Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Robust Control of Partially Observable Failing Systems

Michael Jong Kim

OPERATIONS RESEARCH
Vol. 64, No. 4, July–August 2016, pp. 999–1014
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2016.1495

Robust Control of Partially Observable Failing Systems

Michael Jong Kim
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada,
mikekim@mie.utoronto.ca

This paper is concerned with optimal maintenance decision making in the presence of model misspecification. Specifically, we are interested in the situation where the decision maker fears that a nominal Bayesian model may be miss-specified or unrealistic, and would like to find policies that work well even when the underlying model is flawed. To this end, we formulate a robust dynamic optimization model for condition-based maintenance in which the decision maker explicitly accounts for distrust in the nominal Bayesian model by solving a worst-case problem against a second agent, "nature," who has the ability to alter the underlying model distributions in an adversarial manner. The primary focus of our analysis is on establishing structural properties and insights that hold in the face of model miss-specification. In particular, we prove (i) an explicit characterization of nature's optimal response through an analysis of the robust dynamic programming equation, (ii) convexity results for both the robust value function and the optimal robust stopping region, (iii) a general robustness result for the preventive maintenance paradigm, and (iv) the optimality of a robust control limit policy for the important subclass of Bayesian change point detection problems. We illustrate our theoretical result on a real-world application from the mining industry.

Keywords: failure models; partially observable systems; dynamic programming; model ambiguity; robust control; relative entropy; stochastic dynamic games.
Subject classifications: dynamic programming; games/group decisions: stochastic.
Area of review: Stochastic Models.
History: Received November 2014; revision received September 2015; accepted March 2015. Published online in *Articles in Advance* May 13, 2016.

1. Introduction

Many stochastic control problems that arise in condition-based maintenance applications are concerned with determining the optimal time to repair or replace a system before it fails. A common modeling approach is to represent system degradation as an unobservable Markov process that can only be indirectly or partially observed using stochastically related signals collected through online condition monitoring (see, e.g., Dayanik and Gurler 2002, Dayanik et al. 2008, Kim and Makis 2013, Maillart 2006). From the perspective of real-world implementation, this modeling approach implicitly assumes that the decision maker is able to accurately and completely specify the full joint distribution governing the real-world degradation/failure dynamics as well as the condition-monitoring observations.

In practice however, there are many situations in which this assumption is violated. For example, the model dynamics may contain complex nonstationarities or depend on (possibly unknown) factors (e.g., human, environmental, etc.) in a way that is difficult to identify and model. Alternatively, calibration or model selection errors may arise when there is limited or poor quality data. In such cases, model miss-specification is unavoidable, and consequently the out-of-sample performance of maintenance policies computed on the basis of a miss-specified model may be poor. Motivated by these concerns, the goals of this paper are to (i) formulate a model for condition-based maintenance optimization that explicitly accounts for possible model miss-specification at the decision-making stage, and (ii) characterize the structural form of the optimal policy in the presence of such miss-specification.

In this paper, we consider a failing system whose underlying state is unobservable and can only be inferred using (imperfect) signals obtained through condition monitoring. At each stage, the decision maker acquires condition monitoring data and processes this information using Bayesian learning to obtain a posterior distribution over the underlying system state. A decision is then made to either initiate preventive maintenance or to run the system until the next decision epoch. We formulate this nominal problem as a partially observable Markov decision process. We are interested in situations where the decision maker distrusts the nominal model (e.g., state dynamics, signal distributions, etc.) and fears that it may be miss-specified. To account for such distrust, we formulate a robust version of the nominal problem that explicitly incorporates possible model miss-specification at the decision-making stage.

One novelty of our model is that we account for possible miss-specification in the decision-maker's learning mechanism itself. That is, we allow the decision maker to update his or her belief about the underlying system state via Bayesian learning, while also accounting for the possibility that the resulting posterior distribution may be incorrect. To model this feature, we introduce two new distinct concepts of model miss-specification—positional ambiguity and transitional ambiguity. In this regard, one of the general contributions of the this paper is that it provides a modeling framework in which decision makers can remain robust with respect to model miss-specification while also learning at the same time.

The decision maker accounts for distrust in the nominal model by solving a worst-case problem against an adversary, "nature," who alters the underlying model distributions to maximally damage the decision-maker's expected profit. We formulate this adversarial problem as a zero-sum stochastic dynamic game, and use the concept of relative entropy as a way to measure the distance between the nominal model and the alternative distributions chosen by nature. Under this formulation we derive the robust dynamic programming (Bellman) equation, which in principle prescribes the optimal action of the decision maker as well the optimal worst-case response of nature.

The focus of our analysis is on characterizing the structural form of the optimal robust policy. In this regard, our primary contribution to the maintenance optimization literature is establishing new structural properties and insights for decision making in the presence of model miss-specification. More generally, our findings also make contributions to the robust dynamic optimization literature in which very few structural results characterizing the optimal robust policy have been published.

To this end, we first establish an explicit and intuitive characterization of nature's optimal policy through an analysis of the robust dynamic programming equation using a variational equality from the theory of large deviations. We show how this characterization allows us to transform the robust dynamic programming equation into its so-called certainty equivalent. The certainty equivalent representation is then analyzed to establish structural properties of the decision-maker's optimal policy. In particular, we first prove that even in the presence of model miss-specification, the robust value function preserves convexity with respect to the current posterior distribution. We next prove a general robustness result for the preventive maintenance paradigm. In particular, we show that preventive maintenance is always a worthwhile strategy for managing failure risk, even when the decision maker distrusts the underlying model assumptions. Following this result, we then show that the optimal robust stopping region may in general not necessarily be convex, however it is convex when we impose a condition that moderates nature's ability to select alternative distributions. We show how this convex structure leads to useful geometric and managerial insights.

Finally, we establish structural results for the practical subclass of problems—Bayesian change point detection problems. In particular, we show that the optimal robust policy for this class of problems is characterized by a *robust threshold policy*. To illustrate how this policy can be both visualized and implemented in practice, we apply it to a real-world problem from the mining industry.

In the condition-based maintenance literature, there has been a long and rich history of establishing structural properties of optimal maintenance policies (see, e.g., Ross 1971, Makis and Jardine 1992, White 1978, Elwany et al. 2011, Maillart 2006, Kurt and Kharoufeh 2010, Ulukus et al. 2012). Although a lot of research has been conducted in this area, to our knowledge, this paper constitutes the first paper reporting structural results for maintenance optimization in the presence of model miss-specification.

In the area of dynamic optimization with model miss-specification, broadly speaking, there are two basic approaches found in the literature: the constraint approach and the penalty approach. In the constraint approach, the set of alternative models is represented as a hard constraint, and confidence in the nominal is captured by the size of this uncertainty set (see, e.g., Ben-Tal and Nemirovski 1998, 1999, 2000; Bertsimas and Sim 2004; El Ghaoui and Lebret 1997; Iyengar 2005; Li and Kwon 2013; Nilim and El Ghaoui 2005; Wiesemann et al. 2013). The penalty approach on the other hand expresses confidence in the nominal by penalizing alternative models that deviate too far from the nominal, and does so via a penalty function (soft constraint) that appears in the objective function (see, e.g., Dai Pra et al. 1996, Peterson et al. 2000, Hansen and Sargent 2007, Jain et al. 2010, Kim and Lim 2015, Lim and Shanthikumar 2007). In this paper, we adopt an entropy penalty approach to represent model miss-specification, because it provides a number of advantages from the perspective of characterizing the structure of the optimal robust control policy.

Research on dynamic optimization with model miss-specification—using either the penalty or constraint approach—has mainly focused on *computation* of the optimal robust policy (e.g., Lim and Shanthikumar 2007, Mannor et al. 2012, Wiesemann et al. 2013, Xu and Mannor 2012). As pointed out by Chan and Farias (2009), Brown et al. (2010), and Rogers (2007), even without the added difficulty of accounting for model miss-specification, solving dynamic optimization problems to optimality is difficult and one must often resort to heuristics. As such, other papers in this research area have diverted their efforts to developing methods for bounding the performance of *heuristic policies* relative to the robust optimal (e.g., Caro and Gupta 2015, Kim and Lim 2015). However, very few results regarding the structure of the optimal robust policy have been published in the robust dynamic optimization literature, which is the focus of this paper.

2. The Nominal Model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space on which the following stochastic processes are defined. The nominal model assumes that the state process (X_n: $n \in \mathbb{Z}_+$) follows a discrete-time homogeneous Markov chain with N unobservable states $\mathcal{S} = \{1, \ldots, N\}$. The state transition matrix is denoted by $\mathbf{P} = (p_{ij})_{N\times N}$, where transition probabilities

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad i, j \in \mathcal{S},\ n \in \mathbb{Z}_+. \qquad (1)$$

State 1 represents the "good-as-new" or healthy state, and state N represents the failure state. To model monotonic system deterioration, the state process is assumed to be nondecreasing with probability one, i.e., $p_{ij} = 0$ for all $j < i$. Furthermore, $p_{ii} < 1$ for all $i \neq N$. This implies that the failure state N is absorbing. In this setting, the time to system failure $\xi = \inf\{n \in \mathbb{Z}_+ : X_n = N\}$ is unobservable and can only be detected upon full system inspection.

The system is monitored at each time point, and condition monitoring data obtained at time n are denoted by Y_n, which takes values in some measurable space $(E, \mathcal{E})$. Here, E is the data space and $\mathcal{E}$ is the σ-algebra of subsets of E. The condition monitoring data Y_n are viewed as a noisy observation of the true underlying system state X_n. In particular, under the nominal model, conditional on the state process $X_n = i$, $i \in \mathcal{S}$, observations Y_n are distributed according to some known probability measure μ_i on $(E, \mathcal{E})$. We assume that probability measures $\mu_1, \ldots, \mu_N$ admit densities with respect to some σ-finite measure λ on $(E, \mathcal{E})$, and we denote these densities by $f_1, \ldots, f_N$, i.e., $f_i \triangleq d\mu_i/d\lambda$.

Let $\mathbf{Y} = (\mathcal{Y}_n$: $n \in \mathbb{Z}_+)$ denote the natural filtration generated by the condition monitoring information at each sampling epoch, i.e.,

$$\mathcal{Y}_n = \sigma(Y_1, \ldots, Y_n), \qquad (2)$$

and let $\mathbf{F} = (\mathcal{F}_n$: $n \in \mathbb{Z}_+)$ denote the larger filtration generated by the state and observation process at each sampling epoch, i.e., $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n, X_0, X_1, \ldots, X_n)$.

At each stage n, after collecting a condition monitoring sample Y_n and processing the new information, a decision is made to continue running the system until the next sampling epoch or to stop the system and carry out full system inspection followed by preventive maintenance. We consider the following cost/reward structure: $C_i > 0$ is the operational cost rate in state $i \in \mathcal{S}$, $C_{pi} > 0$ is the preventive maintenance cost in state $i \in \mathcal{S}$, and $R_i > 0$ is the production reward rate in state $i \in \mathcal{S}$. To avoid trivial cases, we assume the profit rate $R_i - C_i$ is nonincreasing in $i \in \mathcal{S}$, and that the boundary cases satisfy $R_1 - C_1 > 0$ and $R_N - C_N < 0$.

The nominal problem can then be formulated as follows. Let $\mathcal{T} \triangleq \mathcal{T}(\mathbf{Y})$ be the class of all $\mathbf{Y}$-stopping times $\tau \in \mathbb{Z}_+ \cup \{\infty\}$. The objective is to find a $\mathbf{Y}$-stopping time $\tau^* \in \mathcal{T}$, if it exists, that maximizes the expected total discounted profit, i.e.,

$$V(\pi_0) \triangleq \sup_{\tau\in\mathcal{T}} \mathbb{E}\left[\sum_{n=0}^{\tau} \alpha^n \sum_{i\in\mathcal{S}} (R_i - C_i) I_{\{X_n = i\}} - \alpha^\tau \sum_{i\in\mathcal{S}} C_{pi} I_{\{X_\tau = i\}} \,\middle|\, \pi_0 = \pi\right], \qquad (3)$$

where $\pi_0 = (\pi_0^{(1)}, \ldots, \pi_0^{(N)})$ is the initial distribution of the system state X_0, i.e., $\pi_0^{(i)} = \mathbb{P}(X_0 = i)$, $i \in \mathcal{S}$, and $\alpha \in (0, 1)$ is the discount factor. We refer to $V(\cdot)$ in (3) as the *nominal value function*.

