Maintenance Optimization for a Markovian Deteriorating System with Population Heterogeneity

Chiel van Oosterom* Hao Peng Geert-Jan van Houtum
School of Industrial Engineering, Eindhoven University of Technology
Eindhoven, PO Box 513, 5600 MB, The Netherlands

Abstract

We develop a partially observable Markov decision process (POMDP) model to incorporate population heterogeneity when scheduling replacements for a deteriorating system. The single-component system deteriorates over a finite set of condition states according to a Markov chain. The population of spare components that is available for replacements is composed of multiple component types that cannot be distinguished by their exterior appearance, but deteriorate according to different transition matrices. This situation arises, for example, if new components and repaired components are mixed into the population without proper records of their repair history. We provide a set of conditions for which we characterize the structure of the optimal policy that minimizes the total expected discounted operating and replacement cost over an infinite horizon. By a numerical experiment, we benchmark the optimal policy against a heuristic policy that neglects population heterogeneity.

Keywords: Replacement optimization, Population heterogeneity, Partially observable Markov decision process, Optimal policy structure

1 Introduction

Capital goods, such as lithography machines in semiconductor fabrication plants, baggage handling systems at airports, or wind turbines in wind farms, are essential for the primary processes of their users. Deterioration of a capital good can lower the efficiency of its operations or cause a quality loss in the products or services it delivers, resulting in increased operating

*Corresponding author, E-mail: c.d.v.oosterom@tue.nl

costs. It can therefore be advantageous to replace deteriorated components of a capital good by non-deteriorated spare components. Therefore, it can be advantageous to replace deteriorated components of a capital good by non-deteriorated spare components. Often, components of capital goods are inspected periodically to obtain observations of their deterioration level, which are then used to make replacement decisions. Such a maintenance strategy falls under the umbrella of condition-based maintenance (CBM).

A common assumption in the literature on CBM models is that after each replacement the newly installed component is subject to the same stochastic deterioration process, i.e., the components form a homogeneous population. It is noticeable that population heterogeneity is taken into consideration in closely related research areas. For example, in degradation modeling, Yuan and Pandey (2009) present a case study on carbon steel pipes in a nuclear power plant. They demonstrate how the accuracy in predictions of the future degradation path improves if random parameters are included in the degradation model to account for population heterogeneity. There exist several causes by which the population of components can be heterogeneous. One cause is a difference in deterioration behavior between newly purchased spare components and spare components that have been returned after a repair. Other causes include variations in the production process and faults during the installation process. If these components cannot be distinguished by their exterior appearance or historical records, the installed components should be modeled as being randomly selected from a heterogeneous population. Despite these facts, little work has been done on CBM models for components that form a heterogeneous population. In this paper, we formulate and analyze a model for scheduling replacements for a single-component, Markovian deteriorating system if the population of spare components consists of multiple component types that cannot be distinguished by their exterior appearance.

In the literature on CBM models, a large number of works is devoted to maintenance optimization models for single-component systems that deteriorate over a finite set of deterioration levels (condition states) according to a Markov chain. One of the earliest works is due to Derman (1963), who studies the problem of scheduling replacements to minimize long-run average cost when corrective replacement of the installed component in the highest deterioration level (i.e., failure) is more costly than preventive replacement in other deterioration levels. He formulates the problem using a Markov decision process (MDP) model and provides sufficient conditions on the cost parameters and the transition matrix for the optimal policy to have a control-limit structure. Numerous variations and extensions to this classical problem are explored in later

research. For example, Kawai et al. (2002) also include operating cost in the model. The model formulation in Bobos and Protonotarios (1978) allows for multiple maintenance activities, e.g., replacement and multiple types of repair. Kurt and Kharoufeh (2010) introduce a limit on the number of repairs. Semi-Markovian deterioration (Kao, 1973) and age-dependent deterioration (Benyamini and Yechiali, 1999) have been considered as generalizations of the deterioration behavior. An excellent review of these variations and extensions is provided in Çekyay and Özekici (2011). The majority of the works focus, just as Derman (1963), on identifying intuitively meaningful conditions for the optimal policy to possess a certain structure—most commonly a control-limit structure. The knowledge that the optimal policy has a certain structure offers managerial insight, and may enable efficient computation. All works assume that each installed component deteriorates according to the same transition matrix, or, equivalently, that the population of components is homogeneous. By contrast, we consider a heterogeneous population of components that consists of multiple component types, each of which is associated with a unique transition matrix.

We formulate the problem of scheduling replacements for a single-component system to minimize the total expected discounted operating and replacement cost under this assumption as a partially observable Markov decision process (POMDP) model and we characterize the structure of the optimal policy.

In contrast, we consider the problem of scheduling replacements for a single-component system to minimize the total expected discounted operating and replacement cost if the population of spare components consists of multiple component types that cannot be distinguished by their exterior appearance, but deteriorate according to different transition matrices. We formulate the problem using a partially observable Markov decision process (POMDP) model and we characterize the structure of the optimal policy.

In research areas that are related to maintenance, a number of efforts have been made to take population heterogeneity into consideration. In reliability analysis, population heterogeneity is modeled by using mixture models for the lifetime distribution of components (Follmann and Goldberg, 1988). For components with such a lifetime distribution, Cha and Finkelstein study optimal burn-in procedures (Cha and Finkelstein, 2010) and investigate the effect of time-based, imperfect, preventive maintenance (Cha and Finkelstein, 2012). In degradation modeling, population heterogeneity is often referred to as unit-to-unit variation. Typically, it is modeled by including one or more random parameters in the degradation model that differ over the

components in a population. Examples are random drift and diffusion parameters in a Wiener degradation process (Peng and Tseng, 2009; Wang, 2010; Bian and Gebraeel, 2012) and random scale parameters in a gamma degradation process (Lawless and Crowder, 2004; Tsai et al., 2012). Most of these works focus on deriving reliability indices such as the failure time distribution or the mean time to failure, assuming that failure is defined in terms of a specified level of degradation. Also, for components with such a degradation process, Xiang et al. study optimal burn-in procedures (Xiang et al., 2011) and optimal joint burn-in and age-based replacement policies (Xiang et al., 2013). Optimal joint burn-in and age-based replacement policies are investigated by Ye et al. (2012) as well.

Population heterogeneity is taken into account in CBM models by Crowder and Lawless (2007) and Elwany et al. (2011), but neither of these works deals with the problem of scheduling replacements to a system that deteriorates according to a Markov chain. Crowder and Lawless (2007) consider a fairly simple maintenance scheme, in which only one inspection can be performed to observe the component's deterioration level before a preventive replacement is scheduled, for components with a gamma process or Wiener process degradation model that includes random parameters. Elwany et al. (2011) consider the problem of scheduling replacements for a system that deteriorates according to a Wiener process degradation model with random parameters. For this specific degradation process, they prove a control-limit structure in the optimal policy with respect to the deterioration level and the age of an installed component. These works do not explicitly examine the importance of accounting for population heterogeneity in the CBM models that they consider.

Elwany et al. (2011) consider the problem of scheduling replacements for components with a geometric Brownian motion degradation process, in which population heterogeneity is modelled by a prior bivariate normal distribution on the intercept and drift parameters. In this model, the posterior distribution on a component's intercept and drift parameters can be fully determined from the latest observed degradation signal and the age of the component.

Following this approach, Elwany et al. (2011) prove a control-limit structure in the optimal policy with respect to the degradation signal and the age of the component.

Our main contributions are summarized as follows. We extend the literature on a classical CBM problem by considering a heterogeneous population of spare components which is composed of multiple component types that cannot be distinguished by their exterior appearance but deteriorate according to different transition matrices. We formulate the sequential decision

process of scheduling replacements as a POMDP model and characterize the structure of the optimal policy. We introduce a new stochastic order to formulate the structural result. Also, we examine in a numerical experiment the importance of accounting for population heterogeneity in the replacement model. We present an adapted version of Hansen's policy iteration algorithm (Hansen, 1998) that is suitable for our model and use it as a solution technique in our numerical study.

This paper is organized as follows. In Section 2, we introduce the POMDP model for the problem of scheduling replacements for a Markovian deteriorating system with population heterogeneity. We derive structural results for the POMDP model in Section 3. Next, in Section 4 a numerical study is presented. We conclude and provide directions for future research are in Section 5.

2 Model Formulation

We consider a single-component system that is operated over an infinite horizon, where time is divided into periods of unit length. The deterioration level of the system evolves as a discrete-time Markov chain on a finite set of deterioration levels, $\mathcal{D} = \{0, 1, \ldots, N\}$. Level 0 corresponds to the best condition and level N indicates that the installed component has failed. We assume that the installed component stems from a heterogeneous population that comprises components of multiple component types, represented by the set $\mathcal{T} = \{1, 2, \ldots, M\}$. The type of the installed component determines the transition probability matrix of the Markov chain that describes the deterioration process on $\mathcal{D}$: each component type $t \in \mathcal{T}$ is associated with a unique transition probability matrix P_t with elements p_{ij}^t, $i, j \in \mathcal{D}$. At any discrete time epoch, the installed component may be either kept operating or replaced by a spare component with deterioration level 0 from the heterogeneous population. The spare components are assumed to be indistinguishable with respect to their component type; therefore, upon a replacement, the type of the newly installed component is random. The probability of it being $t \in \mathcal{T}$ equals the proportion ρ_t that components of component type t constitute in the population.

Two types of costs are incurred, operating costs and replacement costs. The per-period operating cost for a system in deterioration level $i \in \mathcal{D}$ is given by $L_i \geq 0$, independent of the type of the installed component. This cost depends on the deterioration level because the deterioration level may have an impact on the quality loss in the products or services that the system delivers or influence the system's energy consumption. The cost of replacing the

installed component in deterioration level $i \in \mathcal{D}$ is given by $C_i \geq 0$. Because the salvage value for a component may vary in the amount of deterioration, again, this cost depends on the deterioration level. After a replacement, which we assume is instantaneous, the system is immediately put into operation with the newly installed component. This means that the total cost for a period when replacing the installed component in deterioration level $i \in \mathcal{D}$ is $C_i + L_0$. Our aim is to find the replacement policy that minimizes the total expected discounted costs of system operation and replacements over an infinite horizon, where costs are discounted by a factor $\lambda \in [0, 1)$.

