we observe that the role of the robustness parameter is to control the size of these sets. Given the sample paths and choice of the robustness parameter, our approach obtains an approximation of $\text{OPT}$ by solving the following robust optimization problem:

$$\sup_{\mu} \frac{1}{N} \sum_{i=1}^N \inf_{y \in \mathcal{X}^T : \|y - x^i\|_{\infty} \leq 0} g(\tau_{\mu}(y), x^i).$$  \hspace{1cm} (RO)$$

By solving the above robust optimization problem, we obtain the policies $\hat{\mu} \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_T)$. These policies constitute our approximate solution to the stochastic optimal stopping problem $\text{OPT}$.

### 2.3. Background and Motivation

In contrast to traditional robust optimization or distributionally robust optimization, our motivation behind adding adversarial noise to the sample paths in (RO) is not to find stopping rules which have worst-case performance guarantees, are attractive in risk-averse settings, or perform well the presence of an ambiguous probability distribution. Rather, this paper proposes using robust optimization purely as an algorithmic tool for solving stochastic optimal stopping problems of the form $\text{OPT}$ when the joint probability distributions are known. The present section elaborates on this motivation and positions our use of robust optimization within the optimal stopping literature.

For the sake of developing intuition, let us suppose for the moment that the robustness parameter of the uncertainty sets in (RO) was set equal to zero. In this case, for any fixed exercise policies $\mu = (\mu_1, \ldots, \mu_T)$, the expected reward of those exercise policies,

$$J^*(\mu) \triangleq \mathbb{E}[g(\tau_{\mu}(x), x)],$$

would be approximated in (RO) by the sample average approximation:

$$\hat{J}_{N,0}(\mu) \triangleq \frac{1}{N} \sum_{i=1}^N \inf_{y \in \mathcal{X}^T : \|y - x^i\|_{\infty} \leq 0} g(\tau_{\mu}(y), x^i) = \frac{1}{N} \sum_{i=1}^N g(\tau_{\mu}(x^i), x^i).$$

For these fixed exercise policies, we observe that the sample average approximation is a consistent estimator of the expected reward. In other words, for the fixed exercise policies $\mu$, it follows from
the strong law of large numbers under relatively mild assumptions\(^1\) that \( \hat{J}_{N,0}(\mu) \) will converge almost surely to \( J^*(\mu) \) as the number of simulated sample paths is taken to infinity.

However, it is well known in the optimal stopping literature that these desirable asymptotic properties of \( \hat{J}_{N,0}(\mu) \) are generally not retained when considering the problem of optimizing over the space of all exercise policies. For instance, Ciocan and Mišić (2020, EC.1.2) provide simple examples in which the following statements hold almost surely:

\[
\lim_{N \to \infty} \sup_\mu \hat{J}_{N,0}(\mu) \gg \sup_\mu J^*(\mu); \quad \lim_{N \to \infty} J^*(\hat{\mu}_{N,0}) \ll \sup_\mu J^*(\mu). \tag{1}
\]

The asymptotic suboptimality of the optimal objective value and optimal policies for the problem \( \sup_\mu \hat{J}_{N,0}(\mu) \) can be intuitively understood as a type of overfitting. To see why line (1) occurs, we recall for any fixed choice of exercise policies \( \mu \) that the sample average approximation \( \hat{J}_{N,0}(\mu) \) is an unbiased estimate of the expected reward \( J^*(\mu) \). However, when simultaneously considering the space of all exercise policies, there exists for each \( N \in \mathbb{N} \) with high probability a collection of exercise policies which satisfies \( \hat{J}_{N,0}(\mu) \gg J^*(\mu) \). The problem \( \sup_\mu \hat{J}_{N,0}(\mu) \) will thus be biased towards choosing those exercise policies, which in general will be suboptimal for the problem \( \sup_\mu J^*(\mu) \). Because the set of all \( \mu \) is an infinite-dimensional space, the gap between the objective values \( \hat{J}_{N,0}(\mu) \) and \( J^*(\mu) \) does not converge to zero uniformly over the set of all \( \mu \) as the number of sample paths tends to infinity.

To circumvent this overfitting in the context of optimal stopping in line (1), a vast literature has focused on restricting the functional form of exercise policies to a finite-dimensional space, such as Garcia (2003), Andersen (1999), Belomestny (2011a), Gemmrich (2012), Ciocan and Mišić (2020). In this approach, the choice of the parameterization for the space of exercise policies must be made very carefully. On one hand, the effective dimension of the restricted space of exercise policies must be small relative to the number of simulated sample paths to ensure that the sample average approximation problem finds the parametric exercise policies which are ‘best-in-class’ with respect to the stochastic optimal stopping problem (Belomestny 2011a, §3). On the other hand, the parameterization must be chosen appropriately in order for the sample average approximation problem to obtain a good approximation of (OPT). Choosing such an appropriate parameterization “may be counterfactual in some cases”, as explained by Garcia (2003, p. 1859), “since we may not have a good understanding of what the early exercise rule should depend on.”

Our approach, in view of the above discussion, provides an alternative means to circumvent overfitting. The proposed robust optimization problem allows the space of exercise policies to remain general, and thus relieves the decision maker from the need to select and impose a parametric structure on the exercise policies. Moreover, we show in the following section that our use of robust optimization provably overcomes the asymptotic overfitting described in line (1).

### 2.4. Optimality Guarantees

To establish the theoretical justification for the robust optimization method, we make the following relatively mild assumptions on the stochastic process \( x \equiv (x_1, \ldots, x_T) \) and the reward function \( g(t, x) \) in the stochastic optimal stopping problem (OPT):

\(^1\) For example, Assumptions 2 and 4 in §2.4.
Assumption 1. \( \lim_{\epsilon \to 0} \Delta_\epsilon(x) = 0 \) almost surely, where
\[
\Delta_\epsilon(x) \triangleq \min_{t \in \{1, \ldots, T\}} \left\{ \inf_{y \in \mathcal{X}^T : \|y - x\|_\infty \leq \epsilon} g(t, y) - g(t, x) \right\}.
\]

Assumption 2. The stochastic process satisfies \( \mathbb{E}[\exp(\|x\|_\infty)] < \infty \) for some \( a > 1 \).

Assumption 3. There is an optimal \( \mu^* \) to (OPT) such that \( \lim_{\mu \to x} \tau_{\mu^*}(y) = \tau_{\mu^*}(x) \) almost surely.

Assumption 4. The reward function satisfies \( 0 \leq g(t, y) \leq U \) for all \( t \in \{1, \ldots, T\} \) and \( y \in \mathcal{X}^T \).

Let us briefly interpret the above assumptions on the optimal stopping problem. We observe that Assumption 1 holds automatically when the functions \( g(1, \cdot), \ldots, g(T, \cdot) \) are continuous, and it can also hold in important stochastic optimal stopping problems with discontinuous reward functions\(^2\). Assumption 2 is a standard light-tail assumption on the stochastic process which is satisfied, for example, if the stochastic process is bounded or has a multivariate normal distribution. Assumption 3 says that there are optimal exercise policies for the stochastic optimal stopping for which the stochastic process will lie a positive distance away from the margins of the stopping regions with probability one. The assumption of boundedness on the reward function in Assumption 4 leads to a considerably simpler proof and statement of the results, but can generally be relaxed to reward functions bounded above by an integrable function.

Under the above conditions, the following theorems provide justification for using the robust optimization problem as a proxy for the stochastic optimal stopping problem. In a nutshell, the following Theorems 1-3 show that (RO) will, for all sufficiently small choices of the robustness parameter and all sufficiently large choices of the number of simulated sample paths, yield a near-optimal approximation of (OPT). Stated another way, the following theorems show that our use of robust optimization provably overcomes the asymptotic overfitting described in line (1) of the previous section. While the following theorems do not specify how to choose the robustness parameter and number of simulated sample paths for any particular optimal stopping problem \textit{a priori}, we provide guidance (§2.5) and numerical evidence (§5) which suggest that these parameters can be found effectively in practice.

Our first theorem shows that the optimal objective value of the robust optimization problem (RO) will converge almost surely to that of the stochastic problem (OPT) as the robustness parameter tends to zero and the number of sample paths tends to infinity. In the following result, we use the notation \( \hat{J}_{N,\epsilon}(\mu) \) to denote the objective value of the robust optimization problem (RO) corresponding to exercise policies \( \mu \).

**Theorem 1 (Consistency of optimal objective value).** Under the above assumptions,
\[
\lim_{\epsilon \to 0} \lim_{N \to \infty} \sup_{\mu} \hat{J}_{N,\epsilon}(\mu) = \sup_{\mu} J^*(\mu) \quad \text{almost surely.}
\]

\(^2\)To illustrate, consider Robbin’s problem (Bruss 2005), in which the reward functions \( g(t, x) = \text{rank}(x_t; x_1, \ldots, x_T) \) are discontinuous and the probability distribution is \( x_1, \ldots, x_T \overset{iid}{\sim} \text{Uniform}[0, 1] \). To show that Assumption 1 is satisfied, we observe that the random variable \( \bar{\epsilon} = \min_{s < t} |x_s - x_t| \) is strictly positive with probability one, which implies that \( \Delta_\epsilon(x) = 0 \) for all \( \epsilon < \bar{\epsilon} \).
Our second theorem shows that the expected reward of the optimal exercise policies for the robust optimization problem (RO) will converge almost surely to the optimal objective value of the stochastic problem (OPT). We let \( \hat{\mu}_{N,\epsilon} \) denote optimal exercise policies for (RO), and we remark that the existence of optimal exercise policies for the robust optimization problem will be established in §3.

**Theorem 2 (Consistency of optimal policies).** Under the above assumptions,

\[
\lim_{\epsilon \to 0} \liminf_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) = \lim_{\epsilon \to 0} \limsup_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) = \sup_{\mu} J^*(\mu) \quad \text{almost surely.}
\]

Because we will develop algorithms that solve the robust optimization problem approximately as well as exactly, it is imperative for us to have theoretical guarantees which hold for any Markovian stopping rule that can be found by the robust optimization problem. To this end, our third and final theorem of this section shows that the (in-sample) robust objective value will asymptotically provide a low-bias estimate of the (out-of-sample) expected reward, and this bound holds uniformly over all exercise policies. The result yields theoretical assurance, provided that the robustness parameter is sufficiently small and the number of sample paths is sufficiently large, that searching for exercise policies with high robust objective values \( \hat{J}_{N,\epsilon}(\mu) \) will typically result in exercise policies with high expected rewards \( J^*(\mu) \).

**Theorem 3 (Asymptotic low-bias).** Under the above assumptions,

\[
\lim_{\epsilon \to 0} \liminf_{N \to \infty} \inf_{\mu} \left\{ J^*(\mu) - \hat{J}_{N,\epsilon}(\mu) \right\} \geq 0 \quad \text{almost surely.}
\]

### 2.5. Implementation Details

In anticipation of algorithmic techniques for solving the robust optimization problem (RO) in the remainder of the paper, it remains to be specified how the parameters of the robust optimization problem (the number of simulated sample paths \( N \in \mathbb{N} \) and the robustness parameter \( \epsilon \geq 0 \)) should be selected in practice. For the sake of concreteness, we conclude §2 by briefly providing guidance for choosing these parameters and applying the robust optimization method in practice. The procedures described below are formalized in Algorithm 1 and implemented in our numerical experiments in §5.

As described previously, this paper addresses stochastic optimal stopping problems in which the probability distributions are known. Consequently, the decision-maker is granted flexibility in choosing the number of sample paths \( N \) to simulate. On one hand, we have established in the previous section that larger choices of the number of simulated sample paths will generally lead to tighter approximations of the stochastic optimal stopping problem. On the other hand, larger choices of \( N \) require a greater computation cost in performing the Monte-Carlo simulation and creates a robust optimization problem of a larger size. To balance these tradeoffs in particular applications, we recommend using a straightforward procedure of starting out with a small choice of \( N \) and iteratively increasing the number of simulated sample paths until the total computational cost meets the allocated computational budget.

Given a fixed number of sample paths, the choice of the robustness parameter \( \epsilon \geq 0 \) can have a significant impact on the policies produced by the robust optimization problem. To this end,
The Robust Optimization Method for Stochastic Optimal Stopping

Inputs:
- **Sizes of Training Sets**: A collection of integers $\mathcal{N} \triangleq \{N_1, \ldots, N_K\}$ sorted in ascending order.
- **Size of Validation Set**: An integer $\bar{N} \in \mathbb{N}$.
- **Size of Testing Set**: An integer $\tilde{N} \in \mathbb{N}$.
- **Robustness Parameters**: A collection of nonnegative real numbers $\mathcal{E} \triangleq \{\epsilon_1, \ldots, \epsilon_L\}$.

Outputs:
- Exercise policies $\hat{\mu} \equiv (\hat{\mu}_1, \ldots, \hat{\mu}_T)$ for the stochastic optimal stopping problem (OPT).
- An (unbiased) estimate of the expected reward $J^*(\hat{\mu})$ for the exercise policies.

Procedure:
1. Simulate the validation set $\bar{S} \triangleq \{\bar{x}^1, \ldots, \bar{x}^\bar{N}\}$ and testing set $\tilde{S} \triangleq \{\tilde{x}^1, \ldots, \tilde{x}^\tilde{N}\}$.
2. For each $N \in \mathcal{N}$:
   (a) Simulate a training set $S \triangleq \{x^1, \ldots, x^N\}$.
   (b) For each $\epsilon \in \mathcal{E}$:
      i. Obtain exercise policies $\hat{\mu}_{N,\epsilon}$ by solving the robust optimization problem (RO) constructed from the training set $S$ and robustness parameter $\epsilon$.
      ii. Estimate the expected reward for these exercise policies using the validation set:
         \[
         \bar{J}(\hat{\mu}_{N,\epsilon}) \triangleq \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} g(\tau_{\hat{\mu}_{N,\epsilon}}(\bar{x}^i), \bar{x}^i).
         \]
   (c) If the total computation time of steps (2a)-(2b) reached the allowed budget (or if $N = N_K$), go to step (3) with the current value of $N$.
3. Output the exercise policies which maximized the estimated expected reward:
   \[
   \hat{\mu} \leftarrow \arg \max_{\epsilon \in \mathcal{E}} \bar{J}(\hat{\mu}_{N,\epsilon}),
   \]
   and output an unbiased estimate of their expected reward using the testing set:
   \[
   J^*(\hat{\mu}) \approx \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} g(\tau_{\hat{\mu}}(\tilde{x}^i), \tilde{x}^i).
   \]

Algorithm 1: A procedure for selecting the robustness parameter and number of sample paths.
we recommend solving (RO) over a grid of possible choices for the robustness parameter. Because
the probability distribution is known, we can generate a second set of ‘validation’ sample paths to
select the best choice of the robustness parameter. Specifically, for each choice of the robustness
parameter, one solves the robust optimization problem to obtain exercise policies. The expected
reward of the exercise policies is then estimated using the validation set of sample paths. Finally,
we select the value of the robustness parameter (and the corresponding exercise policies) which
maximizes the average reward with respect to the validation set.

In summary, we have described straightforward and easy-to-implement heuristics for choosing
the parameters of the robust optimization problem. Applying the heuristics and solving the robust
optimization problem yields exercise policies for the stochastic optimal stopping problem, and an
unbiased estimate of the expected reward of these exercise policies can similarly be obtained by
simulating a set of ‘testing’ sample paths (see Figure 1). Because the exercise policies obtained
from the robust optimization problem are feasible for the stochastic optimal stopping problem, the
expected reward of these exercise policies is thus a lower bound on the optimal objective value
of the stochastic optimal stopping problem. Finally, we remark that under a stronger assumption
in which one has the ability to perform conditional Monte Carlo simulation, the exercise policies
obtained from solving the robust optimization can be combined with the method of Andersen and
Broadie (2004) to obtain an upper bound on the optimal objective value of the stochastic optimal
stopping problem.

3. Characterization of Optimal Policies

Like the stochastic optimal stopping problem (OPT), the robust optimization problem (RO)
involves searching over the space of all exercise policies. This renders (RO) an infinite-dimensional
optimization problem, and thus is not readily solvable in its current form. As the key step towards
circumventing this computational challenge, this section characterizes the structure of optimal
policies for the robust optimization problem.