Example 1 (Bayesian Change-Point Detection). An important special case is that of Bayesian Change-Point Detection (see, e.g., Dayanik et al. 2008, Makis 2008), where $N = 2$; it has applications in quality control, financial surveillance, and biomedical signal processing, among others. We will revisit this application in our investigation of the structural form of the optimal robust policy in Section 4.

From the theory of partially observable Markov decision processes, it is known that the stochastic control problem (3) can be cast as a fully observable Markov decision process in which the "state" at each decision epoch n is the *posterior distribution* $\pi_n = (\pi_n^{(1)}, \ldots, \pi_n^{(N)})$, where

$$\pi_n^{(j)} \triangleq \mathbb{P}(X_n = j \mid \mathcal{Y}_n), \quad j \in \mathcal{S}, \qquad (4)$$

which is sufficient for optimal decision making.

Bayes' Theorem implies that $\pi_n = (\pi_n^{(1)}, \ldots, \pi_n^{(N)})$ can be expressed recursively as

$$\pi_{n+1}^{(j)} = \frac{f_j(Y_{n+1})\{\sum_{i\in\mathcal{S}} \pi_n^{(i)} p_{ij}\}}{\sum_{l\in\mathcal{S}} f_l(Y_{n+1})\{\sum_{k\in\mathcal{S}} \pi_n^{(k)} p_{kl}\}}, \quad j \in \mathcal{S}. \qquad (5)$$

To make the above dependence on Y_{n+1} and π_n clear, we will sometimes write $\pi_{n+1}^{(j)}(Y_{n+1}, \pi_n)$ and $\pi_{n+1}(Y_{n+1}, \pi_n)$ instead of $\pi_{n+1}^{(j)}$ and π_{n+1}, respectively.

It follows that the *posterior distribution process* (π_n: $n \in \mathbb{Z}_+$) itself constitutes a temporally homogeneous Markov chain taking values in the space

$$\mathcal{P} = \left\{\pi \in [0, 1]^N : \sum_{i\in\mathcal{S}} \pi^{(i)} = 1\right\}, \qquad (6)$$

referred to as the probability simplex. In particular, the posterior distribution π_n makes one-step transitions in $\mathcal{P}$ according to (5). It will be convenient in later analysis to "abstract away" the details of (5) and generically denote the nominal transition probability function of π_n by

$$\begin{aligned}\rho(\pi_n, A) &\triangleq \mathbb{P}(\pi_{n+1} \in A \mid \mathcal{Y}_n) \\ &= \mathbb{P}(\{Y_{n+1} : \pi_{n+1}(Y_{n+1}, \pi_n) \in A\} \mid \pi_n), \\ &\qquad \pi \in \mathcal{P},\ A \subseteq \mathcal{P}, \end{aligned} \qquad (7)$$

where $\pi_{n+1}(Y_{n+1}, \pi_n)$ is given in (5).

If we define functions $h_1(x) = -C_{px}$ and $h_2(x) = R_x - C_x$ for all $x \in \mathcal{S}$, then

$$\mathbb{E}[h_1(X_n) \mid \pi_n = \pi] = -\sum_{i\in\mathcal{S}} C_{pi} \pi^{(i)}, \qquad (8)$$

represents the expected profit generated by initiating preventive maintenance given $\pi_n = \pi$, and

$$\mathbb{E}[h_2(X_n) \mid \pi_n = \pi] = \sum_{i\in\mathcal{S}} (R_i - C_i) \pi^{(i)}, \qquad (9)$$

represents the expected profit when running the system until the next sampling epoch given $\pi_n = \pi$.

For any $\pi \in \mathcal{P}$, standard dynamic programming arguments show that the nominal value function $V(\pi)$ defined in (3) is the unique solution to the dynamic programming (Bellman) equation

$$V(\pi) = \max\{\mathbb{E}_\pi[h_1(X_n)], \mathbb{E}_\pi[h_2(X_n)] + \alpha\,\mathbb{E}_{\rho(\pi,\cdot)}[V(\pi_{n+1})]\} \tag{10}$$

and the optimal control policy can in principle be obtained by solving (10). In particular, for all $\pi \in \mathcal{P}$ such that $V(\pi) = \mathbb{E}_\pi[h_1(X_n)]$, it is optimal to stop the system and carry out full system inspection followed by preventive maintenance. On the other hand, for all $\pi \in \mathcal{P}$ such that $V(\pi) = \mathbb{E}_\pi[h_2(X_n)] + \alpha\,\mathbb{E}_{\rho(\pi,\cdot)}[V(\pi_{n+1})]$, it is optimal to run the system until the next sampling epoch.

3. The Robust Model

In this section, we formulate a robust version of nominal control problem described in Section 2 that explicitly incorporates possible model miss-specification at the decision-making stage. To model possible miss-specification in the nominal model we first introduce the notions of positional and transitional ambiguity.

Positional and Transitional Ambiguity. Recall that under the nominal model assumptions we showed the posterior distribution π_n, which summarizes the decision-maker's current belief about the unobservable system state X_n, is sufficient for optimal decision making at each stage n. Furthermore, the posterior distribution process (π_n: $n \in \mathbb{Z}_+$) constitutes a temporally homogeneous Markov chain with state space $\mathcal{P}$ defined in (6) and transition probability function $\rho(\pi_n, \cdot)$ defined in (7). We let $D(\mathcal{P})$ be the set of all probability measures on $\mathcal{P}$, so that the nominal transition probability $\rho(\pi_n, \cdot)$ can be viewed as an element of $D(\mathcal{P})$.

If the decision maker distrusts the nominal model assumptions (e.g., the state transition matrix $\mathbf{P} = (p_{ij})_{N\times N}$ or the observation densities $f_1, \ldots, f_N$), then (5) suggests this distrust may carry over to both the value of the posterior distribution $\pi_n \in \mathcal{P}$ at each stage n, as well as the law governing its one-step transitions, $\rho(\pi_n, \cdot) \in D(\mathcal{P})$. The decision maker is therefore compelled to weight possible alternative specifications of π_n and $\rho(\pi_n, \cdot)$ when evaluating the quality of a given decision at each stage. We refer to distrust in π_n as *positional ambiguity*, since the current position of the posterior π_n in the probability simplex $\mathcal{P}$ may be miss-specified, and we refer to distrust in the nominal transition probability function $\rho(\pi_n, \cdot)$ as *transitional ambiguity* for obvious reasons.

To account for positional and transitional ambiguity, we introduce a hypothetical second agent, "nature," who is allowed to select an alternative posterior distribution $\upsilon_n \in \mathcal{P}$ and transition probability function $\eta_n \in D(\mathcal{P})$ in an adversarial manner. In effect, nature is introduced to help the decision maker make *robust decisions* that work well even when the nominal model is miss-specified. With these concepts of ambiguity in place, we now formalize the robust control problem as a zero-sum stochastic dynamic game.

Zero-Sum Stochastic Dynamic Game. The robust control problem is formulated as a zero-sum stochastic dynamic game that evolves in discrete time ($n \in \mathbb{Z}_+$), between the decision maker and the adversarial second agent "nature."

Let us consider what happens at an arbitrary stage n. At the beginning of the stage, the decision maker computes the current value of the posterior distribution π_n under the nominal model assumptions, which he or she fears may be miss-specified. The decision maker then "makes the first move" and chooses between stopping the system to initiating preventive maintenance ($a_n = 1$) or running the system until the next sampling epoch ($a_n = 2$). We note that the decision-maker's choice $a_n \in \{1, 2\}$ can only depend on historical observations $Y_1, \ldots, Y_n$ and previous choices $a_0, \ldots, a_{n-1}$. Mathematically, if we let $\mathcal{A}_{n-1} = \sigma(a_0, \ldots, a_{n-1})$ denote the sigma-algebra generated by all previous choices $a_0, \ldots, a_{n-1}$ of the decision maker, then we require that the decision-maker's current choice $a_n \in \{1, 2\}$ be $(\mathcal{Y}_n \vee \mathcal{A}_{n-1})$-measurable.

The decision maker accounts for possible miss-specification in the nominal posterior distribution $\pi_n \in \mathcal{P}$ (positional ambiguity) and nominal transition probability function $\rho(\pi_n, \cdot) \in D(\mathcal{P})$ (transitional ambiguity) by introducing the second agent, nature, who is permitted to choose an alternative posterior distribution $\upsilon_n \in \mathcal{P}$ and an alternative transition probability function $\eta_n \in D(\mathcal{P})$, with the objective of minimizing the decision-maker's expected profit.

In particular, once the decision maker selects a decision $a_n \in \{1, 2\}$, this choice is communicated to nature, who responds by choosing alternatives $\upsilon_n \in \mathcal{P}$ and $\eta_n \in D(\mathcal{P})$ so that the posterior distribution now follows $\mathbb{P}(X_n = j \mid \mathcal{Y}_n) = \upsilon_n^{(j)}$, $j \in \mathcal{S}$, and the posterior transition probability function now follows $\mathbb{P}(\pi_{n+1} \in A \mid \mathcal{Y}_n) = \eta_n(A)$, $A \subseteq \mathcal{P}$. We allow nature to have an informational advantage in the sense that its decision

$$b_n \triangleq (\upsilon_n, \eta_n) \in \mathcal{P} \times D(\mathcal{P}), \tag{11}$$

can depend on all historical observations $Y_1, \ldots, Y_n$ and state transitions $X_0, \ldots, X_n$, as well as all past choices of both the decision maker $a_0, \ldots, a_n$ and nature $b_0, \ldots, b_{n-1}$. Mathematically, if we let $\mathcal{B}_{n-1} = \sigma(b_0, \ldots, b_{n-1})$ denote the sigma-algebra generated by all previous choices $b_0, \ldots, b_{n-1}$ of nature, then we allow nature's current choice $b_n \in \mathcal{P} \times D(\mathcal{P})$ to be $\mathcal{F}_n \vee \mathcal{A}_n \vee \mathcal{B}_{n-1}$-measurable.

The decision-maker's level of confidence (or lack thereof) in the nominal model is captured by limiting the amount by which nature's choices υ_n and η_n can deviate from π_n and $\rho(\pi_n, \cdot)$, respectively, using the notion of *relative entropy*.

Definition 1 (Relative Entropy). Given a measurable space $(M, \mathcal{M})$ and probability measures φ, κ on $(M, \mathcal{M})$

such that φ is absolutely continuous with respect to κ, i.e., $\varphi \ll \kappa$, the *relative entropy* of φ with respect to κ is defined by

$$\mathcal{R}(\varphi \mid \kappa) \triangleq \int_M \log\left(\frac{d\varphi}{d\kappa}\right) d\varphi. \tag{12}$$

If φ is not absolutely continuous with respect to κ, then $\mathcal{R}(\varphi \mid \kappa) \triangleq +\infty$.

Relative entropy $\mathcal{R}(\varphi \mid \kappa)$ is convex in φ, nonnegative, and equals zero if and only if $\varphi = \kappa$, so one can think of $\mathcal{R}(\varphi \mid \kappa)$ as a "distance" from probability measure κ to probability measure φ.

Applying the notion of relative entropy in our context, nature is free to select alternative posteriors υ_n and transition probability functions η_n in an adversarial manner, but does so at a cost

$$\theta_P \mathcal{R}(\upsilon_n \mid \pi_n) + \theta_T \mathcal{R}(\eta_n \mid \rho(\pi_n, \cdot)), \tag{13}$$

which must be paid directly to the decision maker.

The first penalty parameter $\theta_P > 0$ controls the level of positional ambiguity, whereas the second penalty parameter $\theta_T > 0$ controls the level of transitional ambiguity. Specifically, a large value of θ_P (respectively, θ_T) makes it expensive for nature to choose an alternative υ_n (respectively, η_n) that is far from the nominal and corresponds to a high degree of confidence in π_n (respectively, $\rho(\pi_n, \cdot)$), while a small θ_P (respectively, θ_T) corresponds to less confidence in the nominal and forces the decision maker to play against a more powerful adversary since it is less costly for nature to deviate. In this regard, one can view θ_P and θ_T as a tuning parameters that parametrizes a family of policies of increasing robustness with decreasing θ_P and θ_T. In practice, to select *specific* values of the ambiguity parameters using historical data, the parameters can be tuned using out-of-sample tests (see, for example, the recent papers of Kim and Lim 2015 and Gotoh et al. 2016 who consider this issue in detail and propose a number of robust cross-validation procedures to select ambiguity parameters from data).