We model this replacement problem using a POMDP model. The system state is given by (t, i), where $t \in \mathcal{T}$ is the type of the installed component and $i \in \mathcal{D}$ is the deterioration level; accordingly, $\mathcal{S} = \mathcal{T} \times \mathcal{D}$ is the set of system states. We assume that, at every time epoch, a perfect observation is made of the deterioration level but the type of the installed component cannot be directly observed. Partial information on the type of the installed component can be inferred from the history of deterioration levels that have been observed since the last replacement, because the component type determines the transition probability matrix on $\mathcal{D}$. As such, the system state is said to be partially observable (as opposed to completely observable). A policy can prescribe actions based on the observed deterioration level and the partial information with respect to the type of the installed component. The set of available actions is $\mathcal{A} = \{CO, RE\}$, where CO denotes "continue operating" and RE denotes "replace the installed component".

A general result for POMDPs (see, e.g., Monahan, 1982, and references therein) is that, at every time epoch, the information available for decision making can be summarized by a probability distribution over the set of system states, called an information state, which represents a belief about the system state. Here, we can simplify the definition of information states based on the fact that one state variable, namely, the deterioration level, is completely observable—a setting sometimes referred to as mixed observability (Araya-López et al., 2010; Ong et al., 2010). This allows us to define information states as a combination of a (univariate) probability distribution over the set of component types and a single deterioration level. Thus, we define $\Omega = \Pi \times \mathcal{D}$ as the information state space, where

$$\Pi = \{\pi \in \mathbb{R}^M : \textstyle\sum_{t=1}^{M} \pi_t = 1, \pi_t \geq 0 \text{ for all } t \in \mathcal{T}\}$$

is the set of probability mass functions on $\mathcal{T}$. In information state $(\pi, i) \in \Omega$, π is the belief about the type of the installed component and i is the observed deterioration level.

Because the information state is a sufficient statistic of the history, a POMDP can be

expressed as an MDP on the information state space. Suppose $(\pi, i) \in \Omega$ is the current information state. If action CO is taken, it incurs an immediate cost L_i and induces a probability $\sigma(j; \pi, i) = \sum_{t=1}^{M} \pi_t p_{ij}^t$ that the system is in deterioration level j at the next time epoch, for all $j \in \mathcal{D}$. The set of all deterioration levels that can be reached from (π, i) is $\mathcal{O}(\pi, i) = \{j \in \mathcal{D} : \sigma(j; \pi, i) > 0\}$. Having observed a deterioration level $j \in \mathcal{O}(\pi, i)$, the belief π is updated to $\psi(\pi, i, j)$ using Bayes' rule. The updated probability that the type of the installed component is t is obtained as

$$\psi_t(\pi, i, j) = \frac{\pi_t p_{ij}^t}{\sigma(j; \pi, i)},$$

for all $t \in \mathcal{T}$. This results in a new information state $(\psi(\pi, i, j), j)$. Otherwise, if action RE is taken, it incurs an immediate cost C_i and then continues the process from information state $(\pi^{new}, 0)$, where $\pi_t^{new} = \rho_t$ for all $t \in \mathcal{T}$.

The optimal value function $V : \Omega \to \mathbb{R}$, where $V(\pi, i)$ denotes the minimum total expected discounted cost for initial information state $(\pi, i) \in \Omega$, satisfies the optimality equations

$$V(\pi, i) = \min \Bigg\{ L_i + \lambda \sum_{j \in \mathcal{O}(\pi, i)} \sigma(j; \pi, i) V(\psi(\pi, i, j), j),$$
$$C_i + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new}, 0)} \sigma(j; \pi^{new}, 0) V(\psi(\pi^{new}, 0, j), j) \Bigg\}$$

for all $(\pi, i) \in \Omega$ (Monahan, 1982). The first (second) term in the minimization represents the total expected discounted cost achieved by taking action CO (RE) in (π, i) and following the optimal policy thereafter.

3 Structural Results

In this section, we characterize the structure of the optimal policy under a set of sufficient conditions on the cost parameters and the transition probability matrices. The derivation is based on certain monotonicity properties of the optimal value function. In Section 3.1, we first provide preliminary results that are needed to establish these monotonicity properties. We then present our main structural results in Section 3.2. The proofs are given in the Appendix.

3.1 Preliminary Results

To formulate any monotonicity results, a partial order has to be defined on the information states. Whereas the deterioration levels in $\mathcal{D}$ are naturally ordered, it is not clear which stochastic order should be imposed on the probability distributions in Π. Intuitively, this stochastic

order must reflect how strong a component is believed to be. This requires that the strength of component types can be compared, such that one component type is either stronger or weaker than another component type. Therefore, we will need, as a condition for our structural results, that the component types can be totally ordered by their transition probability matrices. Our first task is now to determine suitable orders on the transition probability matrices and the information states.

To begin, we provide definitions and properties of stochastic orders that will be helpful in our analysis. The first two orders are commonly used for comparing probability distributions (equivalently, the respective random variables), and their properties have been extensively studied in the literature (see, e.g., Shaked and Shanthikumar, 2007). Note that the definitions we provide assume discrete probability distributions on some totally ordered finite support $\mathcal{X}$. For more general definitions, we refer to Shaked and Shanthikumar (2007).

Definition 1. Let g and h be probability mass functions. Then g is smaller than h in the usual stochastic order, denoted by $g \preceq_{st} h$, if and only if $\sum_{x\in\mathcal{X}:x\geq y} g_x \leq \sum_{x\in\mathcal{X}:x\geq y} h_x$ for all $y \in \mathcal{X}$.

Definition 2. Let g and h be probability mass functions. Then g is smaller than h in the likelihood ratio order, denoted by $g \preceq_{lr} h$, if and only if $g_y h_x \leq g_x h_y$ for all $x, y \in \mathcal{X}$ such that $x \leq y$.

The same concepts can be used to define a partial order on transition probability matrices. For this, note that each row of a transition probability matrix is a probability mass function for the next state given a current state; hence, stochastic orders can be employed in a row-wise comparison of such matrices. The following definitions assume transition probability matrices on some totally ordered finite state space $\mathcal{X}$. We use $p^{(x)}$ to denote the row of a transition probability matrix P that corresponds to state $x \in \mathcal{X}$.

Definition 3. Let P and Q be transition probability matrices. Then P is smaller than Q in the usual stochastic order, denoted by $P \preceq_{st} Q$, if and only if $p^{(x)} \preceq_{st} q^{(x)}$ for all $x \in \mathcal{X}$.

Definition 4. Let P and Q be transition probability matrices. Then P is smaller than Q in the likelihood ratio order, denoted by $P \preceq_{lr} Q$, if and only if $p^{(x)} \preceq_{lr} q^{(x)}$ for all $x \in \mathcal{X}$.

The following proposition describes the relationship between these orders, which is that the likelihood ratio order implies the usual stochastic order (for part (i), see Theorem 1.C.1 in Shaked and Shanthikumar, 2007, and part (ii) is a direct consequence of part (i)).

Proposition 1.

(i) Let g and h be two probability mass functions. If $g \preceq_{lr} h$, then $g \preceq_{st} h$.

(ii) Let P and Q be two transition probability matrices. If $P \preceq_{lr} Q$, then $P \preceq_{st} Q$.

In addition, we propose a third stochastic order. As will become clear later, having this stochastic order along with the usual stochastic order and the likelihood ratio order enables us to state our structural results under more general conditions in order to widen their applicability.

Definition 5. Let g and h be probability mass functions. Then g is smaller than h in the likelihood ratio order on the left and the usual stochastic order on the very right, denoted by $g \preceq_{lrst} h$, if and only if $g_y h_x \leq g_x h_y$ for all $x, y \in \mathcal{X}$ such that $x \leq y < u$, and $g_u \leq h_u$, where $u = \max \mathcal{X}$.

Definition 6. Let P and Q be transition probability matrices. Then P is smaller than Q in the likelihood ratio order on the left and the usual stochastic order on the very right, denoted by $P \preceq_{lrst} Q$, if and only if $p^{(x)} \preceq_{lrst} q^{(x)}$ for all $x \in \mathcal{X}$.

To gain intuition for the new stochastic order, we examine how it relates to the previous two stochastic orders, thereby complementing the relationship given in Proposition 1. In the following lemma, it is shown that the $\preceq_{lr}$ order implies the $\preceq_{lrst}$ order and the $\preceq_{lrst}$ order implies the $\preceq_{st}$ order.

Lemma 1.

(i) Let g and h be two probability mass functions. If $g \preceq_{lr} h$, then $g \preceq_{lrst} h$.

(ii) Let g and h be two probability mass functions. If $g \preceq_{lrst} h$, then $g \preceq_{st} h$.

(iii) Let P and Q be two transition probability matrices. If $P \preceq_{lr} Q$, then $P \preceq_{lrst} Q$.

(iv) Let P and Q be two transition probability matrices. If $P \preceq_{lrst} Q$, then $P \preceq_{st} Q$.

Any of the three orders on the transition probability matrices could be applied in our replacement problem to compare the strength of component types. If P_s and P_t, $s, t \in \mathcal{T}$, can be ordered by $P_s \preceq_{lr} P_t$, $P_s \preceq_{lrst} P_t$, or $P_s \preceq_{st} P_t$, component type t may be regarded as weaker than component type s: all three orders imply that, for any current deterioration level, the (random) next deterioration level is larger for component type t than for component type s in a certain stochastic sense. The following example highlights the different comparisons that can be made with these orders for an important class of transition probability matrices.

Example 1. Many engineering systems are subject to both gradual deterioration and random shocks that cause sudden failures (Lam and Yeh, 1994). In a continuous-time setting, researchers have modeled such deterioration processes with a continuous-time Markov chain in which, from any deterioration level, transitions can be made to the next deterioration level or to the failed state (Ohnishi et al., 1986; Lam and Yeh, 1994; Chiang and Yuan, 2001). In discrete time, if the deterioration level is monitored at sufficiently small time intervals and gradual deterioration is a process with stationary increments, such deterioration behavior can be captured by a transition probability matrix of the form

$$P = \begin{pmatrix} 1-\alpha-\beta & \alpha & 0 & \cdots & 0 & \beta \\ 0 & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & 1-\alpha-\beta & \alpha & 0 & \beta \\ \vdots & & \ddots & 1-\alpha-\beta & \alpha & \beta \\ \vdots & & & \ddots & 1-\alpha-\beta & \alpha+\beta \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}, \tag{1}$$

parameterized by (α, β). Parameters α and β represent the probability of experiencing a one-level increase in deterioration and the probability of experiencing a sudden failure, respectively.

