IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie21

Maintenance optimization for a Markovian deteriorating system with population heterogeneity

Chiel van Oosterom, Hao Peng & Geert-Jan van Houtum

To cite this article: Chiel van Oosterom, Hao Peng & Geert-Jan van Houtum (2017) Maintenance optimization for a Markovian deteriorating system with population heterogeneity, IISE Transactions, 49:1, 96-109, DOI: 10.1080/0740817X.2016.1205239

To link to this article: http://dx.doi.org/10.1080/0740817X.2016.1205239

Accepted author version posted online: 28 Jun 2016.
Published online: 28 Jun 2016.

IISE TRANSACTIONS
2017, VOL. 49, NO. 1, 96–109
http://dx.doi.org/10.1080/0740817X.2016.1205239

Taylor & Francis
Taylor & Francis Group

Maintenance optimization for a Markovian deteriorating system with population heterogeneity

Chiel van Oosterom[a,b], Hao Peng[c] and Geert-Jan van Houtum[b]

[a]Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands; [b]School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; [c]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

We develop a partially observable Markov decision process 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 probability matrices. This situation may arise, for example, because of variations in the production process of components. 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. In a numerical experiment, we benchmark the optimal policy against a heuristic policy that neglects population heterogeneity.

ARTICLE HISTORY
Received 14 August 2013
Accepted 14 June 2016

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, and medical equipment in hospitals, are essential for the primary processes of their users. Deterioration of a capital good can lower the efficiency of its operations or diminish the quality of the goods or services it delivers, resulting in increased operating costs. Therefore, it can be advantageous to replace deteriorated components of a capital good with non-deteriorated spare components. Often, a Condition-Based Maintenance (CBM) policy is adopted, and information on the deterioration level of components is collected (e.g., via remote monitoring or manual inspections) to support replacement decisions.

A common assumption in 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. However, there exist several causes by which the population of components can be heterogeneous. For example, variability in the production process of components might affect their deterioration behavior. Then if the components are not distinguishable by appearance, the installed components will be randomly selected from a heterogeneous population (cf. Cha and Finkelstein, 2012; Xiang *et al.*, 2014). Faults in the installation of components could be another cause of heterogeneity. In degradation modeling, therefore, population heterogeneity (also often referred to as unit-to-unit variability) is increasingly taken into account. Typically, this is done by including random parameters in the degradation model that differ over the components in a population; e.g., random drift and diffusion parameters in a Brownian motion (Peng and Tseng, 2009; Wang, 2010; Bian and Gebraeel, 2012) or random scale parameters in a gamma process (Lawless and Crowder, 2004; Tsai *et al.*, 2012). Heterogeneity in populations of produced components has also motivated the development of burn-in testing procedures, which aim to reduce the number of weak components released for field operation; see Ye *et al.* (2012) and Xiang *et al.* (2013) for recent work on degradation-based burn-in. Despite these facts, little work has been done on CBM models for components that form a heterogeneous population. In this article, we formulate and analyze a model for scheduling replacements for a single-component, Markovian deteriorating system under population heterogeneity.

In the literature on CBM models, many works are devoted to maintenance optimization for single-component systems that deteriorate according to a Markov chain over a finite set of deterioration levels (condition states). One of the earliest works is due to Derman (1963), who studies the problem of scheduling replacements to minimize the long-run average cost when a corrective replacement of the installed component after it has failed (i.e., reached the highest deterioration level) is more costly than a preventive replacement. Derman formulates the problem using a Markov Decision Process (MDP) model and provides sufficient conditions on the transition probability matrix for the optimal policy to have a control-limit structure. Numerous variations on and extensions to this classic problem have been explored in later research. For example, Kawai *et al.* (2002) include operating cost in the model. Bobos and Protonotarios (1978) propose a model formulation that allows for multiple

CONTACT Chiel van Oosterom vanoosterom@ese.eur.nl

maintenance activities—e.g., replacement and multiple types of repair—and 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. A comprehensive review of these variations and extensions is given in Çekyay and Özekici (2011). In this stream of literature, the population of components is homogeneous, in that each installed component deteriorates according to the same transition probability matrix. The majority of the works focus, like 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 particular structure offers managerial insight and may ease implementation.

The key difference between the CBM model we present here and the aforementioned models is that we consider population heterogeneity by assuming that the population of spare components consists of multiple component types that cannot be distinguished by their exterior appearance but deteriorate according to different transition probability matrices. The type of an installed component is not known; the only information available for making replacement decisions is the history of observed deterioration levels. We use a Partially Observable Markov Decision Process (POMDP) model to formulate the problem of scheduling replacements to minimize the total expected discounted operating and replacement cost over an infinite horizon, and we also derive a result on the structure of the optimal policy. We note that our model can also be applied to scenarios where heterogeneity among the installed components arises at the time of installation; in that case, the different component types should be reinterpreted as different categories of installation faults.

Efforts to incorporate population heterogeneity in CBM models so far have concentrated on models based on continuous-state deterioration processes. Crowder and Lawless (2007) and Zhang *et al.* (2014) study a fairly simple maintenance scheme to cope with population heterogeneity, in which only one inspection can be performed to observe a component's deterioration level before scheduling a preventive replacement. They apply their proposed policy to gamma process and Brownian motion degradation models with random parameters. Xiang *et al.* (2014) develop a joint burn-in and CBM policy for heterogeneous populations; however, in their policy, maintenance decisions for components that are placed into service after surviving burn-in are not adapted to information on component-specific parameters that may be inferred from observations of the deterioration level. More closely related to our work is research by Elwany *et al.* (2011) and Chen *et al.* (2015), who study the problem of adaptively scheduling replacements for components that deteriorate according to a geometric Brownian motion and an inverse Gaussian process with random parameters, respectively. In both works, a conjugate prior distribution is assumed for the random parameters such that the posterior distribution, after having observed a sequence of deterioration levels of a component, is completely determined by (the logarithm of) the latest deterioration level and the component's age. This assumption permits a reduction of the state space of the decision process but also implies a restriction on the composition of the population. By contrast, we make no assumptions on the distribution of component types in the population, and we allow the entire history of observed deterioration levels to contain relevant information about the type of an installed component.

