Structural Properties of a Class of Robust Inventory and Queueing Control Problems

Zeynep Turgay,[1,2] Fikri Karaesmen,[1,2] Egemen Lerzan Örmeci [1,2]

[1] *Migros T.A.Ş. R&D Centre, Ataşehir, İstanbul 34758, Turkey*

[2] *Department of Industrial Engineering, Koç University, Sariyer, İstanbul 34450, Turkey*

Received 14 February 2015; revised 6 April 2017; accepted 17 July 2017
DOI 10.1002/nav.21748
Published online 7 August 2017 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: In standard stochastic dynamic programming, the transition probability distributions of the underlying Markov Chains are assumed to be known with certainty. We focus on the case where the transition probabilities or other input data are uncertain. Robust dynamic programming addresses this problem by defining a min-max game between Nature and the controller. Considering examples from inventory and queueing control, we examine the structure of the optimal policy in such robust dynamic programs when event probabilities are uncertain. We identify the cases where certain monotonicity results still hold and the form of the optimal policy is determined by a threshold. We also investigate the marginal value of time and the case of uncertain rewards. Naval Research Logistics 65: 699–716, 2018

Keywords: stochastic dynamic programming; robust optimization; inventory control; queueing control

1. INTRODUCTION

In many practical optimization problems, the input parameters to the problem are not known with certainty; rather they are either estimated from the existing data or the possible values that they can take can be specified by expert opinions. This uncertainty, if ignored, may cause significant suboptimality or infeasibility for the solution considered. Robust optimization is a specific methodology that addresses this problem and has received much attention lately (see [2] for a comprehensive presentation).

Our focus in this article is on a set of stochastic dynamic problems from inventory and queueing theory where the transition probability distributions of the underlying Markov Chains may be uncertain. The models considered include discrete-time versions of some well-established cases such as service rate control and admission control problems of Lippman [17], the stock rationing problem of Ha [9] and the dynamic revenue management problem of Lautenbacher and Stidham [15].

To address the uncertainty on the probability distributions in such problems, we formulate a robust stochastic dynamic program with a maximin approach. This approach defines a game between the controller (system manager) and Nature. For instance, in the context of demand admission control, the controller's aim is to maximize the expected profit by choosing the allowable actions (for example by admitting a given class of demand or not), whereas Nature tries to minimize the controller's expected profit by choosing the worst-possible parameters (for example arrival probabilities for different classes) and acts upon observing the controller's choice. This formulation is known as the robust counterpart of the standard problem. The robust optimal policy designates the policy which yields the highest expected profit result after minimization by Nature. In this formulation, the controller acts upon the worst-case scenario, which does not always happen. Consequently, his actions can be labeled as the most conservative with respect to maximizing the expected revenue. To overcome this problem, semirobust Markov Decision Process (MDPs) are developed, which allows the controller to have varying degrees of conservatism.

The robust formulation of a MDP with an uncertain transition probability distribution goes back to Satia and Lave [22] who propose a solution by a policy iteration approach. White and Eldeib [24] present value-iteration-based numerical algorithms and bounds for polyhedral uncertainty sets on transition probabilities. Bagnell and coworkers [1] propose a robust value iteration and discuss the related computational complexity issues. Nilim and El Ghaoui [18] and Iyengar [10] simultaneously study robust stochastic dynamic

Correspondence to: E. Lerzan Örmeci (lormeci@ku.edu.tr)

programs and establish the existence of a robust Bellman recursion whose solution yields the robust value function and the corresponding optimal policy. In addition, both papers emphasize that under appropriate choice of uncertainty sets, the additional complexity brought by the robust formulation is reasonable when the standard formulation has a tractable solution. Paschalidis and Kang [19] explore the effect of using stronger uncertainty set formulations and investigate a specific queueing control problem. Delage and Mannor [7] consider chance-constrained MDPs with uncertain parameters and show that some important instances of this class of problems is computationally tractable. Kardeş et al. [12] investigate infinite horizon discounted stochastic games with uncertain transition probabilities, and establish the existence of equilibria and propose a method for computing the equilibria. Finally, Xu and Mannor [25] propose a more sophisticated model of parameter uncertainty which allows multiple nested uncertainty sets. The above papers directly take into account the probability structure of the problem. An alternative approach is to formulate a corresponding deterministic optimization formulation that takes into account the uncertainty. Bertsimas and Thiele [4] consider such a formulation in the context of inventory control. They are able to interpret the solution obtained from the deterministic formulation in terms of a dynamic control policy and obtain approximate solutions to the motivating stochastic inventory problem. In a different context, Bertsimas et al. [3] obtain bounds for queuing network performance measures based on a deterministic formulation.

In this article, we mainly focus on investigating the structural properties of optimal policies in a class of robust and semi-robust MDPs. The queueing and inventory control literature has a strong tradition in exploring the structure of optimal policies. This is in part due to the computational efficiency of structured policies. However, the main motivation for looking for structured policies is that they are usually expressed in a few parameters and tend to be easy to understand and implement. There are known effective techniques to investigate the structure of the solution of a stochastic dynamic program. Event-based dynamic programming (EBDP) proposed by Koole [13] and further extended in Koole [14] streamlines this procedure for a class of queueing control problems. Recently, Çil et al. [6] use EBDP to explore structural properties in a class of production/inventory problems and propose an extension to study the effects of perturbations of input parameters. In this article, we use EBDP to formulate the robust MDPs and investigate the structure of their optimal policies.

There are also a number of recent papers that investigate specific robust MDP problems in queueing or inventory control. Birbil et al. [5] consider a dynamic revenue management problem and show that there is a very efficient solution under an ellipsoidal models of uncertainty. Turgay et al. [26] investigate a similar problem with interval and polyhedral uncertainty sets and establish monotonicity results for nested uncertainty sets. Lim and Shanthikumar [16], and Jain et al.[11] consider robust versions of dynamic pricing problems for queueing systems using entropy-based models of uncertainty. Rusmevichientong and Topaloglu [20] explore a robust assortment optimization problem and show that optimal assortment policy can be computed efficiently by giving a complete characterization of the optimal policy.

All of these papers obtain results on the structure of optimal policies for specific problems by exploiting certain properties of the uncertainty set. In contrast, we consider general models in the EBDP framework with general uncertainty sets to establish structural results that hold for a class of problems.

As the literature shows, the robust dynamic programming problems have received a great interest recently. Our contributions in this area can be summarized as follows: (1) Our results on robust MDPs apply to the EBDP framework with general uncertainty sets on the transition probabilities, (2) Our results extend to the semirobust MDP models with general uncertainty sets on the transition probabilities, (3) A discussion of the robust MDPs with general uncertainty sets on other parameters is included, where the structure of optimal policy in a specific system is derived. By (1), our results are valid for all models that can be generated within the EBDP framework in addition to the models considered in this paper, whereas (2) allows us to compare the performances of controllers with different levels of conservatism in a computationally efficient way. By (3), we observe the difficulties for the robust MDP models when the other parameters are uncertain.

The organization of the paper is as follows; in Section 2, we present the EBDP framework for the nominal MDPs, while Section 3 defines the robust counterparts of the nominal MDPs. Section 4 establishes the structural properties of both nominal and robust MDPs under certain conditions on the underlying operators that form the EBDP framework. In Section 5, we introduce the operators commonly used in literature within EBDP framework, and show that they all have the desired properties that guarantee certain structural properties of the underlying value functions. Section 6 presents models analyzed in literature that fall into the EBDP framework. The purpose here is twofold: (1) to illustrate the intriguing effects of our results, and (2) to show the generality of our results. Section 7 extends our results when the uncertainty on the parameters are represented by two or more uncertainty sets, rather than one as in the regular robust definition. However, we also show that we cannot show monotonicity of the value functions with respect to nested uncertainty sets. Finally, we present our conclusions and research perspectives in Section 8 where we address two important extensions. The first extension explores uncertainty that takes place within a dynamic programming operator rather than in event probabilities. The second extension considers multi-dimensional control problems.

2. NOMINAL EBDP FRAMEWORK

In this section, we describe the EBDP framework introduced by Koole [13], which provides the basis of our analysis. The problems we consider in the scope of this paper are represented in discrete time. In our setting, the stages (time) are denoted by $t = 0, 1, \ldots, T$, where T is the last stage in the horizon. Let $i = 0, 1, 2, \ldots, n$ denote the event indices, where events $1, 2, \ldots, n$ are events which may lead to a state transition (if the appropriate action is selected) while event 0 corresponds to a fictitious event where no observable real event (and therefore no state transition) occurs. The system state x can take values in set X (i.e., $x \in X$) at any stage t, where X is a subset of integers. We let a_i denote the controller action regarding event i, so that an action can be defined by $a_i \in \{0, 1\}$ as in the case admission/rejection or by $a_i \in \Re$ as in the case of pricing, where $\Re$ is the set of real numbers. Note that the controller is allowed to choose her/his actions independently for all (x, t) pairs, and there is no restriction among the actions regarding different events. The action vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ denotes the actions for all possible events. At each stage, depending on the controller action, $R_i(a, x)$ ($C_i(a, x)$) is gained (incurred) as an immediate reward (cost). The randomness is characterized by a transition probability distribution at each stage which is assumed to be independent of prior uncertainties. The probability that event i occurs at stage t is given by $p_{i,t}$, so $\sum_{i=0}^{n} p_{i,t} = 1$ for all t. When event i occurs in state x at stage t, the conditional probability that the next state is y if controller selects action a is denoted by $q_t(a, x, y|i)$, hence $\sum_y q_t(a, x, y|i) = 1$ for all t, a, x, and i. We express the optimal value function $v_t(x)$ according to the following equation when the time left until the end of the horizon is $T - t$:

$$v_t(x) = \max_{\boldsymbol{a}} \sum_{i=0}^{n} p_{i,t} \sum_{y \in X} q_t(a_i, x, y|i)(v_{t+1}(y) + R_i(a_i, x)). \tag{1}$$

Note that the action vector $\boldsymbol{a}$ is a function of x and t, but we suppress this for notational simplicity. The event-based approach allows us to define the value function $v_t(x)$ as a convex combination of operators, T_i, where T_i is defined as follows:

$$T_i v_{t+1}(x) = \max_{a_i} \sum_{y \in X} q_t(a_i, x, y|i)(v_{t+1}(y) + R_i(a_i, x)),$$

hence $v_t(x)$ can be written as:

$$\begin{aligned} v_t(x) &= \sum_{i=0}^{n} p_{i,t} \max_{a_i} \sum_{y \in X} q_t(a_i, x, y|i)(v_{t+1}(y) + R_i(a_i, x)) \\ &= \sum_{i=0}^{n} p_{i,t} T_i v_{t+1}(x). \end{aligned} \tag{2}$$

3. ROBUST COUNTERPART OF NOMINAL FRAMEWORK

This section introduces a robust version of the dynamic program (which will be referred to as the nominal problem) described in Section 2. The robust formulation assumes that a subset of the problem parameters is uncertain, where robustness is captured as a game between the controller and an adversary (Nature) that chooses parameters and takes the min-max approach as in [2]. In this section, we assume that the transition probability distributions are uncertain. Typically, the system controller decides on his actions before observing the uncertain parameters. Once his decisions are taken, Nature selects these parameters from an uncertainty set in order to minimize the expected profit of the system. Hence, the controller has to consider the worst-case scenario in terms of the transition probability distribution when choosing his actions. Our robust formulations for stochastic dynamic programs are based on the maximin approach suggested by Nilim and El Ghaoui [18] and Iyengar [10].