Now consider a population with four component types, $\mathcal{T} = \{1, 2, 3, 4\}$, whose associated transition probability matrices on $\mathcal{D}$ are of the form (1). The parameters of P_1, P_2, P_3, and P_4 are given by: $(\alpha_1, \beta_1) = (0.02, 0.02)$, $(\alpha_2, \beta_2) = (0.05, 0.02)$, $(\alpha_3, \beta_3) = (0.12, 0.02)$, and $(\alpha_4, \beta_4) = (0.25, 0.02)$. Compare these component types with another component type associated with a transition probability matrix of the form (1) on $\mathcal{D}$, $\widehat{P}$, with parameters $(\hat{\alpha}, \hat{\beta}) = (0.10, 0.05)$. This illustrates all possible levels of comparability:

- $P_1 \preceq_{lrst} \widehat{P}$ and $P_1 \preceq_{st} \widehat{P}$, but P_1 is incomparable to $\widehat{P}$ under the $\preceq_{lr}$ order;
- $P_2 \preceq_{lr} \widehat{P}$, $P_2 \preceq_{lrst} \widehat{P}$, and $P_2 \preceq_{st} \widehat{P}$;
- $P_3 \preceq_{st} \widehat{P}$, but P_3 is incomparable to $\widehat{P}$ under the other orders;
- P_4 is incomparable to $\widehat{P}$.

Note that the likelihood ratio order only supports the conclusion that component type 2, not component type 1, is stronger than the alternative component type, even though component type 1 seems stronger than component type 2 as it is less susceptible to gradual deterioration. The reason is that, under the likelihood ratio order, a decrease in α needs to be accompanied by a relatively larger decrease in β to result in an improvement in strength. $\Diamond$

Since $\mathcal{T}$ is finite and totally ordered with the natural order, any of the three stochastic orders could be applied to compare beliefs about the type of the installed component. However, the stochastic orders have a meaningful interpretation only if the natural order on $\mathcal{T}$ corresponds to a ranking of component types according to strength: with component type 1 being the strongest and component type M being the weakest, a larger belief then implies that the component is believed to be weaker. This is also why, to obtain monotonicity in the belief about the type of the installed component, we will need to impose a condition like $P_1 \preceq_{lr} P_2 \preceq_{lr} \ldots \preceq_{lr} P_M$, $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \ldots \preceq_{lrst} P_M$, or $P_1 \preceq_{st} P_2 \preceq_{st} \ldots \preceq_{st} P_M$.

Results under the condition based on the $\preceq_{st}$ order are more general than results under the condition based on the $\prec_{lrst}$ order, and results based on the the $\preceq_{lrst}$ order are more general than results under the condition based on the $\prec_{lr}$ order (see Lemma 1)

Example 1 (Continued). Compare the four component types in $\mathcal{T}$ among each other. Given that $\beta_s = \beta_t$ for $s, t \in \mathcal{T}$, it can be shown that $P_s \preceq_{lr} P_t$ if and only if $\alpha_s = \alpha_t$ and $P_s \preceq_{lrst} P_t$ if and only if $\alpha_s \leq \alpha_t$. Hence, no two component types in $\mathcal{T}$ can be compared under the $\preceq_{lr}$ order, while there is a total ordering $P_1 \preceq_{lrst} P_2 \preceq_{lrst} P_3 \preceq_{lrst} P_4$ under the $\preceq_{lrst}$ order.

This reasoning generalizes to all settings where the transition probability matrices are of the form (1) and the component type has no influence on the occurrence of fatal shocks. In these settings, structural results that require $P_1 \preceq_{lr} P_2 \preceq_{lr} \ldots \preceq_{lr} P_M$ do not apply but, upon a suitable labeling of the component types, structural results that require $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \ldots \preceq_{lrst} P_M$ do. $\Diamond$

In Lemma 2, we show that the usual stochastic order is sufficient to conclude that, upon taking action CO, a belief that the type of the installed component is stronger results in a stochastically larger deterioration level at the next time epoch.

Lemma 2. *Let* $P_1 \preceq_{st} P_2 \preceq_{st} \ldots \preceq_{st} P_M$. *If* $\pi, \hat{\pi} \in \Pi$ *such that* $\pi \preceq_{st} \hat{\pi}$, *then* $\sigma(\cdot\,; \pi, i) \preceq_{st} \sigma(\cdot\,; \hat{\pi}, i)$ *for all* $i \in \mathcal{D}$.

Another important property in proving monotonicity properties of the value function is that if the component is believed to be stronger, it still is believed to be stronger at the next time epoch if the same or a larger deterioration level is observed. This result, presented in Lemma 3, requires more restrictive conditions on the transition probability matrices and beliefs about the type of the installed component.

Lemma 3. *Let $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \ldots \preceq_{lrst} P_M$. If $\pi, \hat{\pi} \in \Pi$ such that $\pi \preceq_{lr} \hat{\pi}$, then $\psi(\pi, i, j) \preceq_{lr} \psi(\hat{\pi}, i, l)$ for all $i \in \mathcal{D}$ and $j \in \mathcal{O}(\pi, i)$, $l \in \mathcal{O}(\hat{\pi}, i)$ such that $j \leq l < N$.*

Lemma 3 leaves the case in which the next deterioration level is N out of consideration. In that case, the component fails and one would expect it to be replaced regardless of its component type so that the updated belief about the type of the installed component is of no importance.

SECOND PART OF THIS SECTION

It still needs to be verified that a larger deterioration level results in a higher future deterioration level and does not allow less information to be gathered about the type of the installed component.

We can establish this result for a rich class of transition probability matrices.

Definition 7. Let $t \in \mathcal{T}$. Transition probability matrix P_t is truncated Toeplitz if and only if there exists a sequence $\{p_l^t\}_{l \in \mathcal{D}}$ such that

$$p_{ij}^t = \begin{cases} p_{j-i}^t, & i \leq j < N, \\ \sum_{l=N-i}^{N} p_l^t, & i \leq j = N, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 4. *Let P_t be truncated Toeplitz for all $t \in \mathcal{T}$. If $\pi \in \Pi$ and $i, k \in \mathcal{D}$ such that $i \leq k$, then*

(i) $\sum_{l=0}^{j} \sigma(l; \pi, i) = \sum_{l=0}^{j+(k-i)} \sigma(l; \pi, k)$ *for all* $j \in \mathcal{D}$ *such that* $j < N - (k - i)$;

(ii) $\psi(\pi, i, j) = \psi(\pi, k, j + (k - i))$ *for all* $j \in \mathcal{O}(\pi, i)$ *such that* $j < N - (k - i)$.

Lemma 4 shows that, if all transition probability matrices are truncated Toeplitz, the distribution of increments remains unchanged as the deterioration level increases. Thus, the next deterioration level shifts to the right, and the updated belief about the type of the installed component depends only on the increment, not on the deterioration level.

This brings us to the final result of this section, which is key to the the inductive proof of the monotonicity properties of the optimal value function. We first define a partial order on the information state space

Definition 8. For $(\pi, i), (\hat{\pi}, k) \in \Omega$, we say $(\pi, i) \preceq (\hat{\pi}, k)$ if and only if $\pi \preceq_{lr} \hat{\pi}$ and $i \leq k$. A function $F \colon \Omega \to \mathbb{R}$ is called nondecreasing on $(\Omega, \preceq)$ if and only if $F(\pi, i) \leq F(\hat{\pi}, k)$ for all $(\pi, i), (\hat{\pi}, k) \in \Omega$ such that $(\pi, i) \preceq (\hat{\pi}, k)$.

We show the expectation of a function of the next information state to be monotone in the current information state.

Lemma 5. *Let P_t be truncated Toeplitz for all $t \in \mathcal{T}$ and satisfy $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \ldots \preceq_{lrst} P_M$. Let $F\colon \Omega \to \mathbb{R}$ be nondecreasing on $(\Omega, \preceq)$ with $F(\pi, N)$ being constant in $\pi \in \Pi$. Define the function $G\colon \Omega \to \mathbb{R}$ by*

$$G(\pi, i) = \sum_{j \in \mathcal{O}(\pi, i)} \sigma(j; \pi, i) F(\psi(\pi, i, j), j)$$

for all $(\pi, i) \in \Omega$. Then, G is nondecreasing on $(\Omega, \preceq)$, with $G(\pi, N)$ being constant in $\pi \in \Pi$.

3.2 Main Results

The results we present in this section assume that the following list of conditions is satisfied:

- **(C1)** L_i is nondecreasing in i;
- **(C2)** C_i is nondecreasing in i;
- **(C3)** $L_i - C_i$ is nondecreasing in i;
- **(C4)** $L_N \geq C_N + L_0$;
- **(C5)** $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \ldots \preceq_{lrst} P_M$;
- **(C6)** P_t is truncated Toeplitz for all $t \in \mathcal{T}$.

We briefly discuss these conditions to ascertain that they are reasonable and allow for realistic problem instances. Conditions (C1)–(C3) have also been applied to derive structural results for the analogous problem with population homogeneity (cf. Kawai et al., 2002, Section 8.1). Conditions (C1) and (C2) require a positive relationship between the deterioration level and the operating and replacement costs to capture the adverse effects of deterioration. Condition (C3) implies that an increase in the deterioration level leads to a more significant increase in operating cost than in replacement cost, which is natural for systems that are essential for the operations of their user. The last condition on the cost parameters, (C4), is to ensure that a failed component is replaced. This is done by requiring that, in deterioration level N, replacement is even myopically preferred over continue operating, which would imply a period of downtime. While this requirement may be relaxed, it is generally satisfied in the domain of capital goods where the cost of downtime is very large. Conditions (C5) and (C6) have been discussed in Section 3.1.

The first result states monotonicity properties of the optimal value function.

Theorem 1. *V is nondecreasing on $(\Omega, \preceq)$, with $V(\pi, N)$ being constant in $\pi \in \Pi$.*

Specifically, Theorem 1 shows that the optimal value function is monotone with respect to the partial order defined in Definition 8. Thus, if the installed component is believed to be weaker in the sense of the likelihood ratio order and the deterioration level is higher, then the minimum total expected discounted cost is larger. Theorem 1 also shows that when the installed component has failed, its component type is irrelevant for the minimum total expected discounted cost.