In summary, at each stage n, if the decision maker selects choice $a_n \in \{1, 2\}$ and nature responds by selecting alternatives $b_n = (\upsilon_n, \eta_n) \in \mathcal{P} \times D(\mathcal{P})$, the payoff to the decision maker is the sum of the immediate reward $\mathbb{E}_{\upsilon_n}[h_{a_n}(X_n)]$ defined in (8)–(9) and the penalty $\theta_P \mathcal{R}(\upsilon_n \mid \pi_n) + \theta_T \mathcal{R}(\eta_n \mid \rho(\pi_n, \cdot))$ on nature, i.e.,

$$\begin{aligned} r(\pi_n, a_n, b_n) \triangleq \mathbb{E}_{\upsilon_n}[h_{a_n}(X_n)] &+ \theta_P \mathcal{R}(\upsilon_n \mid \pi_n) \\ &+ \theta_T \mathcal{R}(\eta_n \mid \rho(\pi_n, \cdot)). \end{aligned}$$

The decision maker and nature choose equilibrium policies to solve a zero-sum stochastic dynamic game with *robust value function* defined by

$$W(\pi) \triangleq \sup_{a \in \mathcal{A}} \inf_{b \in \mathcal{B}} \mathbb{E}^{a,b}\left[\sum_{n=0}^{\sigma} \alpha^n r(\pi_n, a_n, b_n) \,\middle|\, \pi_0 = \pi\right], \tag{14}$$

where $\sigma = \inf\{n\colon a_n = 1\}$ is the first time the decision maker decides to stop the system for preventive maintenance, and the class of admissible policies for the decision maker and nature by

$$\mathcal{A} = \{a = (a_0, a_1, \ldots) \mid a_n \in \{1, 2\} \text{ is } (\mathcal{Y}_n \vee \mathcal{A}_{n-1})\text{-measurable}\},$$
$$\begin{aligned} \mathcal{B} = \{&b = (b_0, b_1, \ldots) \mid b_n \in \mathcal{P} \times D(\mathcal{P}) \\ &\text{is } (\mathcal{F}_n \vee \mathcal{A}_n \vee \mathcal{B}_{n-1})\text{-measurable}\}. \end{aligned}$$

Robust Dynamic Programming Equation. The robust value function defined in (14) satisfies the following dynamic programming equation.

Theorem 1. *At any arbitrary stage $n \in \mathbb{Z}_+$, given the current posterior is $\pi_n = \pi$, the robust value function $W(\pi)$ defined in (14) is the unique solution to*

$$W(\pi) = \max\{H^1(\pi, W), H^2(\pi, W)\}, \tag{15}$$

where

$$\begin{aligned} H^1(\pi, W) &= \inf_{(\upsilon, \eta) \in \mathcal{P} \times D(\mathcal{P})} \{\mathbb{E}_{\upsilon}[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon \mid \pi) \\ &\qquad + \theta_T \mathcal{R}(\eta \mid \rho(\pi, \cdot))\}, \\ H^2(\pi, W) &= \inf_{(\upsilon, \eta) \in \mathcal{P} \times D(\mathcal{P})} \{\mathbb{E}_{\upsilon}[h_2(X_n)] + \theta_P \mathcal{R}(\upsilon \mid \pi) \\ &\qquad + \theta_T \mathcal{R}(\eta \mid \rho(\pi, \cdot)) + \alpha \mathbb{E}_{\eta}[W(\pi_{n+1})]\}. \end{aligned} \tag{16}$$

The action $a_* \in \{1, 2\}$ that maximizes the right-hand side of (15), i.e., $W(\pi) = H^{a_*}(\pi, W)$, prescribes the optimal (robust) choice of the decision maker, and the ordered pair $b_* = (\upsilon_*, \eta_*)$ minimizing the right-hand side of (16) prescribes the optimal response of nature.

Furthermore, the robust value function $W(\pi)$ can be obtained as the limit $W(\pi) = \lim_{k \to \infty} W_k(\pi)$, where $W_k(\pi)$ is the unique solution of the k-stage robust dynamic programming equations

$$\begin{aligned} W_l(\pi) &= \max\{H^1(\pi, W_{l-1}), H^2(\pi, W_{l-1})\}, \quad l = 1, \ldots, k, \\ W_0(\pi) &= 0. \end{aligned} \tag{17}$$

The robust control problem (14) is a special case of the infinite-horizon α-discounted minimax control problem studied in González-Trejo et al. (2003), Section 4. Consequently, Theorem 1 follows as an application of González-Trejo et al. (2003, Theorem 4.2), for appropriately defined action sets for the decision maker and nature.

In the next section, we analyze the robust dynamic program (15) to analyze the structure of the optimal robust policy.

4. Structure of the Optimal Robust Policy

The goal of this section is to investigate the structural form of the optimal robust policy. That is, we would like to characterize the optimal choice of the decision maker at an arbitrary stage in the presence of the adversarial agent nature. To accomplish this, we analyze the robust dynamic programming equation (15) developed in the previous section, which (in principle) prescribes the optimal actions of both the decision maker and nature. The strategy we take is to first characterize the optimal response of nature. We do this by making use of a variational equality from the theory of large deviations. Characterizing nature's optimal response then allows us to transform the robust dynamic programming equation (15) into its so-called certainty equivalent. This representation is then used to establish structural properties of the decision-maker's optimal policy.

4.1. Structure of Nature's Optimal Policy

We begin by first studying nature's optimal policy in detail. Recall, in Theorem 1 we showed that at any stage n, given the current posterior is $\pi_n = \pi$, and the decision maker selects action $a \in \{1, 2\}$, the pair (υ_*, η_*) minimizing the right-hand side of (16) prescribes the optimal response of nature.

Let us first consider nature's response when the decision maker selects action $a = 1$, i.e., stop the system and initiate preventive maintenance. In this case, nature's minimization problem is

$$\begin{aligned} H^1(\pi, W) &= \inf_{(\upsilon,\,\eta)\in\mathcal{P}\times D(\mathcal{P})} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon|\pi) \\ &\qquad\qquad + \theta_T \mathcal{R}(\eta|\rho(\pi,\cdot))\big\} \\ &= \inf_{\upsilon\in\mathcal{P}} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon|\pi)\big\} \\ &\quad + \inf_{\eta\in D(\mathcal{P})} \big\{\theta_T \mathcal{R}(\eta|\rho(\pi,\cdot))\big\} \\ &= \inf_{\upsilon\in\mathcal{P}} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon|\pi)\big\} \\ &\quad + \theta_T \inf_{\eta\in D(\mathcal{P})} \mathcal{R}(\eta|\rho(\pi,\cdot)). \end{aligned} \tag{18}$$

The first term on the right-hand side of (18) captures positional ambiguity, and can be simplified by restricting the space of alternatives $\upsilon \in \mathcal{P}$ without loss of optimality. In particular, we note that nature will only explore alternative posterior distributions $\upsilon \in \mathcal{P}$ that are *absolutely continuous* with respect to the nominal posterior $\pi \in \mathcal{P}$, i.e., $\upsilon^{(i)} = 0$ whenever $\pi^{(i)} = 0$, $i \in \mathcal{S}$. This is because, if υ is not absolutely continuous with respect to π, by the definition of relative entropy (12), $\mathcal{R}(\upsilon\,|\,\pi) = +\infty$, which is clearly a suboptimal response of nature.

The second term on the right-hand side of (18) captures transitional ambiguity and in this case is a trivial minimization problem—nature will simply select $\eta_* \equiv \rho(\pi, \cdot)$ so that $\theta_T \inf_\eta \mathcal{R}(\eta\,|\,\rho(\pi,\cdot)) = \theta_T \mathcal{R}(\rho(\pi,\cdot)\,|\,\rho(\pi,\cdot)) = 0$. This optimal response of nature is intuitive. If the decision maker stops the system, there are no further transitions of the posterior distribution π_n, and therefore nature cannot hurt the decision-maker's expected profit by selecting alternative transition probability functions $\eta \in D(\mathcal{P})$, so it has no incentive to deviate from the nominal $\rho(\pi, \cdot)$.

Using these two observations, nature's minimization problem (18) can now be simplified to

$$\begin{aligned} H^1(\pi, W) &= \inf_{\upsilon\in\mathcal{P}} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon\,|\,\pi)\big\} \\ &\quad + \theta_T \inf_{\eta\in D(\mathcal{P})} \mathcal{R}(\eta\,|\,\rho(\pi,\cdot)) \\ &= \inf_{\upsilon\ll\pi} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon\,|\,\pi)\big\}. \end{aligned} \tag{19}$$

To solve the minimization problem (19), we appeal to the following variational formula from the theory of large deviations (see, e.g., Dupuis and Ellis 1997).

Proposition 1. *Let $(M, \mathcal{M})$ be a measurable space, Z a bounded measurable function mapping M into $\mathbb{R}$, θ a positive constant, and κ a probability measure on $(M, \mathcal{M})$. Then, we have the following variational formula*:

$$\inf_{\varphi\ll\kappa} \left\{ \int_M Z\,d\varphi + \theta \mathcal{R}(\varphi|\kappa) \right\} = -\theta \log \int_E \exp\left(-\frac{Z}{\theta}\right) d\kappa. \tag{20}$$

Furthermore, the infimum in (20) is uniquely attained at probability measure $\varphi_* \ll \kappa$ with Radon-Nikodym derivative

$$\frac{d\varphi_*}{d\kappa}(\omega) = \frac{\exp(-Z(\omega)/\theta)}{\int_M \exp(-Z/\theta)\,d\kappa}, \quad \omega \in M. \tag{21}$$

Applying Proposition 1 to problem (19), we obtain

$$\begin{aligned} H^1(\pi, W) &= \inf_{\upsilon\ll\pi} \big\{\mathbb{E}_\upsilon[h_1(X_n)] + \theta_P \mathcal{R}(\upsilon\,|\,\pi)\big\} \\ &= -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_1(X_n)}{\theta_P}\right)\right]. \end{aligned} \tag{22}$$

Notice that there is no longer any model ambiguity (positional or transitional) in representation (22). In this sense, (22) can be thought of as a *certainty equivalent* of nature's original minimization problem (18). It turns out that representation (22) is easier to analyze and will be used as the basis to establish structural properties of the decision-maker's optimal policy in later analysis.

In addition, it follows from (21) that the worst-case alternative posterior distribution υ_* chosen by nature has Radon-Nikodym derivative

$$\frac{d\upsilon_*}{d\pi}(i) = \frac{\exp(-h_1(i)/\theta_P)}{\mathbb{E}_\pi[\exp(-h_1(X_n)/\theta_P)]}, \quad i \in \mathcal{S}, \tag{23}$$

so that the measure υ_* is given explicitly by

$$\upsilon_*^{(i)} = \frac{\exp(-h_1(i)/\theta_P)}{\mathbb{E}_\pi[\exp(-h_1(X_n)/\theta_P)]} \pi^{(i)}, \quad i \in \mathcal{S}. \tag{24}$$

Nature's optimal response (24) is intuitive. It modifies the nominal posterior distribution π by increasing the probability that the unobservable system state X_n belongs to states with high preventive maintenance costs, i.e., states i in which $-h_1(i) = C_{pi}$ is large, and decreases the probability that X_n belongs to states with low preventive maintenance costs, i.e., states i in which $-h_1(i) = C_{pi}$ is small.