In the robust formulation, the transition probabilities belong to an uncertainty set, rather than being fixed values as in the nominal problem. In the EBDP approach, the transition probabilities consist of two components as observed in Eq. (1): $q_t(a_i, x, y|i)$ and $p_{i,t}$. We assume that $q_t(a_i, x, y|i)$ is known with certainty and $p_{i,t}$ is uncertain and can depend on the chosen action and the system state.

We let $p_{i,t}(x, \boldsymbol{a})$ be the probability of observing event i when the system is in state x and action vector $\boldsymbol{a}$ is chosen at stage t. Note that $\boldsymbol{a} = (a_1, \ldots, a_n)$, so $\boldsymbol{a}$ specifies the actions for all possible events. Naturally, $\boldsymbol{a}$ depends on state x and stage t. However, as in the previous section, we suppress this dependency for notational simplicity. We assume that $p_{i,t}(x, \boldsymbol{a})$ belongs to an uncertainty set $\mathcal{P}_t$, where $\mathcal{P}_t$ represents the available information on the event probability distribution. Our uncertainty model is based on the model proposed by Nilim and El Ghaoui [18], where the so-called rectangularity property is the main condition for obtaining a recursive solution. When the rectangularity property holds, nature can independently select its action for every stage, state and the controller action. This assumption essentially implies that Nature's choice of a particular distribution at time t or in state x does not limit its choices in the future or in other states. In addition to this basic assumption, we also assume that the uncertainty set at each stage t, $\mathcal{P}_t$, is independent of the controller's action vector $\boldsymbol{a}$ as well as the state x. Note that this additional assumption is not very restrictive when handling queueing/inventory problems: The state-dependent event probability distribution is difficult to estimate from limited statistical data. Therefore, it is natural to assume that uncertainty sets depend on common global estimates rather than on state-dependent estimates. Moreover, in typical examples, these probabilities represent demand or processing rates which do not depend on the state of the system or the actions taken.

Now, we let $w_t(x)$ be the robust counterpart of the value function given in (2), and consider the following robust DP equation:

$$w_t(x) = \max_{\boldsymbol{a}} \left\{ \min_{p_t(x,\boldsymbol{a})} \sum_{i=0}^{n} p_{i,t}(x,\boldsymbol{a}) \sum_{y \in X} q_t(a_i, x, y|i) \times (w_{t+1}(y) + R_i(a_i, x)) \right\}, \quad (3)$$

where $\boldsymbol{p}_t(x,\boldsymbol{a}) = (p_{1,t}(x,\boldsymbol{a}), p_{2,t}(x,\boldsymbol{a}), \ldots, p_{n,t}(x,\boldsymbol{a}))$. This problem can be solved recursively due to the above assumptions, as shown by Nilim and El Ghaoui [18] and Iyengar [10].

Our next result, Theorem 1, shows that the optimal action at any stage never depends on the choice of the event probability distribution at that stage. This allows to change the order of maximization and minimization in Eq. (3). Then the robust value functions can be presented analogously to the nominal value functions, as shown in the second part of Theorem 1. In Section 4.2, this result will play a key role in extending the structural properties of the nominal MDPs to the corresponding robust MDPs.

THEOREM 1: (i) The optimal policy of the controller does not depend on the Nature's posteriori decision.

(ii) The robust value function $w_t(x)$ can be expressed as:

$$w_t(x) = \min_{p_t(x)} \left\{ \sum_{i=0}^{n} p_{i,t}(x) T_i w_{t+1}(x) \right\}. \quad (4)$$

PROOF: Before proceeding with the proof, we state the following observation that will be used below: Let's suppose $f(x, y) \in \Re$ is any function. Further suppose that for every $y \in \Re$, $\mathrm{argmax}_x f(x, y) = x^*$ corresponds to a specific value $x^* \in \Re$. Then it is easy to show that $\max_x \{\min_y f(x, y)\} = \min_y f(x^*, y)$.

(i) Let $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{n,t})$ be any transition probability distribution in uncertainty set $\mathcal{P}_t$. If Nature decides to use $\boldsymbol{p}$ before observing the controller's action, the controller's optimal policy is determined by the solution of the following problem:

$$\max_{\boldsymbol{a}} \sum_{i=0}^{n} p_{i,t} \sum_{y \in X} q_t(a_i, x, y|i)(w_{t+1}(y) + R_i(a_i, x)), \quad (5)$$

by Eq. (3). We first observe that the optimal action corresponding to an event does not have any effect on the optimal actions of other events. Then, by the above observation, it is clear that the controller's optimal decision is unaffected by the Nature's choice of $\boldsymbol{p}_t$. This allows us to change the order of maximization and summation in (5):

$$\sum_{i=0}^{n} p_{i,t} \max_{\boldsymbol{a}} \sum_{y \in X} q_t(a_i, x, y|i)(w_{t+1}(y) + R_i(a_i, x)), \quad (6)$$

which can then be written in terms of operators T_i's as follows:

$$\sum_{i=0}^{n} p_{i,t} T_i w_{t+1}(x). \quad (7)$$

The optimal actions of the controller are the outputs of the operators, which are the same for each possible choice of $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{n,t})$. Hence, we can conclude that the optimal policy of the controller does not depend on the Nature's posteriori decision.

(ii) Let $\boldsymbol{a}_t^*(x) = \boldsymbol{a}^* = (a_1^*, \ldots, a_n^*)$ be the optimal action of the controller found as a result of the operators. From part (1), we know that $\boldsymbol{a}_t^*(x)$ determines the controller's optimal policy, regardless of the Nature's posteriori decision. Then, in Eq. (3), it is enough for Nature to find the action which minimizes the expected revenue of the controller for $\boldsymbol{a}_t^*(x)$ only. Hence, Eq. (3) can be written as:

$$w_t(x) = \min_{p_t(x,\boldsymbol{a}^*)} \sum_{i=0}^{n} p_{i,t}(x,\boldsymbol{a}^*) \sum_{y \in X} q_t(a_i^*, x, y|i) \times (w_{t+1}(y) + R_i(a_i^*, x)), \quad (8)$$

But then we have $T_i w_{t+1}(x) = \sum_{y \in X} q_t(a_i^*, x, y|i)(w_{t+1}(y) + R_i(a_i^*, x))$ by definition. Moreover, $\boldsymbol{a}_t^*(x)$ is a function of x and t, so that the original action-dependent probability distribution, $p_{i,t}(x,\boldsymbol{a})$ in (3) can be replaced by $p_{i,t}(x)$. Accordingly, Eq. (8) can be expressed as follows:

$$w_t(x) = \min_{p_t(x)} \sum_{i=0}^{n} p_{i,t}(x) T_i w_{t+1}(x). \quad (9)$$

This completes the proof. □

By Theorem 1, we establish that the controller's action does not depend on the Nature's posteriori action for all models that can be constructed within EBDP framework. This enables us to define the optimality equation as in (4). Consequently, it is enough to solve the inner problem of Nature only for one action of the controller, instead of solving it for all possible actions of the controller, at each stage and state. This not only improves the solution time significantly especially when the controller's action set is large, but also leads to useful structural properties.

Finally, we would like to remark about what this result does not mean: Theorem 1 does not imply that actions $a_t(x)$ does not depend on Nature's choices in all stages. Nature's actions in stages t' with $t' > t$, $p_{i,t'}(x)$, affect the value functions $w_t(x)$, which, in turn, influence actions $a_t(x)$. More specifically, actions $a_t(x)$ depend on both the uncertainty sets and Nature's choices in stages $t+1, t+2, \dots, T$. Moreover, Nature's optimal actions, $p_{i,t}(x)$, depend on the controller's actions, $a_t(x)$, unless some special conditions hold. To summarize: (1) $a_t(x)$ does not depend on $p_{i,t}(x)$, but it depends on $p_{i,t'}(x)$ and $\mathcal{P}_{t'}$ with $t' > t$. (2) $a_t(x)$ does not depend on $p_{i,t}(x)$, but $p_{i,t}(x)$ depends on $a_t(x)$.

4. STRUCTURAL PROPERTIES FOR NOMINAL AND ROBUST MDPS

In this section, we first present the structural properties of nominal MDPs, then extend these to robust MDPs within the EBDP framework. For this purpose, we first define the properties of the value functions that can guarantee the existence of monotone optimal policies. In an EBDP framework, if the operators preserve these properties, then the value functions will also have them due to the construction of EBDP and the value iteration algorithm.

4.1. Structural Properties for Nominal MDPs

We are interested in the following properties of the value function: (1) Nonincreasingness (nondecreasingness) in x refers to $v_t(x) \leq (\geq) v_t(x-1)$ for all $x \geq 1$ and all t, and (2) concavity (convexity) in x refers to: $v_t(x+1) - v_t(x) \leq (\geq) v_t(x) - v_t(x-1)$ for all $x \geq 1$ and all t. Concavity and monotonicity properties of the value functions determine the structure of the optimal policies as well, for example, concavity of the value function leads to the optimality of threshold policies for admission control and optimal base stock policies for inventory control.

Next, we consider the supermodularity/submodularity of the value functions in x and t, which has not been studied in this perspective to our knowledge: Supermodularity (submodularity) (in (x, t)) refers to $v_t(x) - v_t(x-1) \geq (\leq)$ $v_{t+1}(x) - v_{t+1}(x-1)$ for all $x \geq 1$ and all t. These properties ensure that optimal thresholds of structured optimal policies are also monotone in time. The preservation of supermodularity is related to the marginal benefit (MB) of the operator, which is first defined by Çil et al. [6]. MB of an operator indicates the difference of the value functions between the two systems, where one system observes the event that corresponds to the operator and the other system remains in the same state. Hence, the MB of the operator, B_i, is given by:

$$B_i v(x) = T_i v(x) - v(x). \tag{10}$$

Then, Eq. (2) can be written in terms of B_i as follows:

$$v_t(x) = \sum_{i=0}^{n} p_{i,t} B_i v_{t+1}(x) + v_{t+1}(x), \tag{11}$$

since $\sum_{i=0}^{n} p_{i,t} = 1$ for all t. This representation of the value function sets a direct relationship between the MB function and the supermodularity/submodularity properties in the context of these systems. If the MB function of an operator T_i, $B_i v_t(x)$, is nonincreasing (nondecreasing) in x, then submodularity (supermodularity) in x and t is preserved by the corresponding operator, $T_i v_t(x)$.

To summarize, we are interested in the operators which preserve the following properties: nonincreasingness (NI), nondecreasingness (ND), concavity (C). Moreover, the MB functions of the operators should be either nonincreasing (MB-NI) or nondecreasing (MB-ND). Properties NI, ND, and C are well studied in the EBDP framework. In particular, part (i) of Theorem 2 is established by Koole [13], which is included for completeness. Part (ii) of Theorem 2 establishes the link between the monotonicity of MB functions and the submodularity (supermodularity) of the value functions.

THEOREM 2:

(i) If a nominal value function, $v_t(x)$, can be represented as a convex combination of operators with properties NI (ND) and C, then $v_t(x)$ is NI (ND) and C.
(ii) Furthermore, if all the operators that constitute the nominal value function have property MB-ND (MB-NI), then $v_t(x)$ is supermodular (submodular) in x, t.