3.1. An Illustrative Example

Before delving into the main result in the following section, we first develop intuition by analyzing
a simple example. Specifically, we consider an instance of a robust optimization problem which is
constructed from two sample paths \( N = 2 \) over a horizon of three periods \( T = 3 \). The values of
the sample paths are given by

\[
x^1 = (x^1_1, x^1_2, x^1_3) = (8, 7, 6) \quad \text{and} \quad x^2 = (x^2_1, x^2_2, x^2_3) = (3, 4, 3).
\]

The state space is the real numbers \( \mathcal{X} = \mathbb{R}^1 \) and the robustness parameter is equal to two \( \epsilon = 2.0 \). In Figure 1a, we illustrate the uncertainty sets constructed around the two sample paths. In
Figure 1b, we show the following choice of exercise policies overlaid onto the two uncertainty sets:

\[
\mu_1(x_1) = \begin{cases} 
\text{STOP}, & \text{if } x_1 \in [3, 7] \cup [9, 10], \\
\text{CONTINUE}, & \text{otherwise}; 
\end{cases}
\]
\[ \mu_2(x_2) = \begin{cases} \text{STOP,} & \text{if } x_2 \in [5,9], \\ \text{CONTINUE,} & \text{otherwise;} \end{cases} \]

\[ \mu_3(x_3) = \begin{cases} \text{STOP,} & \text{if } x_3 \in [1,6], \\ \text{CONTINUE,} & \text{otherwise.} \end{cases} \]

To evaluate the optimality of these exercise policies, we consider the objective value of the robust optimization problem associated with these exercise policies:

\[ \frac{1}{2} \left( \inf_{y \in U^1} g(\tau_\mu(y), x^1) + \inf_{y \in U^2} g(\tau_\mu(y), x^2) \right). \]

For the first uncertainty set, we observe that the adversary may (but is not required to) stop in the first period. Moreover, the condition \( \mu_2(y_2) = \text{STOP} \) is satisfied for all state realizations \( y_2 \in U^2 \equiv [5,9] \). Therefore, for any trajectory \( y \equiv (y_1, y_2, y_3) \) inside the first uncertainty set \( U^1 \), it follows that the stopping time from the exercise policies satisfies \( \tau_\mu(y) \in \{1,2\} \); see Figure 2a. From this reasoning, we conclude that the worst-case reward over the first uncertainty set is given by

\[ \inf_{y \in U^1} g(\tau_\mu(y), x^1) = \min \{ g(1, x^1), g(2, x^1) \}. \]

We can apply a similar reasoning to the second uncertainty set. Indeed, we observe that the adversary may (but is not required to) stop in the first or second period, and the condition \( \mu_3(y_3) = \text{STOP} \) is satisfied for all state realizations \( y_3 \in U^3 \equiv [1,5] \); see Figure 2b. Therefore, the worst-case reward over the second uncertainty set is given by

\[ \inf_{y \in U^2} g(\tau_\mu(y), x^2) = \min \{ g(1, x^2), g(2, x^2), g(3, x^2) \}. \]

Without further specification of the reward function, it is not possible for us to conclude whether the above exercise policies are optimal for the robust optimization problem. On the other hand, our above analysis reveals that the exercise policies can be improved in a potentially significant way. Specifically, we observe that the current exercise policy on the first period, \( \mu_1 \), only provides the adversaries with more opportunities to find a lower reward. In other words, providing the adversary with the option (but not an obligation) to stop on the first period will only lead to the same or smaller worst-case rewards. This implies that exchanging the original exercise policy \( \mu_1 \) with the new exercise policy \( \mu'_1(y_1) \triangleq \text{CONTINUE} \) for all \( y_1 \in X \) will never lead to a worse stopping rule for the robust optimization problem. We further observe that the exercise policies \( \mu_2 \) equals \( \text{STOP} \) for all realizations in \( U^4 \) prevents the adversary from selecting realizations on the third period. Therefore, we can, without any loss of optimality, set the exercise policies on the third period to \( \text{STOP} \) only for realizations in the interval \([1,5]\).

In summary, the above analysis illustrates a pruning procedure for potentially improving (but never worsening) the robust objective value of a collection of exercise policies. The procedure resulted in the following new exercise policies:

\[ \mu'_1(x_1) = \text{CONTINUE} \]

\[ \mu'_2(x_2) = \begin{cases} \text{STOP,} & \text{if } x_2 \in [5,9], \\ \text{CONTINUE,} & \text{otherwise;} \end{cases} \]
Sturt: A nonparametric algorithm for optimal stopping based on robust optimization

(a) Sample paths and their corresponding uncertainty sets
(b) Exercise policies, $\mu \equiv (\mu_1, \mu_2, \mu_3)$

Figure 1  Robust optimization problem with two sample paths over three periods.

(a) Stopping regions for first uncertainty set
(b) Stopping regions for second uncertainty set

Figure 2  Possible stopping regions for the exercise policies $\mu = (\mu_1, \mu_2, \mu_3)$.

Figure 3  New exercise policies $\mu' = (\mu'_1, \mu'_2, \mu'_3)$
\[ \mu'_3(x_3) = \begin{cases} \text{STOP}, & \text{if } x_3 \in [1, 5], \\ \text{CONTINUE}, & \text{otherwise}. \end{cases} \]

The above exercise policies consist of two stopping regions, which are illustrated in Figure 3. A straightforward analysis shows that the robust objective value associated with these new exercise policies is

\[ \frac{1}{2} \left( \inf_{y \in U^1} g(\tau_{\mu'}(y), x^1) + \inf_{y \in U^2} g(\tau_{\mu'}(y), x^2) \right) = \frac{1}{2} \left( g(2, x^1) + \min \left\{ g(2, x^2), g(3, x^2) \right\} \right). \]

Since our goal is to find exercise policies which maximize the robust objective function, and since the reward function is nonnegative, we observe that these new exercise policies \( \mu'_\equiv (\mu'_1, \mu'_2, \mu'_3) \) are at least as good as the original exercise policies \( \mu \equiv (\mu_1, \mu_2, \mu_3) \), no matter the reward function. For some reward functions, the improvement of the new exercise policies is strict; one such example is \( g(1, x) = 1, g(2, x) = 0, g(3, x) = 0 \). In the following section, we formalize the above intuition and generalize the pruning procedure described above.

### 3.2. Main Result

We now present our main result of this section, Theorem 4, which characterizes the structure of optimal policies for the robust optimization problem. Given the uncertainty sets that comprise (RO), we define the following set of exercise policies:

\[ M \triangleq \left\{ \mu \equiv (\mu_1, \ldots, \mu_T) : \mu_t(y_t) = \begin{cases} \text{STOP}, & \text{if } y_t \in \bigcup_{i : \sigma_i = t} U_t^i, \\ \text{CONTINUE}, & \text{if } y_t \notin \bigcup_{i : \sigma_i = t} U_t^i, \end{cases} \right\}. \]

To develop intuition for the above class of policies, we remark that each of the exercise policies \( \mu \in M \) is parameterized by integers \( \sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\} \). As a result, the number of distinct exercise policies in the above set is finite and upper bounded by \( T^N \). A visualization of the stopping regions generated by the collections of exercise policies for the example in §3.1 is found in Figure 4. Our main result is the following:

**Theorem 4.** There exists \( \mu \in M \) which is optimal for (RO).

**Proof.** Consider any arbitrary exercise policies \( \mu \equiv (\mu_1, \ldots, \mu_T) \). We will show that we can always construct an exercise policy \( \mu' \in M \) which has the same or better objective value for the robust optimization problem. Indeed, since the reward function is nonnegative for all trajectories \( y \equiv (y_1, \ldots, y_T) \in \mathcal{X}^T \) and all periods \( t \in \{1, \ldots, T\} \), we assume without loss of generality that the equality \( \mu_T(y_T) = \text{STOP} \) is satisfied for all \( y_T \in \mathcal{X} \). For each sample path \( i \), define the following integer:

\[ \sigma^i \triangleq \min_{t \in \{1, \ldots, T\}} \left\{ t : \mu_t(y_t) = \text{STOP} \text{ for all } y_t \in U^i_t \right\}. \]
Figure 4  The class of exercise policies $\mathcal{M}$ for the robust optimization problem from §3.1.
We observe that the above minimization problem attains an optimal solution $\sigma^i \in \{1, \ldots, T\}$ due to our assumption that $\mu_T(y_T) = \text{STOP}$ for all $y_T \in X$. Given this construction of the integers $\sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\}$, we consider the following new exercise policy for each period $t$:

$$\mu'_t(y_t) \triangleq \begin{cases} \text{STOP}, & \text{if } y_t \in \bigcup_{i : \sigma^i = t} U_i, \\ \text{CONTINUE}, & \text{if } y_t \notin \bigcup_{i : \sigma^i = t} U_i. \end{cases}$$

Note that these new exercise policies $\mu' \equiv (\mu'_1, \ldots, \mu'_T)$ satisfy $\mu' \in M$. For each sample path $i$, it follows from the construction of the new exercise policies that

$$\{y_t \in U'_i : \mu'_t(y_t) = \text{STOP}\} \subseteq \{y_t \in U'_i : \mu_t(y_t) = \text{STOP}\} \quad \text{for all } t. \quad (2)$$

Additionally, the stopping rules associated with the original and new exercise policies satisfy

$$\max_{y \in U'} \tau_y = \max_{y \in U} \tau_y = \sigma^i. \quad (3)$$

It remains to be shown that the new exercise policies $\mu' \equiv (\mu'_1, \ldots, \mu'_T)$ achieve the same or better objective value as $\mu \equiv (\mu_1, \ldots, \mu_T)$ in the robust optimization problem. Indeed,

$$\frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U} g(\tau_y, x_i) \leq \frac{1}{N} \sum_{i=1}^{N} \min_{t \in \{1, \ldots, \sigma^i\}} \left\{ g(t, x') : \text{there exists } y_t \in U'_i \text{ such that } \mu_t(y_t) = \text{STOP} \right\} \leq \frac{1}{N} \sum_{i=1}^{N} \min_{t \in \{1, \ldots, \sigma^i\}} \left\{ g(t, x') : \text{there exists } y_t \in U'_i \text{ such that } \mu'_t(y_t) = \text{STOP} \right\} = \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U} g(\tau'_y, x_i),$$

where the two equalities follow from line (3), and the inequality follows from line (2). Since $\mu \equiv (\mu_1, \ldots, \mu_T)$ was chosen arbitrarily, our proof is complete. \qed


Theorem 4 implies that we can, without any loss of optimality, restrict the robust optimization problem (RO) to searching over exercise policies in the class $M$. In other words, (RO) can be transformed from an optimization problem over an infinite-dimensional space of exercise policies into a finite-dimensional optimization problem over integer decision variables $\sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\}$. To formulate an equivalent optimization problem over these integers, let the following constants be precomputed for all periods $t \in \{1, \ldots, T\}$ and sample paths $i, j \in \{1, \ldots, N\}$:

$$v^ij \triangleq \begin{cases} g(t, x'), & \text{if } U'_i \cap U'_j \neq \emptyset, \\ \infty, & \text{otherwise}. \end{cases}$$

It then follows from Theorem 4 that the optimal objective value of the robust optimization problem (RO) is equal to the optimal objective value of the following optimization problem:

\[^3\] Under the construction of the uncertainty sets from §2.2, and provided that $X = \mathbb{R}^d$, we observe that the constants $v^ij$ can be precomputed in a total of $O(N^2Td)$ computation time.
Proposition 1. The robust optimization problem (RO) is equivalent to
\[ \begin{array}{c}
\text{maximize} \\ \sigma^1, ..., \sigma^N \in \{1, ..., T\}
\end{array} \frac{1}{N} \sum_{i=1}^{N} \min_{j: \sigma^j \leq \sigma^i} v^{ij}, \quad \text{(IP)} \]

In summary, by solving the optimization problem (IP), we obtain the optimal objective value of (RO). The optimal solution to (IP) can further be transformed into an optimal policy for (RO) through the transformation described in §3.2. Through this optimization problem over integer decision variables, the remainder of the section establishes the computational tractability and develops exact and approximation algorithms for solving (RO).

4.1. Exact Algorithms

We begin our discussion by considering the case of (IP) when there are two periods. Optimal stopping problems with two periods has been studied in the literature as a testbed for understanding the complexity of solving optimal stopping problems, e.g., Glasserman and Yu (2004). We defer the proof of the following result to the end of §4.2.

Proposition 2. (IP) can be solved in polynomial-time when \( T = 2 \).

Continuing with our discussion on the theoretical tractability of (IP), we next consider problems with three or more periods. The following negative result shows that the computational tractability of (IP) in the case of two periods (Proposition 2) does not generally extend to optimal stopping problems with three or more periods. The proof of the following result consists of a reduction from MIN-2-SAT, which is shown to be strongly NP-hard by Kohli et al. (1994).

Theorem 5. (IP) is strongly NP-hard for any fixed \( T \geq 3 \).

Motivated by the above computational complexity result, we proceed to develop an exact reformulation of (IP) as a zero-one bilinear program. The exact reformulation, presented below in Proposition 3, is valuable for two reasons. First, the exact reformulation will provide the foundation for the tractable lower-bound approximation of the robust optimization problem in §4.2. Second, for small robust optimization problems, the exact reformulation can be reformulated as a mixed-integer linear optimization problem and solved by commercial software, and thus provides a useful benchmark for numerically evaluating the suboptimality of the lower bound.

Before presenting the exact reformulation, we first introduce the key decision variables and notation. For each sample path \( i \) and period \( t \), we let \( b^t_i \in \{0, 1\} \) denote a binary decision variable which satisfies \( b^t_i = 1 \) if and only if \( \sigma^i = t \). Hence, provided that the equality \( \sum_{t=1}^{T} b^t_i = 1 \) is satisfied for each sample path \( i \), these decision variables have a one-to-one mapping with exercise policies via the formula \( \sigma^i = \sum_{t=1}^{T} t b^t_i \). Next, for each sample path \( i \), we define the following set:

\[ K^i \triangleq \left\{ \kappa: \text{there exists } \sigma^1, ..., \sigma^N \in \{1, ..., T\} \text{ such that } \min_{j: \sigma^j \leq \sigma^i} v^{ij} = \kappa \right\} \cup \{0\} \]

\[ = \{ \kappa: \text{there exists } t \in \{1, ..., T\} \text{ such that } v^t_i = \kappa \} \cup \{0\}. \]
Moreover, there exists an optimal solution to (BP). Let the elements of \( K^i \) be indexed in ascending order, \( \kappa^i_1 < \cdots < \kappa^i_{|K^i|} \), and let \( L_i \in \{1, \ldots, |K^i|\} \) be defined as the unique index which satisfies \( v^i_{\kappa^i} = \kappa^i_{L_i} \). Using the aforementioned notation, we obtain the following exact reformulation of the robust optimization problem.

**Proposition 3.** (IP) is equivalent to the following zero-one bilinear program:

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{\ell=1}^{L_i-1} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) b^i_{\ell}(1 - w^i_{\ell t}) \\
\text{subject to} & \quad w^i_{t,\ell} \leq w^i_{t+1,\ell} \quad \text{for all } i \in \{1, \ldots, N\}, t \in \{1, \ldots, T-1\}, \ell \in \{1, \ldots, |K^i|\} \\
& \quad w^i_{t,\ell} \leq w^i_{t,\ell+1} \quad \text{for all } i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}, \ell \in \{1, \ldots, |K^i| - 1\} \\
& \quad b^i_{\ell} \leq w^i_{t,\ell+1} \quad \text{for all } i \in \{1, \ldots, N\}, t \in \{1, \ldots, T-1\} \\
& \quad b^i_{\ell} \leq w^i_{t,\ell} \quad \text{for all } i \neq j \in \{1, \ldots, N\} \text{ and } t \in \{1, \ldots, T\} \text{ such that } v^i_{\ell j} = \kappa^i_{\ell} \\
& \quad b^i_{\ell} \in \{0, 1\} \quad \text{for all } i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\} \\
& \quad w^i_{t,\ell} \in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}, \ell \in \{1, \ldots, |K^i|\}.
\end{align*}
\]

Moreover, there exists an optimal solution to (BP) where \( \sum_{t=1}^{T} b^i_{\ell} = 1 \) for all \( i \in \{1, \ldots, N\} \).

We note that the optimization problem (BP) does not explicitly require that each of the decision variables \( w^i_{t,\ell} \) be equal to zero or one. Nonetheless, we readily observe by inspection that there is always an optimal solution to (BP) in which the decision variables satisfy \( w^i_{t,\ell} \in \{0, 1\} \). For this reason, (BP) is referred to as a zero-one bilinear program. We also note that (BP) can be reformulated as a mixed-integer linear optimization problem by introducing auxiliary decision variables; see Remark EC.1 in Appendix D for more details.