The next result demonstrates a threshold-type structure of the optimal policy with respect to the $\preceq$ order.

Theorem 2. *If RE is the optimal action in information state* $(\pi, i) \in \Omega$ *and if* $(\hat{\pi}, k) \in \Omega$ *such that* $(\pi, i) \preceq (\hat{\pi}, k)$*, then RE is also the optimal action in information state* $(\hat{\pi}, k)$*. Furthermore, RE is the optimal action in information state* (π, N) *for all* $\pi \in \Pi$*.*

The intuition is that failed components should be replaced and it is more advantageous to replace a component if it is believed to be weaker and it is more deteriorated. The likelihood ratio order is appropriate to formalize when a component is believed to be weaker.

4 Numerical Study

The optimal policy for the POMDP model adapts replacement decisions to the probabilistic information that the history of observed deterioration levels provides about the component type. Alternatively, a heuristic policy that ignores this information may be easier to implement and has the advantage of avoiding the computational complexity of solving a POMDP. The purpose of this section is to identify factors that suggest a large decrease in total expected discounted cost can be realized by taking population heterogeneity into account—knowledge which can be used to assess a priori, for a given problem instance, whether it is worthwhile to construct a POMDP model and search for the optimal policy. To this end, we compare in a numerical experiment the optimal policy with a heuristic policy that neglects population heterogeneity under different parameter settings.

The heuristic policy is specified in Section 4.1. The solution technique that we use for the POMDP model is described in Section 4.2, where it is also explained how the heuristic policy is evaluated. In Section 4.3, we present an example to illustrate the comparison between the optimal and the heuristic policy. Section 4.4 gives the parameter settings we use in the numerical experiment, and Section 4.5 reports the outcomes.

4.1 Heuristic Policy

The construction of the heuristic policy does not require formulating the POMDP model for incorporating population heterogeneity in the replacement problem. We approximate the deterioration process by making the simplifying assumption that all components deteriorate according to transition probability matrix $\sum_{t=1}^{M} \rho_t P_t$, which reduces the problem into a standard replacement problem with population homogeneity that we can formulate as an MDP. Using standard policy iteration, we obtain the optimal policy for the MDP model, which prescribes an action to select in each deterioration level. Then, we get the heuristic policy for the POMDP model by applying these actions irrespective of the information on the type of the installed component. Note that the construction of the heuristic policy is computationally inexpensive.

4.2 Solution Technique

We wish to compare the total expected discounted cost of the heuristic policy with the minimum total expected discounted cost as attained by the optimal policy. However, in general, the problem of determining the optimal policy for infinite-horizon POMDPs is undecidable (Madani et al., 2003). Therefore, we instead compute an ϵ-optimal policy, i.e., a policy which yields a total expected discounted cost that is at most $\epsilon > 0$ higher than the minimum total expected discounted cost. By setting ϵ to a small value, we obtain tight bounds on the minimum total expected discounted cost, which can then be used to make the desired comparison.

One of the best known methods to compute ϵ-optimal policies for POMDPs is Hansen's policy iteration algorithm (Hansen, 1998). It is not directly applicable to our problem, however, due to the fact that the information states contain a completely observable state variable. Here, we adapt Hansen's policy iteration algorithm to handle such information states so that it can be applied in our numerical experiment. We start by defining a key concept for the algorithm, which we will see is also useful to evaluate the heuristic policy.

4.2.1 Finite-State Controllers

An FSC κ is defined by a triple $\langle (\Gamma_i)_{i\in\mathcal{D}}, \delta, \zeta \rangle$, where Γ_i is a non-empty, finite set of control states for all $i \in \mathcal{D}$; δ is a mapping from control states $\bigcup_{i\in\mathcal{D}} \Gamma_i$ to action space $\mathcal{A}$; and ζ is a mapping from $\{(\gamma, j) : \gamma \in \Gamma_i, i \in \mathcal{D}, j \in \mathcal{O}^{\delta(\gamma)}(i)\}$ to control states $\bigcup_{i\in\mathcal{D}} \Gamma_i$. At every time epoch, if the system is in deterioration level $i \in \mathcal{D}$, the FSC is in a control state from Γ_i. If the FSC is in control state $\gamma \in \Gamma_i$, $i \in \mathcal{D}$, action $\delta(\gamma)$ is selected. Upon observing deterioration

level $j \in \mathcal{O}^{\delta(\gamma)}(i)$ at the next time epoch, the finite-state controller transitions into control state $\zeta(\gamma, j) \in \Gamma_j$. Here, $\mathcal{O}^{CO}(i) = \bigcup_{\pi \in \Pi} \mathcal{O}(\pi, i)$ and $\mathcal{O}^{RE}(i) = \mathcal{O}^{CO}(0)$.

An FSC can be conveniently represented by a cyclic directed graph. In such a graph, the nodes represent control states and are labeled by the action to which it relates via δ. The arcs indicate the successor of a control state for each deterioration level that can be observed at the next time epoch, which is given by ζ. In Figures 1a and 1e we will provide such graphs for policies in the example of Section 4.3.

FSC κ is evaluated by solving a system of linear equations. For initial system state $(t, i) \in \mathcal{S}$, let $v_t^\kappa(\gamma)$ denote the total expected discounted cost when starting in control state $\gamma \in \Gamma_i$. It can be computed from the system of linear equations given by

$$v_t^\kappa(\gamma) = \begin{cases} L_i + \lambda \sum_{j \in \mathcal{O}^{CO}(i)} p_{ij}^t v_t^\kappa(\zeta(\gamma, j)), & \delta(\gamma) = CO, \\ C_i + L_0 + \lambda \sum_{j \in \mathcal{O}^{RE}(i)} \sum_{s=1}^{M} \rho_s p_{0j}^s v_s^\kappa(\zeta(\gamma, j)), & \delta(\gamma) = RE, \end{cases} \tag{2}$$

for all $\gamma \in \Gamma_i$, $(t, i) \in \mathcal{S}$. Consequently, for initial information state $(\pi, i) \in \Omega$, the total expected discounted cost when starting in control state $\gamma \in \Gamma_i$ is given by $\sum_{t=1}^{M} \pi_t v_t^\kappa(\gamma)$. Therefore, $\arg\min_{\gamma \in \Gamma_i} \sum_{t=1}^{M} \pi_t v_t^\kappa(\gamma)$ is used as the starting control state of the FSC, resulting in a total expected discounted cost of $V^\kappa(\pi, i) = \min_{\gamma \in \Gamma_i} \sum_{t=1}^{M} \pi_t v_t^\kappa(\gamma)$.

The above definition of an FSC deviates from the definition that is used in Hansen (1998) by letting an FSC have control states per deterioration level, being the observable factor in the information state space. It can be noticed that, under this definition, the heuristic policy can be represented by an FSC that has one control state per deterioration level. Hence, in our numerical experiment, we can evaluate the heuristic policy by solving a system of $M(N+1)$ linear equations.

4.2.2 Algorithm

In the presentation of our adapted version of Hansen's policy iteration algorithm, we use $v^\kappa(\gamma)$ to denote the $M \times 1$ vector with elements $v_t^\kappa(\gamma)$, $t \in \mathcal{T}$, for all $\gamma \in \Gamma_i$, $i \in \mathcal{D}$; we use e to denote the $M \times 1$ vector with ones; for all $i, j \in \mathcal{D}$, we use $R^{CO}(i, j)$ to denote the $M \times M$ diagonal matrix with diagonal elements $r_{tt}^{CO} = p_{ij}^t$ for all $t \in \mathcal{T}$; and for all $i, j \in \mathcal{D}$, we use $R^{RE}(i, j)$ to denote the $M \times M$ matrix with elements $r_{st}^{RE} = \rho_t p_{0j}^t$ for all $s, t \in \mathcal{D}$. The outline of the algorithm is as follows:

Step 1 *(Initialization)* Construct the heuristic policy and let κ be the corresponding FSC.

Set $\epsilon > 0$.

Step 2 *(Policy evaluation)* Solve the system of linear equations (2) to obtain $v^\kappa(\gamma)$ for all $\gamma \in \Gamma_i$, $i \in \mathcal{D}$.

Step 3 *(Dynamic programming update)* For all $i \in \mathcal{D}$, generate the set of vectors

$$\begin{aligned}\Upsilon_i = \quad & \{L_i e + \lambda \sum_{j \in \mathcal{O}^{CO}(i)} R^{CO}(i,j) v^\kappa(\gamma_j) : \gamma_j \in \Gamma_j \text{ for all } j \in \mathcal{O}^{CO}(i)\} \\ & \cup \{(C_i + L_0)e + \lambda \sum_{j \in \mathcal{O}^{RE}(i)} R^{RE}(i,j) v^\kappa(\gamma_j) : \gamma_j \in \Gamma_j \text{ for all } j \in \mathcal{O}^{RE}(i)\}.\end{aligned}$$

Each vector $v \in \Upsilon_i$ is associated with a pair $(a, (\gamma_j)_{j \in \mathcal{O}^a(i)})$, where $a \in \mathcal{A}$ is an action and $\gamma_j \in \Gamma_j$ is a control state to be selected upon observing deterioration level j, for all $j \in \mathcal{O}^a(i)$. Prune all vectors $v \in \Upsilon_i$ for which there exists no $\pi \in \Pi$ such that $\sum_{t=1}^M \pi_t v_t < \sum_{t=1}^M \pi_t \hat{v}_t$ for all $\hat{v} \in \Upsilon_i / \{v\}$.

Step 4 *(Policy improvement)* Construct the improved FSC $\hat{\kappa} = \langle (\hat{\Gamma}_i)_{i \in \mathcal{D}}, \hat{\delta}, \hat{\zeta} \rangle$ as follows. For all $v \in \Upsilon_i$, $i \in \mathcal{D}$, with its associated pair $(a, (\gamma_j)_{j \in \mathcal{O}^a(i)})$:

(i) If there exists $\gamma \in \Gamma_i$ such that $\delta(\gamma) = a$ and $\zeta(\gamma, j) = \gamma_j$ for all $j \in \mathcal{O}^{\delta(\gamma)}(i)$, then include a copy of γ in $\hat{\kappa}$.