PROOF: We provide only the proof of part (*ii*) when all the operators have property MB-ND, that is, $B_i v_t(x)$ is ND in x for all t, since the other case is similar. Then, we have:

$$\begin{aligned} v_t(x) - v_{t+1}(x) &= \sum_{i=0}^{n} p_{i,t} B_i v_{t+1}(x) \\ &\geq \sum_{i=0}^{n} p_{i,t} B_i v_{t+1}(x-1) \\ &= v_t(x-1) - v_{t+1}(x-1), \end{aligned} \tag{12}$$

where the first and last equalities are due to Eq. (11) and the inequality follows from the assumption that MB functions, $B_i v_t(x)$, are ND in x for all t. This completes the proof. □

REMARK: Theorem 2 establishes concavity under fairly general conditions, so that threshold policies are optimal for all models that can be represented by a convex combination of the operators above. It also ensures the supermodularity (submodularity) property for a general class of models under the limitation that all the operators of a given model should have ND (NI) MB functions. However, this limitation turns out to be more restrictive and problem-dependent. For instance, it is not natural to conceive a queueing system that consists of operators whose MB functions are all NI (or ND). By nature, queueing systems do not possess this property (see the queueing operators in Section 5). There are, however, plausible inventory systems that support the property.

4.2. Structural Properties for Robust MDPs

In this section, we show that the robust value functions have similar properties as nominal value functions for the problems of interest, regardless of the shape of the uncertainty set. Our main result extends Theorem 2 to the robust counterparts of the nominal problems under consideration. More explicitly, for any problem that can be expressed in the form of Eq. (2), all the structural results given in Theorem 2 extend to the value function of the robust counterpart, $w_t(x)$.

THEOREM 3:

(i) If a robust value function, $w_t(x)$, can be represented as a convex combination of operators with properties NI (ND) and C, then $w_t(x)$ is NI (ND) and C.
(ii) Furthermore, if all the operators that constitute the robust value function have property MB-ND (MB-NI), then $w_t(x)$ is supermodular (submodular) in x, t.

PROOF: Let $\boldsymbol{p}_t^*(x-1), \boldsymbol{p}_t^*(x), \boldsymbol{p}_t^*(x+1)$ denote the optimal choices of Nature in states x–1, x and $x+1$, respectively, at stage t, in the rest of the proof.

We first establish concavity. Assume that $w_{t+1}(x)$ is concave in x at stage $t+1$. By assumption, all operators that constitute the value function preserve concavity in x. Then, the convex combination of these operators also preserve this property. Hence, we can write the following inequality:

$$\sum_i p_{i,t}^*(x) T_i w_{t+1}(x-1) + \sum_i p_{i,t}^*(x) T_i w_{t+1}(x+1) \leq 2\left[\sum_i p_{i,t}^*(x) T_i w_{t+1}(x)\right]. \tag{13}$$

We know that:

$$\sum_i p_{i,t}^*(x-1) T_i w_{t+1}(x-1) \leq \sum_i p_{i,t}^*(x) T_i w_{t+1}(x-1),$$

by the optimality of $\boldsymbol{p}_t^*(x-1)$ in state $x-1$, and:

$$\sum_i p_{i,t}^*(x+1) T_i w_{t+1}(x+1) \leq \sum_i p_{i,t}^*(x) T_i w_{t+1}(x+1),$$

by the optimality of $\boldsymbol{p}_t^*(x+1)$ in state $x+1$. Then, the sum of the left-hand-sides (LHSs) of these two inequalities is less than the LHS of inequality (13), which proves that $w_t(x)$ is concave in x at stage t.

Next, we prove that the robust value function, $w_t(x)$ which is represented as a convex combination of operators with properties ND and C is also ND and C. The result for operators with properties NI and C can be proven similarly. We suppose that $w_{t+1}(x)$ is ND in x. Then $T_i w_{t+1}(x)$ is ND in x, since T_i is one of the operators defined in Table 1. Hence, we have:

$$\begin{aligned}\sum_i p_{i,t}^*(x-1) T_i w_{t+1}(x-1) &\leq \sum_i p_{i,t}^*(x) T_i w_{t+1}(x-1) \\ &\leq \sum_i p_{i,t}^*(x) T_i w_{t+1}(x),\end{aligned} \tag{14}$$

where the first inequality is due to the optimality of $\boldsymbol{p}_t^*(x-1)$ in state x–1, and the second inequality follows since $T_i w_{t+1}(x)$ is ND in x. This completes the proof of Part (i).

We prove part (2) only for the case when MB functions of all the operators are ND, since the proof of the other case is similar. First, let us note that by Eqs. (11) and (9), we have:

$$w_t(x) - w_{t+1}(x) = \sum_i p_{i,t}^*(x) B_i w_{t+1}(x).$$

On the other hand:

$$\begin{aligned}\sum_i p_{i,t}^*(x) B_i w_{t+1}(x) &\geq \sum_i p_{i,t}^*(x) B_i w_{t+1}(x-1) \\ &\geq \sum_i p_{i,t}^*(x-1) B_i w_{t+1}(x-1),\end{aligned}$$

where the first inequality follows since $B_i w_{t+1}(x)$ is ND in x, and the second inequality is due to the optimality of $\boldsymbol{p}_t^*(x-1)$ in state x–1, and. Therefore:

$$\begin{aligned}w_t(x) - w_{t+1}(x) &= \sum_i p_{i,t}^*(x) B_i w_{t+1}(x) \\ &\geq \sum_i p_{i,t}^*(x-1) B_i w_{t+1}(x-1) \\ &= w_t(x-1) - w_{t+1}(x-1).\end{aligned}$$

This completes the proof. □

Table 1. Definitions for the operators (Q denotes a queueing operator and I denotes an inventory operator, see Appendix A for more details).

Type	Operator name	Notation	Definition
I	Rationing	$T_{R_i} v(x)$	Special case of T_{BR_i} where $B = 1$
I	Batch rationing	$T_{\mathrm{BR}_i} v(x)$	$\max_{\kappa_i \in \min(x,B)} \{\kappa_i R_i + v(x - \kappa_i)\}$
I	Production rate	$T_{\mathrm{PR}_i} v(x)$	$\max_{\Pi \in [0,1]} \{-C_{\Pi_i} + \Pi_i v(x+1) + (1 - \Pi_i) v(x)\}$
I	Production	$T_{P_i} v(x)$	$\max \{v(x+1) - C_i, v(x)\}$
I	Inventory pricing	$T_{\mathrm{IP}} v(x)$	$\max_R \{\bar{F}_Z(R)[v(x-1) + R] + F_Z(R) v(x)\}$
Q	Admission	$T_{\mathrm{ADM}_i} v(x)$	Special case of T_{BADM_i} where $B = 1$
Q	Batch admission	$T_{\mathrm{BADM}_i} v(x)$	$\max_{\kappa_i} \{\kappa_i R_i + v(x + \kappa_i)\}$
Q	Controlled departure	$T_{\mathrm{CD}_i} v(x)$	$\max \{v(x-1) - C_i, v(x)\}$
Q	Departure rate	$T_{\mathrm{DR}_i} v(x)$	$\max_{\Pi \in [0,1]} \{-C_\Pi + \Pi v(x-1) + (1 - \Pi) v(x)\}$
Q	Queue pricing	$T_{\mathrm{QP}} v(x)$	$\max_R \{\bar{F}_Z(R)[v(x+1) + R] + F_Z(R) v(x)\}$
Q	Uncontrolled arrival	$T_{\mathrm{ARR}} v(x)$	$a(x) v(x+1) + (1 - a(x)) v(x)$
Q	Uncontrolled departure	$T_{\mathrm{DEP}} v(x)$	$b(x) v(x-1) + (1 - b(x)) v(x)$
-	Cost	$T_{\mathrm{COST}} v(x)$	$v(x) - h(x)$
-	Fictitious	$T_{\mathrm{FIC}} v(x)$	$v(x)$

Theorem 3 establishes a general result: the robust value function $w_t(x)$ has the same monotonicity properties with the nominal value function $v_t(x)$, regardless of the uncertainty set in stage t. Particularly, this result implies that the structure of the optimal policy of the robust problem is the same with the nominal problem. Hence, the robust problem could have the same computing cost with its nominal counterpart.

Here, we note the role of the assumption that event probabilities belong to the same uncertainty set for all states x in each period, that is, $p_{i,t}(x) \in \mathcal{P}_t$ for all x for given i and t. Theorem 3 will not be true when this assumption is relaxed. Consider the following counter example to the simplest property ND in the rationing problem defined by Eq. (2), where the controller aims to maximize the total expected reward over a finite horizon, N. We set the value function at the end of the horizon to 0 for all states, that is, $V_N(x) = 0$ for all x. In period N–1, it will be optimal for the controller to satisfy the demand for all types when $x > 0$, so that $T_i w_{N-1}(x) > 0$ for all i and $x > 0$ in (14). Then, Nature will aim to minimize the probability that a type-i demand occurs in period N–1 for all i. Now assume that there are two different uncertainty sets for states x and x–1 for all demand types i: $p_{i,N-1}(x) \in (0, 1)$ and $p_{i,N-1}(x-1) \in (0.4, 0.6)$ for all i. Note that for the nominal problem which uses the expectation over an uncertainty set, all of these sets will have the same demand probability 0.5. Hence, the structure of the nominal problem will not be affected. It is easy to see that Nature will select $p^*_{i,N-1}(x) = 0$ and $p^*_{i,N-1}(x-1) = 0.4$ for all i, which makes the right-hand side of (14) zero whereas the left-hand side will be strictly positive. Similar sets can be found for other properties also. Hence, the results of the theorem are violated when the uncertainty sets are allowed to change with respect to state x.

Finally, let us briefly discuss the infinite horizon extension. Iyengar [10] and Nilim and El Ghaoui [18] establish that the respective controller and nature policies are stationary for the infinite horizon problem. Moreover, Nilim and El Ghaoui [18] show that the optimal value function of the infinite horizon problem with a discounted cost function can be obtained as the unique limit of the finite horizon problem. This implies that the optimal policy structure can be extended to the infinite horizon case.

5. STRUCTURAL PROPERTIES OF SOME QUEUEING AND INVENTORY CONTROL PROBLEMS

In this section, we introduce a number of commonly-used operators for queueing and inventory problems as in Koole [13, 14] and Çil et al. [6]. Table 1 presents type ("I" for inventory and "Q" for queueing), names, notations, and definitions of these operators. Their detailed definitions as well as the structural properties they preserve are summarized in Table 4 in Appendix A.

From Koole [13] and Çil et al. [6], it is known that if an initial value function $v(x)$ is concave in x, then the operators in Table 1 preserve concavity in x. In addition, the queueing operators preserve the NI property whereas the inventory operators preserve the ND property in x. Note that additional conditions are necessary for general forms of T_{ARR} and T_{DEP}, and for T_{COST} to preserve these properties as given in Çil et al. [6]. Çil et al. [6] showed the following properties of the MB function:

PROPERTY 1 (Properties of the MB function):

1. The MB functions of the below operators are nonincreasing (NI) in x.

(a) The queueing operators, admission control $B_{A_i} v(x)$, batch admission $B_{\mathrm{BA}_i} v(x)$, queue pricing $B_{\mathrm{QP}_i} v(x)$, and uncontrolled arrival $B_{\mathrm{UA}_i} v(x)$,

(b) The inventory operators, production control $B_{P_i}v(x)$ and production rate control $B_{\mathrm{PR}_i}v(x)$.