Let us reflect on what the above Proposition 3 aims to achieve. As discussed previously, (IP) provided a formulation for finding an optimal solution to the robust optimization problem as an optimization problem over integer decision variables, as enabled by the characterization of optimal policies from §3. However, (IP) was not computationally viable from the perspective of standard optimization solvers, as its objective function involved decision variables in the subscripts of vectors. To remedy this, Proposition 3 shows that (IP) can be reformulated as a bilinear program with \( \mathcal{O}(NT) \) binary decision variables, \( \mathcal{O}(NT^2) \) continuous decision variables, and \( \mathcal{O}(NT(N + T)) \) constraints. Moreover, we observe that the constraints of the optimization problem are totally unimodular, which implies that every extreme point of the polyhedron defined by the constraints of (BP) is integral (Conforti et al. 2014, §4.2). In the next section, we harness this reformulation to develop a polynomial-time algorithm for approximating the robust optimization problem.

### 4.2. Approximation Algorithm

Building upon the exact reformulation (BP) of the robust optimization problem from the previous section, we now develop and analyze a tractable heuristic algorithm for approximating the robust optimization problem. Our development of the algorithm is split into three steps.
Step 1: We begin by developing an intermediary lower bound on the robust optimization problem. Recall that $T$ is the number of periods in the optimal stopping problem. For each sample path $i$, let $T^i \triangleq \arg \max_{t \in \{1,...,T\}} v_{ii}^t$ be defined as the period in which the sample path $i$ achieves its maximum reward, and if there are multiple optimal solutions, arbitrarily choose the optimal solution which is smallest. To simplify our discussion, let us define $T_i \triangleq \{T^i\} \cup \{T\}$ as the set which contains the period in which sample path $i$ achieves its maximum reward as well as the last period of the optimal stopping period. With this notation, it follows immediately from Proposition 1 that the optimal objective value of the following intermediary optimization problem (LB-1) is a lower bound on the optimal objective value of the robust optimization problem (RO):

$$\max_{\sigma^i \in T_i \forall i} \frac{1}{N} \sum_{i=1}^{N} \min_{j: \sigma^j \leq \sigma^i} v_{ij}^{\sigma_j}. \quad (LB-1)$$

Step 2: We next develop an exact reformulation of the intermediary optimization problem (LB-1) as a zero-one bilinear program. This construction, presented in Lemma 1, will use similar notation and decision variables as those developed for Proposition 3 in the previous §4.1. Indeed, let $b^1, \ldots, b^N \in \{0,1\}$ be binary decision variables which encode, for each sample path $i$, whether the integer $\sigma^i \in T_i$ satisfies $\sigma^i = T^i$ or $\sigma^i = T$. More precisely, the relationship between the decision variables and the exercise policies is the following:

$$b^i = \begin{cases} 1, & \text{if } \sigma^i = T^i, \\ 0, & \text{if } \sigma^i \neq T^i. \end{cases} \quad (4)$$

We also define the following sets for each sample path $i$:

$$S_1^i \triangleq \{ j \in \{1, \ldots, N\} : T_j \leq T^i \text{ and } v_{ij}^{T_j} \leq v_{ii}^{T^i} \};$$

$$S_2^i \triangleq \{ j \in \{1, \ldots, N\} : T_j > T^i \text{ and } v_{ij}^{T_j} \leq v_{ii}^{T_i} \};$$

$$K_i \triangleq \{ \kappa : \text{there exists } \sigma^1 \in T^1, \ldots, \sigma^N \in T^N \text{ such that } \min_{j: \sigma^j \leq \sigma^i} v_{ij}^{\sigma_j} = \kappa \} \cup \{0\}$$

$$= \{ \kappa : \text{there exists } j \in S_1^i \cup S_2^i \text{ such that } \kappa = v_{ij}^{T_j} \} \cup \{v_{ii}^{T_i}\} \cup \{0\}.$$
Lemma 1. (LB-1) is equivalent to the following zero-one bilinear program:

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{\ell=1}^{(|K^i|-1)} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) b^i (1 - w^i_{\ell}) + \sum_{\ell=1}^{L^i-1} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) (1 - b^i) (1 - u^i_{\ell}) \right) \\
\text{subject to} & \quad w^i_{\ell} \leq u^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 1\} \\
& \quad w^i_{\ell+1} \leq w^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, |K^i| - 2\} \\
& \quad u^i_{\ell+1} \leq u^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 2\} \\
& \quad b^i \leq u^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \\
& \quad b^i \leq w^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \text{ and } j \in S^i_1 \text{ such that } v^{ij}_T = \kappa^i_T \\
& \quad b^i \leq u^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \text{ and } j \in S^i_2 \text{ such that } v^{ij}_T = \kappa^i_T \\
& \quad b^i \in \{0, 1\} \quad \text{for all } i \in \{1, \ldots, N\} \\
& \quad w^i_{\ell} \in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, |K^i| - 1\} \\
& \quad u^i_{\ell} \in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 1\}.
\end{align*}
\]

Step 3: We finally develop the main result of this section, Theorem 6, which presents a lower-bound approximation of the robust optimization problem. The proof of the theorem consists of replacing the bilinear objective function of (BP-1) with a lower-bound linear function.

Theorem 6. The optimal objective value of the following mixed-integer linear optimization problem yields a lower bound on the optimal objective value of (RO):

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{\ell=1}^{(|K^i|-1)} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) (b^i - w^i_{\ell}) + \sum_{\ell=1}^{L^i-1} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) (1 - u^i_{\ell}) \right) \\
\text{subject to} & \quad w^i_{\ell} \leq u^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 1\} \\
& \quad w^i_{\ell+1} \leq w^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, |K^i| - 2\} \\
& \quad u^i_{\ell+1} \leq u^i_{\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 2\} \\
& \quad b^i \leq u^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \\
& \quad b^i \leq w^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \text{ and } j \in S^i_1 \text{ such that } v^{ij}_T = \kappa^i_T \\
& \quad b^i \leq u^i_1 \quad \text{for all } i \in \{1, \ldots, N\} \text{ and } j \in S^i_2 \text{ such that } v^{ij}_T = \kappa^i_T \\
& \quad b^i \in \{0, 1\} \quad \text{for all } i \in \{1, \ldots, N\} \\
& \quad w^i_{\ell} \in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, |K^i| - 1\} \\
& \quad u^i_{\ell} \in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, \ell \in \{1, \ldots, L^i - 1\}.
\end{align*}
\]

Combining the above steps, we have arrived at the formulation of an optimization problem (LB), which provides a lower-bound approximation of the robust optimization problem (RO). The optimization problem (LB), in a nutshell, is the main contribution of this section. Specifically, our heuristic for approximating the robust optimization problem (RO) consists of solving the optimization problem (LB) and then converting the optimal solution to (LB) to a Markovian stopping rule by applying the transformation from line (4). This proposed heuristic for the robust optimization problem is particularly convenient from an implementation standpoint: as a mixed-integer
linear optimization problem, (LB) can be easily formulated and solved directly by commercial optimization solvers such as CPLEX or Gurobi.

In comparison to the exact reformulation of the robust optimization problem from the previous section (Proposition 3), the optimization problem (LB) is significantly more tractable from a computational perspective. In particular, the following Proposition 4 shows that this optimization problem can be solved with a computation time that is polynomial in both the number of simulated sample paths as well as the number of time periods. Such a tractability guarantee is ultimately important from a practical perspective: indeed, in the following §5, we will present numerical experiments which show that (LB) can be solved to optimality in seconds on realistic problem sizes with over fifty periods and thousands of sample paths.

**PROPOSITION 4.** (LB) can be solved in $O(N^2T(N+T))$ time.

In certain settings, the optimality gap between the optimization problem (LB) and the robust optimization problem (RO) can be shown to equal zero. To establish these guarantees, we first develop a general upper-bound on the optimality gap of the optimization problem (LB) relative to the intermediary optimization problem (LB-1). In the following, we let $J_{LB}$ and $J_{LB-1}$ denote the optimal objective values of these two problems.

**PROPOSITION 5.** If $\sigma_1 \in T^1, \ldots, \sigma_N \in T^N$ is an optimal solution to (LB-1), then

$$J_{LB-1} \geq J_{LB} \geq J_{LB-1} - \frac{1}{N} \sum_{i: \sigma_i \neq T^i} \max \left\{ v_{T^i} - \min_{j \in S^i_i \mid \sigma_j = T^j} v_{T^j}, 0 \right\}.$$  

Speaking intuitively, we see from the above proposition that the optimality gap between (LB-1) and (LB) will be small when there exists an optimal solution to the former that satisfies either $\sigma^i = T^i$ or $\sigma^i = T$ for many of the sample paths. Harnessing the above upper bound, we obtain the following settings in which the optimal objective value of the optimization problem (LB) is guaranteed to equal to the optimal objective value $J^{RO}$ of the robust optimization problem (RO).

**COROLLARY 1.** If $\sigma_1 = T, \ldots, \sigma_N = T$ is an optimal solution for (IP), then $J^{RO} = J^{LB}$.

**COROLLARY 2.** If $\sigma_1 = T^1, \ldots, \sigma_N = T^N$ is an optimal solution for (IP), then $J^{RO} = J^{LB}$.

**COROLLARY 3.** If $T = 2$, then $J^{RO} = J^{LB}$.

In particular, we observe that the combination of Corollary 3 and Proposition 4 implies that the robust optimization problem (RO) can be solved in $O(N^3)$ time for any optimal stopping problem with two periods, which constitutes our proof of Proposition 2 from §4.1.

Ultimately, the practical value of (LB) lies in its performance in the context of optimal stopping applications. In the following section, we provide numerical evidence that (LB) can find high-quality stopping rules for the stochastic optimal stopping problem (OPT).
5. Numerical Experiments

In this section, we perform numerical experiments to compare the robust optimization method and two state-of-the-art benchmarks from the literature (Longstaff and Schwartz 2001, Ciocan and Mišić 2020). The benchmarks serve as representatives of two classes of approximation methods for stochastic optimal stopping problems (approximate dynamic programming and parametric exercise policies) which, similarly as the robust optimization method, only require the ability to simulate sample paths of the entire sequence of random states. All experiments were conducted on a 2.6 GHz 6-Core Intel Core i7 processor with 16 GB of memory. The robust optimization method was implemented in the Julia programming language and solved using the JuMP library and Gurobi optimization software.

5.1. A Simple Non-Markovian Problem

To demonstrate the value of robust optimization method, we begin by investigating a simple, one-dimensional stochastic optimal stopping problem with a non-Markovian probability distribution. The optimal stopping problem of consideration involves a state space which is equal to the real numbers \( (X = \mathbb{R}^T) \) and a reward function of \( g(t, x) = x_t \) in each period \( t \in \{1, \ldots, T\} \). For any fixed duration \( \Delta \in \mathbb{N} \), the joint probability distribution of the stochastic process is given by

\[
x_t \sim \text{Uniform}[0, 1] + \frac{2\theta}{T} \mathbb{1}\{\theta \leq t \leq \theta + \Delta\}
\]

for all \( t = 1, \ldots, T \), where the random parameter \( \theta \sim \text{Uniform}\{1, 2, \ldots, T - \Delta\} \) is selected once per sample path and is unobserved. Simulated sample paths of this non-Markovian stochastic process \( x = (x_1, \ldots, x_T) \) are visualized in Figure 5. We perform numerical experiments on the following methods:

- **Robust Optimization (RO):** The robust optimization method is used here to approximate \( \text{OPT} \) over the non-Markovian stochastic process \( x \equiv (x_1, \ldots, x_T) \), and thus aims to find the best Markovian stopping rules for the stochastic optimal stopping problem. The method was run with robustness parameters \( \epsilon \in \{0, 0.01, \ldots, 0.3\} \) and solved approximately using the algorithm from Theorem 6.

- **Least-Squares Regression (LS):** We implement the method of Longstaff and Schwartz, which employs least-squares regression to approximate the continuation value function (i.e., the expected reward from not stopping) in each period using backwards recursion. To apply this method, we first transform each non-Markovian sample path \( x^i \equiv (x^i_1, \ldots, x^i_T) \) into an augmented Markovian sample path of the form \( X^i \equiv (X^i_1, \ldots, X^i_T) \) by adding the full history into the state in each period: \( X^i_t \triangleq (x^i_1, \ldots, x^i_t, 0, \ldots, 0) \). The regression step requires a specification of basis functions, and we consider the following relatively generic categories of basis functions.

  - **Full-History:** This category of basis functions uses the entire vector \( X^i_t \) of states observed up to that point. The basis functions that we consider in this category are ONE (the constant function 1), PRICES (the states observed up to that point, \( X^i_t \in \mathbb{R}^T \)), and PRICES2 (the product of each pair of states observed up to that point, \( X^i_t(X^i_t)^\top \in \mathbb{R}^{T \times T} \)).
Markovian: This category of basis functions uses only the current state $x_i$ in each period. The purpose of considering these basis functions is to analyze the performance when, like the robust optimization method, the method of Longstaff-Schwartz is restricted to stopping rules which depend only on the current state in each period. We consider basis functions based on the Laguerre polynomials, where $\text{LAGUERRE-K}$ is the polynomial $\sum_{\ell=0}^{k} \binom{k}{\ell} \frac{(-1)^{\ell}}{\ell!} (x_i)^{\ell}$.

- **Tree Method (Tree):** The method of Ciocan and Mišić approximates the stochastic optimal stopping problem (OPT) by restricting the space of exercise policies to decision trees. Like the robust optimization method, the tree approach is used to find Markovian stopping rules for the non-Markovian optimal stopping problem (OPT). We apply the method with the same information as the robust optimization method at each time period (the current state and the time period) and with a splitting parameter of 0.005.

In our computational experiments, we consider a time horizon of $T = 50$ stopping periods with a duration parameter of $\Delta = 5$. With the exceptions of “RO” and “LS Full-History (one, prices, prices2)”, all methods were run using simulated training datasets of sizes $N \in \{10^2, 10^2.1, \ldots, 10^4.9, 10^5\}$. The two remaining methods, “RO” and “LS Full-History (one, prices, prices2)”, were run on simulated training datasets of sizes $N \in \{10^2, 10^2.1, \ldots, 10^2.9, 10^3\}$. The robustness parameters in the “RO” method were selected using a validation set of size $\bar{N} = 10^3$; see §2.5 for more details. All methods were evaluated on a common and independent testing dataset of $\tilde{N} = 10^5$ sample paths, and experiments were repeated over 10 replications.

Figures 5 and 6 visualize the policies found by the robust optimization method, the expected rewards of the stopping rules obtained by the various methods, and the computation times of the various methods. The results of this experiment show that the robust optimization method outperforms the other approaches, producing stopping rules with an expected reward of approximately 1.62 from training datasets of $N = 10^3$ simulated sample paths. We reflect below on the main differences between our robust optimization method and the benchmark methods.

For the least-squares regression method with full-history, selecting a good choice of basis functions is found to be a first-order challenge. For example, the basis functions (ONE, PRICES) turn out to be insufficiently rich to provide an accurate approximation of the continuation function with the full state history. When the number of sample paths is sufficiently large, we expect that the basis functions (ONE, PRICES, PRICES2) should provide a better approximation of the stochastic optimal stopping problem. However, the computational cost resulting from this rich class of basis functions precluded its practicality on sufficiently large training datasets. In contrast, the robust optimization method only searches for Markovian stopping rules, and in doing so, has a reduction in sample complexity. In addition to achieving a significantly better expected reward in this example, the Markovian stopping rules found by the robust optimization method are considerably more interpretable than those which have full-history dependance, as illustrated in Figure 5.

Compared to the tree method, the complex structure of the Markovian stopping rules found by the robust optimization method demonstrates the value of algorithms that do not impose parametric restrictions on the exercise policies. In theory, decision trees with sufficient depth are capable of approximating the best Markovian stopping rules to the NMOS problem to arbitrary accuracy (Ciocan and Mišić 2020, Theorem 2). However, the greedy heuristic that is proposed by Ciocan and Mišić (2020) to efficiently optimize over decision trees is not able to find a good
Sturt: A nonparametric algorithm for optimal stopping based on robust optimization

Figure 5  The left figure shows simulated sample paths of the non-Markovian stochastic process with $T = 50$ and $\Delta = 5$. The right figure shows the exercise policies obtained from solving the robust optimization problem constructed from a training dataset of $N = 10^3$ simulated sample paths and with robustness parameter $\epsilon = 0.1$.