Our main contributions are summarized as follows. We extend the literature on a classical CBM problem by considering population heterogeneity when scheduling replacements for a single-component, Markovian deteriorating system. We formulate the resulting sequential decision process as a POMDP and obtain a result on the optimal policy structure. In developing our structural results, we introduce a new stochastic order and establish its relationship to existing stochastic orders. We complement the analytical results by conducting a numerical experiment to identify factors that make it especially important to account for population heterogeneity in replacement decisions. To perform this experiment, we adapt Hansen's policy iteration algorithm (Hansen, 1998) such that it can be used as a solution technique for our POMDP model, in which one state variable is completely observable.

The remainder of this article is organized into the following sections. In Section 2, we formulate the POMDP model for our problem of scheduling replacements under population heterogeneity. We derive structural results in Section 3 and present the numerical experiment, including the solution technique used, in Section 4. Finally, in Section 5, we conclude and provide directions for future research.

2. Model formulation

We consider a single-component system that operates 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 the finite set $\mathcal{D} = \{0, 1, \ldots, N\}$. Deterioration level 0 corresponds to the best condition, and deterioration level N indicates that the installed component has failed. We assume that the installed component stems from a heterogeneous population that includes components of multiple 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 either be kept operating or be replaced with a spare component in deterioration level 0 from the heterogeneous population. The spare components are assumed to be indistinguishable with respect to their type; therefore, upon replacement, the type of the newly installed component is random. The probability of it being $t \in \mathcal{T}$ equals the proportion ρ_t of components of type t in the population.

Two kinds 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 since the deterioration level may have an impact on the quality loss in the goods or services that the system produces 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 of a component may vary based on the amount of deterioration, this cost again 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 in a period when the installed component is replaced 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 replacement 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 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 the information state space as $\Omega = \Pi \times \mathcal{D}$, where

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

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.

As 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 causes the process to be continued 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

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

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. Therefore, as a condition for our structural results, we will need that the component types can be totally ordered by their transition probability matrices. Our task is 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 $\sum_{x \in \mathcal{X}: x \geq y} g_x \leq \sum_{x \in \mathcal{X}: x \geq y} h_x$ 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 $g_y h_x \leq g_x h_y$ 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 $p^{(x)} \preceq_{st} q^{(x)}$ 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 $p^{(x)} \preceq_{lr} q^{(x)}$ 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 $g_y h_x \leq g_x h_y$ 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 $p^{(x)} \preceq_{lrst} q^{(x)}$ 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 as $P_s \preceq_{lr} P_t$, $P_s \preceq_{lrst} P_t$, or $P_s \preceq_{st} P_t$, component type s may be regarded as stronger than component type t: all three orders imply that, for any current deterioration level, the (random) next deterioration level is smaller for component type s than for component type t 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 as a continuous-time Markov chain in which, from any deterioration level, transitions can be made to the next-higher 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. □

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 smaller belief then implies that the component is believed to be stronger. This is why, to obtain monotonicity in the belief about the type of the installed component, we will need to impose a condition such as $P_1 \preceq_{lr} P_2 \preceq_{lr} \cdots \preceq_{lr} P_M$, $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$, or $P_1 \preceq_{st} P_2 \preceq_{st} \cdots \preceq_{st} P_M$. From Lemma 1, we know that the condition based on the $\preceq_{st}$ order is the most general and the condition based on the $\preceq_{lrst}$ is more general than the one based on the $\preceq_{lr}$ order. Revisiting Example 1, we see that $P_1 \preceq_{lr} P_2 \preceq_{lr} \cdots \preceq_{lr} P_M$ actually excludes important settings compared with $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$.

Example 1 (continued). Compare the four component types in $\mathcal{T}$ against 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.

In general, when the transition probability matrices are of the form (1) and the component type has no influence on the occurrence of fatal shocks, the above reasoning implies that structural results that require $P_1 \preceq_{lr} P_2 \preceq_{lr} \cdots \preceq_{lr} P_M$ do not apply but, possibly after relabeling the component types, structural results that require $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$ do. □

As the following lemma shows, the usual stochastic order is sufficient to ensure that, for any current deterioration level, the (random) next deterioration level after taking action *CO* is stochastically smaller if the installed component is believed to be stronger.

Lemma 2. *Let $P_1 \preceq_{st} P_2 \preceq_{st} \cdots \preceq_{st} P_M$. 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 $i \in \mathcal{D}$.*

Under more restrictive conditions on the transition probability matrices and the beliefs about the type of the installed component, we also have a result on how, under action *CO*, the current belief and the realization of the next deterioration level relate to the updated belief. It then holds, for any current deterioration level, that if the installed component is believed to be stronger and the next deterioration level is lower, the installed component is believed to be stronger after Bayesian updating. This is established in the following lemma.

Lemma 3. *Let $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$. 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 does not consider the case where the next deterioration level is N. In that case, the component fails and one would expect it to be replaced regardless of its component type, so the updated belief about the type of the installed component is of no importance.