(ii) Else if there exists $\gamma \in \Gamma_i$ such that $v^\kappa(\gamma) \geq v$, then replace γ by a new control state $\hat{\gamma}$ in $\hat{\kappa}$ with $\hat{\delta}(\hat{\gamma}) = a$ and $\hat{\zeta}(\hat{\gamma}, j) = \gamma_j$ for all $j \in \mathcal{O}^a(i)$. If there are multiple such control states in Γ_i, merge them into one single control state.

(iii) Else, add a new control state $\hat{\gamma}$ to $\hat{\kappa}$ with $\hat{\delta}(\hat{\gamma}) = a$ and $\hat{\zeta}(\hat{\gamma}, j) = \gamma_j$ for all $j \in \mathcal{O}^a(i)$.

For all $\gamma \in \Gamma_i$, $i \in \mathcal{D}$, that were not addressed in (i) or (ii) above: include a copy of γ in $\hat{\kappa}$ if there exists a vector $v \in \Upsilon_k$, $k \in \mathcal{D}$, that is associated with a pair $(a, (\gamma_j)_{j \in \mathcal{O}^a(k)})$ in which $\gamma_i = \gamma$.

Step 5 *(Convergence test)* Compute $m = \min_{(\pi,i) \in \Omega} \min_{v \in \Upsilon_i} \left(\sum_{t=1}^M \pi_t v_t - V^\kappa(\pi, i) \right)$. If $-m\lambda/(1-\lambda) < \epsilon$, exit with FSC $\hat{\kappa}$. Else, set κ to $\hat{\kappa}$ and go to Step 2.

It can be shown that, in any iteration of the algorithm, $V^{\hat{\kappa}}(\pi, i) + m\lambda/(1-\lambda) \leq V(\pi, i) \leq V^{\hat{\kappa}}(\pi, i)$ for all $(\pi, i) \in \Omega$ (note that $m \leq 0$). Therefore, the termination criterion guarantees that the algorithm returns an FSC $\hat{\kappa}$ that is ϵ-optimal, as

$$V^{\hat{\kappa}}(\pi, i) - V(\pi, i) \leq V^{\hat{\kappa}}(\pi, i) - \left(V^{\hat{\kappa}}(\pi, i) + m \frac{\lambda}{1-\lambda} \right) = -m \frac{\lambda}{1-\lambda} < \epsilon$$

for all $(\pi, i) \in \Omega$. We let $\underline{V}(\pi, i) = V^{\hat{\kappa}}(\pi, i) + \frac{\lambda}{1-\lambda} m$ and $\overline{V}(\pi, i) = V^{\hat{\kappa}}(\pi, i)$.

It is also possible to derive an ϵ-optimal policy that applies a stationary action rule that maps Ω to $\mathcal{A}$. This policy selects at every time epoch action $\delta(\arg\min_{\gamma \in \Gamma_i} \sum_{t=1}^M \pi_t v_t^\kappa(\gamma))$ in information

state (π, i), for all $(\pi, i) \in \Omega$. In general, this policy is not equivalent to the policy that is represented by FSC $\hat{\kappa}$. The policies are equivalent, however, if $\hat{\kappa}$ represents the optimal policy, which is certain if $m = 0$ at termination of the algorithm.

One difference between the algorithm that is presented here and the original algorithm in Hansen (1998) is that multiple steps of the algorithm contain loops over the deterioration levels to accomodate the modified definition of an FSC. Another difference is that our algorithm is initialized with a well-founded initial FSC: the heuristic policy is computationally inexpensive to construct and evaluate, and it is likely to be closer to the optimal policy than an arbitrary policy. Note that this choice underlines the increase in the computational complexity that is caused by the incorporation of population heterogeneity in the replacement problem.

Remark 1. Interestingly, the structural result in Section 3 also provides knowledge about the FSC that is returned by the adapted version of Hansen's policy iteration algorithm. In the special case that the algorithm returns the optimal policy, if conditions (C1)–(C6) are satisfied, then the FSC satisfies the following properties.

(i) Let $i, k \in \mathcal{D}$ such that $i \leq k$. If there exists $\gamma \in \Gamma_i$ such that $\delta(\gamma) = RE$, then there also exists $\gamma \in \Gamma_k$ such that $\delta(\gamma) = RE$.

(ii) The set Γ_N is a singleton, and $\delta(\gamma) = RE$ for the only control state $\gamma \in \Gamma_N$.

(iii) Let $\gamma \in \Gamma_i$, $i \in \mathcal{D}$, and $j, l \in \mathcal{O}^{\delta(\gamma)}(i)$ such that $j \leq l$. If $\delta(\zeta(\gamma, j)) = RE$, then also $\delta(\zeta(\gamma, l)) = RE$.

Properties (i) and (ii) follow directly from Theorem 2. To derive property (iii), Theorem 2 needs to be combined with Lemma 3, which gives that, for a given information state, a higher deterioration level at the next time epoch also makes it more likely that the component is of a weaker component type.

4.3 Example

Consider a population with three component types, $\mathcal{T} = \{1, 2, 3\}$, each accounting for an equal fraction of the population, $\rho_1 = \rho_2 = \rho_3 = 1/3$. There are four deterioration levels, $\mathcal{D} = \{0, 1, 2, 3\}$, and the cost parameters are $L_0 = L_1 = L_2 = 0$, $L_3 = 500$ and $C_0 = C_1 = C_2 = 100$,

$C_3 = 200$. The transition probability matrices are given by

$$P_1 = \begin{pmatrix} 0.9 & 0.05 & 0 & 0.05 \\ 0 & 0.9 & 0.05 & 0.05 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; P_2 = \begin{pmatrix} 0.6 & 0.25 & 0.05 & 0.1 \\ 0 & 0.6 & 0.25 & 0.15 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } P_3 = \begin{pmatrix} 0 & 0.5 & 0.1 & 0.4 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

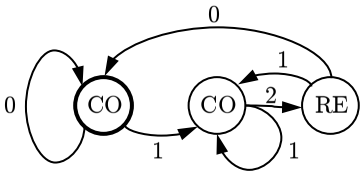
Note that conditions (C1)–(C6) are satisfied. In particular, component type 1 is the strongest and component type 3 is the weakest based on the $\preceq_{lrst}$ order. The discount factor is $\lambda = 0.99$.

In Figure 1a and Figures 1b–1d, we depict the heuristic policy as an FSC and as a stationary action rule that maps Ω to $\mathcal{A}$, respectively. It can be observed that the heuristic policy replaces the installed component if and only if the deterioration level is 2. We denote the total expected discounted cost of the heuristic policy by V_{heu}. For initial information state $(\pi^{new}, 0)$, which corresponds to the situation where the system begins operating with a newly installed component, the total expected discounted cost is evaluated as $V_{heu}(\pi^{new}, 0) = 5846.37$.

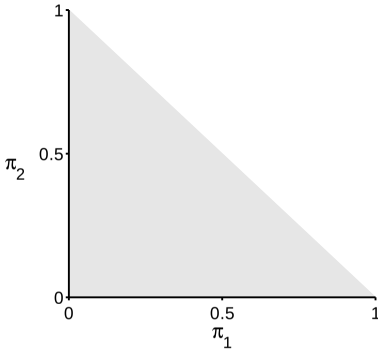
In Figure 1e, we depict the ϵ-optimal FSC that is obtained by the adapted version of Hansen's policy iteration algorithm with $\epsilon = 0.05$. The node with a thicker border corresponds to the control state in which the FSC is started for initial information state $(\pi^{new}, 0)$. Only the control states that are reachable from this starting control state are shown in the figure. The actual FSC is larger and contains 22 control states. It can be observed that the ϵ-optimal policy, in contrast to the heuristic policy, may replace the component if it is in deterioration level 1. More precisely, it replaces the installed component if it reaches deterioration level 1 in three or less periods after it is taken into operation. That would make it likely that the installed component is of component type 2 or 3, and the probability of failure, which would require a corrective replacement at cost C_2, is too high to continue to operate the installed component. For initial information state $(\pi^{new}, 0)$, the algorithm provides bounds $\underline{V}(\pi^{new}, 0) = 4779.86$ and $\overline{V}(\pi^{new}, 0) = 4779.91$ on the minimum total expected discounted cost. In Figures 1f–1h, we depict the ϵ-optimal stationary action rule that can be derived from the ϵ-optimal FSC. As can be observed, also this policy replaces the installed component in deterioration level 1 if it is likely to be of a weak component type.

We evaluate the relative decrease in total expected discounted cost that results from applying the optimal policy instead of the heuristic policy, for initial information state $(\pi^{new}, 0)$, as

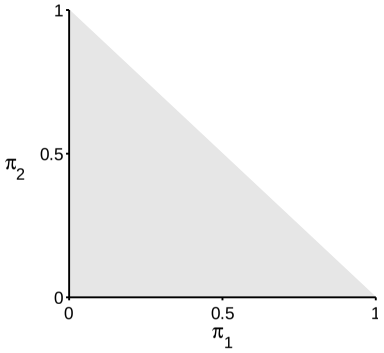
$$S = \frac{V_{heu}(\pi^{new}, 0) - \overline{V}(\pi^{new}, 0)}{\overline{V}(\pi^{new}, 0)} = \frac{5846.37 - 4779.91}{4779.91} = 22.3\%.$$



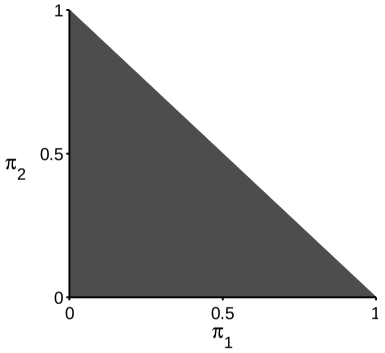
(a) Heuristic policy as an FSC



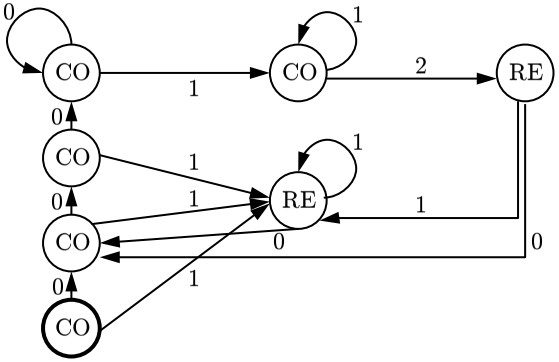
(b) Heuristic policy at deterioration level 0