Figure 6  The left figure shows the expected reward of stopping rules obtained by the various methods, and the right figure shows the total computation time. The results for the robust optimization method are shown using the algorithm from Theorem 6, the displayed computation time is the total time to construct and approximately solve the robust optimization problems over all choices of the robustness parameter $\epsilon \in \{0, 0.01, \ldots, 0.3\}$, and the expected rewards are displayed for the robustness parameter that is chosen using a validation set of size $\bar{N} = 10^3$ for each choice of $N$. The expected rewards for “LS, Full-History (one,prices,prices2)” were below 1.20 for all replications and $N \in \{10^2, 10^{2.1}, \ldots, 10^{2.9}, 10^3\}$, and this method is omitted from the left figure for clarity. All results for all methods are averaged over 10 replications.
approximation of the optimal Markovian stopping rule in this example, even as the number of sample paths is large. We note that this issue is not unique to the decision trees *per se*; the computational intractability of optimizing over parametric spaces of Markovian stopping rules has resulted in heuristics for other settings as well (Glasserman 2013, §8.2). This example provides evidence that our robust optimization method, in conjunction with the proposed approximation from §4.2, can yield high-quality Markovian stopping rules to complex (low-dimensional) non-Markovian optimal stopping problems.

Like the robust optimization and tree methods, “LS, Markovian” aims to find Markovian stopping rules for the non-Markovian optimal stopping problem. In particular, we observe that “LS, Markovian” was implemented with basis functions of Laguerre polynomials with a rather large degree.⁴ The expected reward of “LS, Markovian (Laguerre 0-15)” may thus be interpreted as an estimate for the best Markovian stopping rule that one could achieve using dynamic programming. However, the optimality of backwards induction no longer holds for exercise policies that depend on the current state when the stochastic process is non-Markovian. This explains the inferior performance of this method in this example and motivates the use of methods for addressing non-Markovian optimal stopping problems that optimize directly over the exercise policies in all periods simultaneously. A further discussion on the limitations of dynamic programming in finding Markovian stopping rules for NMOS problems can be found in Appendix A.

In Figure 7, we compare the (in-sample) robust objective values and the (out-of-sample) expected rewards of stopping rules obtained by our lower-bound approximation (LB) and our exact reformulation (BP) of the robust optimization problem. The comparison in Figure 7 is performed on a smaller instance where \(T = 20, \Delta = 2\), and the robust optimization problem is constructed from a training set of \(N = 100\) simulated sample paths. We observe the gap in objective values between the two algorithms is relatively small across choices of the robustness parameter, and the gap is equal to zero when the robustness parameter is set equal to zero. Moreover, the expected rewards of stopping rules obtained from these two algorithms were similarly relatively close across choices of the robustness parameter. We view these results as promising as they suggest, at least for the present example, that the significantly more tractable optimization problem (LB) can provide a close approximation of the robust optimization problem (RO).

In summary, the experiments from this section reveal our first setting in which the robust optimization method is attractive. In non-Markovian optimal stopping problems, there can exist high-quality stopping rules which are Markovian. However, finding such stopping rules is a non-trivial challenge, as the best Markovian stopping rules will not necessarily exhibit a simple structure, and cannot in general be found using dynamic programming. The robust optimization method thus provides a practical nonparametric approach for finding them. In view of these observations, the next section considers a benchmark optimal stopping problem from the options pricing literature.

### 5.2. Pricing Multi-Dimensional Barrier Options

Building upon the previous section, we next consider the well-studied problem of pricing discretely-monitored barrier call options over multiple assets. These are stochastic optimal stopping problems

---

⁴ We were unable to run the “LS, Markovian (Laguerre 0-K)” for degrees \(K > 15\) due to the numerical precision required to encode the coefficients in the basis functions.
Figure 7  The left figure shows the robust objective value of stopping rules obtained from solving the lower-bound approximation (LB) and the exact reformulation (BP) of the robust optimization problem (RO). The right figure shows the expected rewards of the stopping rules obtained by these two algorithms. The expected rewards are estimated by evaluating the stopping rules on a testing dataset of $10^5$ simulated sample paths.

in which the random state at each period, $\xi_t \in \Xi \equiv \mathbb{R}^d$, is comprised of $d$ non-dividend paying assets. Provided that the option has not been ‘knocked-out’, the reward from stopping on period $t$ is essentially an increasing function of the maximum value of the current assets, $\max_{a \in \{1, \ldots, d\}} \xi_{t,a}$. However, if the maximum value of the assets exceeds a prespecified barrier, the option becomes ‘knocked-out’ and the reward becomes equal to zero for the remainder of the time horizon. In this section, we compare various methods on several instances of these stochastic optimal stopping problems in which the barrier threshold changes over time (Kunitomo and Ikeda 1992).

To formalize the problem setting, let the components of the state vector $\xi_t \in \Xi \equiv \mathbb{R}^d$ be referenced through the notation $\xi_t \equiv (\xi_{t,1}, \ldots, \xi_{t,d})$, where $\xi_{t,a} \in \mathbb{R}$ represents the value of asset $a$ at exercise opportunity $t \in \{1, \ldots, T\}$. The exercise opportunities are evenly spaced over a calendar of $Y$ years and thus, defining $\lambda \triangleq Y/T$, it follows that exercise opportunity $t$ occurs at the calendar time $\lambda t$. Let $r \in [0, 1)$ be the annualized discount rate, $K \geq 0$ be the strike price, $B(t) \triangleq B_0 e^{\delta \lambda t}$ be the barrier threshold on exercise opportunity $t$, and $\xi \equiv (\xi_1, \ldots, \xi_T)$ be the sequence of random states. With this notation, the reward function of this stochastic optimal stopping problem is defined as

$$g(t, \xi) = \begin{cases} e^{-r \lambda t} \max \left\{ 0, \max_{a \in \{1, \ldots, d\}} \xi_{t,a} - K \right\}, & \text{if } \max_{a \in \{1, \ldots, d\}} \xi_{s,a} \leq B(s) \text{ for all } s \in \{1, \ldots, t\}, \\ 0, & \text{otherwise.} \end{cases}$$

Under the standard Black-Scholes setup, the sequence of random states $\xi \equiv (\xi_1, \ldots, \xi_T)$ obeys a multidimensional geometric Brownian motion where each asset $a \in \{1, \ldots, d\}$ has an initial value of $\bar{x}_a$, drift equal to the risk-free rate $r$, and annualized volatility equal to $\sigma_a$. The value of asset $a$
at exercise opportunity \( t \) is thus given by \( \xi_{t,a} \triangleq \bar{x}\exp\left(-\frac{\sigma^2}{2}\right)^{at} + \sigma_{xW,a}^{\text{max}} \), where each \( W_{t,a} \) is a standard Brownian motion process and the instantaneous correlation of \( W_{t,a} \) and \( W_{t,a'} \) is equal to \( \rho_{a,a'} \). We perform numerical experiments on the following methods:

- **Robust Optimization (RO):** The robust optimization method is used here to approximate (OPT) over the projected stochastic process \( x \equiv (x_1, \ldots, x_T) \), where \( x_t \triangleq \max_{a \in \{1, \ldots, d\}} \xi_{t,a} \) denotes the maximum value of the assets in each exercise opportunity \( t \). Note that this stochastic process does not include information on whether the option has been knocked-out. The robust optimization method is run with robustness parameters \( \epsilon \in \{0\} \cup \{0.01, \ldots, 0.09\} \cup \{0.1, \ldots, 0.9\} \cup \{1, \ldots, 10\} \) and solved approximately using the algorithm from Theorem 6.

- **Least-Squares Regression (LS):** The method of Longstaff and Schwartz is applied to stochastic optimal stopping problem (OPT) over the Markovian stochastic process \( X \equiv (X_1, \ldots, X_T) \), where the state at each exercise opportunity \( X_t \triangleq (\xi_{t,1}, \ldots, \xi_{t,d}, q_t) \) consists of the values of the assets at the current exercise opportunity as well as an indicator variable which equals one if the option has been knocked-out: \( q_t \triangleq 1 - \mathbb{I}\{\max_{a \in \{1, \ldots, d\}} \xi_{s,a} \leq B(s) \} \) for all \( s \in \{1, \ldots, t\} \). We consider the following basis functions:
  - **ONE:** The constant function, 1.
  - **KOIND:** An indicator variable for whether the option was knocked out, \( q_t \).
  - **PRICES:** The values of the assets, \( \xi_{t,1}, \ldots, \xi_{t,d} \).
  - **PRICESKO:** The assets multiplied by the indicator variable, \( \xi_{t,1}(1-q_t), \ldots, \xi_{t,d}(1-q_t) \).
  - **MAXPRICE:** The maximum asset value, \( x_t \triangleq \max_{a \in \{1, \ldots, d\}} \xi_{t,a} \).
  - **PAYOFF:** The reward of exercising the option, \( (1-q_t) e^{-rT} \max\{0, x_t - K\} \).

- **Tree Method (Tree):** The method of Ciocan and Mišić approximates the stochastic optimal stopping problem (OPT) by restricting the space of exercise policies to decision trees. Like the robust optimization method, the tree approach is used to find Markovian stopping rules for the optimal stopping problem over the projected stochastic process \( x \equiv (x_1, \ldots, x_T) \). We apply the method with the same information as the robust optimization method at each time period (the current state \( x_t \) and the time period \( t \)) and with a splitting parameter of 0.005.

Our experiments and parameter settings closely parallel those of Ciocan and Mišić (2020, §5), albeit with two primary differences. First, our experiments consider a barrier threshold that changes as a function of time. Second, we perform experiments in which the underlying assets have symmetrical as well as asymmetrical annualized volatilities (i.e., \( \sigma_a \neq \sigma_{a'} \) for \( a \neq a' \)). Our motivations behind these differences are to compare the various methods in settings in which the best Markovian stopping rules may not have a simple structure and in which the low-dimensional projections \( x_1, \ldots, x_T \) are not sufficient statistics for the optimal stopping rules (Broadie and Detemple 1997, §7). In other words, our experiments aim to evaluate the performance of the proposed robust optimization method in settings where the best Markovian stopping rules are not guaranteed to be optimal stopping rules for the optimal stopping problem. Similarly as Ciocan and Mišić (2020, §5), we perform experiments on various numbers of assets \( \{d \in \{8, 16, 32\}\} \) and various initial prices for the assets \( \{\bar{x} \in \{90, 100, 110\}\} \). In our implementation of LS, we experimented on a variety of combinations of basis functions and, for the sake of brevity, report the results for a subset of combinations which included those with the best performance. The expected rewards of the stopping rules obtained by the various methods was estimated using a common and independent testing dataset of \( N = 10^5 \) sample paths, and all experiments were repeated over 10 replications.
Tables 1 and 2 present the expected rewards of stopping rules obtained from the various methods in experiments with symmetrical and asymmetrical annualized volatilities. Across the parameter settings, the robust optimization method yields stopping rules with either the best or close-to-best expected reward relative to the alternative state-of-the-art methods. These results are viewed as particularly encouraging given the practical significance of this class of options pricing problems. In Figure 8, we present visualizations of the Markovian stopping rules produced by the robust optimization method. These visualizations provide interpretability to the stopping rules found by the robust optimization, which is not possible for the LS method due to the high-dimensional state space. Further visualizations of the stopping rules obtained by the robust optimization method under additional problem parameters are provided in Appendix E.
Table 2  Barrier Option (Asymmetric).

<table>
<thead>
<tr>
<th>d</th>
<th>Method</th>
<th>Basis functions</th>
<th>Initial Price</th>
<th>Initial Price</th>
<th>Initial Price</th>
<th># of Sample Paths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$x = 90$</td>
<td>$x = 100$</td>
<td>$x = 110$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>RO</td>
<td>maxprice</td>
<td>57.59 (0.18)</td>
<td>67.29 (0.27)</td>
<td>60.30 (0.40)</td>
<td>$2 \times 10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, pricesKO, KOind, payoff</td>
<td>56.44 (0.11)</td>
<td>65.52 (0.16)</td>
<td>68.34 (0.09)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, pricesKO, payoff</td>
<td>56.54 (0.12)</td>
<td>65.74 (0.17)</td>
<td>68.31 (0.08)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>payoff, KOind, pricesKO</td>
<td>56.44 (0.11)</td>
<td>65.52 (0.16)</td>
<td>68.34 (0.09)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>pricesKO, payoff</td>
<td>56.51 (0.12)</td>
<td>65.75 (0.18)</td>
<td>68.31 (0.08)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, prices, payoff</td>
<td>55.56 (0.09)</td>
<td>61.92 (0.10)</td>
<td>60.07 (0.12)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one</td>
<td>45.80 (0.06)</td>
<td>52.84 (0.07)</td>
<td>48.56 (0.08)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, KOind, prices</td>
<td>52.33 (0.09)</td>
<td>61.85 (0.10)</td>
<td>63.88 (0.08)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, prices</td>
<td>48.49 (0.11)</td>
<td>53.03 (0.09)</td>
<td>50.83 (0.06)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>one, pricesKO</td>
<td>51.67 (0.08)</td>
<td>61.49 (0.10)</td>
<td>63.14 (0.09)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>LS</td>
<td>maxprice, KOind, pricesKO</td>
<td>52.63 (0.09)</td>
<td>61.97 (0.10)</td>
<td>64.14 (0.12)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>8</td>
<td>Tree</td>
<td>maxprice, time</td>
<td>55.74 (0.13)</td>
<td>63.17 (0.66)</td>
<td>63.89 (0.48)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>RO</td>
<td>maxprice</td>
<td>72.59 (0.27)</td>
<td>77.45 (0.26)</td>
<td>70.04 (0.37)</td>
<td>$2 \times 10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, pricesKO, KOind, payoff</td>
<td>70.94 (0.08)</td>
<td>76.10 (0.10)</td>
<td>70.59 (0.13)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, pricesKO, payoff</td>
<td>71.06 (0.08)</td>
<td>76.14 (0.08)</td>
<td>70.04 (0.12)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>payoff, KOind, pricesKO</td>
<td>70.94 (0.08)</td>
<td>76.10 (0.10)</td>
<td>70.59 (0.13)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>pricesKO, payoff</td>
<td>71.06 (0.08)</td>
<td>76.14 (0.08)</td>
<td>70.04 (0.12)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, prices, payoff</td>
<td>67.57 (0.08)</td>
<td>67.07 (0.16)</td>
<td>56.36 (0.14)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, KOind, prices</td>
<td>59.09 (0.05)</td>
<td>56.20 (0.11)</td>
<td>45.55 (0.05)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, prices</td>
<td>66.78 (0.06)</td>
<td>71.06 (0.14)</td>
<td>63.54 (0.20)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, pricesKO</td>
<td>59.35 (0.07)</td>
<td>57.68 (0.08)</td>
<td>49.67 (0.07)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, KOind, prices</td>
<td>66.51 (0.05)</td>
<td>70.70 (0.13)</td>
<td>62.53 (0.16)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>LS</td>
<td>one, pricesKO</td>
<td>67.00 (0.06)</td>
<td>71.27 (0.13)</td>
<td>64.04 (0.19)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>16</td>
<td>Tree</td>
<td>maxprice, KOind, pricesKO</td>
<td>68.90 (0.09)</td>
<td>71.02 (0.21)</td>
<td>54.97 (0.07)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>RO</td>
<td>maxprice</td>
<td>84.28 (0.27)</td>
<td>79.13 (0.53)</td>
<td>60.54 (0.50)</td>
<td>$2 \times 10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, pricesKO, KOind, payoff</td>
<td>82.38 (0.09)</td>
<td>79.17 (0.10)</td>
<td>62.92 (0.15)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, pricesKO, payoff</td>
<td>82.45 (0.10)</td>
<td>78.91 (0.10)</td>
<td>62.27 (0.14)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>payoff, KOind, pricesKO</td>
<td>82.38 (0.09)</td>
<td>79.17 (0.10)</td>
<td>62.92 (0.15)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>pricesKO, payoff</td>
<td>82.45 (0.10)</td>
<td>78.91 (0.10)</td>
<td>62.27 (0.14)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, prices, payoff</td>
<td>73.46 (0.19)</td>
<td>62.81 (0.17)</td>
<td>48.79 (0.17)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one</td>
<td>64.26 (0.12)</td>
<td>50.78 (0.09)</td>
<td>44.32 (0.05)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, KOind, prices</td>
<td>77.53 (0.11)</td>
<td>71.06 (0.12)</td>
<td>56.06 (0.21)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, prices</td>
<td>64.63 (0.11)</td>
<td>55.21 (0.11)</td>
<td>46.58 (0.05)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>one, pricesKO</td>
<td>77.36 (0.10)</td>
<td>70.44 (0.16)</td>
<td>55.11 (0.24)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>LS</td>
<td>maxprice, KOind, pricesKO</td>
<td>77.67 (0.11)</td>
<td>71.53 (0.13)</td>
<td>63.53 (0.22)</td>
<td>$10^5$</td>
</tr>
<tr>
<td>32</td>
<td>Tree</td>
<td>maxprice, time</td>
<td>77.33 (0.24)</td>
<td>65.57 (3.18)</td>
<td>50.96 (0.08)</td>
<td>$10^5$</td>
</tr>
</tbody>
</table>