The previous results saw a monotone relationship between the belief about the type of the installed component and the next information state. At this point, we also need to consider the relationship between the deterioration level and the next information state. We first define the following property.

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

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

We refer to the transition probability matrices that satisfy the condition in Definition 7 as truncated Toeplitz, as they may be thought of as a finite section of a larger Toeplitz (i.e., constant-diagonal) matrix where the last column is augmented such that all row sums equal 1 (Böttcher and Silbermann, 1999). The class of truncated Toeplitz matrices is rich. It encompasses the transition probability matrices of the form (1) as a special case, in which $p_0^t = 1 - \alpha_t - \beta_t$, $p_1^t = \alpha_t$, $p_N^t = \beta_t$, and $p_l^t = 0$ for all $l \in \mathcal{D}$ such that $2 \leq l \leq N-1$, but it also allows for multi-level increases in deterioration. Note that upper triangularity is a necessary condition for a transition probability matrix to be truncated Toeplitz.

Under the condition that P_t is truncated Toeplitz for all $t \in \mathcal{T}$, when action *CO* is taken, the probability of a given increment in deterioration does not depend on the current deterioration level, as long as the next deterioration level is smaller than N. Two consequences are, for any belief about the type of the installed component, that the distribution of the next deterioration level shifts to the right (in a linear fashion) as the current deterioration level increases and that, if the component does not fail, the updated belief is completely determined by the deterioration increment. This result is captured in the following lemma.

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 $j < N - (k-i)$.*

We are now ready to state the final result of this section. We define a partial order on the information state space and establish conditions under which the expectation of a function of the (random) next information state after taking action *CO* is monotone in the current information state. This result is key to the inductive proof of the monotonicity properties of the optimal value function.

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

Lemma 5. *Let P_t be truncated Toeplitz for all $t \in \mathcal{T}$ and satisfy $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$. Let $F : \Omega \to \mathbb{R}$ be nondecreasing on $(\Omega, \preceq)$ with $F(\pi, N)$ being constant in $\pi \in \Pi$. Define the function $G : \Omega \to \mathbb{R}$ 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) $P_1 \preceq_{lrst} P_2 \preceq_{lrst} \cdots \preceq_{lrst} P_M$;
(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) to (C3) have also been used to derive structural results for the analogous problem with population homogeneity (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 continued operation, which would imply a period of downtime. Although 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 demonstrates that the optimal value function is monotone with respect to the partial order we defined in Definition 8: the minimum total expected discounted cost is higher when the system is in a higher deterioration level and the installed component is believed to be weaker. In addition, this result asserts that if the installed component has failed, its component type is irrelevant to the minimum total expected discounted cost.

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

The next result concerns the optimal policy structure. With Theorem 1 in hand, we can show that the optimal policy has a threshold-type structure.

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 behind Theorem 2 is that it becomes more advantageous to replace a component when it is more deteriorated and believed to be weaker. For a component that has failed, replacement is always optimal.

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 that can be used for a given problem instance to assess *a priori* 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. We illustrate the comparison between the optimal policy and the heuristic policy by means of an example in Section 4.3. We then perform the numerical experiment. Section 4.4 gives the parameter settings used, and Section 4.5 reports the outcomes.

4.1. Heuristic policy

The construction of the heuristic policy does not require formulating the POMDP model to incorporate 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 to a standard replacement problem with population homogeneity that we can formulate as an MDP. Using standard policy iteration, we compute the optimal policy for the MDP model, which prescribes for each deterioration level an action to take. We then obtain the heuristic policy for the POMDP model by applying these actions irrespective of the information on the type of the installed component. Note that constructing 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 obtained using 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 that 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 for computing 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. The key concept, which we will see is also useful for evaluating the heuristic policy, is that of a finite-state controller.

4.2.1. Finite-State Controllers

Despite the fact that information states encapsulate all information necessary for optimal decision making, it is not practical to perform policy iteration for POMDPs by searching over policies that prescribe actions based on the information state. The problem is that no general method is known for exactly evaluating policies defined as a mapping from the information state space to the action space. Hansen's policy iteration algorithm circumvents this difficulty by searching in an alternative policy space, consisting of policies that are expressed as a Finite-State Controller (FSC). An FSC uses a finite number of control states, on which all histories of actions and observations are mapped, as a basis for decision making (as opposed to the uncountably infinite number of information states). This makes policy evaluation simply a matter of solving a system of linear equations. Importantly, although the optimal policy might not have a representation as an FSC, the existence of an ϵ-optimal FSC is guaranteed (see Theorem 2 in Hansen (1998)).

Given that the deterioration level is completely observable, we formally define an FSC κ for our POMDP model as 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 $i \in \mathcal{D}$; $\delta : \bigcup_{i\in\mathcal{D}} \Gamma_i \to \mathcal{A}$ is a mapping prescribing that action $\delta(\gamma) \in \mathcal{A}$ is taken in control state $\gamma \in \Gamma_i$, $i \in \mathcal{D}$; and $\zeta : \{(\gamma, j) : \gamma \in \Gamma_i, i \in \mathcal{D}, j \in \mathcal{O}^{\delta(\gamma)}(i)\} \to \bigcup_{i\in\mathcal{D}} \Gamma_i$ is a mapping specifying that, given a current control state $\gamma \in \Gamma_i$, $i \in \mathcal{D}$, after observing $j \in \mathcal{O}^{\delta(\gamma)}(i)$ as the next deterioration level, the successor control state is $\zeta(\gamma, j) \in \Gamma_j$. Here, we use the notation $\mathcal{O}^{CO}(i) = \bigcup_{\pi\in\Pi} \mathcal{O}(\pi, i)$ and $\mathcal{O}^{RE}(i) = \mathcal{O}^{CO}(0)$, for all $i \in \mathcal{D}$. The transition mechanism between the control states warrants that whenever the system is in deterioration level $i \in \mathcal{D}$, the FSC is in a control state from Γ_i.