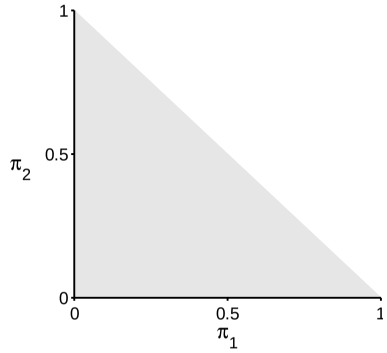
(c) Heuristic policy at deterioration level 1



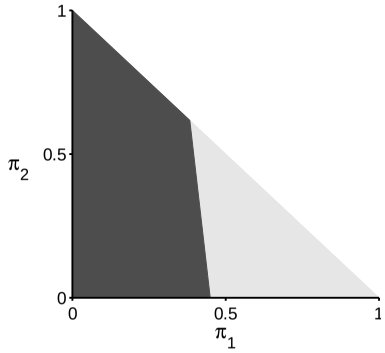
(d) Heuristic policy at deterioration level 2



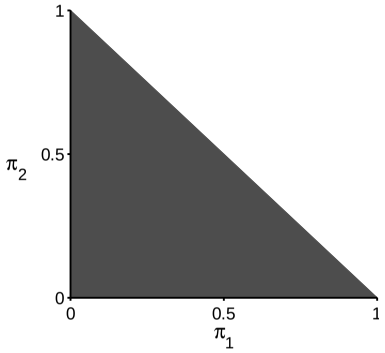
(e) ϵ-optimal FSC



(f) ϵ-optimal stationary action rule at deterioration level 0



(g) ϵ-optimal stationary action rule at deterioration level 1



(h) ϵ-optimal stationary action rule at deterioration level 2

Figure 1: Heuristic and ϵ-optimal policies for the example in Section 4.3

4.4 Parameter Settings

In our numerical experiment, we study settings with $\mathcal{T} = \{1, 2\}$ where component type 1 is stronger than component type 2. To generate a set of problem instances, we vary the population composition, the number of deterioration levels, the transition probability matrices, and the cost parameters.

For the fraction $\rho_1 = 1-\rho_2$, which determines the composition of the population, we consider values of 0.5 and 0.8. We consider values of 3, 5, and 10 for the number of deterioration levels, $N+1$. The transition probability matrices are assumed to be of the form (1). We set (α_1, β_1), the parameters of P_1, at $(0.15, 0.03)$ and consider values of $(0.4, 0.2)$ and $(0.7, 0.1)$ for (α_2, β_2), the parameters of P_2. A cost structure is assumed such that the replacement cost is given by

$$C_i = \begin{cases} C, & i < N, \\ aC, & i = N, \end{cases}$$

with $C > 0$ and $a > 1$, and the operating cost is given by

$$L_i = \begin{cases} \frac{i}{N-1} bC, & i < N, \\ 2aC, & i = N, \end{cases}$$

with $b \geq 0$. Thus, the cost of preventively replacing a component does not depend on the deterioration level and is lower than the cost of a corrective replacement. The operating cost is linearly increasing for deterioration levels in $\{0, \ldots, N-1\}$ and sufficiently high in deterioration level N to ensure that the optimal policy replaces the installed component if it has failed. Here, C is the cost of a preventive replacement, a is the factor by which the corrective replacement cost is a multiple of the preventive replacement cost, and b is the ratio between the operating and replacement cost at deterioration level $N-1$. The parameters a, b, and C completely determine the cost structure, with C merely acting as a scale factor. We set C at 100, consider values of 2, 5, 10, and 20 for a, and consider values of 0, 0.1, and 0.5 for b. We assume a discount factor $\lambda = 0.99$. By taking all combinations of the parameter values that we consider, a test bed of in total 144 instances is created.

4.5 Results

We set $\epsilon = 0.05$ in our adapted version of Hansen's policy iteration algorithm and compute V_{heu}, $\underline{V}$, and $\overline{V}$ for all instances. We evaluate the decrease in total expected discounted cost by

applying the optimal policy instead of the heuristic policy by

$$S = \frac{V_{heu}(\pi^{new}, 0) - \overline{V}(\pi^{new}, 0)}{\overline{V}(\pi^{new}, 0)},$$

where we assume that the initial information state is given by $(\pi^{new}, 0)$. We find that the average value of S over all instances is 3.66%. Because we are interested in parameter settings for which the optimal policy achieves a large decrease in total expected discounted cost, we report in Table 1 the 20 instances with the highest values of S.

Table 1: The 20 highest ranked instances (out of 144) in our numerical experiment based on the value of S.

Rank	ρ_1	$N+1$	(α_2,β_2)	a	b	$\underline{V}(\pi^{new},0)$	$\overline{V}(\pi^{new},0)$	$V_{heu}(\pi^{new},0)$	S (%)
1	0.5	10	(0.7,0.1)	20	0	7626.13	7626.17	9267.00	21.52
2	0.5	10	(0.7,0.1)	20	0.1	7875.65	7875.68	9569.83	21.51
3	0.5	10	(0.4,0.2)	20	0.5	11381.94	11381.98	13784.42	21.11
4	0.5	10	(0.4,0.2)	20	0.1	10487.18	10487.22	12286.48	17.16
5	0.5	10	(0.4,0.2)	20	0	10253.45	10253.49	12011.46	17.15
6	0.5	10	(0.7,0.1)	10	0.1	4350.37	4350.41	5019.47	15.38
7	0.5	10	(0.7,0.1)	10	0	4099.91	4099.96	4716.64	15.04
8	0.5	10	(0.7,0.1)	20	0.5	8792.39	8792.43	10082.53	14.67
9	0.5	5	(0.4,0.2)	20	0.5	13197.45	13197.45	15051.20	14.05
10	0.5	10	(0.4,0.2)	10	0.5	6496.18	6496.22	7404.44	13.98
11	0.5	10	(0.4,0.2)	10	0.1	5578.92	5578.97	6316.15	13.21
12	0.5	10	(0.4,0.2)	10	0	5342.77	5342.81	6041.13	13.07
13	0.5	5	(0.4,0.2)	20	0.1	12418.20	12418.20	13832.65	11.39
14	0.5	5	(0.7,0.1)	20	0.5	9792.90	9792.90	10880.80	11.11
15	0.5	5	(0.4,0.2)	20	0	12221.58	12221.58	13559.75	10.95
16	0.5	3	(0.7,0.1)	2	0	2897.20	2897.21	3181.11	9.80
17	0.5	5	(0.7,0.1)	20	0.1	8892.91	8892.91	9740.06	9.53
18	0.5	5	(0.4,0.2)	10	0.5	7594.63	7594.64	8314.41	9.48
19	0.5	10	(0.7,0.1)	10	0.5	5185.07	5185.10	5668.61	9.32
20	0.5	5	(0.7,0.1)	20	0	8667.05	8667.05	9454.87	9.09

From Table 1 several observations can be made. Most notably, $\rho_1 = 0.5$ among all 20 instances that give the highest values of S. To explain why the heuristic policy performs relatively worse when $\rho_1 = 0.5$, note that the uncertainty in the random type of a newly installed component is higher than when $\rho_1 = 0.8$ (more formally, the entropy is higher). Therefore, the observed deterioration levels contain more information about the type of the component and neglecting that information is more disadvantageous.

From Table 1 several observations can be made. One observation is that the highest values of S occur in settings with $\rho_1 = 0.5$. For settings with $\rho_1 = 0.5$ the uncertainty in the type of a newly installed component is higher than for settings with $\rho_1 = 0.8$ so that the observed deterioration levels contain more information about the type of the component. Formally, the entropy of the random type of a newly installed component is higher for $\rho_1 = 0.5$ than it is for $\rho_1 = 0.8$. This explains why the heuristic policy, which neglects the information on the type of the installed component that is provided by observations on its deterioration level, performs worse for $\rho_1 = 0.5$.

A next observation is that high values of S occur more often in settings with $N + 1 = 10$. That is because a higher number of deterioration levels has the effect that component lifetimes are longer, which the optimal policy can exploit to gather more information about the type of the installed component in order to improve replacement decisions. Also, we observe that higher values of a result in higher values of S. For an explanation, note that the heuristic policy can only avoid failures due to gradual deterioration. This is achieved by replacing the installed component if it is in deterioration level $N - 1$. By replacing the installed component if it is likely to be weak, the optimal policy can also reduce the occurrence of sudden failures. In this way, the optimal policy can be more successful in avoiding failures and reducing the need for corrective maintenance at cost $C_N = aC$, which would cause the performance gap between the optimal policy and the heuristic policy to increase in a.

In view of the previous observation, it is remarkable that a has a low value in the only instance in Table 1 with three deterioration levels. Indeed, in settings with $N+1 = 3$, there is no real difference between the optimal policy and the heuristic policy if the corrective replacement cost is very high. To avoid failures, both policies replace the installed component in deterioration level 1 regardless of its component type. That is because $\alpha_1 + \beta_1 > \rho_1\beta_1 + \rho_2\beta_2$ for all values of (α_2, β_2) that we consider, so that even for the strong component type the probability of failure in deterioration level 1 is higher than the probability of failure for a newly installed component. Furthermore, for initial information state $(\pi^{new}, 0)$, neither of the policies replaces an installed component in deterioration level 0. Clearly, the optimal policy does not replace a newly installed component or a component that is still in deterioration level 0 at a later time epoch, which only increases the likelihood that the component is of the strong component type.

No clear relation is observed between (α_2, β_2) and S. For settings with $(\alpha_2, \beta_2) = (0.4, 0.2)$, in which, in particular, the difference between β_1 and β_2 is large, the optimal policy benefits

more from its ability to replace the installed component if it is likely to be of the weak component type. For settings with $(\alpha_2, \beta_2) = (0.7, 0.1)$, in which, in particular, the difference between α_1 and α_2 is large, it is more probable that a component is correctly identified as being of the weak component type. Neither of these two effects is dominant. The effect of b on S is ambiguous as well. An increase in b leads to an increase in the costs that are common to the heuristic policy and the optimal policy, namely,

the operating costs in the first periods after installing a new component in which the heuristic policy and the optimal policy still coincide. On the other hand, it also increases the operating costs that the optimal policy can save afterwards.