2. The MB functions of the below operators are nondecreasing (ND) in x.

(a) The queueing operators, controlled departure $B_{\mathrm{CD}_i}v(x)$, departure rate $B_{\mathrm{DR}_i}v(x)$, uncontrolled departure $B_{\mathrm{UD}_i}v(x)$,
(b) The inventory operators, rationing $B_{R_i}v(x)$, batch rationing $B_{\mathrm{BR}_i}v(x)$ and inventory pricing $B_{\mathrm{IP}_i}v(x)$.

Property 1 shows that these operators have all the properties that satisfy the assumptions of Theorems 2 and 3, so that all the nominal and robust MDP models that can be represented by a combination of these operators have monotone and concave value functions. Moreover, if all the operators in a certain model have MB-NI (MB-ND) property, then the corresponding value functions are also supermodular (submodular) in x and t. However, note that, as remarked above, it is not possible to represent a meaningful queueing system with only the operators having MB-NI or MB-ND properties. Hence, queueing systems do not possess supermodularity or submodularity. There are, however, plausible inventory systems that support the property.

6. ILLUSTRATION OF THE FRAMEWORK

The aim of this section is to illustrate our results in well-known models. The illustrative example in Section 6.1 points out why our results are interesting even in a simple setting. Section 6.2, on the other hand, presents well-known problems in the literature that can be analyzed within our framework.

6.1. An Illustrative Example: Robust Capacity Control

To illustrate the type of results that we seek in the sequel, let us take the example of single-resource capacity control problem from revenue management Lautenbacher and Stidham [15], where the objective is to admit or reject demands from customer classes with different rewards to maximize the expected reward until the end of the horizon. This problem is modeled using multiple inventory rationing operators. The nominal value function can be expressed as:

$$v_t(x) = \sum_{i=1}^{n} p_{i,t} T_{R_i} v_{t+1}(x) + p_{0,t} v_{t+1}(x) \quad \text{for } x > 0$$

where x denotes the number of available inventory (seats) and $p_{i,t}$ is the probability of a class-i arrival with a corresponding reward of R_i in period t ($p_{0,t}$ corresponds to the probability that there are no arrivals in period t). Note that a similar problem in revenue management is analyzed by Turgay et al. [23].

Following the results in Cil et al. [6] and Theorem 2, it is easily seen that $v_t(x)$ is concave in x for all t and is supermodular in (x, t). Using these properties, we conclude that a threshold (protection level) policy is optimal due to concavity and that the optimal thresholds are monotone over time due to supermodularity. This formulation naturally assumes that the event probabilities $p_{i,t}$ do not depend on the state of the system.

To obtain numerical results, let us focus on a particular case of the above problem with two customer classes where the arriving batch size is equal to one (single arrivals) and $p_{1,t} = p_1$, $p_{2,t} = p_2$ for all t. For each demand class $a_i = 1$ denotes the action that admits the arriving demand, and $a_i = 0$ corresponds to rejecting the customer. The nominal problem is then represented by:

$$\begin{aligned} v_t(x) = p_1 \max\{a_1(v_{t+1}(x-1) + R_1) + (1-a_1)v_{t+1}(x)\} \\ + p_2 \max\{a_2(v_{t+1}(x-1) + R_2) + (1-a_2)v_{t+1}(x)\} \\ + p_3 v_{t+1}(x) \text{ for } x = 1, 2, ..., Q. \end{aligned}$$

with $v_T(x) = 0$, for all x, and $v_t(0) = 0$ for all t.

Let us next consider the robust version where the uncertainty set $\mathcal{P}$ consists of two vectors: $\mathbf{p}_1 = (p_1, p_2, p_3) = (0.48, 0.2, 0.32)$ and $\mathbf{p}_2 = (p_1, p_2, p_3) = (0.45, 0.55, 0)$. There are clearly four admissible actions as $\mathbf{a} = (a_1, a_2)$. In summary, Nature is allowed to choose the probability distributions at each stage, state and action from $\mathcal{P}$: $(p_{1,t}(x, \boldsymbol{a}), p_{2,t}(x, \boldsymbol{a}), p_{3,t}(x, \boldsymbol{a})) \in \mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2\}$.

Then the robust value function is given by the following recursion.

$$\begin{aligned} w_t(x) = \max_{\boldsymbol{a}} \Bigg\{ \min_{p_t(x,\boldsymbol{a}) \in \mathcal{P}} p_{1,t}(x, \boldsymbol{a}) \\ \times (a_1(w_{t+1}(x-1) + R_1) + (1-a_1)w_{t+1}(x)) \\ + p_2(x, \boldsymbol{a})\,(a_2(w_{t+1}(x-1) + R_1) \\ + (1-a_2)w_{t+1}(x)) + p_3(x, \boldsymbol{a}) w_{t+1}(x) \Bigg\}. \qquad (15) \end{aligned}$$

with $w_T(x) = 0$, for all x, and $w_t(0) = 0$ for all t.

Let us further assume that $T = 20$, $R_1 = 10$, $R_2 = 1$ and that the starting inventory level is 12. We solve the resulting robust MDP numerically. Let $\mathbf{a}_t^*(x)$ denote the optimal action selected by controller at time t and state x and $\mathbf{p}_t^*(x)$ denote the optimal event probability distribution selected by Nature for that action. Table 2 reports $\mathbf{p}_t^*(x)$ and $\mathbf{a}_t^*(x)$ for $t = 5$, 10, 15, and $x = 1, 2, \ldots, 12$.

The numerical solution reveals an interesting property for this example. It is observed from Table 2 that both the controller's and Nature's optimal policies are state and time dependent. In particular, for Nature, different probability distributions may be optimal at different states and at different

Table 2. Controller's and Nature's optimal policies for the example.

$t \downarrow \mid x \rightarrow$	1	2	3	4	5	6
5	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$
10	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$
15	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$
$t \downarrow \mid x \rightarrow$	7	8	9	10	11	12
5	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$
10	$(1,0), \mathbf{p}_2$	$(1,0), \mathbf{p}_2$	$(1,1), \mathbf{p}_2$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$
15	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$	$(1,1), \mathbf{p}_1$

times. Regardless, as established by Theorem 3, the robust value function is concave in x for all t and is supermodular in x and t just like the nominal value function. As a result, the optimal policy of the controller is of threshold type. For each t, there is a protection level for Class-2 demand arrivals below which they are rejected. In addition, these protection levels are nonincreasing in t. This is an interesting result because the robust problem has the same optimal policy structure as the nominal problem even though the event probability distributions chosen by Nature do depend on t and on x in the robust case.

6.2. Illustrations from the Literature

In this section, we present some results on the structure of optimal policies for robust versions of some well-known examples from the literature using the earlier results.

We first consider an extended version of the single-resource capacity control problem introduced in Section 6.1 by adding dynamic pricing. In this case, rationing operators and dynamic pricing operators are used together in order to model a special customer segment (class $n+1$ that is offered a spot price). A typical value function is then given by:

$$v_t(x) = \sum_{i=1}^{n} p_{i,t} T_{\mathrm{BR}_i} v_{t+1}(x) + p_{n+1,t} T_{\mathrm{IP}} v_{t+1}(x) + p_{0,t} v_{t+1}(x),$$

By Theorem 3, the value function of the robust version is concave, so that optimal admission policies are of threshold type, and the thresholds are monotone over time. Moreover, optimal prices to charge to class $n+1$ are nonincreasing in x and in t.

Apart from discrete-time models, our results apply for continuous-time models under certain assumptions. In particular, a class of continuous-time problems can be converted to equivalent discrete time problems using uniformization [17]. Uniformization converts the transition rates of a continuous-time model to the transition probabilities of an equivalent discrete-time model. The basic assumption needed in the uniformization technique is the existence of a bound on the total potential transition rates in all states and time periods of the continuous-time model under consideration. To be able to use uniformization in a robust setting, the total potential transition rate should continue to be bounded over all the uncertainty sets in all time periods. If the uncertainty in the model can be represented after the conversion in discrete time, the results continue to apply directly. As an example, let us consider the uniformized version of a typical admission control problem to a single-server Markovian queue [13]:

$$v_t(x) = -h(x) + p_{1,t} T_{\mathrm{ADM}} v_{t+1}(x) + p_{2,t} T_{\mathrm{DEP}} v_{t+1}(x) + p_{0,t} v_{t+1}(x),$$

where x denotes the number of customers in the system, $p_{1,t}$ is the probability of an arrival, $p_{2,t}$ is the probability of a service completion and $h(x)$ is a nondecreasing and convex holding cost function.

From Theorem 3, the optimal admission control policy of the robust version of the above problem is a threshold policy but optimal thresholds are not necessarily monotone over time because the MBs of the two operators have monotonicity properties in opposing directions.

Now, let us consider a dynamic pricing and production control problem for a make-to-stock queue with lost sales (see Gayon et al. [8] for example). After uniformization, the value function is expressed as:

$$v_t(x) = -h(x) + p_{1,t} T_{IP} v_{t+1}(x) + p_{2,t} T_P v_{t+1}(x) + p_{0,t} v_{t+1}(x),$$

where x denotes the available inventory, $p_{1,t}$ is the probability of a potential demand arrival, $p_{2,t}$ is the probability of production completion and $h(x)$ is a nondecreasing and convex holding cost function.

Using Theorem 3, we observe that the optimal production policy is determined by a threshold and that the optimal prices are nondecreasing in x for the robust version of this problem.

7. NESTED UNCERTAINTY SETS: STRUCTURE OF OPTIMAL POLICIES

In this section, we aim to investigate the conditions under which the monotonicity results shown in the previous section can be extended. It has been shown by several authors that less conservative robust policies are more adaptive to volatile conditions than both nominal and absolute robust policies. Examples can be found in [19, 23, 25]. We focus on a special formulation -S-robust policy-, an extended robust formulation recently proposed by Xu and Mannor [25], and explore the structure of the robust value function for nested uncertainty sets. In this formulation, the uncertainty is not represented by a single uncertainty set, but by a number of uncertainty sets that have a nested structure. Each uncertainty set corresponds to a probabilistic guarantee for a different confidence level. Then, the corresponding optimal policy, which is called S-robust, must take into account these different probabilistic guarantees. In this section, we show that, for the class of problems treated in this paper, the structural properties of the standard and robust problem are retained for the S-robust counterpart.