Graphically, FSCs can be represented using a state diagram. In this representation, nodes represent control states and are labeled by the actions prescribed by δ. Arcs represent transitions between control states and are labeled by the next deterioration level for which these transitions are made, as specified by ζ. We use this method to depict FSCs for the example in Section 4.3 (see Figs. 1 and 3).

To evaluate an FSC κ, we compute the total expected discounted cost $v_t^\kappa(\gamma)$ of starting in control state γ if the (partially observable) initial system state is (t, i), for all $\gamma \in \Gamma_i$, $(t, i) \in \mathcal{S}$. This is done by solving the following system of linear equations:

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

for all $\gamma \in \Gamma_i$, $(t, i) \in \mathcal{S}$. The total expected discounted cost of starting in control state $\gamma \in \Gamma_i$ given an initial information state $(\pi, i) \in \Omega$ is then obtained as $\sum_{t=1}^{M} \pi_t v_t^\kappa(\gamma)$. Using the rule to start an FSC in the control state that optimizes it, we have that the total expected discounted cost of FSC κ for initial information state $(\pi, i) \in \Omega$ is $V^\kappa(\pi, i) = \min_{\gamma\in\Gamma_i} \sum_{t=1}^{M} \pi_t v_t^\kappa(\gamma)$.

One example of a policy that can be expressed as an FSC is the heuristic policy. It has a representation as an FSC with one control state per deterioration level; therefore, we can obtain the total expected discounted cost of the heuristic policy for initial information state (π, i), denoted by $V_{heu}(\pi, i)$, for all $(\pi, i) \in \Omega$ by solving a system of $M(N + 1)$ linear equations.

4.2.2. Algorithm

We first introduce some notation that is useful for the presentation of our adapted version of Hansen's policy iteration algorithm. For any FSC κ, let $v^\kappa(\gamma)$ denote the $M \times 1$ vector with elements $v_t^\kappa(\gamma)$, $t \in \mathcal{T}$, for all $\gamma \in \Gamma_i$, $i \in \mathcal{D}$. Let $R^{CO}(i, j)$ denote the $M \times M$ diagonal matrix with diagonal elements $r_{tt}^{CO}(i, j) = p_{ij}^t$, $t \in \mathcal{T}$, for all $i, j \in \mathcal{D}$. Let $R^{RE}(i, j)$ denote the $M \times M$ matrix with elements $r_{ts}^{RE}(i, j) = \rho_s p_{0j}^s$, $s, t \in \mathcal{T}$, for all $i, j \in \mathcal{D}$. Let e denote the $M \times 1$ vector with ones.

The outline of the algorithm is as follows:

Step 1 *(Initialization).* Set $\epsilon > 0$. Construct the heuristic policy; let κ be the corresponding FSC.

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

$$\Upsilon_i = \left\{ L_i e + \lambda \sum_{j\in\mathcal{O}^{CO}(i)} R^{CO}(i, j) v^\kappa(\gamma_j) : \gamma_j \in \Gamma_j \text{ for all } j \in \mathcal{O}^{CO}(i) \right\}$$

$$\cup \left\{ (C_i + L_0)e + \lambda \sum_{j\in\mathcal{O}^{RE}(i)} R^{RE}(i, j) v^\kappa(\gamma_j) : \gamma_j \in \Gamma_j \text{ for all } j \in \mathcal{O}^{RE}(i) \right\},$$

where each vector $v \in \Upsilon_i$ is associated with a pair $(a, (\gamma_j)_{j\in\mathcal{O}^a(i)})$: $a \in \mathcal{A}$ is an action to be taken 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 $\hat{v} \in \Upsilon_i \setminus \{v\}$.

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 their associated pairs $(a, (\gamma_j)_{j\in\mathcal{O}^a(i)})$:

(i) If there exists a $\gamma \in \Gamma_i$ 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 a $\gamma \in \Gamma_i$ such that $v_t^\kappa(\gamma) \geq v_t$ for all $t \in \mathcal{T}$, 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 $j \in \mathcal{O}^a(i)$.

In addition, include a copy in $\hat{\kappa}$ of all $\gamma \in \Gamma_i$, $i \in \mathcal{D}$, not addressed in (i) or (ii) above for which there exists

a vector $\upsilon \in \Upsilon_k$, $k \in \mathcal{D}$, that is associated with a pair $(a, (\gamma_j)_{j\in\mathcal{O}^a(k)})$ such that $\gamma_i = \gamma$.

Step 5 *(Convergence test).* Set

$$m = \max_{(\pi,i)\in\Omega} \max_{\upsilon\in\Upsilon_i} \left(V^{\kappa}(\pi, i) - \sum_{t=1}^{M} \pi_t \upsilon_t \right).$$

If $m < \frac{1-\lambda}{\lambda}\epsilon$, exit with FSC $\hat{\kappa}$. Else, set κ to $\hat{\kappa}$ and go to Step 2.

It can be shown that the algorithm terminates with an ϵ-optimal FSC $\hat{\kappa}$; it holds that $\underline{V}(\pi, i) = V^{\hat{\kappa}}(\pi, i) - \frac{\lambda}{1-\lambda}m$ and $\overline{V}(\pi, i) = V^{\hat{\kappa}}(\pi, i)$ are lower and upper bounds on $V(\pi, i)$ such that $\overline{V}(\pi, i) - \underline{V}(\pi, i) < \epsilon$, for all $(\pi, i) \in \Omega$.