Optimal is indicated in bold for each number of assets $d \in \{8, 16, 32\}$ and initial price $\bar{x} \in \{90, 100, 110\}$. The number of sample paths for the robust optimization method (RO) is equal to the number of training sample paths ($N = 10^3$) plus the number of validation sample paths ($N = 10^3$), see §2.5 for additional details. The robust optimization method was run with robustness parameters $\epsilon \in \{k \times 10^p : k \in \{1, \ldots, 10\}, p \in \{0.01, 0.1, 1\}\}$ and solved approximately using the algorithm from Theorem 6. Problem parameters are $T = 54$, $Y = 3$, $r = 0.05$, $K = 100$, $B_0 = 150$, $\delta = 0.25$, $\sigma_a = 0.1 + s \bar{x}^{0.5}$, $p_{a,a'} = 0$ for all $a \neq a'$. Figure 9 illustrates the impact of the number of simulated training sample paths on the performance and computation times of the various methods. We observe that the robust optimization method requires significantly fewer sample paths compared to alternative methods to find high-quality stopping rules. The smaller number of sample paths can have considerable benefits in practice; for example, we note that the computation times on the right half of Figure 9 capture neither the computation times nor the memory (RAM) requirements for generating and manipulating the simulated sample paths, the latter of which was a key bottleneck when running the LS method in our experiments. We also note that the recorded computation times for the robust optimization method are the sum of the computation times over all considered choices of the robustness parameter. In this sense, our implementation of the robust optimization method is rather naive, as the
Figure 8 Each plot shows the exercise policies obtained from solving a robust optimization problem constructed from a training dataset of size $N = 10^3$ and with the robustness parameter selected using a validation set of size $\tilde{N} = 10^3$. The problem parameters are $\bar{x} = 100$ and $d = 16$. The remaining parameters in Figures 8a and 8b are the same as those shown in Tables 1 and 2, respectively. The thick black rectangles are the stopping regions and the thin black line shows the barrier threshold.
Figure 9  The plots show the expected rewards (left) and computation times (right) for various methods as a function of the number of simulated training sample paths. The problem parameters are $\bar{x} = 100$ and $d = 16$, and the remaining parameters are the same as those shown in Table 2.

Figure 10  Each plot shows performance metrics of the robust optimization problems constructed from a training dataset of size $N = 10^3$ as a function of the robustness parameter $\epsilon$. The left plot shows the robust objective value and expected reward of policies obtained by solving the robust optimization problem. The right plot shows the computation time for solving the robust optimization problem. In both plots, the robust optimization problem is solved approximately using the algorithm from Theorem 6. The problem parameters are $\bar{x} = 100$ and $d = 16$, and the remaining parameters are the same as those shown in Table 2.
set of robustness parameters was chosen \textit{a priori}. This observation is particularly salient, as we see in Figure 10 that the computation time of solving the robust optimization problem is highly dependent on the robustness parameter; indeed, most of the computation times of the robust optimization problem comes from solving the robust optimization problem with unnecessarily large choices of the robustness parameter. For this reason, a dynamic search over the space of robustness parameters will reduce the computation cost significantly. Additional numerical experiments which show the relationship between the robustness parameter and the robust optimization method are provided in Appendix E.

6. Conclusion

Over the past two decades, dynamic robust optimization has experienced a surge of algorithmic advances. Until now, these advances from robust optimization have not been harnessed to develop algorithms for stochastic dynamic optimization problems with known probability distributions. In this paper, we showed the value of bridging these traditionally separate fields of research, in application to the classical and widely-studied problem of optimal stopping. In this context, we devised new and theoretically-justified algorithms for non-Markovian optimal stopping problems and highlighted the performance of these new algorithms on stylized and well-studied problems from options pricing. Along the way, we also developed novel theoretical and computational results for solving dynamic robust optimization problems which average over multiple uncertainty sets. We believe this work takes a meaningful step towards broadening the impact of robust optimization to address stochastic dynamic optimization problems of importance to industry.

References


Technical Proofs and Additional Results

Appendix A: Limitations of DP for Finding Markovian Stopping Rules in Non-Markovian Optimal Stopping Problems

In the context of non-Markovian optimal stopping problems, a subtle but important challenge is that the exercise policies which define the best Markovian stopping rule cannot be found in general using backwards recursion. Intuitively, this problem arises because Bellman’s dynamic programming equations no longer hold when the stochastic process is non-Markovian. This fact is illustrated numerically in §5.1 and motivates the development of methods which optimize over the exercise policies in all time periods simultaneously. For the sake of completeness, we provide the following example in which there is a Markovian stopping rule that is optimal for the non-Markovian stopping problem but is not obtained using backwards recursion.

Example EC.1. Consider a three-period optimal stopping problem with a one-dimensional non-Markovian stopping process that obeys the following probability distribution:

\[
P(x_1 = 3, x_2 = 2, x_3 = 1) = \frac{2}{3}; \quad P(x_1 = 1, x_2 = 2, x_3 = 3) = \frac{1}{3}.
\]

Let the reward of stopping on each period \( t \) be equal to the current state \( x_t \), and recall that our goal is to find a stopping rule which maximizes the expected reward. We observe that there is an optimal stopping rule for this non-Markovian stopping problem that is a Markovian stopping rule, defined by exercise policies \( \mu^*_1(x_1) = \text{STOP} \) if and only if \( x_1 = 3 \), \( \mu^*_2(x_2) = \text{CONTINUE} \) for all \( x_2 \), and \( \mu^*_3(x_3) = \text{STOP} \) for all \( x_3 \). Applying this stopping rule yields an expected reward of \( \frac{2}{3} \times 3 + \frac{1}{3} \times 2 = 3.66 \).

We now show that the best Markovian stopping rule which is obtained using backwards recursion will have a strictly lower expected reward. Indeed, assume that the reward from not stopping on any period is equal to zero. Then, starting on the last period, it is clear from the dynamic programming principle that the optimal exercise policy for the last period is \( \mu^\text{DP}_3(x_3) = \text{STOP} \). Hence, conditioned on \( x_2 = 2 \), the expected reward from stopping on the third period is \( 1 \times P(x_3 = 1 \mid x_2 = 2) + 3 \times P(x_3 = 3 \mid x_2 = 2) = 1 \times \frac{2}{3} + 3 \times \frac{1}{3} = \frac{5}{3} \). Because the reward from stopping on the second period (2) is greater than the conditional expected reward of not stopping on the second period (\( \frac{5}{3} \)), the dynamic programming principle says that the exercise policy in the second period should be chosen to satisfy \( \mu^\text{DP}_2(2) = \text{STOP} \). Finally, unfolding to the first period, we conclude that the exercise policy obtained from dynamic programming is \( \mu^\text{DP}_1(x_1) = \text{STOP} \) if and only if \( x_1 = 3 \). All together, the Markovian stopping rule for the non-Markovian optimal stopping problem that is obtained using backwards recursion yields an expected reward of \( \frac{2}{3} \times 3 + \frac{1}{3} \times 2 = 2.66 \). □
Appendix B: Proofs from §2.4

We begin by presenting some preliminary notation and results. Recall that $x \equiv (x_1, \ldots, x_T)$, $x^1 \equiv (x^1_1, \ldots, x^1_T)$, $x^2 \equiv (x^2_1, \ldots, x^2_T), \ldots \in \mathcal{X}^T \subseteq \mathbb{R}^d$ are sample paths drawn independently from an identical joint probability distribution. We will make use of the following additional notation:

\[
\delta_N \triangleq N^{- \max(3, T+1)}; \quad \mathcal{U}^i(\delta_N) \triangleq \left\{ y \in \mathcal{X}^T : \|y - x^i\|_\infty \leq \delta_N \right\}; \quad M_N \triangleq N^{(1/T + 2) \log N}.
\]

We will also utilize the following two lemmas from the literature which have been adapted to the notation of the present paper.

**Lemma EC.1** (Theorem 2 of Bertsimas et al. (2020)). Let Assumption 2 hold. Then there exists a finite $\bar{N} \in \mathbb{N}$, almost surely, such that the following inequality holds for all $N \geq \bar{N}$ and all measurable functions $f : \mathbb{R}^{Td} \rightarrow \mathbb{R}$:

\[
\mathbb{E}\left[f(\bar{x}) \mathbb{I}\{x \in \bigcup_{i=1}^{N} \mathcal{U}^i(\delta_N)\}\right] \geq \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in \mathcal{U}^i(\delta_N)} f(y) - M_N \sup_{y \in \bigcup_{i=1}^{N} \mathcal{U}^i(\delta_N)} |f(y)|.
\]

**Lemma EC.2.** Let Assumption 4 hold. Then for all $\epsilon > 0$,

\[
\liminf_{N \rightarrow \infty} \sup_{\mu} \tilde{J}_{N, \epsilon}(\mu) = \limsup_{N \rightarrow \infty} \sup_{\mu} \tilde{J}_{N, \epsilon}(\mu) \quad \text{almost surely.}
\]

**Proof.** Consider any fixed $\epsilon > 0$, and, for notational convenience, define the following function:

\[
h_\epsilon(x^1, \ldots, x^N) \triangleq \sup_{\mu} \tilde{J}_{N, \epsilon}(\mu).
\]

We will utilize the following intermediary claim:

**Claim 1:** The function $h_\epsilon : \mathcal{X}^T \times \cdots \times \mathcal{X}^T \rightarrow \mathbb{R}$ has the following ‘bounded differences’ property: for all $\bar{x}^1, \ldots, \bar{x}^N \in \mathcal{X}^T$ and $\bar{x}^1, \ldots, \bar{x}^N \in \mathcal{X}^T$ that differ only on the $j$th coordinate ($\bar{x}^i = \bar{x}^j$ for $i \neq j$),

\[
|h_\epsilon(\bar{x}^1, \ldots, \bar{x}^N) - h_\epsilon(\bar{x}^1, \ldots, \bar{x}^N)| \leq \frac{U}{N}.
\]

**Proof of Claim 1:** For any arbitrary $\eta > 0$, let the exercise policies $\tilde{\mu}^n$ be chosen to satisfy

\[
\frac{1}{N} \sum_{i=1}^{N} \inf_{y \in \mathcal{X}^T : \|y - \bar{x}^i\|_\infty \leq \epsilon} g(\tau_{\tilde{\mu}^n}(y), \bar{x}^i) \geq h_\epsilon(\bar{x}^1, \ldots, \bar{x}^N) - \eta.
\]  

(EC.1)

We observe that

\[
h_\epsilon(\bar{x}^1, \ldots, \bar{x}^N) - h_\epsilon(\bar{x}^1, \ldots, \bar{x}^N)
\]

\[
\leq \left( \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in \mathcal{X}^T : \|y - \bar{x}^i\|_\infty \leq \epsilon} g(\tau_{\tilde{\mu}^n}(y), \bar{x}^i) + \eta \right) - \left( \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in \mathcal{X}^T : \|y - \bar{x}^i\|_\infty \leq \epsilon} g(\tau_{\tilde{\mu}^n}(y), \bar{x}^i) \right)
\]

\[
= \frac{1}{N} \inf_{y \in \mathcal{X}^T : \|y - \bar{x}\|_\infty \leq \epsilon} g(\tau_{\tilde{\mu}^n}(y), \bar{x}) - \frac{1}{N} \inf_{y \in \mathcal{X}^T : \|y - \bar{x}\|_\infty \leq \epsilon} g(\tau_{\tilde{\mu}^n}(y), \bar{x}) + \eta
\]

\[
\leq \frac{U}{N} + \eta.
\]
Indeed, the first inequality holds because of line (EC.1) and because $\hat{\mu}^n$ is a feasible but possibly suboptimal solution to the optimization problem $\sup_{\mu} \frac{1}{N} \sum_{i=1}^N \inf_{x \in \mathcal{X}} \|x - \hat{x}\|_\infty \leq g(\tau_{\mu}(y), \hat{x}) \equiv h_x(\hat{x}, \ldots, \hat{x})$, and the second inequality follows from Assumption 4. Because $\eta > 0$ was chosen arbitrarily, we have shown that

$$h_x(\hat{x}, \ldots, \hat{x}) - h_x(\hat{x}, \ldots, \hat{x}) \leq \frac{U}{N}.$$  

It follows from symmetry that

$$h_x(\hat{x}, \ldots, \hat{x}) - h_x(\hat{x}, \ldots, \hat{x}) \leq \frac{U}{N},$$

which concludes our proof of Claim 1.

Because the above Claim 1 holds, it follows from McDiarmid’s inequality that

$$\mathbb{P}\left( |h_x(x^1, \ldots, x^N) - \mathbb{E}[h_x(x^1, \ldots, x^N)] | > \eta \right) \leq 2\exp\left( \frac{-2\eta^2 N}{U^2} \right) \quad \forall \eta > 0.$$  

It follows from the above line that

$$\sum_{N=1}^{\infty} \mathbb{P}\left( |h_x(x^1, \ldots, x^N) - \mathbb{E}[h_x(x^1, \ldots, x^N)] | > \eta \right) < \infty \quad \forall \eta > 0,$$

and so the Borel-Cantelli lemma implies that

$$\lim_{N \to \infty} |h_x(x^1, \ldots, x^N) - \mathbb{E}[h_x(x^1, \ldots, x^N)]| = 0 \quad \text{almost surely.} \quad (\text{EC.2})$$

We observe from identical reasoning as in the proof of Shapiro et al. (2014, Proposition 5.6) that $\mathbb{E}[h_x(x^1, \ldots, x^N)]$ is monotonically decreasing with respect to $N \in \mathbb{N}$. Since the random variables $h_x(x^1, \ldots, x^N)$ are also contained in the interval $[0, U]$ for all $N \in \mathbb{N}$, we conclude that $\lim_{N \to \infty} \mathbb{E}[h_x(x^1, \ldots, x^N)]$ exists, and thus it follows from line (EC.2) that $\lim_{N \to \infty} h_x(x^1, \ldots, x^N)$ exists almost surely. This concludes the proof of Lemma EC.2. \qed

In view of the above notation and lemmas, we are now ready to present the proofs of the theorems from §2.4.