Similarly to Xu and Mannor [25], we define the nested uncertainty sets as follows $\mathcal{P}_t^1 \subseteq \mathcal{P}_t^2 \subseteq \cdots \subseteq \mathcal{P}_t^K$ for all t, so that the uncertain transition probability distribution belongs to set $\mathcal{P}_t^k$ with probability λ_k, where $\lambda_k \geq \lambda_{k-1}$ and $\sum_{k=1}^K(\lambda_k - \lambda_{k-1}) = 1$ by setting $\lambda_0 = 0$. In this context, a policy is said to be S-robust if it satisfies the following condition:

$$w_t(x) = \max_{\boldsymbol{a}}\left\{ \sum_{k=1}^{K}(\lambda_k - \lambda_{k-1})\left(\min_{\boldsymbol{p}_t(x,\boldsymbol{a})\in\mathcal{P}_t^k} \sum_{i=0}^{n} p_{i,t}(x,\boldsymbol{a}) \sum_{y\in X} q_t(a_i,x,y|i)(w_{t+1}(y) + R_i(a_i,x))\right)\right\}, \quad (16)$$

Eq. (16) is a convex combination of K expressions with the same structural properties. It is clear that Theorem 3 is also valid for this equation. The S-robust value function $w_t(x)$ can then be represented as follows:

$$w_t(x) = \sum_{k=1}^{K}(\lambda_k - \lambda_{k-1})\left(\min_{\boldsymbol{p}_t^k(x)\in\mathcal{P}_t^k} \left\{ \sum_i p_{i,t}^k(x) T_i w_{t+1}(x) \right\}\right), \quad (17)$$

where $\boldsymbol{p}_t^k(x) = (p_{1,t}^k(x), \ldots, p_{n,t}^k(x))$ is the transition probability distribution that Nature will select from uncertainty set $\mathcal{P}_t^k$. Note that (17) can also be expressed as:

$$w_t(x) = \sum_{k=1}^{K}(\lambda_k - \lambda_{k-1})\left(\min_{\boldsymbol{p}_t^k(x)\in\mathcal{P}_t^k} \left\{ \sum_i p_{i,t}^k(x) B_i w_{t+1}(x) \right\}\right) + w_{t+1}(x).$$

REMARK: We note the extension of the structural results to S-robust policies has not used the assumption that the sets are nested. Consequently, the structural results will still be valid when there are K different uncertainty sets $\mathcal{P}_t$ for all t, where $\mathcal{P}_t$'s are completely arbitrary. We let γ_k be the probability that uncertainty set k is used in any period, where $\sum_{k=1}^K \gamma_k = 1$. Then replacing $(\lambda_k - \lambda_{k-1})$ with γ_k in Eq. (17) will give the desired result.

REMARK: S-robust policies can be used to reflect varying degrees of conservatism. As a simple example, consider the case where the controller believes that a certain probability distribution (nominal parameter) will occur with probability λ_1, and with probability $1 - \lambda_1$ the probability distribution will be chosen from an uncertainty set (which includes the nominal parameter) by Nature in order to minimize the controller's expected revenue. Then, taking $\lambda_1 = 0$ corresponds to the robust problem, where the controller is the most conservative, while $\lambda_1 = 1$ corresponds to the nominal problem, where the controller is the least conservative. Hence, increasing values of λ_1 shows decreasing degrees of conservatism. By using more than one λ_1, we can model even more complicated conservatism types and degrees. With this approach, we can define a set of comparable robust strategies systematically and evaluate their performance with respect to other parameters such as variance, fill rates, service level etc.

Now we investigate the possibility of having some monotonicity results with respect to uncertainty sets. In this case, we consider two similar robust MDP models, identical except for their uncertainty sets which are nested, that is, $\mathcal{P}_t^1 \subseteq \mathcal{P}_t^2$ for all t. Let $w_t^k(x)$ denote the robust value function corresponding to $\mathcal{P}_t^k$. Intuitively, we expect that the value function of the system with a bigger uncertainty set should be greater, that is, $w_t^1(x) \leq w_t^2(x)$ for all x and t. This can be formally proven by induction as in Paschalidis and Kang [19].

On the other hand, establishing monotonicity of optimal policies, which require second order comparisons turns out to be much more challenging, see Turgay [26]. In particular, nested uncertainty sets do not imply monotone control policies in general. However, it is possible to show such monotonicity holds for special cases of uncertainty sets, see Turgay et al. [23].

8. CONCLUSIONS AND PERSPECTIVES ON FUTURE RESEARCH DIRECTIONS

We considered robust versions of a class of event-based DPs frequently encountered in queueing and inventory control. Under event probability uncertainty, we were able to show that robust optimal policies have the same structure as nominal optimal policies. This is very appealing from a practical point of view because simple policies are easier to communicate, parameterize and adjust.

In this section, we look into two challenging directions for extending our results. Under other uncertain parameters than event probabilities, general results appear difficult to obtain and simple policies may no longer be optimal. Section 8.1 considers such a case when the uncertainty affects the rewards of the model.

Multidimensional problems in queueing and inventory control bring significant additional challenges because establishing structural properties usually requires verifying multiple conditions that should hold simultaneously. Investigating the structure of robust optimal policies for multi-dimensional problems is an important avenue for further research. In Section 8.2, we present a first step to tackle this kind of problems.

8.1. Uncertainty on the Parameters of the Operators

So far we have considered the uncertainty affecting transition probability distributions, for which we are able to show that the action of the controller does not depend on the Nature's posteriori action. However, this is not necessarily true when problem parameters that appear in the operators, such as costs or rewards, are subject to uncertainty. Consequently, uncertainty on these parameters leads to a more challenging situation, and our previous results do not generalize easily.

In this section, we illustrate one way of modeling uncertainties affecting other parameters, which may lead to a more systematic analysis in the future. Specifically, we consider an uncertainty on the reward parameter of the rationing operator in a model which consists of a number of such operators, and assume that the corresponding transition probability distributions are fixed and known. We first observe that if we define a separate uncertainty set on the reward for each operator, then Nature will choose the minimal reward, so that the controller can solve the problem for only these parameters. Consequently, this case leads to a trivial solution and is not interesting. However, an intriguing situation arises when Nature has to decide on the reward for a group of operators simultaneously within a given uncertainty set. As before, let us assume that the uncertainty sets for rewards do not depend on the state and the controller's action. This is plausible when the rewards fluctuate according to state-independent external factors. However, there might be other interesting cases where this assumption may not hold. For instance, if there are discounts or surcharges that apply to customers depending on the inventory level or queue length in addition to the state-independent reward fluctuations, the below results do not apply.

To represent this case, we define a super-rationing operator as follows:

$$\mathbf{T}_R w_{t+1}(x) = \max_{a} \left\{ \min_{R_t(x,\boldsymbol{a})} \sum_{i=1}^{n_R} p_{i,t}(a_i R_{i,t}(x,\boldsymbol{a}) + w_{t+1}(x - a_i)) \right\}, \tag{18}$$

where n_R is the number of rationing operators, $\boldsymbol{a} = (a_1, \ldots, a_{n_R})$ with $a_i = 1$ if an incoming type-i demand is satisfied, and $a_i = 0$ otherwise, and $\boldsymbol{R}_t(x,\boldsymbol{a}) = (R_{1,t}(x,\boldsymbol{a}), \ldots, R_{n_R,t}(x,\boldsymbol{a}))$. Hence, Nature will decide on a set of rewards once he sees the actions of the controller. In other words, $\boldsymbol{R}_t(\boldsymbol{a},x)$ is the response of Nature to controller's action $\boldsymbol{a}$ in state x at stage t.

Our second observation is that $w_{t+1}(x)$ is not a coefficient in Nature's objective function. Hence, Nature's response depends only on the controller's action $\boldsymbol{a}$, which allows us to modify Nature's decision as $R_{i,t}(\boldsymbol{a})$:

$$\mathbf{T}_R w_{t+1}(x) = \max_{a} \left\{ \min_{R_t(\boldsymbol{a})} \sum_{i=1}^{n_{R_i}} p_{i,t}(a_i R_{i,t}(\boldsymbol{a}) + w_{t+1}(x - a_i)) \right\}. \tag{19}$$

Finally, we define MB of the super-rationing operator as follows:

$$\mathbf{B}_R w(x) = \mathbf{T}_R w(x) - \sum_{i=1}^{n_R} p_i w(x).$$

Now we are ready to present the main result of this section:

LEMMA 4: If *w(x)* is ND and concave in x for all x at stage $t+1$, then the super-rationing operator $\mathbf{T}_R w(x)$ has the following properties:

1. $\mathbf{T}_R w(x)$ *is ND in x,*
2. $\mathbf{T}_R w(x)$ *is concave in x,*
3. $\mathbf{B}_R w(x)$ *is ND in x.*

PROOF: The proof is given in Appendix B.

Lemma 4 implies that the super-rationing operator $\mathbf{T}_{\boldsymbol{R}} w_{t+1}(x)$ can be incorporated in any robust value function $w_t(x)$ consisting of the operators given in Table 5 with certain (or uncertain) event probabilities without violating the structural properties of $w_t(x)$. Hence, in the presence of this super-operator, the structure of the optimal policy corresponding to the regular operators remains the same. However, Lemma 4 does not guarantee that the optimal policy corresponding to the super-operator has a monotone structure, as the example given in Appendix C demonstrates. In this example, we demonstrate that the optimal rationing policy is not always monotone even though the value function $w(x)$ has the same mathematical properties with the nominal value function $v(x)$.

In this section, we propose a specific way to represent an uncertainty on rewards through a so-called super operator which consists of a number of the same operators. This representation is a novel idea, which presents a vast set of possibilities through combining different types of operators by defining appropriate super operators. Another approach can be accounting for uncertainties which affect the parameters of the operators and the transition probabilities at the same time. These approaches can be explored further in future research.

8.2. Monotonicity Properties in Multidimensional Systems

In this section, we point out another research direction with the aim of extending our results on uncertain transition probabilities to settings with multi-dimensional states. The systems that can be represented in this setting include, but are not limited to, tandem queues, queueing networks and parallel queues. We consider some of the first and second-order inequalities considered by Koole [14]. In this section, we prove that an operator which preserves these properties in the EBDP framework will continue to preserve them in the robust EBDP framework. The first half of the section presents the formal definitions of the properties that are considered, while the second half illustrates our results through specific operators.

Let $\mathbf{x} = (x_1, \ldots, x_m)$ denote the state of the system, where x_i is the number of type-i customers in the system. The customer types can be associated with a set of parameters or with the position of customers in the system, such as the number of customers at a certain node of a network. We set $\mathbf{e}_i$ as the i^{th} unit vector, and $\mathbf{e}_0$ as the zero vector. Now, we are ready to present the formal definitions of the properties that are adopted from [14] when the objective is maximization, as opposed to minimization in [14]:

Property 8.1:

1. A function f has nonincreasing property (NI) if for all $1 \leq j < m$:

$$f(\mathbf{x} + \mathbf{e}_j) \leq f(\mathbf{x}). \tag{20}$$

2. A function f has upstream-nonincreasing property (UNI) if for all $1 \leq j < m$:

$$f(\mathbf{x} + \mathbf{e}_j) \leq f(\mathbf{x} + \mathbf{e}_{j+1}). \tag{21}$$

3. A function f has Schur concavity property *(SC)* if for all $\boldsymbol{x}$ and k, j with $k \neq j$ and $x_k \leq x_j$:

$$f(\mathbf{x} + \mathbf{e}_j) \leq f(\mathbf{x} + \mathbf{e}_k), \tag{22}$$

and for all $\boldsymbol{x}$ and k, j with $k \neq j$ and $x_k = x_j$, and all $b > 0$:

$$f(\mathbf{x} + b\mathbf{e}_k) = f(\mathbf{x} + b\mathbf{e}_j). \tag{23}$$

4. A function f has asymmetric Schur concavity property (ASC) if for all $\boldsymbol{x}$ and k, j with $k < j$ and $x_k \leq x_j$:

$$f(\mathbf{x} + \mathbf{e}_j) \leq f(\mathbf{x} + \mathbf{e}_k), \tag{24}$$

and for all $\boldsymbol{x}$ and k, j with $k < j$ and $x_k = x_j$, and all $b > 0$:

$$f(\mathbf{x} + b\mathbf{e}_j) \leq f(\mathbf{x} + b\mathbf{e}_k). \tag{25}$$

5. A function f has componentwise-concavity property (CCv) if for all $1 \leq j < m$:

$$f(\mathbf{x} + 2\mathbf{e}_j) + f(\mathbf{x}) \leq 2f(\mathbf{x} + \mathbf{e}_j). \tag{26}$$

The following theorem represents the main result of this section, as it extends the EBDP framework from a nominal setting to robust setting:

THEOREM 5: Let $\mathcal{T} = \{T_i\}_{i=1}^n$ be a set of operators that preserve Property 8.1.ℓ for $\ell = 1, \ldots, 5$. Then a robust MDP model that consists of any combination of operators in $\mathcal{T}$ will induce a robust value function $w_t(\mathbf{x})$ that has Property 8.1.ℓ.