From $\hat{\kappa}$, it is also possible to extract an ϵ-optimal policy defined as a mapping from the information state space to the action space by taking action $\delta(\arg\min_{\gamma\in\Gamma_i} \sum_{t=1}^{M} \pi_t \upsilon_t^{\hat{\kappa}}(\gamma))$ in information state (π, i), for all $(\pi, i) \in \Omega$. This policy is equivalent to FSC $\hat{\kappa}$ if $\hat{\kappa}$ represents the optimal policy, which is certain if $m = 0$, but in general the two policies are different.

The algorithm presented here deviates from the original algorithm in Hansen (1998) in that multiple steps contain loops over the deterioration levels to accommodate the modified definition of FSCs. Another difference is that we initialize the algorithm with a well-founded initial FSC: the heuristic policy is computationally inexpensive to construct and evaluate and is likely to be closer to the optimal policy than an arbitrary policy.

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) to (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 $\lambda = 0.99$.

We first construct and evaluate the heuristic policy. Figures 1 and 2 depict the heuristic policy as an FSC and as a mapping from the information state space to the action space. It can be seen that the installed component is replaced if and only if the deterioration level is level 3. For initial information state $(\pi^{new}, 0)$, which corresponds to the scenario that the system begins operating with a newly installed component, the total expected discounted cost of the heuristic policy is $V_{heu}(\pi^{new}, 0) = 2496.40$.

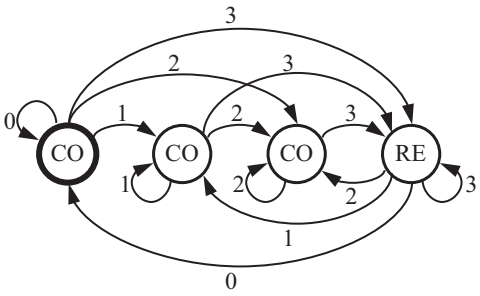


Figure 1. FSC representation of the heuristic policy for the example in Section 4.3. The node with the thickest border corresponds to the control state in which the FSC is started for initial information state $(\pi^{new}, 0)$.

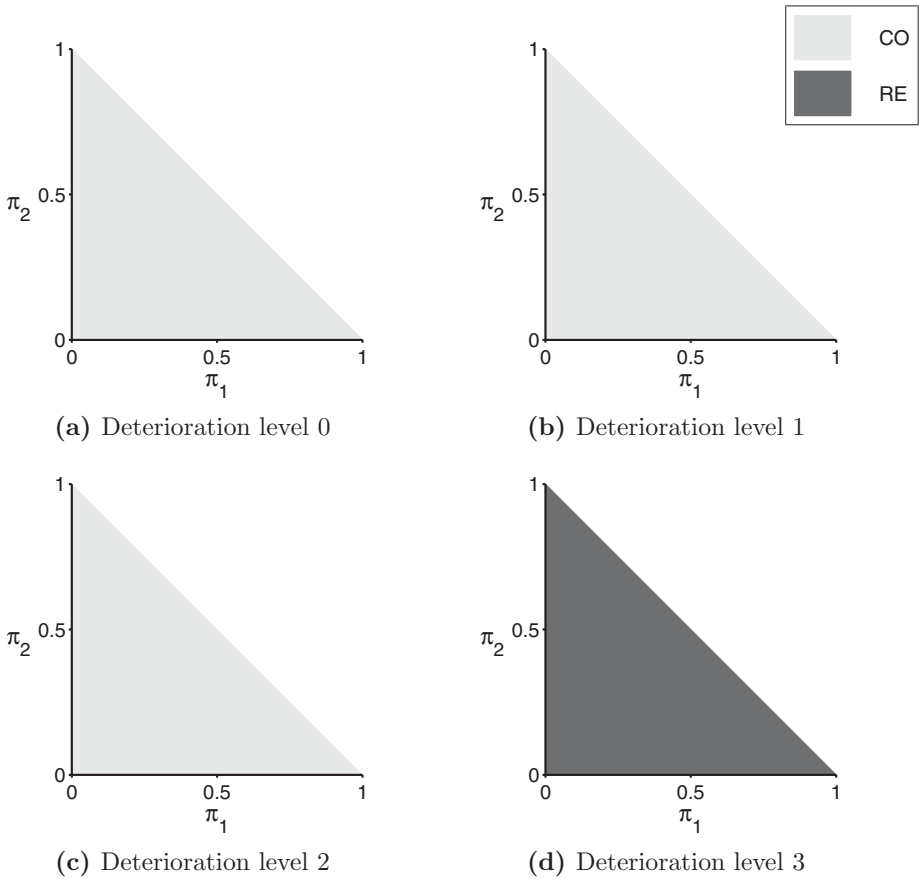
We then compute an ϵ-optimal FSC using the adapted version of Hansen's policy iteration algorithm, taking $\epsilon = 0.05$ and starting from Step 2, and derive an ϵ-optimal mapping from the information state space to the action space. Figure 3 depicts the ϵ-optimal FSC. Apart from replacing the installed component in deterioration level 3, the FSC also replaces the installed component if it is in deterioration level 2 within six periods after it has been taken into operation or a transition is made directly from deterioration level 0 to deterioration level 2. Although not apparent from Fig. 3 (because the relevant control states cannot be reached given initial information state $(\pi^{new}, 0)$), the installed component may even be replaced in deterioration level 0 or 1. The ϵ-optimal mapping from the information state space to the action space, depicted in Fig. 4, exhibits similar characteristics: in deterioration levels 0, 1, and 2, the installed component is replaced if it is likely to be of a weak component type. For initial information state $(\pi^{new}, 0)$, the bounds on the minimum total expected discounted cost are $\underline{V}(\pi^{new}, 0) = 2327.43$ and $\overline{V}(\pi^{new}, 0) = 2327.46$, indicating that a 7.3% cost savings can be achieved by taking population heterogeneity into account.