**Proof of Theorem 3.** Consider any arbitrary choice of the robustness parameter $\epsilon > 0$, and recall that the reward function satisfies $g(\infty, y) = 0$ (see §2) and $0 \leq g(1, y), \ldots, g(T, y) \leq U$ for all trajectories $y \in \mathcal{X}^T$ (Assumption 4). With this notation, we observe that

$$\lim_{N \to \infty} \inf_{\mu} \left\{ J^*(\mu) - \tilde{J}_{N, \epsilon}(\mu) \right\} = \lim_{N \to \infty} \inf_{\mu} \left\{ \mathbb{E}[g(\tau_{\mu}(y), x)] - \tilde{J}_{N, \epsilon}(\mu) \right\} \quad (\text{EC.3})$$

$$\geq \lim_{N \to \infty} \inf_{\mu} \left\{ \mathbb{E}[g(\tau_{\mu}(y), x) \mathbb{I}\{x \in \bigcup_{i=1}^N \mathcal{U}_i(\delta_N)\}] - \tilde{J}_{N, \epsilon}(\mu) \right\} \quad (\text{EC.4})$$

$$\geq \lim_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^N \inf_{y \in \mathcal{U}_i(\delta_N)} g(\tau_{\mu}(y), y) - M_N U - \tilde{J}_{N, \epsilon}(\mu) \right\} \quad \text{almost surely} \quad (\text{EC.5})$$

$$= \lim_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^N \inf_{y \in \mathcal{U}_i(\delta_N)} g(\tau_{\mu}(y), y) - \tilde{J}_{N, \epsilon}(\mu) \right\} \quad (\text{EC.6})$$

$$\geq \lim_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^N \inf_{y \in \mathcal{U}_i(\delta_N)} g(\tau_{\mu}(y), y) - \tilde{J}_{N, \epsilon}(\mu) \right\}. \quad (\text{EC.7})$$
Note that all of the above limits exist because \( \lim_{N \to \infty} \) and \( \limsup_{N \to \infty} \) hold because the reward function is nonnegative; (EC.5) follows from Lemma EC.1 and the boundedness of the reward function; (EC.6) holds because \( M_N \to 0 \); (EC.7) holds because \( \delta_N \to 0 \) implies, for any arbitrary \( \epsilon > 0 \), that the inequality \( \inf_{y \in \mathcal{U}} g(\tau_\mu(y), y) \leq \inf_{y \in \mathcal{U}(\delta_N)} g(\tau_\mu(y), y) \) is satisfied for all \( i \in \mathbb{N} \), for all \( \mu \), and for all large \( N \in \mathbb{N} \). Moreover:

\[
(\text{EC.3}) = \liminf_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left( \inf_{y \in \mathcal{U}} g(\tau_\mu(y), x^i) + \inf_{y \in \mathcal{U}} g(\tau_\mu(y), y) - g(\tau_\mu(y), x^i) \right) - \hat{J}_{N,\epsilon}(\mu) \right\} \quad (\text{EC.8})
\]

\[
\geq \liminf_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left( \inf_{y \in \mathcal{U}} g(\tau_\mu(y), x^i) + \inf_{y \in \mathcal{U}} g(\tau_\mu(y), y) - g(\tau_\mu(y), x^i) \right) - J_{N,\epsilon}(\mu) \right\} \quad (\text{EC.9})
\]

\[
\geq \liminf_{N \to \infty} \inf_{\mu} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left( \inf_{y \in \mathcal{U}} g(\tau_\mu(y), x^i) + \Delta_\epsilon(x^i) \right) - \hat{J}_{N,\epsilon}(\mu) \right\} \quad (\text{EC.10})
\]

\[
= \liminf_{N \to \infty} \inf_{\mu} \left\{ \hat{J}_{N,\epsilon}(\mu) - \hat{J}_{N,\epsilon}(\mu) \right\} + \liminf_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \Delta_\epsilon(x^i) \quad (\text{EC.11})
\]

\[
= \mathbb{E}[\Delta_\epsilon(x)] \quad \text{almost surely.} \quad (\text{EC.12})
\]

(\text{EC.8}), (\text{EC.9}), and (\text{EC.10}) follow from algebra; (\text{EC.11}) follows from the definition of \( \Delta_\epsilon(x^i) \) (see Assumption 1); (\text{EC.12}) follows from the definition of \( \hat{J}_{N,\epsilon}(\mu) \); (\text{EC.13}) follows from the strong law of large numbers.

Since \( \epsilon > 0 \) was chosen arbitrarily, we have shown that

\[
\lim_{\epsilon \to 0} \liminf_{N \to \infty} \inf_{\mu} \left\{ J^*(\mu) - \hat{J}_{N,\epsilon}(\mu) \right\} \geq \lim_{\epsilon \to 0} \mathbb{E}[\Delta_\epsilon(x)] \geq 0 \quad \text{almost surely},
\]

where the first inequality holds almost surely from lines (\text{EC.3})-(\text{EC.13}), and the second inequality holds almost surely due to the dominated convergence theorem and Assumption 1. Note that the above limits exist because \( \epsilon \mapsto \liminf_{N \to \infty} \inf_{\mu} \{ J^*(\mu) - \hat{J}_{N,\epsilon}(\mu) \} \) and \( \epsilon \mapsto \mathbb{E}[\Delta_\epsilon(x)] \) are monotonic functions. This concludes the proof of Theorem 3. \( \square \)

\textbf{Proof of Theorem 1.} We first show that the optimal objective value of (RO) is an asymptotic lower bound on the optimal objective value of (OPT). Indeed,

\[
0 \geq \lim_{\epsilon \to 0} \limsup_{N \to \infty} \sup_{\mu} \left\{ \hat{J}_{N,\epsilon}(\mu) - J^*(\mu) \right\} \quad \text{almost surely}
\]

\[
\geq \lim_{\epsilon \to 0} \limsup_{N \to \infty} \left( \sup_{\mu} \hat{J}_{N,\epsilon}(\mu) - \sup_{\mu} J^*(\mu) \right)
\]

\[
= - \sup_{\mu} J^*(\mu) + \lim_{\epsilon \to 0} \limsup_{N \to \infty} \sup_{\mu} \hat{J}_{N,\epsilon}(\mu),
\]

where the first inequality follows from Theorem 3 and the remaining lines follow from algebra. Note that all of the above limits exist because \( \epsilon \mapsto \limsup_{N \to \infty} \sup_{\mu} \{ \hat{J}_{N,\epsilon}(\mu) - J^*(\mu) \} \) and \( \epsilon \mapsto \limsup_{N \to \infty} \sup_{\mu} \hat{J}_{N,\epsilon}(\mu) \) are monotonic functions.
We next show that the optimal objective value of (RO) provides an asymptotic upper bound on the optimal objective value of (OPT):

\[
\begin{align*}
\lim_{\epsilon \to 0} \liminf_{N \to \infty} \sup_{\mu} \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U^i} g(\tau_{\mu}(y), x^i) &\geq \lim_{\epsilon \to 0} \liminf_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U^i} g(\tau_{\mu^*}(y), x^i) \\
&= \lim_{\epsilon \to 0} \mathbb{E} \left[ \inf_{y \in X^T : \|y - x\| \leq \epsilon} g(\tau_{\mu^*}(y), x) \right] \text{ almost surely} (EC.15) \\
&= \mathbb{E} \left[ g(\tau_{\mu^*}(x), x) \right] (EC.16)
\end{align*}
\]

(EC.14) holds because \(\mu^*\) is a feasible but possibly suboptimal solution to the optimization problems \(\sup_{\mu} \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U^i} g(\tau_{\mu}(y), x')\); (EC.15) follows from the strong law of large numbers; (EC.16) follows from the dominated convergence theorem and Assumption 3.

Combining the above results, we have shown that

\[
\sup_{\mu} \mathbb{E}[g(\tau_{\mu}(x), x)] \leq \lim \liminf_{N \to \infty} \sup_{\mu} \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U^i} g(\tau_{\mu}(y), x^i) \leq \lim \limsup_{N \to \infty} \sup_{\mu} \frac{1}{N} \sum_{i=1}^{N} \inf_{y \in U^i} g(\tau_{\mu}(y), x^i) \leq \sup_{\mu} \mathbb{E}[g(\tau_{\mu}(x), x)] \text{ almost surely},
\]

which implies that

\[
\lim \liminf_{N \to \infty} \sup_{\mu} \hat{J}_{N,\epsilon}(\mu) = \lim \limsup_{N \to \infty} \hat{J}_{N,\epsilon}(\mu) = \sup_{\mu} J^*(\mu) \quad \text{almost surely.}
\]

Combining the above line with Lemma EC.2, our proof of Theorem 1 is complete. \(\square\)

**Proof of Theorem 2.** We observe from Theorems 1 and 3 that for every arbitrary \(\eta > 0\), there exists a finite \(\bar{\epsilon}(\eta) > 0\) almost surely such that the following statements hold for all \(0 < \epsilon < \bar{\epsilon}(\eta)\):

\[
\begin{align*}
\lim_{N \to \infty} \hat{J}_{N,\epsilon}(\hat{\mu}_{N,\epsilon}) - \sup_{\mu} J^*(\mu) &\leq \eta \quad \text{almost surely; (EC.17)} \\
\liminf_{N \to \infty} \left( J^*(\hat{\mu}_{N,\epsilon}) - \hat{J}_{N,\epsilon}(\hat{\mu}_{N,\epsilon}) \right) &\geq -\eta \quad \text{almost surely. (EC.18)}
\end{align*}
\]

Therefore,

\[
\sup_{\mu} J^*(\mu) \geq \limsup_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) \geq \liminf_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) \geq \liminf_{N \to \infty} \hat{J}_{N,\epsilon}(\hat{\mu}_{N,\epsilon}) - \eta \geq \sup_{\mu} J^*(\mu) - 2\eta,
\]

where the first inequality holds because each \(\hat{\mu}_{N,\epsilon}\) is a feasible but possibly suboptimal solution to (OPT), the second inequality is obvious, the third inequality follows from (EC.18), and the final inequality follows from (EC.17). Rearranging the above line, we have shown that the following statements hold for all \(0 < \epsilon < \bar{\epsilon}(\eta)\):

\[
\begin{align*}
\left| \limsup_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) - \sup_{\mu} J^*(\mu) \right| &\leq 2\eta \quad \text{almost surely; (EC.19)} \\
\left| \liminf_{N \to \infty} J^*(\hat{\mu}_{N,\epsilon}) - \sup_{\mu} J^*(\mu) \right| &\leq 2\eta \quad \text{almost surely.}
\end{align*}
\]

Since \(\eta > 0\) was chosen arbitrarily, our proof of Theorem 2 is complete. \(\square\)
Appendix C: Proof of Theorem 5

Our proof of the computational complexity of (IP) consists of a reduction from MIN-2-SAT, which is shown to be strongly NP-hard by Kohli et al. (1994):

\[ v^{\text{MIN-2-SAT}} \triangleq \min_{b,z} \sum_{k=1}^{K} z_k \]

subject to

\[
\begin{align*}
z_k &\geq b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^+ \\
z_k &\geq 1 - b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^- \\
b_\ell &\in \{0, 1\} & \forall \ell \in \{1, \ldots, L\},
\end{align*}
\]

where the given sets \(I_k^+, I_k^- \subseteq \{1, \ldots, L\}\) satisfy \(|I_k^+| + |I_k^-| = 2\) for each \(k \in \{1, \ldots, K\}\).

Note that the following equality is obtained by replacing each decision variable \(z_k\) with \(1 - z_k\):

\[ v^{\text{MIN-2-SAT}} = N - \max_{b,z} \sum_{k=1}^{K} z_k \]

subject to

\[
\begin{align*}
z_k &\leq 1 - b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^+ \\
z_k &\leq b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^- \\
b_\ell &\in \{0, 1\} & \forall \ell \in \{1, \ldots, L\}.
\end{align*}
\]

We now show that any instance of the above maximization problem can be equivalently reformulated as polynomially-size instance of (IP) with \(T = 3\) periods.

**Proof of Theorem 5.** Consider any arbitrary instance of the binary linear optimization problem

\[
\begin{align*}
\max_{b,z} \sum_{k=1}^{K} z_k \\
\text{subject to} & \\
& z_k \leq 1 - b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^+ \\
& z_k \leq b_\ell & \forall k \in \{1, \ldots, K\}, \forall \ell \in I_k^- \\
& b_\ell \in \{0, 1\} & \forall \ell \in \{1, \ldots, L\}, \quad (-\text{MIN-2-SAT})
\end{align*}
\]

and let \(e_\ell \in \mathbb{R}^{L+1}\) denote the \(\ell\)-th column vector of the identity matrix. We construct an instance of (IP) defined as follows:

- The number of periods is \(T = 3\).
- The state space is \(\mathcal{X} = \mathbb{R}^{L+1}\).
- The reward function for each period \(t \in \{1, 2, 3\}\) is \(g(t, y) = y_t \cdot e_{T+1} + K\).
- The robustness parameter in the uncertainty sets is \(\epsilon = \frac{2}{3}\).
- The number of sample paths is \(N \triangleq L + K\), and the sample paths are defined as follows:
  - For each \(\ell \in \{1, \ldots, L\}\), let \(x_1^\ell = x_2^\ell = e_\ell\) and \(x_3^\ell = -K e_{L+1}\).
  - For each \(k \in \{1, \ldots, K\}\), let \(x_1^{L+k} = \frac{1}{2} \sum_{\ell \in I_k^+} e_\ell, x_2^{L+k} = \frac{1}{2} \sum_{\ell \in I_k^-} e_\ell\), and \(x_3^{L+k} = e_{L+1}\).  

\[ e_{\ell} \in \mathbb{R}^{L+1} \text{ denote the } \ell\text{-th column vector of the identity matrix. We construct an instance of (IP) defined as follows:} \]

- The number of periods is \(T = 3\).
- The state space is \(\mathcal{X} = \mathbb{R}^{L+1}\).
- The reward function for each period \(t \in \{1, 2, 3\}\) is \(g(t, y) = y_t \cdot e_{T+1} + K\).
- The robustness parameter in the uncertainty sets is \(\epsilon = \frac{2}{3}\).
- The number of sample paths is \(N \triangleq L + K\), and the sample paths are defined as follows:
  - For each \(\ell \in \{1, \ldots, L\}\), let \(x_1^\ell = x_2^\ell = e_\ell\) and \(x_3^\ell = -K e_{L+1}\).
  - For each \(k \in \{1, \ldots, K\}\), let \(x_1^{L+k} = \frac{1}{2} \sum_{\ell \in I_k^+} e_\ell, x_2^{L+k} = \frac{1}{2} \sum_{\ell \in I_k^-} e_\ell\), and \(x_3^{L+k} = e_{L+1}\).  

\[ e_{\ell} \in \mathbb{R}^{L+1} \text{ denote the } \ell\text{-th column vector of the identity matrix. We construct an instance of (IP) defined as follows:} \]
In the remainder of the proof, we show that the above instance of (IP) is equivalent to (−MIN-2-SAT). Indeed, it follows immediately from the above construction that the values of $v_i^\ell$ for each sample path $i$ and period $\ell$ are:

$$
\begin{align*}
    v_1^\ell & = K; \\
    v_2^\ell & = K; \\
    v_3^\ell & = 0, \\
    \forall \ell \in \{1, \ldots, L\}, \\
    v_{1+k,L+k}^1 & = K; \\
    v_{2+k,L+k}^2 & = K; \\
    v_{3+k,L+k}^3 & = K + 1, \\
    \forall k \in \{1, \ldots, K\}.
\end{align*}
$$

We require two intermediary claims:

Claim 2: There exists an optimal solution for (IP) which satisfies $\sigma^1, \ldots, \sigma^L \in \{1, 2\}$ and $\sigma^{L+1} = \cdots = \sigma^{L+K} = 3$.

Proof of Claim 2: We observe that the optimal objective value of (IP) is greater than or equal to $K$, since this objective value would be achieved by setting $\sigma^1 = \cdots = \sigma^N = 1$. Now consider any solution to (IP) where $\sigma^{\ell'} = 3$ for some $\ell' \in \{1, \ldots, L\}$. For that solution,

$$
\frac{1}{N} \sum_{i=1}^N \min_{j: \sigma^i \leq \sigma^j} v_{\sigma^j}^{i,j} = \frac{1}{N} \left( \sum_{\ell=1}^L \min_{j: \sigma^\ell \leq \sigma^j} v_{\sigma^j}^{\ell,j} + \min_{j: \sigma^{\ell'} \leq \sigma^j} v_{\sigma^j}^{\ell,j} + \sum_{k=1}^K \min_{j: \sigma^L+k \leq \sigma^j} v_{\sigma^j}^{L+k,j} \right) \\
\leq \frac{1}{N} \left( \sum_{\ell=1}^L K \right) + 0 + \left( \sum_{k=1}^K (K+1) \right) \\
= K.
$$

Because the objective value associated with this solution is never better than the objective value obtained by the solution $\sigma^1 = \cdots = \sigma^N = 1$, we have shown that there exists an optimal solution for (IP) that satisfies $\sigma^1, \ldots, \sigma^L \in \{1, 2\}$.

Consider any arbitrary solution $\sigma^1, \ldots, \sigma^N \in \{1, 2, 3\}$ that satisfies $\sigma^1, \ldots, \sigma^L \in \{1, 2\}$, and suppose that $\sigma^{L+k} \in \{1, 2\}$ for some $k \in \{1, \ldots, K\}$. To perform an exchange argument, we construct an alternative solution $\tilde{\sigma}^1, \ldots, \tilde{\sigma}^N \in \{1, 2, 3\}$ defined as

$$
\tilde{\sigma}^i \triangleq \begin{cases} 
    \sigma^i, & \text{if } i \neq L+k; \\
    3, & \text{if } i = L+k.
\end{cases}
$$

We observe that $\{j : \sigma^j \leq \sigma^i\} \supseteq \{j : \tilde{\sigma}^j \leq \tilde{\sigma}^i\}$ holds for all $i \in \{1, \ldots, N\} \setminus \{L+k\}$. Moreover, $\min_{j: \sigma^j \leq \sigma^{L+k}} v_{\sigma^j}^{L+k,j} = K$ and $\min_{j: \tilde{\sigma}^j \leq \tilde{\sigma}^{L+k}} v_{\tilde{\sigma}^j}^{L+k,j} \in \{K, K+1\}$ also hold. Therefore,

$$
\frac{1}{N} \sum_{i=1}^N \min_{j: \sigma^i \leq \sigma^j} v_{\sigma^j}^{i,j} \leq \frac{1}{N} \sum_{i=1}^N \min_{j: \tilde{\sigma}^i \leq \tilde{\sigma}^j} v_{\tilde{\sigma}^j}^{i,j}.
$$

Because $\sigma^1, \ldots, \sigma^N$ was chosen arbitrarily, we conclude that there exists an optimal solution for (IP) which satisfies $\sigma^1, \ldots, \sigma^L \in \{1, 2\}$ and $\sigma^{L+1} = \cdots = \sigma^{L+K} = 3$. This concludes our proof of Claim 2.