From the numerical experiment, we conclude that (1) a high uncertainty in the type of a newly installed component and (2) a large number of deterioration levels in combination with a high cost for corrective replacement provide indications for a significant decrease in total expected discounted cost if population heterogeneity is taken into account in replacement decisions.

5 Conclusion

In this paper, we developed a POMDP model for the problem of scheduling replacements for a single-component, Markovian deteriorating system if the population of components is heterogeneous. This work extends the literature on a classical CBM problem. All previous works assume that each installed component deteriorates according to the same transition matrix, i.e., that the population of components is homogeneous. In contrast, we considered a population of spare components which is composed of multiple component types that cannot be distinguished by their exterior appearance, but deteriorate according to different transition matrices.

We established a structural result for our POMDP model. Under intuitively meaningful conditions on the cost parameters and the transition matrices, we showed monotonicity properties of the value function and derived the structure of the optimal policy. We introduced a new stochastic order, which we denote as the $\preceq_{lrst}$ order, to formulate the structural result. This stochastic order may have applications in other domains as well. Further, we performed a numerical experiment to benchmark the optimal policy against a heuristic policy that neglects population heterogeneity. Results demonstrated that (1) a high uncertainty in the type of a newly installed component and (2) a large number of deterioration levels in combination with a high cost for corrective replacement are indicators for a significant decrease in total

expected discounted cost if replacement decisions take population heterogeneity into account. This knowledge can be used to assess a priori, for a given problem instance, whether it is worth the complexity of formulating and solving a POMDP model.

In our model, we assume that costless, perfect observations on the deterioration level are available at every time epoch. There are also works that consider costly or imperfect inspections in maintenance optimization problems (Maillart, 2006; Kim and Makis, 2013). In future research, it will be interesting to study the implications of population heterogeneity in these contexts. Then, the partial observability of both the deterioration level and the type of the installed component may require to define information states as bivariate probability distributions. Another direction for future research is to relax the assumption that the type of the installed component is independent of the types of the remaining spare components. This assumption is not valid, for example, if the components originate from one production batch that shares the same unknown component type.

Acknowledgements

We thank the Department Editor and three anonymous referees for their time and effort. Their comments have resulted in significant improvements in this article.

References

M. Araya-López, V. Thomas, O. Buffet, and F. Charpillet. A closer look at MOMDPs. In *Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2:197–204, 2010.

Z. Benyamini and U. Yechiali. Optimality of control limit maintenance policies under nonstationary deterioration. *Probability in the Engineering and Informational Sciences*, 13(1):55–70, 1999.

L. Bian and N. Gebraeel. Computing and updating the first-passage time distribution for randomly evolving degradation signals. *IIE Transactions*, 44(11):974–987, 2012.

A.G. Bobos and E.N. Protonotarios. Optimal systems for equipment maintenance and replacement under Markovian deterioration. *European Journal of Operational Research*, 2(4):257–264, 1978.

B. Çekyay and S. Özekici. Condition-based maintenance under Markovian deterioration. In J. Cochran, A. Cox, P. Keskinocak, J.P. Kharoufeh, and J.C. Smith, editors, *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., 2011.

J.H. Cha and M. Finkelstein. Stochastically ordered subpopulations and optimal burn-in procedure. *IEEE Transactions on Reliability*, 59(4):635–643, 2010.

J.H. Cha and M. Finkelstein. Stochastic analysis of preventive maintenance in heterogeneous populations. *Operations Research Letters*, 40(5):416–421, 2012.

J.H. Chiang and J. Yuan. Optimal maintenance policy for a Markovian system under periodic inspection. *Reliability Engineering and System Safety*, 71(2):165–172, 2001.

M. Crowder and J. Lawless. On a scheme for predictive maintenance. *European Journal of Operational Research*, 176(3):1713–1722, 2007.

C. Derman. On optimal replacement rules when changes of state are Markovian. In R.E. Bellman, editor, *Mathematical optimization techniques*, 201–210. Cambridge University Press, 1963.

A.H. Elwany, N.Z. Gebraeel, and L.M. Maillart. Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations Research*, 59(3):684–695, 2011.

D.A. Follmann and M.S. Goldberg. Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, 30(4):389–396, 1988.

E.A. Hansen. An improved policy iteration algorithm for partially observable MDPs. In *Advances in Neural Information Processing Systems*, 1015–1021, 1998.

E.P.C. Kao. Optimal replacement rules when changes of state are semi-Markovian. *Operations Research*, 21(6):1231–1249, 1973.

H. Kawai, J. Koyanagi, and M. Ohnishi. Optimal maintenance problems for Markovian deteriorating systems. In S. Osaki, editor, *Stochastic Models in Reliability and Maintenance*, 193–218. Springer-Verlag, 2002.

M.J. Kim and V. Makis. Joint optimization of sampling and control of partially observable failing systems. *Operations Research*, 61(3):777–790, 2013.

M. Kurt and J.P. Kharoufeh. Optimally maintaining a Markovian deteriorating system with limited imperfect repairs. *European Journal of Operational Research*, 205(2):368–380, 2010.

C.T. Lam and R.H. Yeh. Optimal maintenance-policies for deteriorating systems under various maintenance strategies. *IEEE Transactions on Reliability*, 43(3):423–430, 1994.

J. Lawless and M. Crowder. Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, 10(3):213–227, 2004.

O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147(1–2):5–34, 2003.

L.M. Maillart. Maintenance policies for systems with condition monitoring and obvious failures. *IIE Transactions*, 38(6):463–475, 2006.

G.E. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

M. Ohnishi, H. Kawai, and H. Mine. An optimal inspection and replacement policy for a deteriorating system. *Journal of Applied Probability*, 23(4):973–988, 1986.

S.C.W. Ong, S.W. Png, D. Hsu, and W.S. Lee. Planning under uncertainty for robotic tasks with mixed observablity. *The International Journal of Robotics Research*, 29(8):1053–1068, 2010.

C.Y. Peng and S.T. Tseng. Mis-specification analysis of linear degradation models. *IEEE Transactions on Reliability*, 58(3):444–455, 2009.

M. Shaked and J.G. Shanthikumar. *Stochastic orders*. Springer Science+Business Media, 2007.

C.-C. Tsai, S.-T. Tseng, and N. Balakrishnan. Optimal design for degradation tests based on gamma processes with random effects. *IEEE Transactions on Reliability*, 61(2):604–613, 2012.

X. Wang. Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis*, 101(2):340–351, 2010.

Y. Xiang, D.W. Coit, and Q. Feng. Optimal burn-in for n-subpopulations with stochastic degradation. In *Proceedings of the International Conference on Quality, Reliability, Risk, Maintenance and Safety Engineering (ICQR2SME)*, 148–152, 2011.

Y. Xiang, D.W. Coit, and Q. Feng. n Subpopulations experiencing stochastic degradation: reliability modeling, burn-in, and preventive replacement optimization. *IIE Transactions*, 45(4):391–408, 2013.

Z.S. Ye, Y. Shen, and M. Xie. Degradation-based burn-in with preventive maintenance. *European Journal of Operational Research*, 221(2):360–367, 2012.

X.-X. Yuan and M.D. Pandey. A nonlinear mixed-effects model for degradation data obtained from in-service inspections. *Reliability Engineering and System Safety*, 94(2):509–519, 2009.

Appendix

Proof of Lemma 1. We prove parts (i) and (ii). Parts (iii) and (iv) directly follow by applying respectively parts (i) and (ii) to each pair of rows in the row-wise comparison of P and Q.

(i) Suppose that $g \preceq_{lr} h$. By the definition of the $\preceq_{lr}$ order, $g_y h_x \leq g_x h_y$ for all $x, y \in \mathcal{X}$ such that $x \leq y < u$. Also, since Proposition 1(i) establishes that $g \preceq_{st} h$, it is true that $g_u = \sum_{x \in \mathcal{X}: x \geq u} g_x \leq \sum_{x \in \mathcal{X}: x \geq u} h_x = h_u$.

(ii) Suppose that $g \preceq_{lrst} h$. From the definition of the $\preceq_{lrst}$ order, we find that

$$\begin{aligned}(1-g_u)\sum_{x\in\mathcal{X}:x<y} h_x &= \left(\sum_{z\in\mathcal{X}:z<y} g_x\right)\left(\sum_{x\in\mathcal{X}:x<y} h_x\right) + \left(\sum_{z\in\mathcal{X}:y\le z<u} g_x\right)\left(\sum_{x\in\mathcal{X}:x<y} h_x\right) \\ &\le \left(\sum_{x\in\mathcal{X}:x<y} g_x\right)\left(\sum_{z\in\mathcal{X}:z<y} h_x\right) + \left(\sum_{z\in\mathcal{X}:y\le z<u} h_x\right)\left(\sum_{x\in\mathcal{X}:x<y} g_x\right) \\ &= (1-h_u)\sum_{x\in\mathcal{X}:x<y} g_x\end{aligned}$$

for all $y \in \mathcal{X}$, and also $g_u \le h_u$. If $g_u < 1$, then this implies that $\sum_{x\in\mathcal{X}:x\ge y} g_x \le \sum_{x\in\mathcal{X}:x\ge y} h_x$ for all $y \in \mathcal{X}$. Clearly, the same holds if $g_u = 1$, because then $\sum_{x\in\mathcal{X}:x\ge y} g_x = \sum_{x\in\mathcal{X}:x\ge y} h_x = 1$ for all $y \in \mathcal{X}$. Hence, we conclude that $g \preceq_{st} h$. □

The proof of Lemma 2 relies on the following characterization of the usual stochastic order (cf. Shaked and Shanthikumar, 2007, Section 1.A.1).