PROOF: We prove the statement for properties *(2)*, *(4)* and *(5)*, where the others can be proven similarly.

(2) Since all operators preserve UNI property, the following inequality holds for any element $p \in \mathcal{P}_t$, in particular, for Nature's choice in state $\mathbf{x} + e_{j+1}$, $p_t^i(\mathbf{x} + e_{j+1})$:

$$\sum_{i=1}^{n} p_t^i(\mathbf{x} + e_{j+1}) T_i w_t(\mathbf{x} + e_j) \leq \sum_{i=1}^{n} p_t^i(\mathbf{x} + e_{j+1}) T_i w_t(\mathbf{x} + e_{j+1}).$$

However, Nature's choice in $\mathbf{x} + e_j$, $p_t^i(\mathbf{x} + e_j)$ will achieve a lower value for the LHS of this inequality:

$$\sum_{i=1}^{n} p_t^i(\mathbf{x} + e_j) T_i w_t(\mathbf{x} + e_j) \leq \sum_{i=1}^{n} p_t^i(\mathbf{x} + e_{j+1}) T_i w_t(\mathbf{x} + e_j),$$

Table 3. Definitions for multi-dimensional operators, where R_j is the reward of admitting a type-j job, and $a^+ = \max\{0, a\}$. Note that $\mu \le 1$ in operator T_{MS} and $c_{j+1} \le c_j$ in operator T_C.

Operator name	Notation	Definition
Admission	$T_{A_j} f(\mathbf{x})$	$\max\left\{f(\mathbf{x}), f(\mathbf{x}+\mathbf{e}_j)+R_j\right\}$
Uncontrolled arrival	$T_{UA_j} f(\mathbf{x})$	$f(\mathbf{x}+\mathbf{e}_j)$
Uncontrolled departure	$T_{UD_j} f(\mathbf{x})$	$f\left((\mathbf{x}-\mathbf{e}_j)^+\right)$
Moving server	$T_{MS} f(\mathbf{x})$	$\max_{\left\{j:x_j>0\right\}}\left\{\mu\left((\mathbf{x}-\mathbf{e}_j)^+\right)+(1-\mu)f(\mathbf{x})\right\}$
Routing	$T_R f(\mathbf{x})$	$\max_{1\le j\le m}\left\{f\left(\mathbf{x}+\mathbf{e}_j\right)\right\}$
Cost	$T_C f(\mathbf{x})$	$f(\mathbf{x})-c(\mathbf{x})$, where $c(\mathbf{x})=\sum_j c_j x_j$

which proves that the robust value function $w_t(\mathbf{x})$ that has Property 8.1. *(2)*.

(4) Let k, j and $\mathbf{x}$ be such that $k<j$ and $x_k \le x_j$. Since all operators preserve *ASC* property, the following inequality holds for any element $p \in \mathcal{P}_t$, in particular, for Nature's choice in state $\mathbf{x}+\mathbf{e}_k$, $p_t^i(\mathbf{x}+\mathbf{e}_k)$:

$$\sum_{i=1}^{n} p_t^i(\mathbf{x}+\mathbf{e}_k)T_i w_t(\mathbf{x}+\mathbf{e}_j) \le \sum_{i=1}^{n} p_t^i(\mathbf{x}+\mathbf{e}_k)T_i w_t(\mathbf{x}+\mathbf{e}_k).$$

But Nature's choice in state $\mathbf{x}+e_j$, $p_t^i(\mathbf{x}+e_j)$ will achieve a lower value for the LHS of this inequality:

$$\sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x}+e_j) \le \sum_{i=1}^{n} p_t^i(\mathbf{x}+\mathbf{e}_k)T_i w_t(\mathbf{x}+e_j),$$

which proves the first inequality required for Property 8.1. *(4)*.

Now we consider the second inequality. We let k, j and $\mathbf{x}$ be such that $k<j$ and $x_k=x_j$, and $b>0$. Then we have:

$$\begin{aligned}&\sum_{i=1}^{n} p_t^i(\mathbf{x}+b\mathbf{e}_j)T_i w_t(\mathbf{x}+b\mathbf{e}_j)\\ &\quad\le \sum_{i=1}^{n} p_t^i(\mathbf{x}+b\mathbf{e}_k)T_i w_t(\mathbf{x}+b\mathbf{e}_j)\\ &\quad\le \sum_{i=1}^{n} p_t^i(\mathbf{x}+b\mathbf{e}_k)T_i w_t(\mathbf{x}+b\mathbf{e}_k),\end{aligned}$$

where the first inequality is due to the optimality of Nature's choice in state $\mathbf{x}+b\mathbf{e}_j$, and the other holds since all operators preserve *ASC* property.

(5) All operators preserve *CCv* property, so that the following inequality holds for any element $p \in \mathcal{P}_t$, in particular, for Nature's choice in state $\mathbf{x}+e_j$, $p_t^i(\mathbf{x}+e_j)$:

$$\sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x}+2e_j) + \sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x})$$

$$\le 2\sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x}+e_j).$$

Nature's choice in states $\mathbf{x}+2e_j$ and $\mathbf{x}$, $p_t^i(\mathbf{x}+2e_j)$ and $p_t^i(\mathbf{x})$, respectively, will achieve a lower value for the LHS of this inequality:

$$\begin{aligned}&\sum_{i=1}^{n} p_t^i(\mathbf{x}+2e_j)T_i w_t(\mathbf{x}+2e_j) + \sum_{i=1}^{n} p_t^i(\mathbf{x})T_i w_t(\mathbf{x})\\ &\quad\le \sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x}+2e_j) + \sum_{i=1}^{n} p_t^i(\mathbf{x}+e_j)T_i w_t(\mathbf{x}),\end{aligned}$$

which proves that the robust value function $w_t(\mathbf{x})$ has Property 8.1. *(5)*. □

Now we illustrate the impact of these results by considering some of the operators analyzed for the nominal MDP models by Koole [14]. We present the operators, which are updated according to the objective of maximizing profits in Table 3. A system which can be represented as a combination of these operators can serve m customer types, where type j can be characterized by its holding cost (c_j) and instantaneous reward (R_j).

In the admission operator T_{A_j}, the controller decides whether to admit an incoming type-j job and earn an instantaneous reward R_j or reject the customer. The uncontrolled arrival operator T_{UA_j} admits all incoming customers, whereas the uncontrolled departure operator T_{UD_j} serves a type-j customer. The moving server operator T_{MS} chooses the customer type to be served next. In this setting, service preemption is allowed, so the departure operators may interrupt an ongoing service if necessary. Note that the operators are defined in a slightly more general manner in Koole [14], here we assume that the holding costs are linear, that is, $c(\mathbf{x}) = \sum_j c_j x_j$, and the service rate is the same for all types of customers. Koole [14] shows that the above operators preserve almost all of the properties defined in Property 8.1 when $c_{i+1} \le c_i$ (please see Koole [14] for specific results). Then, Theorem 12 guarantees that these operators will continue preserving these properties in the robust setting. Accordingly, all systems that can be represented by these operators will have their value functions satisfy the relevant properties of Property 8.1 in the corresponding robust counterpart.

For instance, take the dynamic scheduling problem of a server that serves multiple classes of customers in their separate queues. Typical models involve the operators T_{UA_j}, T_{MS} and T_C. By Theorem 12, it is better to be in state $\mathbf{x} + \mathbf{e}_{j+1}$ rather than in $\mathbf{x} + \mathbf{e}_j$, which immediately translates to the well-known $c\mu$ rule through operator T_{MS}: The $c\mu$ rule gives service priority to type-*j* customers whenever the product of the holding cost c_j and the service rate μ_j for type *j* has the largest value among all types. In this section, we assume equal service rates for all types and $c_{j+1} \leq c_j$, so Theorem 12 guarantees the optimality of the $c\mu$ rule in all models that can be represented by the above operators. As an example of going further, we can get additional results if we add the admission control operator T_{A_j}, then because of componentwise concavity, there is an admission threshold for queue *j* whenever all other queue lengths are constant.

Schur Concavity and Almost Schur Concavity is about the preference for more balanced queueing loads across different queues. They are typically useful in admission control problems. For instance, combining T_R, T_{UA_j} and T_C, we can model a routing problem to multiple parallel queues. Thanks to SC and increasingness, the optimal policy routes to the shortest queue whenever the service rates in the parallel queues are identical. For asymmetric service rates, ASC along with increasingness establishes that it is preferable to route to the faster server if it has a shorter queue.

We consider only a subset of the operators and the properties to be preserved by these operators. Extending different monotonicity results of the value functions to a multi-dimensional setting for a larger set of operators is an interesting and challenging problem even for the standard setting (see [21] for some recent results). Obviously, this presents a much bigger challenge in a robust formulation.

APPENDIX A: STRUCTURAL PROPERTIES OF THE NOMINAL PROBLEM

In this section, we give the following monotonicity properties of the operators introduced in Table 1.

A.1. Inventory Control Operators

Batch Rationing Operator

The batch rationing operator represents the choice of the number of customers to be admitted from an arriving batch of class-*i* customers with batch size B_i in inventory systems. Some of the customers in a batch can be admitted while the remaining ones are rejected, which is defined as partial acceptance. κ_i is the number of class-*i* customers admitted from this batch, and R_i is the reward obtained by admitting one class-*i* customer.

Definition of the Operator

$$T_{BR_i} v(x) = \max_{\kappa_i \in \min(x,B)} \{\kappa_i R_i + v(x - \kappa_i)\},$$
$$T_{BR_i} v(x) = \max_{\kappa_i \in \min(x,B)} \{v(x - \kappa_i) - v(x) + \kappa_i R_i\} + v(x)$$

Rationing Operator

The rationing operator is a special case of the batch rationing operator where the batch size B is exactly 1. However, we provide separate proofs for this operator as well in this section.

Definition of the Operator
$T_{R_i} v(x) = \max\{R_i + v(x-1), v(x)\}$ for $x > 0$,
$T_{R_i} v(x) = \{v(x-1) - v(x) + R_i\}^+ + v(x)$ for $x > 0$, $T_{R_i} v(x) = v(x)$ for $x = 0$

Production Rate Operator

Production rate operator represents the choice of optimal service rate in production-inventory systems for production unit *i*. If the system uses Π_i portion of the service rate, then a nonnegative cost of C_{Π_i} is incurred.

