Structured Learning and Decision Making for Maintenance

Structured Learning and Decision Making for Maintenance

This thesis is part of the PhD thesis series of the Beta Research School for Operations Management and Logistics (onderzoeksschool-beta.nl) in which the following universities cooperate: Eindhoven University of Technology, Ghent University, Maastricht University, Tilburg University, University of Twente, Vrije Universiteit Amsterdam, Wageningen University & Research, Katholieke Universiteit Leuven and Universiteit Hasselt.

Structured Learning and Decision Making for Maintenance

A catalogue record and digital copies are available from the Eindhoven University of Technology Library.
ISBN: 978-94-6423-857-0

Printed by ProefschriftMaken || www.proefschiftmaken.nl
Cover Design by Bregje Jaspers || www.proefschiftontwerp.nl

Structured Learning and Decision Making for Maintenance

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen
op vrijdag 1 juli 2022 om 13:30 uur

door

Collin Drent

geboren te Geleen

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. J.J. Lukkien
1e promotor:	prof.dr.ir. O.J. Boxma
2e promotor:	prof.dr.ir. G.J.J.A.N. van Houtum
copromotor:	dr. S. Kapodistria
leden:	prof.dr.ir. I.J.B.F. Adan
	dr.ir. J.J. Arts (University of Luxembourg)
	prof.dr. N.Z. Gebraeel (Georgia Institute of Technology)
	prof.dr. R.H. Teunter (Rijksuniversiteit Groningen)
adviseur:	dr. J.A.C. Resing

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

To my dearest Veerle

Acknowledgments

John Bardeen[1] once said that science is a collaborative effort, and – being on the verge of completing my PhD research – I wholeheartedly agree with this. While this PhD thesis has only one author, it is indeed the product of countless interactions with colleagues, friends, and family members, who to varying degrees provided support, inspiration, encouragement, and guidance, and without whom I would not have succeeded. I would like to take this opportunity to thank them.

I was in the very fortunate position to be supervised by Onno Boxma, Geert-Jan van Houtum, and Stella Kapodistria. Their guidance, continuous support, and immeasurable knowledge have been imperative to completing this PhD research, and for that I am extremely grateful. Stella, your unprecedented research supervision, career advisory, and life guidance have shaped me tremendously as a researcher and, above all, as a person, and I am deeply indebted to you for that. We have shared memorable moments of struggle and frustration, as well as joy, laughter, success and pride in the past years, and I look forward to many more to come. Onno, I genuinely feel privileged to have worked with and to be supervised by such an extremely knowledgeable, caring and wise person. You have taught me how to approach mathematical problems with clarity and rigor, while at the same time to enjoy that process. Your guidance and support are exemplary, and I deeply appreciate it. Geert-Jan, your passion and enthusiasm for research are really admirable and have been the triggers for me to pursue PhD research in the first place. You have taught me the importance of grasping the *big picture* and framing your research problems. I am very grateful for your brilliant mentorship and valuable lessons. Our collaboration started already before this PhD project, and I look forward to continuing it long after.

A special word of thanks for Joachim Arts, for hosting my two visits to the University of Luxembourg, which were both truly enjoyable and productive. Particularly thanks to you and our co-author Melvin Drent for the fruitful collaboration that led to the work contained in Chapter 2, and to you for our second collaboration that resulted in the work contained in Chapter 4. I am glad that you agreed to be part of my committee, and I thank you for all your valuable feedback on my thesis.

Next, I am thankful to Jacques Resing for our wonderful collaboration on the *toy model* (which proved to be not such a toy model after all) that ultimately led to the work contained in Chapter 7. I also thank you for being part of my committee and for your valuable comments on my thesis.

[1] John Bardeen was an American engineer and physicist, and the only person to be awarded the Nobel Prize in Physics twice.

I am grateful to Nagi Gebraeel for his time to share his valuable advice and feedback in Friday afternoon Teams meetings and for being part of my committee. I would also like to thank Nagi for making it initially possible for me to spend a research visit at the Georgia Institute of Technology, which was unfortunately made impossible by COVID-19. Nonetheless, I am looking forward to collaborating with you in the future.

Finally, I would like to extend my gratitude to the other members of my committee, Ivo Adan and Ruud Teunter, for taking the time to read and review my thesis, and providing me with valuable feedback.

Joining a Mathematics department to pursue PhD research can be a daunting task, even more so when you arrive as an *Industrial Engineer*. Fortunately, it never felt like that because of the extremely wonderful and supportive atmosphere that prevails in the group that I joined. I would like to thank all members of our group for creating that environment. In particular, I thank Dennis, Peter, Mayank, Rik, Ellen, Mark, Rowel, Ivo, Alberto, and Marta. Special thanks goes to Richard, for all the enjoyable conversations we had, and the unforgettable memory that we share as paranymphs at Youri's defense (*and* at the party thereafter). Next, I especially thank my (former) office mates: Youri, Joost, Mona, and Fiona. In the pre-COVID-19 era, working in the office with you was always a pleasure. The atmosphere we had in our office was terrific; one that was deeply missed during COVID-19. I have also great memories from our dinners and the one time we participated in a pub-quiz (where we, to our surprise, were actually *not* that good). Youri, Joost; I really enjoyed the bike rides during COVID-19, which allowed us to have again some social interaction. Mona; like Youri (see Raaijmakers, 2021), I have no idea how you coped with three Dutch guys and their lame humor in the office, but you did indeed an excellent job! Youri, thank you for the unforgettable memories, in particular the ones from *Brisbaan*.