Proposition A1. *Let g and h be two probability mass functions. Then $g \preceq_{st} h$ if and only if $\sum_{x\in\mathcal{X}} g_x F(x) \le \sum_{x\in\mathcal{X}} h_x F(x)$ for every nondecreasing function $F\colon \mathcal{X} \to \mathbb{R}$.*

Proof of Lemma 2. Let $\pi, \hat{\pi} \in \Pi$ such that $\pi \preceq_{st} \hat{\pi}$ and $i \in \mathcal{D}$. For all $l \in \mathcal{D}$, because $P_1 \preceq_{st} P_2 \preceq_{st} \ldots \preceq_{st} P_M$ ascertains $\sum_{j=l}^{N} p_{ij}^t$ is nondecreasing in t, Proposition A1 can be applied to show $\sum_{j=l}^{N} \sigma(j;\pi,i) = \sum_{t=1}^{M} \pi_t \sum_{j=l}^{N} p_{ij}^t \le \sum_{t=1}^{M} \hat{\pi}_t \sum_{j=l}^{N} p_{ij}^t = \sum_{j=l}^{N} \sigma(j;\hat{\pi},i)$. □

Proof of Lemma 3. Let $\pi, \hat{\pi} \in \Pi$ such that $\pi \preceq_{lr} \hat{\pi}$, $i \in \mathcal{D}$, and $j \in \mathcal{O}(\pi,i)$, $l \in \mathcal{O}(\hat{\pi},i)$ such that $j \le l < N$ (if such deterioration levels exist). Then, for all $s, t \in \mathcal{T}$ such that $s \le t$,

$$\psi_t(\pi,i,j)\psi_s(\hat{\pi},i,l) = \frac{\pi_t p_{ij}^t}{\sigma(j;\pi,i)} \frac{\hat{\pi}_s p_{il}^s}{\sigma(l;\hat{\pi},i)} \le \frac{\pi_s p_{ij}^s}{\sigma(j;\pi,i)} \frac{\hat{\pi}_t p_{il}^t}{\sigma(l;\hat{\pi},i)} = \psi_s(\pi,i,j)\psi_t(\hat{\pi},i,l),$$

where the inequality follows from the definitions of the $\preceq_{lr}$ order and the $\preceq_{lrst}$ order. □

Proof of Lemma 4. The result directly follows from the definition of the truncated Toeplitz property. □

Proof of Lemma 5. We use a coupling argument to prove that G is nondecreasing on $(\Omega, \preceq)$. Let

$$\xi[u;\pi,i] = \min\{l \in \mathcal{D} : \sum_{j=0}^{l} \sigma(j;\pi,i) \ge u\}$$

for all $(\pi,i) \in \Omega$ and $u \in [0,1]$; note that $\xi[u;\pi,i] \in \mathcal{O}(\pi,i)$. Now fix the information states $(\pi,i), (\hat{\pi},k) \in \Omega$ such that $(\pi,i) \preceq (\hat{\pi},k)$. Letting U be a uniform $[0,1]$ random variable, define

two random variables

$$X \equiv F(\psi(\pi, i, \xi[U; \pi, i]), \xi[U; \pi, i]),$$
$$Y \equiv F(\psi(\hat{\pi}, k, \xi[U; \hat{\pi}, k]), \xi[U; \hat{\pi}, k]).$$

It is easy to see that $E[X] = G(\pi, i)$ and $E[Y] = G(\hat{\pi}, k)$. We will show that, in addition, these random variables are such that $P(X \leq Y) = 1$ by establishing

$$F(\psi(\pi, i, \xi[u; \pi, i]), \xi[u; \pi, i]) \leq F(\psi(\hat{\pi}, k, \xi[u; \hat{\pi}, k]), \xi[u; \hat{\pi}, k]) \tag{A1}$$

for all $u \in [0, 1]$. We distinguish two cases.

Case (i), $u > \sum_{j=0}^{N-1} \sigma(j; \hat{\pi}, k)$. Then $\xi[u; \hat{\pi}, k] = N$, and (A1) is immediate from the properties of F as $F(\psi(\pi, i, \xi[u; \pi, i]), \xi[u; \pi, i]) \leq F(\psi(\pi, i, \xi[u; \pi, i]), N) = F(\psi(\hat{\pi}, k, N), N)$.

Case (ii), $u \leq \sum_{j=0}^{N-1} \sigma(j; \hat{\pi}, k)$. Then $\xi[u; \hat{\pi}, k] < N$. Because Lemma 2, which we can apply by Proposition 1(i) and Lemma 1(iv), gives that $\sum_{j=l}^{N} \sigma(j; \hat{\pi}, k) \geq \sum_{j=l}^{N} \sigma(j; \pi, k)$ for all $l \in \mathcal{D}$ or, equivalently, $\sum_{j=0}^{l} \sigma(j; \hat{\pi}, k) \leq \sum_{j=0}^{l} \sigma(j; \pi, k)$ for all $l \in \mathcal{D}$, we get $\xi[u; \pi, k] \leq \xi[u; \hat{\pi}, k] < N$. Consequently, we can use Lemma 3 to obtain $\psi(\pi, k, \xi[u; \pi, k]) \preceq_{lr} \psi(\hat{\pi}, k, \xi[u; \hat{\pi}, k])$. Furthermore, given that $\xi[u; \pi, k] < N$, it follows from Lemma 4(i) that $\xi[u; \pi, i] + (k - i) = \xi[u; \pi, k]$; therefore, by Lemma 4(ii), $\psi(\pi, i, \xi[u; \pi, i]) = \psi(\pi, k, \xi[u; \pi, k])$. Putting everything together, we have $(\psi(\pi, i, \xi[u; \pi, i]), \xi[u; \pi, i]) \preceq (\psi(\hat{\pi}, k, \xi[u; \hat{\pi}, k]), \xi[u; \hat{\pi}, k])$. Hence, (A1) follows from the monotonicity of F.

This shows $P(X \leq Y) = 1$, which implies that $E[X] \leq E[Y]$. We conclude that $G(\pi, i) \leq G(\hat{\pi}, k)$, which completes the proof that G is nondecreasing on $(\Omega, \preceq)$.

It remains to prove that $G(\pi, N) = G(\hat{\pi}, N)$ for all $\pi, \hat{\pi} \in \Pi$. With P_t being truncated Toeplitz for all $t \in \mathcal{T}$, we have for all $\pi \in \Pi$ that $\sigma(j; \pi, N) = 1$ if $j = N$, $\sigma(j; \pi, N) = 0$ if $j \neq N$, and $\psi(\pi, N, N) = \pi$. This means $G(\pi, N) = F(\pi, N)$ for all $\pi \in \Pi$. The result follows by the properties of F. □

Proof of Theorem 1. Let $V_n(\pi, i)$ be the minimum total expected discounted cost-to-go with n periods remaining for initial information state $(\pi, i) \in \Omega$, where we define $V_0(\pi, i) = 0$. We prove by induction on the number of remaining periods that V_n is nondecreasing on $(\Omega, \preceq)$, with $V_n(\pi, N)$ being constant in $\pi \in \Pi$, for all $n \in \mathbb{N}_0$. It is clear that V_0 satisfies these properties. Assume that, for $m \in \mathbb{N}_0$, V_m is nondecreasing on $(\Omega, \preceq)$, with $V_m(\pi, N)$ being constant in

$\pi \in \Pi$. Using a dynamic programming recursion, V_{m+1} can be expressed by

$$V_{m+1}(\pi, i) = \min \Bigg\{ L_i + \lambda \sum_{j \in \mathcal{O}(\pi,i)} \sigma(j; \pi, i) V_m(\psi(\pi, i, j), j), \\ C_i + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new},0)} \sigma(j; \pi^{new}, 0) V_m(\psi(\pi^{new}, 0, j), j) \Bigg\}$$

for all $(\pi, i) \in \Omega$. In the minimum operator, the first term can be seen to be nondecreasing on $(\Omega, \preceq)$ by combining condition (C1), the induction hypothesis, and Lemma 5. Note that Lemma 5 relies on conditions (C5) and (C6). By condition (C2), the second term is also nondecreasing on $(\Omega, \preceq)$. As a minimum of two such terms, V_{m+1} is nondecreasing on $(\Omega, \preceq)$ as well. Further, it follows by condition (C4), the induction hypothesis, and Lemma 5 that

$$\begin{aligned} & L_N + \lambda \sum_{j \in \mathcal{O}(\pi,N)} \sigma(j; \pi, N) V_m(\psi(\pi, N, j), j) \\ = \; & L_N + \lambda \sum_{j \in \mathcal{O}(\pi^{new},N)} \sigma(j; \pi^{new}, N) V_m(\psi(\pi^{new}, N, j), j) \\ \geq \; & C_N + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new},0)} \sigma(j; \pi^{new}, 0) V_m(\psi(\pi^{new}, 0, j), j) \end{aligned}$$

for all $\pi \in \Pi$. Hence, $V_{m+1}(\pi, N) = C_N + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new},0)} \sigma(j; \pi^{new}, 0) V_m(\psi(\pi^{new}, 0, j), j)$, which is constant in $\pi \in \Pi$. By induction, we conclude that V_n satisfies these properties for all $n \in \mathbb{N}_0$. Finally, the result is obtained from $V(\pi, i) = \lim_{n \to \infty} V_n(\pi, i)$ for all $(\pi, i) \in \Omega$. □

Proof of Theorem 2. Let $(\pi, i) \in \Omega$ be an information state in which *RE* is the optimal action and let $(\hat{\pi}, k) \in \Omega$ such that $(\pi, i) \preceq (\hat{\pi}, k)$. Condition (C3), Theorem 1, and Lemma 5 imply that

$$\begin{aligned} & L_k + \lambda \sum_{j \in \mathcal{O}(\hat{\pi},k)} \sigma(j; \hat{\pi}, k) V(\psi(\hat{\pi}, k, j), j) \\ \geq \; & C_k + L_i - C_i + \lambda \sum_{j \in \mathcal{O}(\pi,i)} \sigma(j; \pi, i) V(\psi(\pi, i, j), j) \\ \geq \; & C_k + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new},0)} \sigma(j; \pi^{new}, 0) V(\psi(\pi^{new}, 0, j), j). \end{aligned}$$

It can be concluded that *RE* is the optimal action in information state $(\hat{\pi}, k)$ as well.

Next, let $\pi \in \Pi$ and consider information state (π, N). By condition (C4), Theorem 1, and

Lemma 5 it can be seen that

$$\begin{aligned}
&L_N + \lambda \sum_{j \in \mathcal{O}(\pi, N)} \sigma(j; \pi, N) V(\psi(\pi, N, j), j) \\
&= L_N + \lambda \sum_{j \in \mathcal{O}(\pi^{new}, N)} \sigma(j; \pi^{new}, N) V(\psi(\pi^{new}, N, j), j) \\
&\geq C_N + L_0 + \lambda \sum_{j \in \mathcal{O}(\pi^{new}, 0)} \sigma(j; \pi^{new}, 0) V(\psi(\pi^{new}, 0, j), j).
\end{aligned}$$

Hence, RE is the optimal action in information state (π, N) for all $\pi \in \Pi$. □