Claim 3: If $\sigma^{L+1} = \cdots = \sigma^{L+K} = 3$, then the following equality holds for each $k \in \{1, \ldots, K\}$:

$$
\min_{j: \sigma^j \leq \sigma^i} v_{\sigma^j}^{L+k,j} = K + \mathbb{I}\{\sigma^\ell = 2 \text{ for all } \ell \in I_k^+ \text{ and } \sigma^\ell = 1 \text{ for all } \ell \in I_k^-\}
$$
Proof of Claim 3: For each $\ell \in \{1, \ldots, L\}$ and $k \in \{1, \ldots, K\}$, we observe that
\[
\|x_1^\ell - x_1^{L+k}\|_\infty = \left\| e_\ell - \frac{1}{2} \sum_{\ell' \in I_k^+} e_{\ell'} \right\|_\infty = \begin{cases} \frac{1}{2}, & \text{if } \ell \in I_k^+, \\ 1, & \text{otherwise.} \end{cases}
\]
\[
\|x_2^\ell - x_2^{L+k}\|_\infty = \left\| e_\ell - \frac{1}{2} \sum_{\ell' \in I_k^-} e_{\ell'} \right\|_\infty = \begin{cases} \frac{1}{2}, & \text{if } \ell \in I_k^-, \\ 1, & \text{otherwise.} \end{cases}
\]
This implies that the set
\[
U_1^k \cap U_1^{L+k} = \left\{ y_1 \in \mathbb{R}^{L+1} : \|y_1 - x_1^\ell\|_\infty \leq \frac{2}{3} \right\} \cap \left\{ y_1 \in \mathbb{R}^{L+1} : \|y_1 - x_1^{L+k}\|_\infty \leq \frac{2}{3} \right\}
\]
is nonempty if and only if $\ell \in I_k^+$, and the set
\[
U_2^k \cap U_2^{L+k} = \left\{ y_2 \in \mathbb{R}^{L+1} : \|y_2 - x_2^\ell\|_\infty \leq \frac{2}{3} \right\} \cap \left\{ y_2 \in \mathbb{R}^{L+1} : \|y_2 - x_2^{L+k}\|_\infty \leq \frac{2}{3} \right\}
\]
is nonempty if and only if $\ell \in I_k^-$. Consequently,
\[
v_1^{L+k, \ell} = \begin{cases} K, & \text{if } \ell \in I_k^+, \\ \infty, & \text{otherwise,} \end{cases} \quad v_2^{L+k, \ell} = \begin{cases} K, & \text{if } \ell \in I_k^-, \\ \infty, & \text{otherwise.} \end{cases}
\]
Since $\sigma^{L+1} = \cdots = \sigma^{L+K} = 3$, we conclude that the following equalities hold for all $k \in \{1, \ldots, K\}$:
\[
\min_{j, \sigma \leq L+k} v_{j, \sigma}^{L+k,j} = \min \left\{ \min_{\ell \in \{1, \ldots, L\}} v_1^{L+k, \ell}, \min_{\ell \in \{1, \ldots, L\} : \sigma' = 2} v_2^{L+k, \ell}, v_3^{L+k, L+k} \right\}
\]
\[
= \min \left\{ \min_{\ell \in I_k^+: \sigma' = 1} K, \min_{\ell \in I_k^- : \sigma' = 2} K, K + 1 \right\}
\]
\[
= K + \mathbb{I} \left\{ \sigma' = 2 \text{ for all } \ell \in I_k^+ \text{ and } \sigma' = 1 \text{ for all } \ell \in I_k^- \right\}
\]
This concludes the proof of Claim 3.

We now combine Claims 2 and 3 to complete our proof of Theorem 5:

\[
\text{(IP)} = \max_{\sigma^1, \ldots, \sigma^N \in \{1, 2, 3\}} \left\{ \frac{1}{N} \sum_{i=1}^N \min_{j, \sigma \leq L} v_{ij, \sigma}^j \right\}
\]
\[
= \max_{\sigma^1, \ldots, \sigma^L \in \{1, 2, 3\}} \left\{ \frac{1}{L+K} \left( \sum_{\ell=1}^L \min_{j: \sigma' \leq \ell} v_{ij, \sigma}^j + \sum_{k=1}^K \min_{j: \sigma' \leq L+k} v_{ij, \sigma}^{L+k,j} \right) \right\} \tag{EC.19}
\]
\[
= \max_{\sigma^1, \ldots, \sigma^L \in \{1, 2, 3\}} \left\{ \frac{1}{L+K} \left( \sum_{\ell=1}^L K + \sum_{k=1}^K \min_{j: \sigma' \leq L+k} v_{ij, \sigma}^{L+k,j} \right) \right\} \tag{EC.20}
\]
\[
= \max_{\sigma^1, \ldots, \sigma^L \in \{1, 2, 3\}} \left\{ \frac{LK}{L+K} + \frac{1}{L+K} \sum_{k=1}^K \left( K + \mathbb{I} \left\{ \sigma' = 2 \text{ for all } \ell \in I_k^+ \text{ and } \sigma' = 1 \text{ for all } \ell \in I_k^- \right\} \right) \right\}
\]
\[
= K + \left( \frac{1}{L+K} \right) \max_{\sigma^1, \ldots, \sigma^L \in \{1, 2, 3\}} \left\{ \sum_{k=1}^K \mathbb{I} \left\{ \sigma' = 2 \text{ for all } \ell \in I_k^+ \text{ and } \sigma' = 1 \text{ for all } \ell \in I_k^- \right\} \right\}
\]
\[
= K + \left( \frac{1}{L+K} \right) \neg \text{MIN-2-SAT}. \tag{EC.22}
\]
Indeed, (EC.19) follows from Claim 2; (EC.20) holds because \( v_t^{\ell} = v_2^{\ell} = K \) for all \( \ell \in \{1, \ldots, L\} \); (EC.21) follows from Claim 3; (EC.22) follows from algebra and setting \( b' = 0 \) if and only if \( \sigma^t = 2 \).

We have thus shown that any instance of MIN-2-SAT can be reduced to solving a polynomially-sized instance of (IP) with \( T = 3 \), which concludes our proof of Theorem 5.  \( \square \)

**Appendix D: Proofs from §4 (except for Theorem 5)**

**Proof of Proposition 1.** Consider any arbitrary choice of integers \( \sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\} \), and define the exercise policy

\[
\mu_i(y_t) \triangleq \begin{cases} 
\text{STOP,} & \text{if } y_t \in \bigcup_{i' \neq i} \mathcal{U}_{i'}^t, \\
\text{CONTINUE,} & \text{if } y_t \notin \bigcup_{i' \neq i} \mathcal{U}_{i'}^t
\end{cases}
\]

to all periods \( t \in \{1, \ldots, T\} \) and all \( y_t \in \mathcal{X} \). We observe that the exercise policies \( \mu \equiv (\mu_1, \ldots, \mu_T) \) are contained in \( \mathcal{M} \). Moreover, for each sample path \( i \in \{1, \ldots, N\} \), we observe that

\[
\sigma^i \geq \bar{\sigma}^i \triangleq \min_{t \in \{1, \ldots, T\}} \left\{ t : \mu_i(y_t) = \text{STOP} \text{ for all } y_t \in \mathcal{U}_i^t \right\}.
\]

Therefore,

\[
\frac{1}{N} \sum_{i = 1}^N \min_{j : \sigma^j \leq \sigma^i} v^{ij}_{\sigma^i} = \frac{1}{N} \sum_{i = 1}^N \min_{t \in \{1, \ldots, \sigma^i\}} \left\{ g(t, x^t) : \text{there exists } j \text{ such that } \sigma^j = t \text{ and } \mathcal{U}_i^t \cap \mathcal{U}_j^t \neq \emptyset \right\}
\]

\[
= \frac{1}{N} \sum_{i = 1}^N \min_{t \in \{1, \ldots, \sigma^i\}} \left\{ g(t, x^t) : \text{there exists } y_t \in \mathcal{U}_i^t \text{ such that } \mu_i(y_t) = \text{STOP} \right\}
\]

\[
\leq \frac{1}{N} \sum_{i = 1}^N \min_{t \in \{1, \ldots, \sigma^i\}} \left\{ g(t, x^t) : \text{there exists } y_t \in \mathcal{U}_i^t \text{ such that } \mu_i(y_t) = \text{STOP} \right\}
\]

\[
= \frac{1}{N} \sum_{i = 1}^N \inf_{y_t \in \mathcal{U}_i^t} g(\tau_\mu(y), x^t)
\]

The first equality follows from the definition of \( v^{ij}_{\sigma^i} \); the second equality from our construction of \( \mu \); the inequality follows from \( \sigma^i \geq \bar{\sigma}^i \); and the final equality follows from our construction of \( \bar{\sigma}^i \).

Since \( \sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\} \) were chosen arbitrarily and since \( \mu \in \mathcal{M} \), we have shown that

\[(IP) = \max_{\sigma^1, \ldots, \sigma^N \in \{1, \ldots, T\}} \frac{1}{N} \sum_{i = 1}^N \min_{j : \sigma^j \leq \sigma^i} v^{ij}_{\sigma^i} \leq \sup_{\mu \in \mathcal{M}} \frac{1}{N} \sum_{i = 1}^N \inf_{y_t \in \mathcal{U}_i^t} g(\tau_\mu(y), x^t) = (RO),\]

where the final equality follows from Theorem 4. To show the other direction, let \( \tilde{\mu} \in \mathcal{M} \) be optimal for \( (RO) \) and define

\[\tilde{\sigma}^i \triangleq \min_{t \in \{1, \ldots, T\}} \left\{ t : \tilde{\mu}_i(y_t) = \text{STOP} \text{ for all } y_t \in \mathcal{U}_i^t \right\} .\]

Then it follows from identical reasoning as above that

\[(RO) = \frac{1}{N} \sum_{i = 1}^N \inf_{y_t \in \mathcal{U}_i^t} g(\tau_{\tilde{\mu}}(y), x^t) = \frac{1}{N} \sum_{i = 1}^N \min_{j : \tilde{\sigma}^j \leq \tilde{\sigma}^i} v^{ij}_{\tilde{\sigma}^i} \leq (IP),\]

which completes the proof.  \( \square \)
Proof of Proposition 2. This follows immediately from Corollary 3 and Proposition 4. □

Proof of Proposition 3. It follows from construction that each $\kappa^i_{i+1} - \kappa^i_i$ is strictly positive. Therefore, we observe that there exists an optimal solution to (BP) where each decision variable $w_{it}^i$ satisfies

$$w_{it}^i = \begin{cases} 
0, & \text{if } b_s^i = 0 \text{ for all } s < t \\
 b_s^i = 0 \text{ for all } j \neq i \text{ and all } s \leq t \text{ which satisfy } v_{ij}^s < \kappa^i_t, \\
1, & \text{otherwise.}
\end{cases}$$

(EC.23)

Consider any optimal solution to (BP) which satisfies equality (EC.23) for each decision variable $w_{it}^i$. Since each $\kappa^i_{i+1} - \kappa^i_i$ is strictly positive, we also observe that each decision variable $b_s^i$ can in principle be set equal to one without decreasing the objective value; therefore, we can and will henceforth assume that we are considering an optimal solution which satisfies $\sum_{t=1}^T b_t^i \geq 1$ for each $i \in \{1, \ldots, N\}$.

For each $i \in \{1, \ldots, N\}$, let $\sigma^i$ be defined as the unique integer which satisfies $b_{\sigma^i}^i = 1$ and $b_t^i = 0$ for all $t < \sigma^i$. Then,

$$\sum_{t=1}^T \sum_{t=1}^{L_{i+1}^i - 1} (\kappa_{i+1}^i - \kappa_i^i) b_t^i (1 - w_{it}^i) = \sum_{t=\sigma^i}^T \sum_{t=1}^{L_{i+1}^i - 1} (\kappa_{i+1}^i - \kappa_i^i) b_t^i (1 - w_{it}^i)$$

$$= \sum_{\alpha^i}^{L_{i+1}^i - 1} (\kappa_{i+1}^i - \kappa_i^i) (1 - w_{\sigma^i t}^i).$$

(EC.24)

Indeed, the first equality holds because $b_t^i = 0$ for all $t < \sigma^i$. The second equality holds because $b_{\sigma^i}^i = 1$, and (EC.23) implies that $w_{it}^i = 1$ for all $t > \sigma^i$.

We observe that $\alpha^i$ is not impacted by the values of the decision variables $b_{\sigma^i+1}^i, \ldots, b_T^i$. Moreover, if $b_t^i = 1$ is changed to $b_t^i = 0$ for any $t \in \{\sigma^i + 1, \ldots, T\}$, then $\alpha^i$ will not decrease for any $j \neq i$. Therefore, we assume without loss of generality that $b_t^i = 0$ for all $t \in \{\sigma^i + 1, \ldots, T\}$. In particular, we have thus shown that there is an optimal solution to (BP) which satisfies $\sum_{t=1}^T b_t^i = 1$ for all $i \in \{1, \ldots, N\}$.

For each $i \in \{1, \ldots, N\}$, let $\hat{\ell}^i$ be the smallest integer such that there exists $j \in \{1, \ldots, N\}$ which satisfies $\sigma^j \leq \sigma^i$ and $v_{ij}^s \leq \kappa_{\hat{\ell}^i}^i$. Then,

$$\frac{1}{N} \sum_{i=1}^N \sum_{i=1}^T \sum_{t=1}^{L_{i+1}^i - 1} (\kappa_{i+1}^i - \kappa_i^i) b_t^i (1 - w_{it}^i) = \frac{1}{N} \sum_{i=1}^N \sum_{i=1}^{L_{\sigma^i+1}^i - 1} (\kappa_{\sigma^i+1}^i - \kappa_{\sigma^i}^i) (1 - w_{\sigma^i t}^i)$$

$$= \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^{\hat{\ell}^i - 1} (\kappa_{\hat{\ell}^i+1}^i - \kappa_{\hat{\ell}^i}^i)$$

$$= \frac{1}{N} \sum_{i=1}^N \kappa_{\hat{\ell}^i}^i$$

$$= \frac{1}{N} \sum_{i=1}^N \min_{j: \sigma^j \leq \sigma^i} v_{ij}^s,$$

where the first equality follows from (EC.24), the second equality follows from (EC.23), the third equality follows from $\kappa_1^1 = 0$, and the final equality follows from the definition of $\hat{\ell}^i$. This completes the proof. □
To strengthen this linear relaxation of (BP), we also add the valid constraints:

\[ f^i_{lt} \leq b^i_t \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, L^i_t - 1\} \]

\[ f^i_{lt} \leq w^i_{t,t+1} - w^i_{lt} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, L^i_t - 1\} \]

and replacing the objective function of (BP) with

\[
\max_{b, w, f} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{\ell=1}^{L^i_t-1} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) f^i_{lt}.
\]

To strengthen this linear relaxation of (BP), we also add the valid constraints:

\[
\begin{align*}
    w^i_{1,t_1} &= 0 \quad \text{for all } i \in \{1, \ldots, N\} \\
    b^i_t + w^i_{t,t+1} &= w^i_{t+1,t_1} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T-1\} \\
    b^i_T + w^i_{T,t_1} &= 1 \quad \text{for all } i \in \{1, \ldots, N\}.
\end{align*}
\]