Definition of the Operator
$T_{\mathrm{PR}_i} v(x) = \max_{\Pi \in [0,1]} \left\{-C_{\Pi_i} + \Pi_i v(x+1) + (1 - \Pi_i) v(x)\right\}$,
$T_{\mathrm{PR}_i} v(x) = \max_{\Pi_i \in [0,1]} \left\{\Pi_i \{v(x+1) - v(x)\} - C_{\Pi_i}\right\} + v(x)$

Production Operator

The production operator is a special case of production rate operator where $\Pi_i = \{0, 1\}$ and $C_{i_0} = 0$.

Definition of the Operator
$T_{P_i} v(x) = \max\{v(x+1) - C_i, v(x)\}$,
$T_{P_i} v(x) = \{v(x+1) - v(x) - C_i\}^+ + v(x)$

Inventory Pricing Operator

The inventory pricing operator represents the optimal price to be charged for the arriving customers in inventory systems. $F_Z(.)$ is the cumulative distribution function of the reservation price of an arriving customer, where R is the maximum price a customer is willing to pay.

Definition of the Operator
$T_{\mathrm{IP}} v(x) = \max_R \left\{\bar{F}_Z(R)[v(x-1) + R] + F_Z(R) v(x)\right\}$ for $x > 0$,
$T_{\mathrm{IP}} v(x) = max_R \bar{F}_Z(R) \{v(x-1) - v(x) + R\} + v(x)$ for $x > 0$,
$T_{\mathrm{IP}} v(x) = v(x)$ for $x = 0$.

A.2. Queueing Operators

The queueing operators we consider throughout the article are given in the following. It is important to note that the waiting room is infinite.

Batch Admission

The batch admission operator represents the choice of the number of the customers to be admitted from an arriving batch of class-i customers with batch size B in queueing systems. Some of the customers in a batch can be admitted while the remaining ones are rejected, which is defined as partial acceptance κ_i is the number of class-*i* customers admitted from this batch, and R_i is the reward obtained by admitting one class-*i* customer.

Table 4. Monotonicity results.

Operator name	Supermodularity in (x, t)	$Bv(x)$	$Tv(x)$
Rationing	Supermodular	ND. in x	ND. in x
Batch rationing	Supermodular	ND. in x	ND. in x
Production	Submodular	NI. in x	ND. in x
Production rate	Submodular	NI. in x	ND. in x
Inventory pricing	Supermodular	ND. in x	ND. in x
Admission	Submodular	NI. in x	NI. in x
Batch admission	Submodular	NI. in x	NI. in x
Controlled departure	Supermodular	ND. in x	NI. in x
Departure rate	Supermodular	ND. in x	NI. in x
Queue pricing	Submodular	NI. in x	NI. in x
Uncontrolled arrival to a queue	Submodular	NI. in x	NI. in x
Uncontrolled departure from a queue	Supermodular	ND. in x	NI. in x

Definition of the Operator

$$T_{\mathrm{BADM}_i} v(x) = \max_{\kappa_i \in \min(x,B)} \{\kappa_i R_i + v(x+\kappa_i\}\,.$$
$$T_{\mathrm{BADM}_i} v(x) = \max_{\kappa_i \in \min(x,B)} \{v(x+\kappa_i) - v(x) + \kappa_i R_i\} + v\,(x)$$

Admission

The admission operator is a special case of the batch admission operator where the batch size B is exactly 1.

Definition of the Operator
$T_{\mathrm{ADM}_i} v_t(x) = \max\{R_i + v_{t+1}(x+1), v_{t+1}(x)\}\,,$
$T_{\mathrm{ADM}_i} v_t(x) = \{v_{t+1}(x+1) - v_{t+1}(x) + R_i\}^+ + v_{t+1}(x).$

Departure Rate Operator

The departure rate operator represents the choice of the best service rate in queueing systems. If the system uses Π portion of the service rate, then a nonnegative cost of C_Π is incurred.

Definition of the Operator
$T_{\mathrm{DR}_i} v(x) = \max_{\Pi\in[0,1]} \{-C_\Pi + \Pi v(x-1) + (1-\Pi)v(x)\}$ for $x>0$,
$T_{\mathrm{DR}_i} v(x) = \max_{\Pi\in[0,1]} \{\Pi\{v(x-1) - v(x)\} - C_\Pi\} + v(x)$ for $x>0$,
$T_{\mathrm{DR}_i} v(x) = v(x)$ for $x=0$

Controlled Departure Operator

The controlled departure operator is a special case of the departure rate operator where $\Pi = \{0, 1\}$ and $C_0 = 0$.

Definition of the Operator
$T_{\mathrm{CD}_i} v(x) = \max\{v(x-1) - C_i, v(x)\}$ for $x>0$,
$T_{\mathrm{CD}_i} v(x) = \{v(x-1) - v(x) - C_i\}^+ + v(x)$ for $x>0$,
$T_{\mathrm{CD}_i} v(x) = v(x)$ for $x=0$.

Queue Pricing Operator

The queue pricing operator represent the optimal price to be charged for the arriving customers in queueing systems. $F_Z(.)$ is the cumulative distribution function of the reservation price of an arriving customer, where R is the maximum price a customer is willing to pay.

Definition of the Operator

$$T_{\mathrm{QP}} v(x) = \max_R \left\{\bar{F}_Z(R)[v(x+1) + R] + F_Z(R)v(x)\right\},$$

$$T_{\mathrm{QP}} v(x) = \max_R \bar{F}_Z(R)\{v(x+1) - v(x) + R\} + v(x)$$

Uncontrolled Arrival to a Queue

The uncontrolled arrival operator represents the arrival process to a queueing system. The function $a(x)$ is, the probability that an arriving customer joins the system when there are x customers, which we refer to as the joining probability. We assume that $a(x)$ is NI in x. When a is constant, arrival operator models a system where customers enter the system with a fixed probability, independent of the state, or choose not to enter the system with probability $1-a$. We will call this type of arrivals as regular arrivals, since they do not depend on the state of the system.

Definition of the Operator

$$T_{\mathrm{ARR}} v(x) = a(x)v(x+1) + (1-a(x))v(x),$$
$$T_{\mathrm{ARR}} v(x) = a(x)\{v(x+1) - v(x)\} + v(x)$$

Uncontrolled Departure to a Queue

The uncontrolled departure operator represents the departure of an existing customer from the system, where the service rate may depend on the state of the system. The function $b(x)$ corresponds to the probability of a service completion when the system has x customers. We assume that $b(x)$ is an ND function of x.

Definition of the Operator
$T_{\mathrm{DEP}} v(x) = b(x)v(x-1) + (1-b(x))v(x)$ for $x>0$,
$T_{\mathrm{DEP}} v(x) = b(x)\{v(x-1) - v(x)\} + v(x)$ for $x>0$,
$T_{\mathrm{DEP}} v(x) = v(x)$ for $x>0$.

APPENDIX B: PROOF OF LEMMA 4

LEMMA 4: If $w(x)$ is ND and concave in x for all x at stage $t+1$, then the super-rationing operator $\mathbf{T}_R w(x)$ has the following properties:

1. $\mathbf{T}_R w(x)$ is ND in x,
2. $\mathbf{T}_R w(x)$ is concave in x,
3. $\mathbf{B}_R w(x)$ is ND in x.

PROOF: We give the details of the proof for concavity, the proof of ND is similar and simpler.

Table 5. Demonstration of concavity.

Case	$a_i(x-1)$	$a_i(x+1)$	Inequality
I	0	0	$w(x)+w(x)\geq w(x-1)+w(x+1)$
II	1	1	$w(x-1)+R_i(\boldsymbol{a}^{-1})+w(x-1)+R_i(\boldsymbol{a}^1)\geq$ $w(x-2)+R_i(\boldsymbol{a}^{-1})+w(x)+R_i(\boldsymbol{a}^1)$
III	0	1	$w(x-1)+w(x)+R_i(\boldsymbol{a}^1)\geq w(x-1)+w(x)+R_i(\boldsymbol{a}^1)$
IV	1	0	$w(x-1)+R_i(\boldsymbol{a}^{-1})+w(x)\geq w(x-2)+R_i(\boldsymbol{a}^{-1})+w(x+1)$ $w(x-1)-w(x-2)\geq w(x)-w(x-1)\geq w(x+1)-w(x)$

For notational simplicity, suppose that we define $\boldsymbol{a}^j$ as the optimal action in state $x+j$ with $j=-1,0,1$, where $\boldsymbol{a}^j=(a_1^j,\dots,a_{n_R}^j)$. We note that the corresponding optimal action of Nature depends only on the action of the controller $\boldsymbol{a}$, that is, it does not depend on state x. Accordingly, we can let $\boldsymbol{R}(\boldsymbol{a}^j)=(R_1(\boldsymbol{a}^j),R_2(\boldsymbol{a}^j),\dots,R_{n_R}(\boldsymbol{a}^j))$ for $j=-1,0,1$ denote Nature's corresponding optimal decisions in states $x-1$, x and $x+1$, respectively.

We first observe that replacing the optimal pair $(\boldsymbol{a}^0,\boldsymbol{R}(\boldsymbol{a}^0))$, denoting the controller's action and Nature's choice in state x, by the pairs $(\boldsymbol{a}^{-1}\boldsymbol{R}(\boldsymbol{a}^{-1}))$ and $(\boldsymbol{a}^1,\boldsymbol{R}(\boldsymbol{a}^1))$ will lead to a smaller revenue, so that:

$$2\sum_{i=1}^{n_R}a_i^0p_i(w(x-1)+R_i(\boldsymbol{a}^0))+p_i(1-a_i^0)w(x)\geq$$
$$\sum_{i=1}^{n_R}a_i^{-1}p_i(w(x-1)+R_i(\boldsymbol{a}^{-1}))+p_i(1-a_i^{-1})w(x)+$$
$$\sum_{i=1}^{n_R}a_i^1p_i(w(x-1)+R_i(\boldsymbol{a}^1))+p_i(1-a_i^1)w(x). \tag{27}$$

Now we would like to show that the right-hand side of (27) satisfies the following inequality, which establishes the concavity of the value functions:

$$\sum_{i=1}^{n_R}a_i^{-1}p_i(w(x-1)+R_i(\boldsymbol{a}^{-1}))+p_i(1-a_i^{-1})w(x)+$$
$$\sum_{i=1}^{n_R}a_i^1p_i(w(x-1)+R_i(\boldsymbol{a}^1))+p_i(1-a_i^1)w(x)\geq$$
$$\sum_{i=1}^{n_R}a_i^{-1}p_i(w(x-2)+R_i(\boldsymbol{a}^{-1}))+p_i(1-a_i^{-1})w(x-1)+$$
$$\sum_{i=1}^{n_R}a_i^1p_i(w(x)+R_i(\boldsymbol{a}^1))+p_i(1-a_i^1)w(x+1).$$

We analyze the four different cases for each operator i separately, where Table 5 presents the results. In Cases I and II, the inequality is satisfied due to the concavity of w, whereas it is obvious in Case III. In Case IV, concavity is used twice, once for state $x-1$ and then for x.

We note that concavity does not imply the following argument:

$$\left[w(x-1)+R_i(\boldsymbol{a}^0)\right]\leq w(x)\Rightarrow\left[w(x-2)+R_i(\boldsymbol{a}^{-1})\right]\leq w(x-1).$$

In other words, although the value functions are concave in x, the corresponding optimal actions are not necessarily monotone, see the counter example in the next section.