We next consider nature's optimal response when the decision maker selects $a = 2$, i.e., continue running the system until the next sampling epoch. Using similar arguments to the previous case ($a = 1$), in this case nature's minimization problem can be simplified to

$$\begin{aligned} &H^2(\pi, W) \\ &= \inf_{(\upsilon,\,\eta)\in\mathcal{P}\times D(\mathcal{P})} \big\{\mathbb{E}_\upsilon[h_2(X_n)] + \theta_P \mathcal{R}(\upsilon|\pi) \\ &\qquad\qquad + \theta_T \mathcal{R}(\eta|\rho(\pi,\cdot)) + \alpha \mathbb{E}_\eta[W(\pi_{n+1})]\big\} \end{aligned}$$

$$
\begin{aligned}
&= \inf_{v\in\mathcal{P}}\{\mathbb{E}_v[h_2(X_n)]+\theta_P\mathcal{R}(v\,|\,\pi)\}\\
&\quad+\inf_{\eta\in D(\mathcal{P})}\{\alpha\mathbb{E}_\eta[W(\pi_{n+1})]+\theta_T\mathcal{R}(\eta\,|\,\rho(\pi,\cdot))\}\\
&= \inf_{v\ll\pi}\{\mathbb{E}_v[h_2(X_n)]+\theta_P\mathcal{R}(v\,|\,\pi)\}\\
&\quad+\inf_{\eta\ll\rho(\pi,\cdot)}\{\alpha\mathbb{E}_\eta[W(\pi_{n+1})]+\theta_T\mathcal{R}(\eta\,|\,\rho(\pi,\cdot))\}. \qquad (25)
\end{aligned}
$$

Notice that in this case, transitional ambiguity, captured by the second term on the right-hand side of (25), persists. Again, we apply Proposition 1 to (25), which gives

$$
\begin{aligned}
H^2(\pi, W) &= -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right]\\
&\quad-\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right]. \qquad (26)
\end{aligned}
$$

We summarize the above findings in the following result.

Theorem 2. *At any arbitrary stage n, given the current posterior is $\pi_n=\pi$, the robust value function $W(\pi)$ defined in* (14) *is the unique solution to*

$$W(\pi)=\max\{H^1(\pi,W),H^2(\pi,W)\}, \qquad (27)$$

where

$$
\begin{aligned}
H^1(\pi, W) &= -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_1(X_n)}{\theta_P}\right)\right],\\
H^2(\pi, W) &= -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right] \qquad (28)\\
&\quad-\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right].
\end{aligned}
$$

The action $a_*\in\{1,2\}$ that maximizes the right-hand side of (27), *i.e., $W(\pi)=H^{a_*}(\pi,W)$, prescribes the optimal (robust) choice of the decision maker.*

Remark 1. Observe that as ambiguity parameters θ_P, $\theta_T\to\infty$, the robust dynamic programming equations (27) and (28) reduce to the nominal dynamic programming equation (10).

Remark 2. One implication of Theorem 2 is that from a complexity perspective, the robust control problem (14) is no harder to solve than the nominal control problem defined in Section 2. While such problems are in general PSPACE-hard (see, e.g., Papadimitriou and Tsitsiklis 1987, Theorem 6), we will show that by deriving structural properties of the optimal robust policy the computational complexity can be reduced substantially (see Section 4.4).

Relation to Dynamic Measures of Risk. The certainty equivalent given in Equation (28) points toward a close relationship between the robust control problem considered in this paper and risk-averse control problems involving dynamic measures of risk (see, for example, the excellent recent papers of Çavuş and Ruszczyński 2014a, b, Ruszczyński 2010 for the coherent case, and Ruszczyński and Shapiro 2006 for the convex case). In particular, for any bounded measurable function $g\colon \mathcal{P}\to\mathbb{R}$ and $\pi\in\mathcal{P}$, if we define

$$\varphi(g,\pi,\rho(\pi,\cdot)) = -\theta_T\log\mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{g(X)}{\theta_T}\right)\right], \qquad (29)$$

then the sequence $\varphi(g,\pi_n,\rho(\pi_n,\cdot))$, $n\geqslant 0$, can be interpreted as one-step conditional risk mappings for *profit maximization* as defined, for example, in Ruszczyński and Shapiro (2006). The mappings are also Markov and stationary in that their dependence on the past is carried only via the current posterior π_n as opposed to the entire history $\mathcal{Y}_n$ (see, e.g., Çavuş and Ruszczyński 2014a, b; Ruszczyński 2010 for a similar construction in the coherent case). Building on this relation, we denote for short $\varphi_n(g(\pi_{n+1}))=\varphi(g,\pi_n,\rho(\pi_n,\cdot))$ and consider the following risk-averse control problem:

$$
\begin{aligned}
J(\pi_0) = \sup_{a\in\mathcal{A}}\{&r(\pi_0,a_0)+\varphi_0(\alpha(r(\pi_1,a_1)\\
&+\varphi_1(\alpha(r(\pi_2,a_2)+\cdots+\varphi_{\sigma-1}(\alpha r(\pi_\sigma,a_\sigma))\cdots))))\},
\end{aligned}
\qquad (30)
$$

where $\mathcal{A}$ is the class of admissible policies given in (3), $\sigma=\inf\{n\colon a_n=1\}$ is the first time the decision maker decides to stop the system for preventive maintenance, and the "reward" sequence

$$r(\pi_n,a_n) = -\theta_P\log\mathbb{E}_{\pi_n}\left[\exp\left(-\frac{h_{a_n}(X_n)}{\theta_P}\right)\right], \quad n\geqslant 0. \qquad (31)$$

If we define the dynamic programming operator

$$T(f)(\pi)=\max\{r(\pi,1),r(\pi,2)+\varphi(\alpha f,\pi,\rho(\pi,\cdot))\}, \qquad (32)$$

then, by adapting similar arguments given in Ruszczyński (2010, Section 7) or Bertsekas and Shreve (1978, Section 4) it can be shown that the operator T is a contraction mapping and the risk-averse value function $J(\pi)$ given in (30) uniquely satisfies $J(\pi)=T(J)(\pi)$. On the other hand, the dynamic programming equation $J(\pi)=T(J)(\pi)$ precisely coincides with the robust dynamic programming equations (27)–(28), so that Theorem 2 implies $J(\pi)=W(\pi)$, where $W(\pi)$ is the robust value function defined in (14).

We next analyze the certainty equivalent Equations (27)–(28) to establish structural properties of the optimal robust policy of the decision maker.

4.2. Structure of the Decision-Maker's Optimal Policy

In the maintenance optimization literature, one of the key structural properties to establish is convexity of the value function (see, e.g., Dayanik et al. 2008, Kim and Makis 2012, Maillart 2006, Makis 2008). In the current setting however, it is not clear that the robust value function $W(\pi)$ defined in (14) will preserve such a convex structure in the presence of positional and transitional model ambiguity. This is because the functions $H^1(\pi,W)$ and $H^2(\pi,W)$

are themselves optimization problems, which makes establishing convexity significantly more complex.

However, by exploiting the structure of nature's optimal policy given in the previous subsection, we are able to show that convexity of the robust value function is indeed preserved under model miss-specification. The proof is long, but we believe it contains useful techniques, and have therefore chosen to keep it in the main body of the paper.

THEOREM 3. *The robust value function $W(\pi)$ is convex in $\pi \in \mathcal{P}$.*

PROOF. By Equation (27) of Theorem 2, since the "max" operator preserves convexity, it suffices to show that both $H^1(\pi, W)$ and $H^2(\pi, W)$ given in (28) are convex functions of π. We start with the first function $H^1(\pi, W)$ given by

$$H^1(\pi, W) = -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_1(X_n)}{\theta_P}\right)\right]. \quad (33)$$

Since the term inside the "log" can be expressed as

$$\mathbb{E}_\pi\left[\exp\left(-\frac{h_1(X_n)}{\theta_P}\right)\right] = \sum_{i\in\mathcal{S}} \exp\left(-\frac{h_1(i)}{\theta_P}\right)\pi^{(i)}, \quad (34)$$

it is linear in π, and therefore also concave in π. Furthermore, since the mapping $z \mapsto -\theta_P \log(z)$ is convex and nonincreasing, it follows that the composition $H^1(\pi, W)$ given in (33) is convex in π.

We next study the second function $H^2(\pi, W)$ given by

$$H^2(\pi, W) = -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right] - \theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right]. \quad (35)$$

The first summand on the right-hand side of (35) is structurally equivalent to $H^1(\pi, W)$, so using the same reasoning as above, it is a convex function of π. Showing that the second summand is also convex cannot be argued so cleanly; we proceed by means of mathematical induction.

For $k = 1$, $H^1(\pi, W_{k-1}) = H^1(\pi, W_0)$ remains the same as in (33). On the other hand, $H^2(\pi, W_{k-1}) = H^2(\pi, W_0)$ simplifies to

$$H^2(\pi, W_0) = -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right], \quad (36)$$

since $W_0 \equiv 0$. Therefore, both $H^1(\pi, W_0)$ and $H^2(\pi, W_0)$ are convex in π (as argued above), and hence $W_1(\pi) = \max\{H^1(\pi, W_0), H^2(\pi, W_0)\}$ is convex in π.

Assume now that for some $k \geqslant 1$, $W_k(\pi)$ is a convex function of π. We want to show that $W_{k+1}(\pi)$ is also convex in π. Again, $H^1(\pi, W_k)$ remains the same as in (33), whereas $H^2(\pi, W_k)$ is given by

$$H^2(\pi, W_k) = -\theta_P \log \mathbb{E}_\pi\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right] - \theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W_k(\pi_{n+1})}{\theta_T}\right)\right]. \quad (37)$$

Since the first summand on the right-hand side of (37) has already been shown to be convex in π, it suffices to show that the second summand is also convex in π, which we denote by

$$\mathscr{C}(\pi) \triangleq -\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W_k(\pi_{n+1}(Y_{n+1}, \pi))}{\theta_T}\right)\right], \quad (38)$$

where $\pi_{n+1}(Y_{n+1}, \pi)$ is given after Equation (5).

Fix any two posteriors $\gamma_1, \gamma_2 \in \mathcal{P}$, and for any constant $\beta \in [0, 1]$ define $\gamma_\beta = \beta\gamma_1 + (1-\beta)\gamma_2$. Suppose the current posterior $\pi_n = \gamma_\beta$, then by Equation (5) for each $j \in \mathcal{S}$,

$$\begin{aligned}\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_\beta) &= \frac{f_j(Y_{n+1})\sum_{i\in\mathcal{S}} \gamma_\beta^{(i)} p_{ij}}{\sum_{l\in\mathcal{S}} f_l(Y_{n+1})\sum_{k\in\mathcal{S}} \gamma_\beta^{(k)} p_{kl}} \\ &= \hat{\beta}\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_1) \\ &\quad + (1-\hat{\beta})\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_2), \end{aligned} \quad (39)$$

where

$$\hat{\beta} \triangleq \beta \frac{\sum_{l\in\mathcal{S}} f_l(Y_{n+1})\sum_{k\in\mathcal{S}} \gamma_1^{(k)} p_{kl}}{\sum_{l\in\mathcal{S}} f_l(Y_{n+1})\sum_{k\in\mathcal{S}} \gamma_\beta^{(k)} p_{kl}}. \quad (40)$$

It is not difficult to see that $\hat{\beta}$ defined above is bounded in the unit interval, i.e., $\hat{\beta} \in [0, 1]$.