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

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



(a) Deterioration level 0 (b) Deterioration level 1

(c) Deterioration level 2 (d) Deterioration level 3

Figure 2. Representation of the heuristic policy for the example in Section 4.3 as a mapping from the information state space to the action space, broken down by deterioration level.

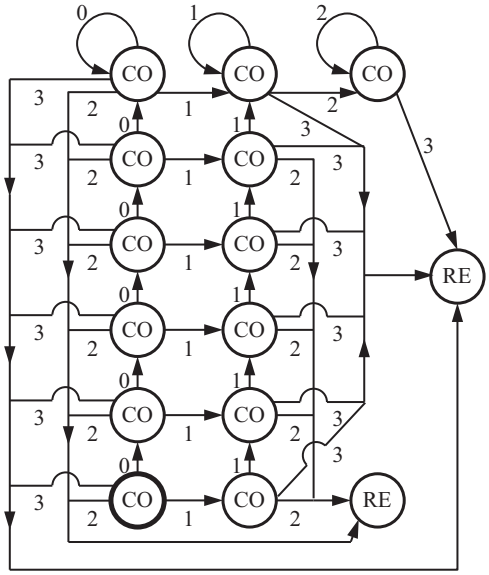


Figure 3. An ϵ-optimal FSC for the example in Section 4.3. The node with the thickest border corresponds to the control state in which the FSC is started for initial information state $(\pi^{new}, 0)$. Only control states that are reachable from this starting control state are depicted (the actual FSC is larger and consists of 154 control states). To reduce clutter, the transitions from the control states in which action *RE* is taken are omitted; they are the same as those for the starting control state.

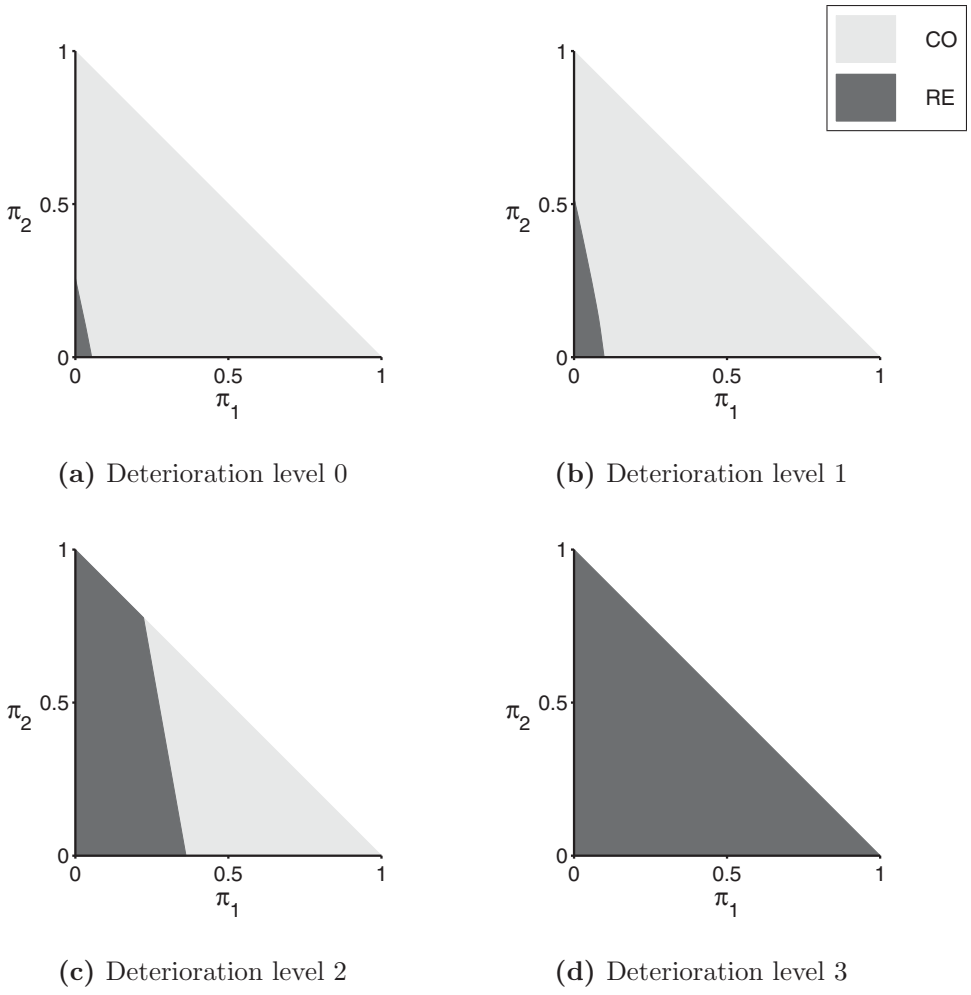
with $C > 0$ and $a > 1$, and the operating cost is given by

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

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, we create a test bed of in total 144 instances.