Now we prove that $\mathbf{B}_Rw(x)$ is ND in x. We first observe that $\mathbf{B}_Rw(x)=\mathbf{T}_Rw(x)-w(x)$ can be written as follows:

$$\sum_{i=1}^{n_R}\left(a_i^0p_i(w(x-1)+R_i(\boldsymbol{a}^0))+p_i(1-a_i^0)w(x)\right)-w(x)$$
$$=\sum_{i=1}^{n_R}a_i^0p_i(R_i(\boldsymbol{a}^0)+w(x-1)-w(x)).$$

Then we need to show that the following inequality holds:

$$\sum_{i=0}^{n_R}p_ia_i[R_i(\boldsymbol{a}^0)-\Delta w(x)]\geq\sum_{i=0}^{n_R}p_ia_i[R_i(\boldsymbol{a}^{-1})-\Delta w(x-1)].$$

The left-hand side of this inequality is always larger when the optimal pair $(\boldsymbol{a}^0,\boldsymbol{R}(\boldsymbol{a}^0))$ is replaced by the optimal pair in state $x-1$, $(\boldsymbol{a}^{-1}\boldsymbol{R}(\boldsymbol{a}^{-1}))$. Hence, the result will be established if the following inequality holds:

$$\sum_{i=0}^{n_R}p_ia_i[R_i(\boldsymbol{a}^{-1})-\Delta w(x)]\geq\sum_{i=0}^{n_R}p_ia_i[R_i(\boldsymbol{a}^{-1})-\Delta w(x-1)],$$

But this is always true by the concavity of w. □

APPENDIX C: COUNTER EXAMPLES

C.1 Counter Example for Monotonicity of Thresholds

We consider a single-resource revenue management system with 4 customer classes and compare two systems. The admission reward of each class is as follows: $R_1=40$, $R_2=35$, $R_3=30$, $R_4=20$ and is identical for the two systems.

The uncertainty set defining probabilities are denoted as $\mathcal{P}$ and defined as follows:

$$\mathcal{P}=\{(p_1,p_2,p_3,p_4): 4p_1+2p_2+3p_3+p_4\geq 1.65,$$
$$0.1\leq p_i\leq 0.2 i=1,\dots,4\}.$$

Now consider the second system. The uncertainty set $\mathcal{P}'$ of that system includes the uncertainty set of the first system $\mathcal{P}'\supseteq\mathcal{P}$. The definition of $\mathcal{P}'$ is as follows:

$$\mathcal{P}'=\{(p_1,p_2,p_3,p_4): 4p_1+2p_2+3p_3+p_4\geq 1.65,$$
$$0.1\leq p_{i,t}\leq 0.2 i=1,\dots,4, t\neq T-5,$$
$$0.05\leq p_{i,t}\leq 0.325 i=1,2,\ 0.1\leq p_{i,t}\leq 0.2 i=3,4\ t=T-5\}.$$

If we denote the value function and probabilities of the second system with $w_t'(x)$ and $p_{i,t}'(x)$, it is apparent that $w_t'(x)=w_t(x)$ and $p_t'(x)=p_t(x)$ for $t=T-1,\dots,T-4$.

We then numerically calculate the values for $w_{T-5}(x)$ and $w_{T-5}'(x)$ for $x=1$, 2. The results are as follows where the greater values between the systems are shown with boldface letters:

x	$w_{T-5}(x)$	$w'_{T-5}(x)$	$\Delta w_{T-5}(x)$	$\Delta w'_{T-5}(x)$
1	**33.68**	33.54	**33.68**	33.54
2	**61.53**	61.40	27.85	**27.86**

Therefore, in the above example, neither of the $\Delta w^{\mathcal{P}} \geq (\leq) \Delta w^{\mathcal{P}'}$ statements are true.

C.2. Counter Example: Optimal Action Depends on Nature's Decision When Rewards Are Uncertain

We consider a revenue management problem with a five customer classes and represent it by a super-rationing operator $\mathbf{T}_R$. Suppose each class has an arrival probability of 0.10 and the uncertainty set for rewards consists of two distinct points over the horizon, that is, $\mathcal{R} = \{\boldsymbol{R}^1, \boldsymbol{R}^2\}$ and $\boldsymbol{R}^1 = (60, 40, 45, 10, 10)$ and $\boldsymbol{R}^2 = (60, 40, 42.5, 20, 10)$. Consider stage $T-1$, when it is optimal to accept all classes regardless of Nature's posteriori decision, so that the optimal solution $\boldsymbol{a}_{T-1}(\boldsymbol{R}^1, x)$ and $\boldsymbol{a}_{T-1}(\boldsymbol{R}^2, x)$ are both $(1, 1, 1, 1, 1)$ and the corresponding Nature choice is $\boldsymbol{R}^1$. We obtain the optimal solution by enumerating all possible actions, and compute $w_{T-1}(x) = 16.5, \Delta w_{T-1}(x) = 16.5$ for all $x > 0$.

Now consider the next stage and suppose the inventory level is 1. We obtain $w_{T-2}(x)$ similarly by enumerating all possible actions $\boldsymbol{a}$. The optimal actions of the controller and Nature are $(1, 1, 1, 0, 0)$ and $\boldsymbol{R}^2$, respectively, with the corresponding value function 25.8. The optimal actions of the controller $\boldsymbol{a}_{T-2}(\boldsymbol{R}^1, 1)$ and $\boldsymbol{a}_{T-2}(\boldsymbol{R}^2, 1)$ are different for $\boldsymbol{R}^1$ and $\boldsymbol{R}^2$: If Nature selects $\boldsymbol{R}^1$, the optimal action $\boldsymbol{a}_{T-2}(\boldsymbol{R}^1, 1)$ is $(1, 1, 1, 0, 0)$, whereas if Nature selects $\boldsymbol{R}^2$, the optimal action is $\boldsymbol{a}_{T-2}(\boldsymbol{R}^2, 1)$ is $(1, 1, 1, 1, 0)$. Finally, note that the fourth class with a reward $R_4 = 20$ is rejected although $20 > w_{T-1}(x) = 16.5$.

C.3. Counter Example for Optimality of Threshold Policy When Rewards are Uncertain

We, again, consider a revenue management problem with a five customer classes, represented by a super-rationing operator $\mathbf{T}_R$. Now we set the time horizon as $T = 15$, where the first decision epoch is $t = 0$. The uncertainty affects only the rewards generated by the customer classes, so that the parameters regarding the event probabilities are certain. Moreover, we assume that the rewards at stages $t = 3, \ldots, 15$ are also known and fixed. The parameters of the model are given in the following table:

Operator (i)	$R_{i,t}$	$p_{i,t}$
1	60	0.1
2	50	0.1
3	45	0.1
4	23	0.13
5	20	0.1

Now suppose that at $t = 2$ the rewards may take values according to the four different scenarios given below without any change in the transition probabilities:

Scenario	R_1	R_2	R_3	R_4	R_5
1	45	50	40	46	50
2	60	50	42	31	20
3	50	55	40	27	30
4	50	60	45	23	20

At $t = 1$, the optimal admission policy of the controller and the optimal scenario chosen by the Nature at $x = 4$ and $x = 5$ are given as follows, which shows that optimal policy does not have a threshold structure.

x	Admission Policy	Scenario
4	(11101)	2
5	(11110)	3

Finally, we note the following: If a production operator were added to the above model, the resulting optimal production policy would still be a base-stock policy for all t, since the robust value functions are concave in the presence of the super-rationing operator.

REFERENCES

[1] J. Bagnell, A. Ng, and J. Schneider, Solving uncertain Markov decision problems, Technical Report, CMU-RI-TR-01-25, Robotics Institute, Carnegie Mellon University, 2001.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, Robust optimization, Princeton University Press, Princeton, USA, 2009.

[3] D. Bertsimas, D. Gamarnik, and A.A. Rikun, Performance analysis of queueing networks via robust optimization, Oper Res 59 (2011), 455–466.

[4] D. Bertsimas and A. Thiele, A robust optimization approach to inventory theory, Oper Res 54 (2006), 150–168.

[5] S.I. Birbil, J.B.G. Frenk, J.A.S. Gromicho, and S. Zhang, The role of robust optimization in single-leg airline revenue management, Manage Sci 55 (2009), 148.

[6] E.B. Cil, E.L. Ormeci, and F. Karaesmen, Effects of system parameters on the optimal policy structure in a class of queueing control problems, Queue Syst 61 (2009), 273–304.

[7] E. Delage and S. Mannor, Percentile optimization for Markov decision processes with parameter uncertainty, Oper Res 58 (2010), 203–213.

[8] J.P. Gayon, I. Talay Degirmenci, F. Karaesmen, and E.L. Örmeci, Dynamic pricing and replenishment in a production/inventory system with Markov-modulated demand, Probab Eng Inform Sci 33 (2009), 205–230.

[9] A.Y. Ha, Inventory rationing in a make-to-stock production system with several demand classes and lost sales, Manage Sci 43 (1997), 1093–1103.

[10] G.N. Iyengar, Robust dynamic programming, Math Oper Res 30 (2005), 257.

[11] A. Jain, A.E.B. Lim, and J.G. Shanthikumar, On the optimality of threshold control in queues with model uncertainty, Queue Syst 65 (2010), 157–174.

[12] E. Kardeş, F. Ordóñez, and R.W. Hall,Discounted robust stochastic games and an application to queueing control, Oper Res 59 (2011), 365–382.

[13] G. Koole, Structural results for the control of queueing systems using event-based dynamic programming, Queue Syst 30 (1998), 323–339.

[14] G. Koole, Monotonicity in Markov reward and decision chains: Theory and applications. Now Publishers Inc., Boston-Delft, 2006.

[15] C.J. Lautenbacher and S. Stidham, Jr., The underlying Markov decision process in the single-leg airline yield-management problem, Transp Sci 33 (1999), 136.

[16] A.E.B. Lim and J.G. Shanthikumar, Relative entropy, exponential utility, and robust dynamic pricing, Oper Res 55 (2007), 198–214.
[17] S.A. Lippman, Applying a new device in the optimization of exponential queuing systems, Oper Res 23 (1975), 687–710.
[18] A. Nilim and L.E. Ghaoui, Robust control of Markov decision processes with uncertain transition matrices, Oper Res 53 (2005), 780–798.
[19] I.C. Paschalidis and S.C. Kang, A robust approach to Markov decision problems with uncertain transition probabilities, Proc IFAC World Congress (2008), 408–413.
[20] P. Rusmevichientong and H. Topaloglu, Robust assortment optimization in revenue management under the multinomial logit choice model, Oper Res 60 (2012), 865–882.
[21] S. Vercraene, J.P. Gayon, and F. Karaesmen, Effects of system parameters on the optimal cost and policy in a class of multi-dimensional queueing control problems, Oper Res (2017) (submitted).
[22] J.K. Satia and R.E. Lave, Jr., Markovian decision processes with uncertain transition probabilities, Oper Res 21 (1973), 728–740.
[23] Z. Turgay, F. Karaesmen, and E.L. Örmeci, A dynamic inventory rationing problem with uncertain demand and production rates, Ann Oper Res 231 (2015), 207–228.
[24] C.C. White and H.K. Eldeib, Markov decision processes with imprecise transition probabilities, Oper Res 42 (1994), 739–749.
[25] H. Xu and S. Mannor, Distributionally robust Markov decision processes, Math Oper Res 37 (2012), 288–300.
[26] Z. Z. Turgay, Robust Optimization of a Class of Dynamic Problems, PhD Thesis in Industrial Engineering and Operations Management, Koc University, 2012.