Then, applying representation (39), $\mathscr{C}(\gamma_\beta)$ defined in Equation (38) satisfies

$$\begin{aligned}\mathscr{C}(\gamma_\beta) &= -\theta_T \log \mathbb{E}_{\rho(\gamma_\beta,\cdot)}\left[\exp\left(-\frac{\alpha W_k(\pi_{n+1}(Y_{n+1}, \gamma_\beta))}{\theta_T}\right)\right] \\ &\leqslant -\theta_T \log \mathbb{E}_{\rho(\gamma_\beta,\cdot)}[\exp(-(\hat{\beta} W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_1)) \\ &\qquad + (1-\hat{\beta}) W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_2))) \cdot (\theta_T/\alpha)^{-1})],\end{aligned}$$

where the last inequality follows since $\hat{\beta} \in [0, 1]$ and we have assumed convexity of $W_k(\cdot)$. Also, since the mapping $z \mapsto \exp(-z)$ is convex, it follows that $\mathscr{C}(\gamma_\beta)$ can be further bounded as

$$\begin{aligned}\mathscr{C}(\gamma_\beta) &\leqslant -\theta_T \log \mathbb{E}_{\rho(\gamma_\beta,\cdot)}\left[\hat{\beta}\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_1))}{\theta_T/\alpha}\right)\right. \\ &\qquad \left. + (1-\hat{\beta})\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_2))}{\theta_T/\alpha}\right)\right] \\ &= -\theta_T \log\left\{\beta \mathbb{E}_{\rho(\gamma_1,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_1))}{\theta_T/\alpha}\right)\right]\right. \\ &\qquad \left. + (1-\beta)\mathbb{E}_{\rho(\gamma_2,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_2))}{\theta_T/\alpha}\right)\right]\right\},\end{aligned}$$

where the last equality above follows from the definition of $\hat{\beta}$. To see this, from (40) we have

$$\begin{aligned}&\mathbb{E}_{\rho(\gamma_\beta,\cdot)}\left[\hat{\beta}\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1}, \gamma_1))}{\theta_T/\alpha}\right)\right] \\ &\quad = \int_{y\in E} \hat{\beta}\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(y, \gamma_1))}{\theta_T/\alpha}\right)\end{aligned}$$

$$\cdot\left(\sum_{l\in\mathcal{S}} f_l(y)\sum_{k\in\mathcal{S}}\gamma_\beta^{(k)}p_{kl}\right)\lambda(dy)$$
$$=\beta\int_{y\in E}\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(y,\gamma_1))}{\theta_T/\alpha}\right)$$
$$\cdot\left(\sum_{l\in\mathcal{S}} f_l(y)\sum_{k\in\mathcal{S}}\gamma_1^{(k)}p_{kl}\right)\lambda(dy)$$
$$=\beta\,\mathbb{E}_{\rho(\gamma_1,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1},\gamma_1))}{\theta_T/\alpha}\right)\right]. \quad (41)$$

Similarly, it can be shown that

$$\mathbb{E}_{\rho(\gamma_\beta,\cdot)}\left[(1-\hat{\beta})\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1},\gamma_2))}{\theta_T/\alpha}\right)\right]$$
$$=(1-\beta)\,\mathbb{E}_{\rho(\gamma_2,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1},\gamma_2))}{\theta_T/\alpha}\right)\right],$$

so that the last equality in (41) follows.

Lastly, using the fact that the mapping $z\mapsto -\theta_T\log(z)$ is convex, it follows that $\mathscr{C}(\gamma_\beta)$ can be further bounded above as

$$\mathscr{C}(\gamma_\beta)\leqslant -\beta\theta_T\log\mathbb{E}_{\rho(\gamma_1,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1},\gamma_1))}{\theta_T/\alpha}\right)\right]$$
$$-(1-\beta)\theta_T$$
$$\cdot\log\mathbb{E}_{\rho(\gamma_2,\cdot)}\left[\exp\left(-\frac{W_k(\pi_{n+1}^{(j)}(Y_{n+1},\gamma_2))}{\theta_T/\alpha}\right)\right]$$
$$=\beta\mathscr{C}(\gamma_1)+(1-\beta)\mathscr{C}(\gamma_2).$$

Therefore, the second summand $\mathscr{C}(\pi)$ of $H^2(\pi,W_k)$ is convex in π, which implies $H^2(\pi,W_k)$ is convex in π, which implies further that

$$W_{k+1}(\pi)=\max\{H^1(\pi,W_k),H^2(\pi,W_k)\}, \quad (42)$$

is convex in π. Thus, by mathematical induction, for each $k\geqslant 1$, $W_k(\pi)$ is a convex function of π. Therefore, $W(\pi)=\lim_{k\to\infty}W_k(\pi)$ is also convex in π, which completes the proof. □

We now make the following definition.

Definition 2. The *optimal robust stopping region* is defined by the following subset of the probability simplex $\mathcal{P}$:

$$\Sigma\triangleq\{\pi\in\mathcal{P}\colon H^1(\pi,W)>H^2(\pi,W)\}, \quad (43)$$

where $H^1(\pi,W)$ and $H^2(\pi,W)$ are defined in (16).

The set Σ represents all posteriors $\pi\in\mathcal{P}$, such that whenever the current posterior $\pi_n=\pi$ at an arbitrary stage n, the optimal choice for the decision maker (in the presence of nature) is to stop the system and initiate preventive maintenance. We next show that the failure point $\pi_\xi\triangleq(0,\ldots,0,1)\in\mathcal{P}$ belongs to this optimal robust stopping region.

Lemma 1. *If at any stage n, $\pi_n=\pi_\xi$, it is optimal to stop the system to initiate preventive maintenance, i.e., $\pi_\xi\in\Sigma$.*

A proof of Lemma 1 is given in Appendix A. Notice that Lemma 1 is not strong enough to assert that preventive maintenance will *always* be employed in the presence of model ambiguity. For example, if it turned out that the optimal robust stopping region were the singleton set $\Sigma=\{\pi_\xi\}$, it is not difficult to see that the posterior π_n will never enter Σ in finite time, and consequently the decision maker would never initiate preventive maintenance.

In the next result, we show that this situation is not possible. In particular, we show that the policy that never stops the system for preventive maintenance is not optimal. In this sense, the next result can be viewed as a robustness result for the preventive maintenance paradigm. That is, preventive maintenance is always a worthwhile strategy for managing failure risk, even when the decision maker distrusts the underlying model assumptions.

Proposition 2. *The optimal robust stopping region Σ is an open set in the subspace topology (induced topology) on the probability simplex $\mathcal{P}\subset\mathbb{R}^N$. Consequently, if we let $\sigma_\Sigma\triangleq\inf\{n\in\mathbb{Z}_+\colon \pi_n\in\Sigma\}$ be the hitting time of $\Sigma\subseteq\mathcal{P}$, then for any $\pi\in\mathcal{P}$,*

$$\mathbb{P}(\sigma_\Sigma<+\infty\mid\pi_0=\pi)=1. \quad (44)$$

In words, the policy that never stops the system for preventive maintenance is not optimal.

Proof. We first show that Σ is an open set in the subspace topology on $\mathcal{P}$. Define the difference function $f\colon\mathcal{P}\to\mathbb{R}$ by

$$f(\pi)\triangleq H^1(\pi,W)-H^2(\pi,W),$$

for all $\pi\in\mathcal{P}$. Since we have shown in the proof of Theorem 3 that both $H^1(\pi,W)$ and $H^2(\pi,W)$ are convex in π, it follows that f is a continuous function. Therefore, from general point-set topology (see, e.g., Munkres 2000, p. 102), it follows that the preimage of the open interval $(0,+\infty)$ under f, i.e., $f^{-1}((0,+\infty))=\Sigma$, is an open set of $\mathcal{P}$.

We next use the fact that Σ is open to show that the hitting time

$$\sigma_\Sigma=\inf\{n\in\mathbb{Z}_+\colon \pi_n\in\Sigma\},$$

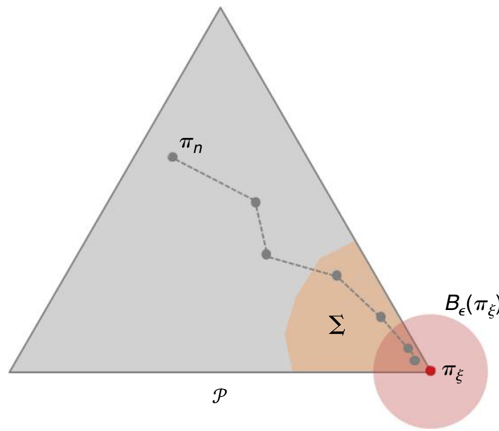
is finite almost surely, i.e., $\mathbb{P}(\sigma_\Sigma<+\infty\mid\pi_0=\pi)=1$ for any $\pi\in\mathcal{P}$. Recall, from Lemma 1 that $\pi_\xi\in\Sigma$. Since Σ is open in the subspace topology on $\mathcal{P}\subset\mathbb{R}^N$, it follows by definition of an open set (see, e.g., Munkres 2000, p. 88) that there exists $\epsilon>0$ such that the open ball $B_\epsilon(\pi_\xi)=\{\pi\in\mathbb{R}^N\colon\|\pi-\pi_\xi\|_2<\epsilon\}$ satisfies

$$B_\epsilon(\pi_\xi)\cap\mathcal{P}\subseteq\Sigma.$$

Since we have assumed that the underlying state process X_n is nondecreasing with probability 1, it follows that for any $\pi\in\mathcal{P}$,

$$\mathbb{P}(\pi_n\in\Sigma\mid\pi_0=\pi)\geqslant\mathbb{P}(\pi_n\in B_\epsilon(\pi_\xi)\cap\mathcal{P}\mid\pi_0=\pi)\to 1,$$

Figure 1. (Color online) A picture proof of Proposition 2: The optimal robust stopping region $\Sigma \subseteq \mathcal{P}$ contains an open ball $B_\epsilon(\pi_\xi)$ centered at failure point π_ξ.



Note. Since the posterior probability process $\pi_n \in \mathcal{P}$ enters this open ball in finite time almost surely, it must also enter the optimal robust stopping region in finite time.

as $n \to \infty$. See Figure 1 for an illustration. Therefore, for any $0 < \delta < 1$, there exists a nonnegative integer $M(\delta) \in \mathbb{Z}_+$ such that $\mathbb{P}(\pi_{M(\delta)} \in \Sigma \mid \pi_0 = \pi) \geqslant \delta$. We now construct a new (related) two-state Markov chain $Z_n \in \{1, 2\}$ with transition probability matrix

$$\mathbf{Q} = \begin{pmatrix} 1-\delta & \delta \\ 0 & 1 \end{pmatrix}.$$

If we let $\tau = \inf\{n \in \mathbb{Z}_+\colon Z_n = 2\}$, then by construction it follows that

$$\begin{aligned}\mathbb{E}[\sigma_\Sigma \mid \pi_0 = \pi] &\leqslant \mathbb{E}[M(\delta)\cdot\tau \mid Z_0 = 1]\\ &= M(\delta)\cdot\mathbb{E}[\tau \mid Z_0 = 1] = \frac{M(\delta)}{\delta} < +\infty.\end{aligned}$$

Hence, $\mathbb{P}(\sigma_\Sigma < +\infty \mid \pi_0 = \pi) = 1$, which completes the proof. ☐

We have shown in Lemma 1 and Proposition 2 that the optimal robust stopping region is an open set in the subspace topology on the probability simplex, which also contains the failure point π_ξ. One may hope to show that this stopping region is also a *convex* subset of the probability simplex. However, as the following reasoning demonstrates, this cannot be established generally.

From the proof of Theorem 3, we have shown that $H^1(\pi, W)$ and $H^2(\pi, W)$ defined in (16) can be expressed as

$$\begin{aligned} H^1(\pi, W) &= -\theta_P \log \mathcal{L}^1(\pi),\\ H^2(\pi, W) &= -\theta_P \log \mathcal{L}^2(\pi) + \mathcal{C}(\pi),\end{aligned} \tag{45}$$

where $\mathcal{L}^1(\pi)$ and $\mathcal{L}^2(\pi)$ are linear functions in π given by

$$\mathcal{L}^i(\pi) = \mathbb{E}_\pi\left[\exp\left(-\frac{h_i(X_n)}{\theta_P}\right)\right], \quad i = 1, 2, \tag{46}$$

and $\mathcal{C}(\pi)$ is a convex function in π given by

$$\mathcal{C}(\pi) = -\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right]. \tag{47}$$

The optimal robust stopping region Σ can therefore be expressed as

$$\Sigma = \left\{\pi \in \mathcal{P}\colon \theta_P \log\left(\frac{\mathcal{L}^1(\pi)}{\mathcal{L}^2(\pi)}\right) + \mathcal{C}(\pi) < 0\right\}, \tag{48}$$

which is a sublevel set of the function

$$h(\pi) = \theta_P \log\left(\frac{\mathcal{L}^1(\pi)}{\mathcal{L}^2(\pi)}\right) + \mathcal{C}(\pi). \tag{49}$$

Therefore, the question of convexity of Σ can be posed as a question of *quasiconvexity* of the function $h(\pi)$.

Definition 3. A function $h\colon \mathbb{R}^N \to \mathbb{R}$ is called *quasiconvex* if both its domain and all of its sublevel sets

$$S_\gamma(h) \triangleq \{x \in \operatorname{dom}(h)\colon h(x) \leqslant \gamma\}, \tag{50}$$

for $\gamma \in \mathbb{R}$, are convex.

Although it can be shown that the two summands $\theta_P \log(\mathcal{L}^1(\pi)/\mathcal{L}^2(\pi))$ and $\mathcal{C}(\pi)$ of $h(\pi)$ are individually both quasiconvex in π (see, e.g., Boyd and Vandenberghe 2004), the sum of two quasiconvex functions is not necessarily quasiconvex (see, e.g., Greenberg and Pierskalla 1971). Therefore, the function $h(\pi)$ is not necessarily quasiconvex, which implies the optimal robust stopping region Σ is not necessarily a convex subset of $\mathcal{P}$.