4.5. Results

We evaluate the decrease in total expected discounted cost by applying the optimal policy instead of the heuristic policy as a percentage savings. For all instances, we compute $V_{heu}(\pi^{new}, 0)$



(a) Deterioration level 0

(b) Deterioration level 1

(c) Deterioration level 2

(d) Deterioration level 3

Figure 4. An ϵ-optimal policy mapping the information state space to the action space for the example in Section 4.3.

and, with $\epsilon = 0.05$ in the adapted version of Hansen's policy iteration algorithm, $\underline{V}(\pi^{new}, 0)$ and $\overline{V}(\pi^{new}, 0)$. We then calculate the percentage savings by

$$S = \frac{V_{heu}(\pi^{new}, 0) - \overline{V}(\pi^{new}, 0)}{\overline{V}(\pi^{new}, 0)} \times 100\%.$$

The average value of S over all instances is found to be 3.66%. Because we are most 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.

Several observations can be made. Most notably, $\rho_1 = 0.5$ in all 20 instances with the highest values of S. To explain why the savings are larger 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 incorporating that information is more advantageous.

Another observation is that higher values of a generally result in higher values of S. For an explanation, note that the optimal policy can be more successful in avoiding failures than the heuristic policy, which becomes more important when the cost $C_N = aC$ of corrective replacement is higher. The heuristic policy can only avoid failures due to gradual deterioration by replacing the installed component in deterioration level $N - 1$; the optimal policy can also reduce the occurrence of sudden failures by replacing the installed component if it is likely to be weak. Also, we observe that high values of S occur more often in settings with $N + 1 = 10$. This 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. It should be mentioned, however, that there also exist settings with $N + 1 = 3$ for which S is high, as the instance ranked 16th in Table 1 shows. For a large savings when there are three deterioration levels, the value of a has to be such that in deterioration level 1 the decision whether to replace or to continue operating (and allow failure due to gradual deterioration) depends on the belief about the type of the installed component.

Table 1 lists both instances with $(\alpha_2, \beta_2) = (0.4, 0.2)$ and instances with $(\alpha_2, \beta_2) = (0.7, 0.1)$. In settings with $(\alpha_2, \beta_2) = (0.4, 0.2)$, in particular the difference between β_1 and β_2 is large, and the optimal policy's ability to adapt decisions to information about the type of the installed component is more valuable. In settings with $(\alpha_2, \beta_2) = (0.7, 0.1)$, in particular the difference between α_1 and α_2 is large, and it is more probable that strong indications of the type of the installed component will

Table 1. The 20 highest-ranked instances based on the value of S (out of 144)

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	11 381.94	11 381.98	13 784.42	21.11
4	0.5	10	(0.4, 0.2)	20	0.1	10 487.18	10 487.22	12 286.48	17.16
5	0.5	10	(0.4, 0.2)	20	0	10 253.45	10 253.49	12 011.46	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	10 082.53	14.67
9	0.5	5	(0.4, 0.2)	20	0.5	13 197.45	13 197.45	15 051.20	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	12 418.20	12 418.20	13 832.65	11.39
14	0.5	5	(0.7, 0.1)	20	0.5	9792.90	9792.90	10 880.80	11.11
15	0.5	5	(0.4, 0.2)	20	0	12 221.58	12 221.58	13 559.75	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

be obtained. Neither effect seems to dominate the other. Also, no clear relation is observed between b and S. The effect of an increase in b is ambiguous: it increases the operating costs that the optimal policy can save over the heuristic policy by replacing an installed component early if it is likely to be weak, but it also increases the operating costs that are common to both policies.

In summary, the numerical experiment demonstrates that taking population heterogeneity into account can lead to a significant cost reduction. The largest savings are achieved when the uncertainty in the type of a newly installed component is high. Furthermore, the savings are generally larger when the number of deterioration levels is higher and corrective replacement is more costly.

5. Conclusions

We have presented a POMDP model for the problem of scheduling replacements for a single-component, Markovian deteriorating system under population heterogeneity. Under intuitively meaningful conditions on the cost parameters and the transition probability matrices associated with the different component types, we have established monotonicity properties of the optimal value function and derived the structure of the optimal policy. The new stochastic order that we introduced to develop these structural results, denoted by $\preceq_{lrst}$, may have applications in other domains as well. We have also performed a numerical experiment to benchmark the optimal policy against a heuristic policy that neglects population heterogeneity. The results indicate that it is particularly important to account for population heterogeneity when there is a high uncertainty in the type of a newly installed component, the corrective replacement cost is high, and the number of deterioration levels is large.

In our model, we assumed that costless, perfect observations on the deterioration level are available at every time epoch. There are also works that study maintenance optimization problems with costly or imperfect inspections (e.g., 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 information states to be defined as bivariate probability distributions. We suspect this will make it difficult to obtain structural results on the optimal policy. Another direction for future research is to relax the assumption that the type of the installed component is independent 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. Future research may also consider the effect of strategic or tactical decisions that influence the composition of the population—e.g., ordering new spare components—on the optimal replacement policy.

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. We are grateful to Bowei Chen (Hygea Medical Technology) and Jop Paauw (Van Oord Dredging and Marine Contractors) for discussions about the practical relevance of our work.

Notes on contributors

Chiel van Oosterom is an assistant professor in the Econometric Institute at Erasmus University Rotterdam, while being in the final phase of his Ph.D. studies conducted at the School of Industrial Engineering, Eindhoven University of Technology. His research interests are in the area of sequential decision making under uncertainty, with a particular interest in stochastic decision processes in which the (un)availability and quality of information plays an important role. He is a member of INFORMS.