Indeed, the validity of the above constraints for (BP) follows from the fact that there is an optimal solution to this zero-one bilinear program which satisfies \(\sum_{t=1}^{T} b^i_t = 1\) for each sample path \(i\). In summary, this linearization procedure transforms (BP) into the following equivalent mixed-integer linear optimization problem:

\[
\max_{b, w, f} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{\ell=1}^{L^i_t-1} (\kappa^i_{\ell+1} - \kappa^i_{\ell}) f^i_{lt}
\]

subject to

\[
\begin{align*}
    f^i_{lt} &\leq b^i_t \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, L^i_t - 1\} \\
    f^i_{lt} &\leq w^i_{t,t+1} - w^i_{lt} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, L^i_t - 1\} \\
    w^i_{t,\ell} &\leq w^i_{t+1,\ell} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T-1\}, \ \ell \in \{1, \ldots, |S^i_t|\} \\
    w^i_{t,\ell} &\leq w^i_{t,\ell+1} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, |S^i_t| - 1\} \\
    w^i_{1,t_1} &= 0 \quad \text{for all } i \in \{1, \ldots, N\} \\
    b^i_t + w^i_{t,t+1} &= w^i_{t+1,t_1} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T-1\} \\
    b^i_T + w^i_{T,t_1} &= 1 \quad \text{for all } i \in \{1, \ldots, N\} \\
    b^i_t &\leq w^i_{t,\ell} \quad \text{for all } i \neq j \in \{1, \ldots, N\} \text{ and } t \in \{1, \ldots, T\} \text{ such that } v^i_{t,j} = \kappa^i_{\ell} \\
    b^i_t &\in \{0, 1\} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\} \\
    w^i_{t,\ell} &\in \mathbb{R} \quad \text{for all } i \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\}, \ \ell \in \{1, \ldots, |S^i_t|\}.
\end{align*}
\]

\[\square\]

**Proof of Lemma 1.** Consider any arbitrary choice of the decision variables \(b^1, \ldots, b^N \in \{0, 1\}\) in the optimization problem (BP-1). Because the inequalities \(\kappa^i_{\ell+1} - \kappa^i_{\ell} > 0\) are satisfied, the optimal values for the remaining decision variables of the optimization problem are:

\[
w^i_t = \begin{cases} 1, & \text{if there exists } j \in S^i_t \text{ and } s \in \{1, \ldots, \ell - 1\} \text{ such that } v^i_{t,s} = \kappa^i_s \text{ and } b^j = 1, \\ 0, & \text{otherwise}, \end{cases} \quad (EC.25)
\]
Then the following equalities hold for each sample path $i \in \{1, \ldots, N\}$ such that $v_{ij}^{ij} = \kappa^i_s$ and $b^i = 1$, $u^i = \begin{cases} 1, & \text{if there exists } j \in S^i_1 \cup S^i_2 \text{ and } s \in \{1, \ldots, \ell - 1\} \text{ such that } v_{ij}^{ij} = \kappa^i_s \text{ and } b^i = 1, \\ 1, & \text{if } b^i = 1, \\ 0, & \text{otherwise.} \end{cases}$ (EC.26)

Now for each sample path $i \in \{1, \ldots, N\}$, define the integer

$$
\sigma^i \triangleq \begin{cases} T^i, & \text{if } b^i = 1, \\ T, & \text{if } b^i = 0. \end{cases}
$$

Then the following equalities hold for each sample path $i \in \{1, \ldots, N\}$:

$$
\frac{1}{|K^i| - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i) b^i (1 - w_{ij}^i) + \frac{1}{\ell^i - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(1 - b^i)(1 - u_{ij}^i)
$$

$$
= \begin{cases} \frac{1}{|K^i| - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(1 - w_{ij}^i), & \text{if } b^i = 1, \\ \frac{1}{\ell^i - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(1 - u_{ij}^i), & \text{if } b^i = 0 \\ \min_{j \in S^i_1 : b^i = 1} v_{ij}^{ij}, & \text{if } b^i = 1, \\ \min_{j \in S^i_1 \cup S^i_2 : b^i = 1} \left\{ \min_{v_{ij}^{ij}, v_{ij}^{ij}} \right\}, & \text{if } b^i = 0 \\ \min_{j : \sigma^i \leq \sigma^i} v_{ij}^{ij}, & \text{if } b^i = 0 \\ 
\end{cases}
$$

Indeed, the first equality follows from algebra, the second equality follows from lines (EC.25) and (EC.26), and the third equality follows from line (EC.27). Since the decision variables $b^i, \ldots, b^N \in \{0, 1\}$ were chosen arbitrarily, we have shown that (LB-1) and (BP-1) are equivalent. \(\square\)

**Proof of Theorem 6.** We readily observe that the following equality holds for all feasible decision variables to the optimization problem (BP-1):

$$
\frac{1}{\ell^i - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(1 - b^i)(1 - u_{ij}^i) = \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(1 - u_{ij}^i).
$$

Moreover, we observe that the following inequality is satisfied for all feasible decision variables to the optimization problem (BP-1):

$$
\frac{1}{|K^i| - 1} \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)b^i(1 - w_{ij}^i) \geq \sum_{\ell = 1}^{\ell^i - 1} (\kappa_{\ell + 1}^i - \kappa_{\ell}^i)(b^i - w_{ij}^i).
$$

Combining the above two observations, we have shown that the optimal objective value of (LB) is a lower bound on the optimal objective value of (BP-1).

To finish the proof of Theorem 6, it remains to be shown that the optimal objective value of (LB) is a lower bound on the optimal objective value of (RO). Indeed, let $J^{\text{RO}}, J^{\text{IP}}, J^{\text{LB-1}}, J^{\text{BP-1}},$ and $J^{\text{LB}}$ denote the optimal objective values of the optimization problems (RO), (IP), (LB-1), (BP-1), and (LB). We recall from Proposition 1 and Lemma 1 that the equalities $J^{\text{RO}} = J^{\text{IP}}$ and $J^{\text{LB-1}} = J^{\text{BP-1}}$ hold. Moreover, we showed in §4.2 that the inequality $J^{\text{IP}} \geq J^{\text{LB-1}}$ holds as well. Combining these relationships, we see that

$$
J^{\text{RO}} = J^{\text{IP}} \geq J^{\text{LB-1}} = J^{\text{BP-1}} \geq J^{\text{LB}},
$$

which concludes our proof of Theorem 6. \(\square\)
Proof of Proposition 4. We observe that \( LB \) is equivalent to a binary linear optimization problem with \( \mathcal{O}(NT) \) binary decision variables and \( \mathcal{O}(N^2 + NT) \) constraints, with each constraint of the form \( \lambda_i \geq \lambda_j \). It thus follows from Picard (1976, §3) that the optimization problem \( LB \) is equivalent to a problem of computing the maximal closure of a directed graph with \( \mathcal{O}(NT) \) nodes and \( \mathcal{O}(N^2 + NT) \) edges. Furthermore, Picard (1976, §4) shows that any maximal closure problem can be solved by computing the maximum flow in an augmented graph of identical size. Applying the algorithm of Orlin (2013) to compute the maximum flow in this augmented graph, we obtain an \( \mathcal{O}(N^2 T(N + T)) \) algorithm for solving \( LB \). \( \square \)

Proof of Proposition 5. It is shown in the proof of Theorem 6 that the inequality \( J^{LB-1} \geq J^{LB} \) holds. Next, let \( \sigma^1 \in \mathcal{T}^1, \ldots, \sigma^N \in \mathcal{T}^N \) be an optimal solution to \( (LB-1) \), and define the following variables:

\[
b^i = \begin{cases} 1, & \text{if } \sigma^i = T^i, \\ 0, & \text{if } \sigma^i \neq T^i, \end{cases}
\]

\[
w^i = \begin{cases} 1, & \text{if there exists } j \in S^i_1 \text{ and } s \in \{1, \ldots, \ell - 1\} \text{ such that } v_{ij}^s = \kappa^i_s \text{ and } b^i = 1, \\ 0, & \text{otherwise}, \end{cases}
\]

\[
u^i = \begin{cases} 1, & \text{if } b^i = 1, \\ 0, & \text{otherwise}. \end{cases}
\]

We observe that \( b, w, u \) is a feasible but possibly suboptimal solution to \( LB \), and it follows from the proof of Lemma 1 and the above construction of \( b, w, u \) that the following equality holds for each sample path \( i \in \{1, \ldots, N\} \):

\[
\min_{j: \sigma^i \leq \sigma^j} v_{ij}^s = \sum_{\ell=1}^{\kappa^i_{\ell+1} - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)(1 - w^\ell_i) + \sum_{\ell=1}^{L^i - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)(1 - b^i)(1 - u^i). \quad (EC.28)
\]

Therefore,

\[
J^{LB} \geq \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{\ell=1}^{\kappa^i_{\ell+1} - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)(1 - w^\ell_i) + \sum_{\ell=1}^{L^i - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)(1 - b^i) \right)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{\ell=1}^{\kappa^i_{\ell+1} - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)b^i(1 - w^\ell_i) + \sum_{\ell=1}^{L^i - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)(1 - u^i) \right) + \sum_{\ell=1}^{L^i - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)w^i(b^i - 1)
\]

\[
= J^{LB-1} - \frac{1}{N} \sum_{i=1}^{N} \sum_{\ell=1}^{\kappa^i_{\ell+1} - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)w^\ell_i(1 - b^i) + \sum_{\ell=1}^{L^i - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)w^i(b^i - 1)
\]

\[
= J^{LB-1} - \frac{1}{N} \sum_{i:b^i=0} \sum_{\ell=1}^{\kappa^i_{\ell+1} - 1} (\kappa^i_{\ell+1} - \kappa^i_\ell)w^\ell_i, \quad (EC.29)
\]

where the inequality holds because \( b, w, u \) is a feasible but possibly suboptimal solution to \( LB \), the first equality follows from algebra, the second equality holds because of line (EC.28) and because \( \sigma^1 \in \mathcal{T}^1, \ldots, \sigma^N \in \mathcal{T}^N \) is an optimal solution for \( (LB-1) \), and (EC.29) follows from algebra.
It also follows from our construction of $w$ that the following equalities are satisfied for each sample path $i \in \{1, \ldots, N\}$:

$$- \sum_{\ell=1}^{|\mathcal{K}|-1} (\kappa_{\ell+1}^i - \kappa_\ell^i) w_{\ell}^i = \sum_{\ell=1}^{|\mathcal{K}|-1} (\kappa_{\ell+1}^i - \kappa_\ell^i)(1 - w_{\ell}^i) - \sum_{\ell=1}^{|\mathcal{K}|-1} (\kappa_{\ell+1}^i - \kappa_\ell^i)$$

$$= \min \left\{ \min_{\ell:w_\ell^i=1} \kappa_\ell^i, \kappa_{|\mathcal{K}|}^i \right\} - \kappa_{|\mathcal{K}|}^i$$

$$= \min \left\{ \min_{j \in S_i:b_j=1} v_{T,j}^{i}, v_{T,T}^i \right\} - v_{T,T}^i$$

$$= - \max \left\{ v_{T,T}^i - \min_{j \in S_i:s_j=T} v_{j,T}^i, 0 \right\}. \quad (EC.30)$$

Combining lines (EC.29) and (EC.30) with our construction of $b$, we have shown that

$$J^{LB} \geq J^{LB-1} - \frac{1}{N} \sum_{i:b_i=0} \sum_{\ell=1}^{|\mathcal{K}|-1} (\kappa_{\ell+1}^i - \kappa_\ell^i) w_{\ell}^i$$

$$\geq J^{LB-1} - \frac{1}{N} \sum_{i:b_i=0} \max \left\{ v_{T,T}^i - \min_{j \in S_i:s_j=T} v_{j,T}^i, 0 \right\}$$

$$= J^{LB-1} - \frac{1}{N} \sum_{i:s_i \neq T} \max \left\{ v_{T,T}^i - \min_{j \in S_i:s_j=T} v_{j,T}^i, 0 \right\},$$

which concludes our proof. □

**Proof of Corollary 1.** If $\sigma^1 = T, \ldots, \sigma^N = T$ is an optimal solution to (RO), then

$$J^{RO} = J^{IP} = J^{LB-1} = J^{LB}.$$ 

Indeed, the first equality follows from Proposition 1. The second inequality holds because $\sigma^1 = T, \ldots, \sigma^N = T$ is a feasible solution to (LB-1), the corresponding objective values of (IP) and (LB-1) are identical, and the inequality $J^{IP} \geq J^{LB-1}$ always holds. The final equality follows from Proposition 5. □

**Proof of Corollary 2.** The proof is identical to that of Corollary 1. □

**Proof of Corollary 3.** Consider any instance of the robust optimization problem (RO) in which the number of periods is equal to two, and let $\sigma^1, \ldots, \sigma^N \in \{1, 2\}$ be any arbitrary solution for the corresponding instance of (IP). For each sample path $i \in \{1, \ldots, N\}$, let us define

$$\bar{\sigma}^i \triangleq \begin{cases} \sigma^i, & \text{if } T^i = 1, \\ 2, & \text{if } T^i = 2. \end{cases}$$

We observe that $\bar{\sigma}^1, \ldots, \bar{\sigma}^N \in \{1, 2\}$ is a feasible solution for (LB-1) and that

$$\frac{1}{N} \sum_{i=1}^{N} \min_{j: \sigma^j \leq \sigma^i} v_{j \sigma^i} \leq \frac{1}{N} \sum_{i=1}^{N} \min_{j: \sigma^j \leq \bar{\sigma}^i} v_{j \bar{\sigma}^i}.$$ 

Because $\sigma^1, \ldots, \sigma^N \in \{1, 2\}$ were chosen arbitrarily, we have shown that (RO) is a lower bound on (LB-1). Since (LB-1) is always a lower bound on (RO), we conclude that (RO) and (LB-1) are equivalent for any optimal stopping problem with two periods.
Now consider any instance of (LB-1) with two periods, and let $\sigma^1 \in T^1, \ldots, \sigma^N \in T^N$ be an optimal solution to (LB-1). For each sample path $i$, we observe that $\sigma^i \neq T^i$ if and only if the equalities $\sigma^i = 2$ and $T^i = 1$ are both satisfied. Therefore, for any sample path $i$ that satisfies $\sigma^i \neq T^i$, it must be the case that
\[
\min_{j \in S^i_1: \sigma^j = T^j} v_{T^j}^{i,j} = \min_{j: \sigma^j = 1} v_{T^j}^{i,j} \geq v_{T^i}^{i,i}.
\]

We have thus shown that the equality $\sum_{i: \sigma^i \neq T^i} \max\{v_{T^i}^{i,i} - \min_{j \in S^i_1: \sigma^j = T^j} v_{T^j}^{i,j}, 0\} = 0$ holds for the optimal solution to (LB-1). Therefore, Proposition 5 implies that (LB-1) and (LB) are equivalent for any optimal stopping problem with two periods. This concludes our proof. \qed

Appendix E: Additional Numerical Results

In this appendix, we present numerical results for additional parameter settings which were omitted from §5.2 due to length considerations.
Figure EC.1  Barrier Option (Symmetric) - Visualization of Robust Optimization Stopping Rules.

Note. Each plot shows the exercise policies obtained from solving the optimization problem (LB) constructed from a training dataset of size $N = 10^3$ and with the robustness parameter selected using a validation set of size $\tilde{N} = 10^3$. The problem parameters are the same as those shown in Table 1.
Figure EC.2  Barrier Option (Asymmetric) - Visualization of Robust Optimization Stopping Rules.

Note. Each plot shows the exercise policies obtained from solving the optimization problem (LB) constructed from a training dataset of size $N = 10^3$ and with the robustness parameter selected using a validation set of size $\tilde{N} = 10^3$. The problem parameters are the same as those shown in Table 2.
Figure EC.3  Barrier Option (Symmetric) - Impact of Robustness Parameter on Reward.

Note. Each plot shows the robust objective value and expected reward of policies obtained by solving the optimization problem (LB) constructed from training datasets of size $N = 10^3$. The problem parameters are the same as those shown in Table 1.
Figure EC.4  Barrier Option (Asymmetric) - Impact of Robustness Parameter on Reward.

Note. Each plot shows the robust objective value and expected reward of policies obtained by solving the optimization problem (LB) constructed from training datasets of size $N = 10^3$. The problem parameters are the same as those shown in Table 2.
Figure EC.5  Barrier Option (Symmetric) - Impact of Robustness Parameter on Computation Time.

Note. Each plot shows the computation times from solving the optimization problem (LB) constructed from training datasets of size $N = 10^3$. The problem parameters are the same as those shown in Table 1.
Figure EC.6  Barrier Option (Asymmetric) - Impact of Robustness Parameter on Computation Time.

Note. Each plot shows the computation times from solving the optimization problem (LB) constructed from training datasets of size $N = 10^3$. The problem parameters are the same as those shown in Table 2.