The following condition will however lead to convexity of the stopping region Σ.

Assumption 1 (Semi-Weakened Nature). *Assume that nature still has the ability to select alternative transition probability functions $\eta_n \in D(\mathcal{P})$ in an adversarial manner, but can no longer select alternative posterior distributions $v_n \in \mathcal{P}$, i.e., $v_n = \pi_n$. Mathematically, we impose this assumption by setting positional ambiguity penalty parameter $\theta_P = +\infty$.*

In other words, we weaken nature in the sense that we allow for transitional ambiguity, but we do not allow for positional ambiguity.

Proposition 3. *Under Assumption 1, the optimal robust stopping region Σ defined in (43) is a convex subset of the probability simplex $\mathcal{P}$.*

A proof of Proposition 3 is given in Appendix B. The structure of the decision-maker's optimal policy can be sharpened further when we specialize the above results to the robust Bayesian change point detection problem (i.e., the case where $N = 2$). In this case, the optimal robust policy is a simple threshold policy.

Theorem 4 (Optimality of Robust Threshold Policy). *Under Assumption 1, the optimal policy for the robust Bayesian change point detection problem* ($N = 2$) *is a threshold policy. Specifically, there exists* $\bar{\pi} \in [0, 1)$ *such that at any stage n, whenever the posterior distribution* $\pi_n = (\pi_n^{(1)}, \pi_n^{(2)})$ *is such that* $\pi_n^{(2)} > \bar{\pi}$, *it is optimal to stop the system and initiate preventive maintenance* ($a = 1$). *Otherwise, if* $\pi_n^{(2)} \leqslant \bar{\pi}$, *it is optimal to run the system until the next sampling epoch* ($a = 2$).

A proof of Theorem 4 is given in Appendix C. Theorem 4 has both managerial insights and practical value—it shows that the optimal robust control policy of the decision maker can be visualized and implemented as a *control chart*, which monitors the conditional probability $\pi_n^{(2)}$ that the system is in a failure state (see Figure 2).

4.3. Consistency of the Robust Model and Related Data Issues

Let us denote for short the nominal model as $\mathbb{P} = \{(p_{ij}), f_1, \ldots, f_N\}$, and the posterior state vector as $\pi_n(Y_{(n)})$ given a data set $Y_{(n)} = (Y_1, \ldots, Y_n)$. In this section, we investigate the situation in which data $\tilde{Y}_{(n)} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$ are actually generated from a data generating model $\tilde{\mathbb{P}} = \{(\tilde{p}_{ij}), \tilde{f}_1, \ldots, \tilde{f}_N\}$ that may be different than $\mathbb{P}$. We first study the behavior of the posterior state vector $\pi_n = \pi_n(\tilde{Y}_{(n)})$ and its associated transition probability function $\rho_n = \rho(\pi_n, \cdot)$ constructed under the nominal model $\mathbb{P}$ when data $\tilde{Y}_{(n)}$ are generated from $\tilde{\mathbb{P}}$. We assume that the data generating model $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ in that for each $i \in \mathcal{S}$, $\tilde{p}_{i\cdot} \ll p_{i\cdot}$ and $\tilde{f}_i \ll f_i$. We also make the following assumption on $\tilde{\mathbb{P}}$.

Assumption 2 (Distinguishable Failure Signals). *For each* $i \in \mathcal{S} \backslash \{N\}$, $\tilde{\mathbb{P}}$ *satisfies*

$$\mathbb{E}_{\tilde{f}_N}\left[\log\left(\frac{f_N(Y)}{f_i(Y)}\right)\right] > 0. \tag{51}$$

To understand the intuition behind Assumption 2, consider the case where $N = 2$, and suppose $\tilde{f}_2 = f_2$. In this case, the quantity in (51) is nothing more than $\mathcal{R}(f_2 \mid f_1)$ the relative entropy of f_2 with respect to f_1 defined in (12), which is positive whenever $f_1 \neq f_2$. On the other hand, if $\tilde{f}_2 \neq f_2$, we can express Equation (51) as

$$\begin{aligned}\mathbb{E}_{\tilde{f}_2}&\left[\log\left(\frac{f_2(Y)}{f_1(Y)}\right)\right]\\ &= \int \log\left(\frac{f_2(y)}{f_1(y)}\right)\tilde{f}_2(y)\, d\lambda(y)\\ &= \int \log\left(\frac{f_2(y)}{f_1(y)}\right)\left(\frac{\tilde{f}_2(y)}{f_2(y)}\right)f_2(y)\, d\lambda(y).\end{aligned} \tag{52}$$

From the above equation, we see that the term in (51) can be thought of as a modification (change of measure) to standard relative entropy $\mathcal{R}(f_2 \mid f_1)$ by the Radon-Nikodym derivative $\tilde{f}_2/f_2$. In this regard, Assumption 2 is quite mild as it effectively states that the only restriction placed on $\tilde{\mathbb{P}}$ is that the resulting "modified" relative entropy must remain positive, i.e., the failure density must remain statistically distinguishable from operational-state densities under this change of measure. An example of a data generating model $\tilde{\mathbb{P}}$ that would not satisfy Equation (51) is as follows. Suppose $\tilde{f}_1 = f_2$ and $\tilde{f}_2 = f_1$, i.e., the operational and failure signal densities are swapped. In this case, one would not expect to be able to distinguish between operational and failure states since data coming from a failure state will induce a belief that the system is still operational, and vice versa. Indeed, in this case, $\mathbb{E}_{\tilde{f}_2}[\log(f_2(Y)/f_1(Y))] = \mathbb{E}_{f_1}[\log(f_2(Y)/f_1(Y))] = -\mathcal{R}(f_1 \mid f_2) < 0$, so that Equation (51) is not satisfied.

If we now let $\tilde{\pi}_n = \tilde{\pi}_n(\tilde{Y}_{(n)})$ and $\tilde{\rho}_n = \tilde{\rho}(\tilde{\pi}_n, \cdot)$ be the posterior state vector and transition probability function under the data generating model $\tilde{\mathbb{P}}$, then it is natural to ask whether the error between π_n and $\tilde{\pi}_n$ (or ρ_n and $\tilde{\rho}_n$) will accumulate or diminish over time. The next result shows that the error is not only bounded over time, but tends to zero ($\tilde{\mathbb{P}}$-a.s.).

Proposition 4. *Suppose model* $\tilde{\mathbb{P}}$ *satisfies Assumption 2. Then, for each* $i \in \mathcal{S}$,

$$|\pi_n^{(i)} - \tilde{\pi}_n^{(i)}| \to 0, \quad \tilde{\mathbb{P}}\text{-a.s.}, \tag{53}$$

and for each measurable subset $A \subset \mathcal{P}$,

$$|\rho_n(A) - \tilde{\rho}_n(A)| \to 0, \quad \tilde{\mathbb{P}}\text{-a.s.} \tag{54}$$

A proof of Proposition 4 is given in Appendix D. Taking this discussion one step further, we next establish a similar consistency result for the worst-case posterior $v_{*,n}$ and transition probability function $\eta_{*,n}$ constructed in Section 4.1 with respect to the (incorrect) nominal model $\mathbb{P}$. In particular, we ask whether the error between the worst-case posterior $v_{*,n}$ and $\tilde{\pi}_n$ (or $\eta_{*,n}$ and $\tilde{\rho}_n$) will accumulate or diminish over time. The answer to this question is not obvious since one implicit assumption made in the robust model is that the ambiguity parameters θ_P and θ_T are fixed constants that do not change with time. This effectively means that nature's power/ability to select alternative worst-case measures $v_{*,n}$ and $\eta_{*,n}$ does not diminish over time, so it is conceivable that consistency can never be achieved. The next result shows that in spite of this, the error still tends to zero.

Proposition 5. *Suppose model* $\tilde{\mathbb{P}}$ *satisfies Assumption 2. Then, for each* $i \in \mathcal{S}$,

$$|v_{*,n}^{(i)} - \tilde{\pi}_n^{(i)}| \to 0, \quad \tilde{\mathbb{P}}\text{-a.s.}, \tag{55}$$

and for each measurable subset $A \subset \mathcal{P}$,

$$|\eta_{*,n}(A) - \tilde{\rho}_n(A)| \to 0, \quad \tilde{\mathbb{P}}\text{-a.s.} \tag{56}$$

A proof of Proposition 5 is given in Appendix E. We conclude this section with a numerical investigation of the robust Bayesian change point detection problem applied to a real-world condition-based maintenance problem coming from mining industry.

4.4. Application to Bayesian Change Point Detection

Consider the following real-world application from the mining industry, which will constitute our nominal model. The system can be in one of two unobservable states—a healthy state 1 and a failure state 2. The system state $X_n \in \{1, 2\}$ follows a Markov chain with nominal state transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.97 & 0.03 \\ 0 & 1 \end{bmatrix}.$$

In this application, each time step represents 600 operational hours. At the beginning of each stage, oil samples are collected and spectrometric analysis is carried out, which provides a d-dimensional vector of concentration levels $Y_n \in \mathbb{R}^d$ (measured in parts per million) of wear metals stochastically related to the underlying system state (see, e.g., Stachowiak et al. 2004). Here, we will consider $d = 2$ wear metals, where the first wear metal is copper (Cu) and the second wear metal is iron (Fe). The nominal model assumes that these two wear metal concentrations $Y_n \in \mathbb{R}^2$ follow a bivariate normal distribution $\mathcal{N}(\mu_1, \Sigma_1)$ when the system is in a healthy state 1, and $\mathcal{N}(\mu_2, \Sigma_2)$ when the system is in a failure state 2, where observation parameters are given by

$$\mu_1 = \begin{bmatrix} 1.1 \\ 1.9 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 4.1 \\ 5.5 \end{bmatrix},$$
$$\Sigma_1 = \begin{bmatrix} 7.2 & 2.0 \\ 2.0 & 3.6 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 7.6 & 1.0 \\ 1.0 & 3.2 \end{bmatrix}.$$

Cost parameters for this application are $C_{p1} = 450$, $C_{p2} = 1{,}000$, $C_1 = 10$, $C_2 = 100$, $R_1 = 100$, $R_2 = 10$, and the discount factor is $\alpha = 0.95$.

We note that in this particular application, the historical data set used to select and calibrate the above nominal model was relatively small and, in some parts, of poor quality. Accordingly, one of the primary concerns with using the above nominal model as the basis for decision making is that it may be miss-specified. We now illustrate how the theoretical results developed in the previous section can be implemented in practice to help come up with decisions that are robust with respect to possible miss-specification in the nominal model.

Suppose now that the initial prior distribution over the system state is $\pi_0 = (\pi_0^{(1)}, \pi_0^{(2)}) = (0.95, 0.05)$. Theorem 4 shows that the optimal robust policy for the decision maker is characterized by a threshold policy. This implies that the policy can be visualized and implemented as a simple *control chart*, in which one plots the posterior probability $\pi_n^{(2)}$ that the system is in the failure state, as it evolves over time according to Bayes' Theorem. Once the posterior probability $\pi_n^{(2)}$ exceeds a fixed threshold $\bar{\pi}$, the decision maker stops the system to initiate preventive maintenance. Accordingly, the problem reduces to finding the optimal value of the threshold $\bar{\pi} \in [0, 1)$, which considerably simplifies the computational complexity of the problem.

In particular, the optimal policy is now identified with a univariate parameter $\bar{\pi}$ contained in the unit interval, so that rather than searching over all possible robust policies, we need only search over the unit interval to determine the optimal robust policy. More specifically, the optimal threshold can be efficiently computed by discretizing the state space $[0, 1)$, i.e., restricting the search to thresholds of the form $\bar{\pi} = m/M$, for $M \in \mathbb{N}$ large and $m = 0, 1, \ldots, M-1$. Under this discretization, we can adapt the computational algorithm developed in Kim and Makis (2012, Section 5), which has complexity $O(M^4)$. This complexity comes from the fact that for each of the M possible values of the threshold $\bar{\pi}$, the algorithm run Gaussian elimination on m variables, which has complexity $O(M^3)$.