Hao Peng is an assistant professor in the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. She received a Ph.D. degree in Industrial Engineering from the University of Houston, Houston, Texas, in 2010. She received her bachelor's degree in Industrial Engineering from Tsinghua University, Beijing, China (2006). Her research interests are maintenance optimization and quality and reliability engineering. Her research is supported by the President Fund of the Academy of Mathematics and Systems Science and the one-hundred plan (class C) of the Chinese Academy of Sciences. She is a member of INFORMS and IISE.

Geert-Jan van Houtum is a professor of maintenance and reliability at the School of Industrial Engineering, Eindhoven University of Technology. He is the scientific director of the BETA Research School for Operations Management and Logistics. His research is focused on spare parts management, production-inventory systems, maintenance and availability management

of capital goods, and the effect of design decisions on the total cost of ownership of capital goods.

References

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

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

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

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

Böttcher, A. and Silbermann, B. (1999) *Introduction to Large Truncated Toeplitz Matrices*, Springer, New York, NY.

Çekyay, B. and Özekici, S. (2011) Condition-based maintenance under Markovian deterioration, in J. Cochran, A. Cox, P. Keskinocak, J. Kharoufeh and J. Smith (eds.), *Wiley Encyclopedia of Operations Research and Management Science*, John Wiley & Sons, New York, NY, pp. 911–922.

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

Chen, N., Ye, Z.S., Xiang, Y. and Zhang, L. (2015) Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research*, **243**(1), 190–199.

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

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

Derman, C. (1963) On optimal replacement rules when changes of state are Markovian, in R. Bellman (ed.), *Mathematical Optimization Techniques*, Cambridge University Press, pp. 201–210.

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

Hansen, E.A. (1998) An improved policy iteration algorithm for partially observable MDPs, in M. Jordan, M. Kearns and S. Solla (eds.), *Advances in Neural Information Processing Systems*, Cambridge, MA, pp. 1015–1021.

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

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

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

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

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

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

Lovejoy, W.S. (1991) A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, **28**(1), 47–66.

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

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

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

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

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

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

Shaked, M. and Shanthikumar, J.G. (2007) *Stochastic Orders*, Springer Science+Business Media, New York, NY.

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

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

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

Xiang, Y., Coit, D.W. and Feng, Q. (2014) Accelerated burn-in and condition-based maintenance for n-subpopulations subject to stochastic degradation. *IIE Transactions*, **46**(10), 1093–1106.

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

Zhang, M., Ye, Z. and Xie, M. (2014) A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering*, **77**, 103–114.

Appendix

Proof of Lemma 1. We prove parts (i) and (ii). Parts (iii) and (iv) directly follow by applying parts (i) and (ii), respectively, 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, $g_y h_x \le g_x h_y$ for all $x, y \in \mathcal{X}$ such that $x \le y < u$. Also, since Proposition 1(i) establishes that $g \preceq_{st} h$, it is true that $g_u = \sum_{x\in\mathcal{X}:x\ge u} g_x \le \sum_{x\in\mathcal{X}:x\ge u} h_x = h_u$.

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

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

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 (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 $\sum_{x\in\mathcal{X}} g_x F(x) \leq \sum_{x\in\mathcal{X}} h_x F(x)$ for every nondecreasing function $F : \mathcal{X} \to \mathbb{R}$.*

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 $P_1 \preceq_{st} P_2 \preceq_{st} \cdots \preceq_{st} P_M$ ascertains $\sum_{j=l}^{N} p_{ij}^t$ is nondecreasing in t, Proposition A1 can be applied to show $\sum_{j=l}^{N} \sigma(j;\pi,i) = \sum_{t=1}^{M} \pi_t \sum_{j=l}^{N} p_{ij}^t \leq \sum_{t=1}^{M} \hat{\pi}_t \sum_{j=l}^{N} p_{ij}^t = \sum_{j=l}^{N} \sigma(j;\hat{\pi},i)$. □

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 $j \leq l < N$ (if such deterioration levels exist). Then, for all $s, t \in \mathcal{T}$ such that $s \leq t$,

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

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

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

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 $(\pi,i) \preceq (\hat{\pi},k)$. Let U be a uniform $(0,1)$ random variable and define two random variables

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

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

$$\begin{aligned}&F(\psi(\pi,i,\xi[u;\pi,i]),\xi[u;\pi,i]) \\ &\quad\leq F(\psi(\hat{\pi},k,\xi[u;\hat{\pi},k]),\xi[u;\hat{\pi},k]) \qquad \text{(A1)}\end{aligned}$$

for all $u \in (0,1)$. We distinguish two cases.

Case (i). $u > \sum_{l=0}^{N-1} \sigma(l;\hat{\pi},k)$. Then $\xi[u;\hat{\pi},k] = N$, and Equation (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). $u \leq \sum_{l=0}^{N-1} \sigma(l;\hat{\pi},k)$. Then $\xi[u;\hat{\pi},k] < N$. Because Lemma 2, which we can apply by Proposition 1(i) and Lemma 1(iv), gives that $\sum_{l=j}^{N} \sigma(l;\hat{\pi},k) \geq \sum_{l=j}^{N} \sigma(l;\pi,k)$ for all $j \in \mathcal{D}$ or, equivalently, $\sum_{l=0}^{j} \sigma(l;\hat{\pi},k) \leq \sum_{l=0}^{j} \sigma(l;\pi,k)$ for all $j \in \mathcal{D}$, 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, Equation (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

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

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. Furthermore, it follows by condition (C4), the induction hypothesis, and Lemma 5 that

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

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 $(\pi,i) \in \Omega$ (see, e.g., Lovejoy (1991)). □

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 $\pi \in \Pi$. □