For example, suppose we choose an ambiguity parameter $\theta_T = 10$, using MATLAB we obtain an optimal robust threshold $\bar{\pi} = 0.74$. To compute this optimal robust threshold value, it took 4.66 seconds on an Intel Core i5-2400, 3.10-GHz with 4.00 GB RAM. Figure 2 illustrates the use of this optimal robust threshold policy on a sample data history.

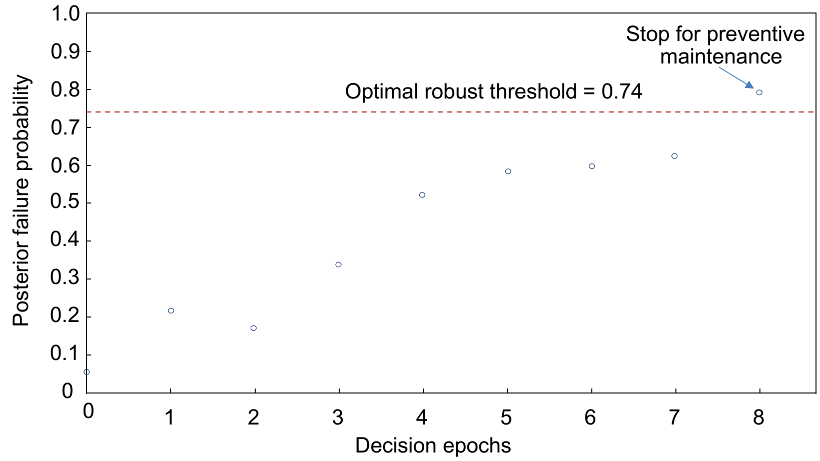
In the current context, one may be interested to know how the optimal robust threshold $\bar{\pi}$ changes as a function of the ambiguity parameter θ_T. Let $\bar{\pi}(\theta_T)$ be the optimal robust threshold value when using ambiguity parameter value θ_T. For example, in Figure 2, we found that $\bar{\pi}(10) = 0.74$. We now compute $\bar{\pi}(\theta_T)$ over a wide range of θ_T and plot the results in Figure 3.

From Figure 3, we make a few interesting observations. First, the optimal robust threshold $\bar{\pi}(\theta_T)$ is nondecreasing in the value of the ambiguity parameter θ_T. The intuition behind this result is as follows. At the core, the primary cause of concern (risk) for the decision maker is the cost due to unobservable failure. When the ambiguity parameter is small, nature has more power to deviate from the nominal model and select alternatives that magnify this failure risk. As a result, the decision maker will naturally lower the threshold value $\bar{\pi}(\theta_T)$ to mitigate such risk. In short, smaller values of the ambiguity parameter leads to smaller values of optimal robust threshold, which explains the monotonicity result of Figure 3. The second interesting observation is that in this application we can clearly see that the optimal threshold approaches a limit $\bar{\pi}_\infty$ as $\theta_T \to \infty$. Moreover, from the previous sections, we know what this limit is; $\bar{\pi}_\infty$ is nothing more than the nominal optimal threshold value $\bar{\pi}_{\text{nom}}$ obtained from solving the nominal model without model miss-specification. In this particular application, $\bar{\pi}_\infty = \bar{\pi}_{\text{nom}} = 0.87$.

5. Conclusions

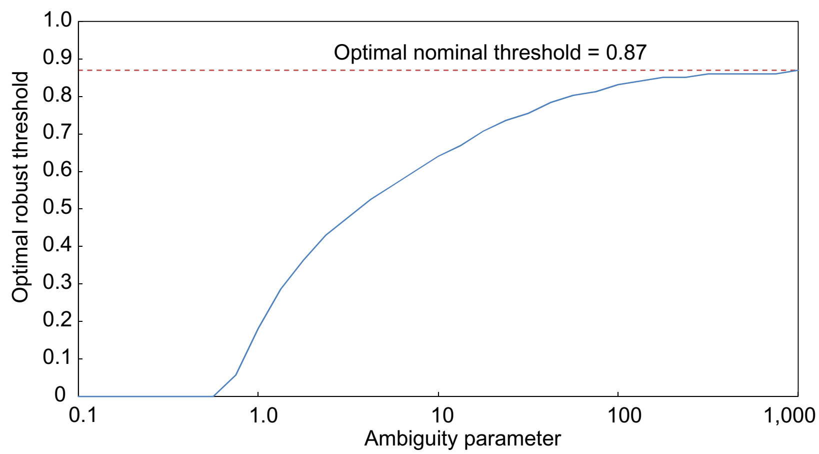
In this paper, we considered a robust modeling approach for maintenance decision making in the presence of model miss-specification. Under this formulation we derived the

Figure 2. (Color online) The control chart plots the posterior probability $\pi_n^{(2)}$ that the system is in the failure state, with starting value $\pi_0^{(2)} = 0.05$.



Note. In this realization of the posterior trajectory, $\pi_n^{(2)}$ exceeds the optimal robust threshold $\bar{\pi} = 0.74$ at the 8th decision epoch.

Figure 3. (Color online) The optimal robust threshold $\bar{\pi}(\theta_T)$ as a function of the ambiguity parameter θ_T.



robust dynamic programming equation, and through its analysis explicitly characterized nature's optimal policy. We then used this characterization of nature's optimal policy to establish structural properties of the decision-maker's optimal policy. In particular, we proved that even in the presence of model miss-specification, the robust value function preserves convexity with respect to the current posterior distribution. We next proved a general robustness result for the preventive maintenance paradigm, which showed that preventive maintenance is always a worthwhile strategy for managing failure risk, even when the decision maker distrusts the underlying model assumptions. We also established the optimality of a robust threshold policy for the practical subclass of Bayesian change point detection problems.

Acknowledgments

The author thanks the two referees, the associate editor, and the department editor for their helpful comments and feedback. A special thanks goes to Dragan Banjević for helpful discussions on Section 4.3. This research was partially supported by the NSERC Discovery [Grant RGPIN-2015-04019]. The opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Natural Sciences and Engineering Research Council.

Appendix A. Proof of Lemma 1

If at an arbitrary stage n, $\pi_n = \pi_\xi$, by Equation (28) of Theorem 2, $H^1(\pi_n, W)$ simplifies to

$$\begin{aligned} H^1(\pi_\xi, W) &= -\theta_P \log \mathbb{E}_{\pi_\xi}\left[\exp\left(-\frac{h_1(X_n)}{\theta_P}\right)\right] \\ &= h_1(N) = -C_{pN}, \end{aligned} \tag{A1}$$

since $\pi_\xi = (0, \ldots, 0, 1)$, i.e., it is concentrated at the failure state $X_n = N$. Similarly,

$$\begin{aligned} H^2(\pi_\xi, W) &= -\theta_P \log \mathbb{E}_{\pi_\xi}\left[\exp\left(-\frac{h_2(X_n)}{\theta_P}\right)\right] \\ &\quad - \theta_T \log \mathbb{E}_{\rho(\pi_\xi, \cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right] \end{aligned}$$

$$= h_2(N) + \alpha W(\pi_\xi)$$
$$= (R_N - C_N) + \alpha W(\pi_\xi) < W(\pi_\xi), \tag{A2}$$

where the second last equality follows since π_ξ is an absorbing state so that $\pi_n = \pi_\xi$ implies $\pi_{n+1} = \pi_\xi$, and the last inequality follows since we have assumed in Section 2 that in the failure state N, the running profit $R_N - C_N < 0$. It therefore follows that

$$\begin{aligned} W(\pi_\xi) &= \max\{H^1(\pi_\xi, W), H^2(\pi_\xi, W)\} \\ &= H^1(\pi_\xi, W) = -C_{pN}. \end{aligned} \tag{A3}$$

In other words, whenever $\pi_n = \pi_\xi$, it is optimal to stop the system and initiate preventive maintenance, i.e., $\pi_\xi \in \Sigma$, which completes the proof. □

Appendix B. Proof of Proposition 3

Under Assumption 1, it can be shown that $H^1(\pi, W)$ and $H^2(\pi, W)$ given in (28) simply to

$$\begin{aligned} H^1(\pi, W) &= \mathbb{E}_\pi[h_1(X_n)], \\ H^2(\pi, W) &= \mathbb{E}_\pi[h_2(X_n)] \\ &\quad -\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right]. \end{aligned} \tag{B1}$$

Furthermore, from the proof of Theorem 3, we have shown that the second summand of $H^2(\pi, W)$,

$$\mathscr{C}(\pi) = -\theta_T \log \mathbb{E}_{\rho(\pi,\cdot)}\left[\exp\left(-\frac{\alpha W(\pi_{n+1})}{\theta_T}\right)\right], \tag{B2}$$

is convex in $\pi \in \mathcal{P}$. If we let $\mathscr{L}^1(\pi) = \mathbb{E}_\pi[h_1(X_n)]$ and $\mathscr{L}^2(\pi) = \mathbb{E}_\pi[h_2(X_n)]$, then the optimal robust stopping region Σ simplifies to

$$\Sigma = \{\pi \in \mathcal{P}\colon (\mathscr{L}^2 - \mathscr{L}^1)(\pi) + \mathscr{C}(\pi) < 0\}, \tag{B3}$$

which is a sublevel set of the function $(\mathscr{L}^2 - \mathscr{L}^1)(\pi) + \mathscr{C}(\pi)$. Since both $\mathscr{L}^1(\pi)$ and $\mathscr{L}^2(\pi)$ are linear functions in π, the difference function $(\mathscr{L}^2 - \mathscr{L}^1)(\pi)$ is also linear, which implies the function $(\mathscr{L}^2 - \mathscr{L}^1)(\pi) + \mathscr{C}(\pi)$ is convex in π. It follows that the optimal robust stopping region Σ is a convex subset of $\mathcal{P}$, which completes the proof. □

Appendix C. Proof of Theorem 4

For the $N = 2$ version of the robust problem, the simplex $\mathcal{P}$ defined in (6) takes the form

$$\mathcal{P} = \{(1-\pi, \pi) \in \mathbb{R}^2\colon 0 \leqslant \pi \leqslant 1\}, \tag{C1}$$

which is the line segment in $\mathbb{R}^2$ connecting points $e_1 = (1, 0)$ and $e_2 = (0, 1)$. By Lemma 1 and Propositions 2 and 3, the optimal robust stopping region Σ defined in (43) is an open convex subset of $\mathcal{P}$ containing the failure point $\pi_\xi = (0, 1) = e_2$, i.e.,

$$\Sigma = \{(1-\pi, \pi) \in \mathbb{R}^2\colon \bar{\pi} < \pi \leqslant 1\}, \tag{C2}$$

for some $0 \leqslant \bar{\pi} < 1$, from which the threshold policy immediately follows. □

Appendix D. Proof of Proposition 4

We start by studying the conditional probability $\pi_n^{(N)}(\tilde{Y}_{(n)})$. If we denote an arbitrary state sequence $\theta_{(n)} = (\theta_0, \ldots, \theta_n) \in \mathcal{S}^{n+1}$, then from Bayes' Theorem (5), we can express $\pi_n^{(N)}(\tilde{Y}_{(n)})$ as

$$\begin{aligned} \pi_n^{(N)}(\tilde{Y}_{(n)}) &= \mathbb{P}(X_n = N \mid \tilde{Y}_1, \ldots, \tilde{Y}_n) \\ &= \Bigg(\sum_{\substack{\theta_{(n)} \in \mathcal{S}^{n+1} \\ \theta_n = N}} \prod_{m=1}^n p_{\theta_{m-1}\theta_m} f_{\theta_m}(\tilde{Y}_m) \pi_0^{(\theta_0)}\Bigg) \\ &\quad \cdot \Bigg(\sum_{\substack{\theta_{(n)} \in \mathcal{S}^{n+1} \\ \theta_n \neq N}} \prod_{m=1}^n p_{\theta_{m-1}\theta_m} f_{\theta_m}(\tilde{Y}_m) \pi_0^{(\theta_0)} \\ &\qquad + \sum_{\substack{\theta_{(n)} \in \mathcal{S}^{n+1} \\ \theta_n = N}} \prod_{m=1}^n p_{\theta_{m-1}\theta_m} f_{\theta_m}(\tilde{Y}_m) \pi_0^{(\theta_0)}\Bigg)^{-1}. \end{aligned} \tag{D1}$$

If we let

$$\begin{aligned} R_n &= \Bigg(\sum_{\substack{\theta_{(n)} \in \mathcal{S}^{n+1} \\ \theta_n \neq N}} \prod_{m=1}^n p_{\theta_{m-1}\theta_m} \left(\frac{f_{\theta_m}(\tilde{Y}_m)}{f_N(\tilde{Y}_m)}\right) \pi_0^{(\theta_0)}\Bigg) \\ &\quad \cdot \Bigg(\sum_{\substack{\theta_{(n)} \in \mathcal{S}^{n+1} \\ \theta_n = N}} \prod_{m=1}^n p_{\theta_{m-1}\theta_m} \left(\frac{f_{\theta_m}(\tilde{Y}_m)}{f_N(\tilde{Y}_m)}\right) \pi_0^{(\theta_0)}\Bigg)^{-1}, \end{aligned} \tag{D2}$$

then

$$\pi_n^{(N)}(\tilde{Y}_{(n)}) = (1 + R_n)^{-1}. \tag{D3}$$

We now study the numerator and denominator of R_n separately. We start with the numerator, which we will show tends to zero $\tilde{\mathbb{P}}$-a.s. To see this, first note that the state process under the nominal model $\mathbb{P}$ enters the failure state N in finite time $\mathbb{P}$-a.s. Since the sum in the numerator is taken over all state sequences $\theta_{(n)}$ for which $\theta_n \neq N$, to show the numerator converges to zero, it suffices to show that the product $\prod_{m=1}^n f_{\theta_m}(\tilde{Y}_m)/f_N(\tilde{Y}_m)$ is finite $\tilde{\mathbb{P}}$-a.s. We will in fact show the product converges to zero. To see this, we now denote for each $i \in \mathcal{S} \backslash \{N\}$,

$$\gamma_i = \mathbb{E}_{\tilde{f}_N}\left[\log\left(\frac{f_N(Y)}{f_i(Y)}\right)\right] > 0, \tag{D4}$$

which is positive by Assumption 2. Since the alternative model $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, the underlying state process $\tilde{X}_n \to N$, $\tilde{\mathbb{P}}$-a.s., so that the Strong Law of Large Numbers implies

$$\frac{1}{n} \sum_{m=1}^n \log\left(\frac{f_i(\tilde{Y}_m)}{f_N(\tilde{Y}_m)}\right) \to -\gamma_i < 0, \quad \tilde{\mathbb{P}}\text{-a.s.} \tag{D5}$$

It follows that the product

$$\begin{aligned} \prod_{m=1}^n \frac{f_{\theta_m}(\tilde{Y}_m)}{f_N(\tilde{Y}_m)} &= \exp\left(\sum_{m=1}^n \log\left(\frac{f_{\theta_m}(\tilde{Y}_m)}{f_N(\tilde{Y}_m)}\right)\right) \\ &= \exp\left(n\left\{\frac{1}{n}\sum_{m=1}^n \log\left(\frac{f_{\theta_m}(\tilde{Y}_m)}{f_N(\tilde{Y}_m)}\right)\right\}\right) \\ &\to 0, \quad \tilde{\mathbb{P}}\text{-a.s.} \end{aligned} \tag{D6}$$

Therefore, the numerator in R_n tends to zero $\tilde{\mathbb{P}}$-a.s. Since the sum in the denominator is taken over all state sequences $\theta_{(n)}$ for which $\theta_n = N$ and the state process under the nominal model $\mathbb{P}$ enters the failure state N in finite time $\mathbb{P}$-a.s., it is clear that the denominator of R_n converges $\tilde{\mathbb{P}}$-a.s. to a constant that *does not* equal to zero. It follows that the ratio $R_n \to 0$, $\tilde{\mathbb{P}}$-a.s., so that $\pi_n^{(N)}(\tilde{Y}_{(n)}) = (1+R_n)^{-1} \to 1$, $\tilde{\mathbb{P}}$-a.s. Therefore, for each $i \in \mathcal{S}$,

$$|\pi_n^{(i)} - \tilde{\pi}_n^{(i)}| \to 0, \quad \tilde{\mathbb{P}}\text{-a.s.} \tag{D7}$$

For any measurable subset of the probability simplex $A \subset \mathcal{P}$, Equation (7) implies that the mapping $\pi \mapsto \rho(\pi, A)$ is continuous. It follows that, for any measurable set $A \subset \mathcal{P}$,

$$\lim_{n\to\infty} |\rho(\pi_n, A) - \tilde{\rho}(\tilde{\pi}_n, A)|$$
$$= |\rho(\pi_\xi, A) - \tilde{\rho}(\pi_\xi, A)| = |\delta_{\pi_\xi}(A) - \delta_{\pi_\xi}(A)| = 0, \quad \tilde{\mathbb{P}}\text{-a.s.}, \tag{D8}$$

which completes the proof. □

Appendix E. Proof of Proposition 5

Recall from Equation (24) that if $\pi_n = \pi$, the worst-case posterior $v_{*,n}$ at stage n is given by

$$v_{*,n}^{(i)}(\pi) = v_{*,n}^{(i)} = \frac{\exp(-h_1(i)/\theta_P)}{\mathbb{E}_\pi[\exp(-h_1(X)/\theta_P)]}\pi^{(i)}, \quad i \in \mathcal{S}. \tag{E1}$$

Since the above mapping $\pi \mapsto v_*(\pi)$ is continuous, it follows from the proof of Proposition 4 that for each $i \in \mathcal{S}$,

$$\lim_{n\to\infty} |v_{*,n}^{(i)}(\pi_n) - \tilde{\pi}_n^{(i)}| = |v_{*,n}^{(i)}(\lim_{n\to\infty}\pi_n) - \lim_{n\to\infty}\tilde{\pi}_n^{(i)}|$$
$$= |v_{*,n}^{(i)}(\pi_\xi) - \pi_\xi| = |\pi_\xi - \pi_\xi| = 0, \quad \tilde{\mathbb{P}}\text{-a.s.}$$

Similarly, since the worst-case transition probability function $\eta_{*,n}(\cdot)$ can also be viewed as a continuous mapping in the current posterior π_n, Equation (56) follows using the same argument above, which completes the proof. □

References

Ben-Tal A, Nemirovski A (1998) Robust convex optmization. *Math. Oper. Res.* 23(4):769–805.
Ben-Tal A, Nemirovski A (1999) Robust solutions to uncertain programs. *Oper. Res. Lett.* 25:1–13.
Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Program.* 88: 411–424.
Bertsekas DP, Shreve SE (1978) *Stochastic Optimal Control: The Discrete Time Case* (Academic Press, NY).
Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1): 35–53.
Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).
Brown DB, Smith JE, Sun P (2010) Information relaxations and duality in stochastic dynamic programs. *Oper. Res.* 58(4):785–801.
Caro F, Gupta AD (2015) Robust control of the multi-armed bandit problem. *Ann. Oper. Res.*, ePub ahead of print August 21, 2015, http://dx.doi.org/10.1007/s10479-015-1965-7.
Çavuş Ö, Ruszczyński A (2014a) Computational methods for risk-averse undiscounted transient Markov models. *Oper. Res.* 62(2):401–417.
Çavuş Ö, Ruszczyński A (2014b) Risk-averse control of undiscounted transient Markov models. *SIAM J. Control Optim.* 52:3935–3966.
Chan CW, Farias VF (2009) Stochastic depletion problems: Effective myopic policies for a class of dynamic optimization problems. *Math. Oper. Res.* 34(2):333–350.
Dai Pra P, Meneghini L, Runggaldier WJ (1996) Connections between stochastic control and dynamic games. *Math. Control Signal* 9: 303–326.
Dayanik S, Gurler U (2002) An adaptive Bayesian replacement policy with minimal repair. *Oper. Res.* 50(3):552–558.
Dayanik S, Goulding C, Poor VH (2008) Bayesian sequential change diagnosis. *Math. Oper. Res.* 33(2):475–496.
Dupuis P, Ellis RS (1997) *A Weak Convergence Approach to the Theory of Large Deviations* (John Wiley & Sons, NY).
El Ghaoui L, Lebret H (1997) Robust solutions to least-square problems to uncertain data matrices. *SIAM J. Matrix Anal. Appl.* 18:1035–1064.
Elwany AH, Gebraeel NZ, Maillart LM (2011) Structured replacement policies for components with complex degradation processes and dedicated sensors. *Oper. Res.* 59(3):684–695.
González-Trejo JI, Hernández-Lerma O, Hoyos-Reyes LF (2003) Minimax control of discrete-time stochastic systems. *SIAM J. Control Optim.* 41:1626–1659.
Gotoh J, Kim MJ, Lim AEB (2016) Robust empirical optimization is almost the same as mean variance optimization. Working paper, Chuo University, Tokyo.
Greenberg HJ, Pierskalla WP (1971) A review of quasiconvex functions. *Oper. Res.* 19(1):1553–1570.
Hansen LP, Sargent TJ (2007) Robust estimation and control without commitment. *J. Econom. Theory* 136:1–27.
Iyengar GN (2005) Robust dynamic programming. *Math. Oper. Res.* 30(2):257–280.
Jain A, Lim AEB, Shanthikumar JG (2010) On the optimality of threshold control in queues with model uncertainty. *Queueing Sy.* 65:157–174.
Kim MJ, Lim AEB (2015) Robust multi-armed bandit problems. *Management Sci.* 62:264–285.
Kim MJ, Makis V (2012) Optimal control of a partially observable failing system with costly multivariate observations. *Stoch. Model.* 28: 584–608.
Kim MJ, Makis V (2013) Joint optimization of sampling and control of partially observable failing systems. *Oper. Res.* 61(3):777–790.
Kurt M, Kharoufeh JP (2010) Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment. *Oper. Res. Lett.* 38:273–279.
Li JY, Kwon RH (2013) Portfolio selection under model uncertainty: A penalized moment-based optimization approach. *J. Global Optim.* 56:131–164.
Lim AEB, Shanthikumar JG (2007) Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res.* 55(2):198–214.
Maillart LM (2006) Maintenance policies for systems with condition monitoring and obvious failures. *IIE Trans.* 38:463–475.
Makis V (2008) Multivariate Bayesian control chart. *Oper. Res.* 56(2):487–496.
Makis V, Jardine AKS (1992) Optimal replacement in the proportional hazards model. *INFOR* 30:172–183.
Mannor S, Mebel O, Xu H (2012) Lightning does not strike twice: Robust MDPs with coupled uncertainty. Langford J, Pineau J, eds. *Proc. 29th Internat. Conf. Machine Learn.* (ACM, New York), 385–392.
Munkres JR (2000) *Topology* (Prentice Hall, Upper Saddle River, NJ).
Nilim A, El Ghaoui L (2005) Robust control of Markov decision process with uncertain transition matrix. *Oper. Res.* 53(5):780–798.
Papadimitriou CH, Tsitsiklis JN (1987) The complexity of Markov decision processes. *Math. Oper. Res.* 12(3):441–450.
Peterson IR, James MR, Dupuis P (2000) Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE T. Automat. Contr.* 45:398–412.
Rogers LCG (2007) Pathwise stochastic optimal control. *SIAM J. Control Optim.* 46:1116–1132.
Ross SM (1971) Quality control under Markovian deterioration. *Management Sci.* 17:587–596.
Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. *Math. Program., Ser. B* 125:235–261.
Ruszczyński A, Shapiro A (2006) Conditional risk mappings. *Math. Oper. Res.* 31(3):544–561.

Stachowiak GW, Batchelor AW, Stachowiak GB (2004) *Experimental Methods in Tribology* (Elsevier, Amsterdam).

Ulukus MY, Kharoufeh JP, Maillart LM (2012) Optimal replacement policies under environment-driven degradation. *Probab. Eng. Inform. Sci.* 26:405–424.

Wiesemann W, Kuhn D, Rustem B (2013) Robust Markov decision processes. *Math. Oper. Res.* 38(1):153–183.

White CC-III (1978) Optimal inspection and repair of a production process subject to deterioration. *J. Oper. Res. Soc.* 29:235–243.

Xu H, Mannor S (2012) Distributionally robust Markov decision processes. *Math. Oper. Res.* 37(2):288–300.

Michael Jong Kim is an assistant professor of industrial engineering at the University of Toronto. His research focuses on the interplay between statistical learning and dynamic optimization, with applications to problems that arise in operations research and business analytics.