I thank Chantal and Ellen for their administrative support and helping out on other practical issues. Chantal, thank you for the great conversations we had, especially the ones we regularly had via Teams during COVID-19. (Talks about) Daan and Ollie (and other topics) were very pleasant and funny distractions from work. Liron, Mark, and Patty; it has been a pleasure to organize the 13th YEQT workshop with you.

While my PhD research took place in the SPOR group, I was also a regular visitor of the OPAC group. Those visits were always a pleasure and something to look forward to, and for that I thank all group members of OPAC. In particular thanks to İpek, for our regular, fruitful discussions about our research topics, and to Claudine, for the administrative support and enjoyable conversations every now and then.

Since my research was supported by the Data Science Flagship framework, I had

the opportunity to interact with people from Philips to learn about the intricacies of maintenance optimization in practice, for which I am grateful. In particular I thank Mauro Barbieri, Jan Korst, and Verus Pronk.

My research visits to the Luxembourg Centre for Logistics and Supply Chain Management (LCL) would not have been as enjoyable as it has been without the many coffee breaks, table tennis matches, lunches, dinners, and other interactions with the LCL members. In particular I thank Joachim, Melvin, Benny, Çağil, Nils, Anne, You, Sarah, Laurens, Roozbeh, Bikey, Nicole, Neeraj, Rishikesh, and Jackie. I gratefully acknowledge the Netherlands Organisation for Scientific Research for financially supporting these visits.

PhD research can be incredibly challenging and strenuous in which escapades can be most welcome. I would like to thank my friends for providing many of such escapades throughout the years. Starting with *de groep*, which formed even long before I started this PhD project. Since most of us are now spread all over the country, we were deemed to only yearly reunions over the past years, but they were always very memorable.. Water polo (especially the social activities after matches and trainings) has been another source of many escapades, for which I thank *de mannen.* Mats and Thom, thank you for the trips, dinners, parties, and other good times together; they were very welcome distractions from the hard work. Marc, during our studies, we have intensively *and* successfully worked together in literally each course that we followed. I thoroughly enjoyed that and looking back, that is where the foundation for this PhD research has been laid and I thank you for that. Special thanks to my friend (and former flat mate) Koen. I can vividly recall supporting each other during *our* long days in the Flux during our Master thesis research, and I thank you for supporting me in *my* long days during my PhD research. I have great memories from when we both lived at the Pisanostraat (Oude Rechtbank, Just Dance, Aït-Ben-Haddou, to name a few), which always brought the right amount of distraction. Finally, I am grateful to my dear friend Bart. It is very special how we support each other through thick and thin, even now that the Atlantic Ocean is between us. In the past years, it seemed that our paths would cross multiple times, even for prolonged times, but each time COVID-19 had other plans. Perhaps it is also for the better, as I am not sure where a five-month get-together would have taken us – perhaps back to San Juan del Sur (?), but most likely *not* to my PhD defense. Nevertheless, your friendship is truly one of a kind and I find tremendous support in it.

Melvin, Youri; thank you for being my ceremonial assistants during my PhD defense. I am sure that your *first-hand* experiences with such events will come in very handy.

I would like to thank my (soon-to-be) in-laws, who have always been incredibly

supportive during my PhD research, and far beyond. Brigitte, Nico, Jesse, and Naomi, I am very fortunate to have such a beautiful *second* family.

I am deeply indebted to my family for all their love and support, especially to my mom, dad, and two brothers. Levi, thank you for always being there for me no matter what; I know I can always count on you and I am forever thankful for that. Melvin, I have thanked you already for being a co-author and paranymph, but I would also like to thank you for always being by my side; our bond is truly special and I find great support in it. Mam & Pap, you have been extremely encouraging, not only during the last years, but throughout my *whole* life. Without your selfless and unconditional support and love, I would not be who I am and where I am today.

Lastly, but most importantly, I would like to express my deepest gratitude to my beautiful fiancee, for her support in all but mathematics. Veerle, you have been an inexhaustible source of support, comfort, encouragement, and, above all, love throughout this journey. I am incredibly lucky to have you in my corner, encouraging me to grow in all of my favorite directions – even when this means training for an Ironman in parallel to completing this PhD research. Thank you for being my *safe haven*. I dedicate this PhD thesis to you.

Collin Drent
Stein, May 2022

Contents

Chapter 1

Introduction

Advanced technical systems are critical for the smooth operation of public services such as public transport (e.g., aircraft, rolling stocks), utilities (e.g., power plants, wind turbines), and health care (e.g., MRI scanners, interventional X-ray machines) as well as for the primary processes of companies (e.g., assembly lines, lithography machines). Unavailability – especially when unplanned – and failures of these systems have severe consequences, both from a *societal* and *financial* point of view.

Recent examples that highlight the impact on society are manifold. For example, the failure of a deteriorating propeller blade led to the crash of a KC-130T aircraft with 16 casualties in Mississippi in the summer of 2017 (Insinna and Ziezuliwicz, 2018). Another example from the aircraft industry concerns the crash of Lion Air flight 610 in Indonesia resulting in 189 casualties, which was partly caused by poor maintenance to the Boeing 737 (Austin and Muktita, 2019; Jasper, 2019). In January 2020, a train of the Dutch Railways (NS in Dutch) derailed in one of the densest areas of the Dutch rail network. Fortunately, as opposed to the examples with aircraft, there were no casualties. Nonetheless, the rail infrastructure was heavily disrupted for several days. Later that year an investigation concluded that poor maintenance was the root cause of this derailment, specifically to deteriorated brake pads of the train (NOS, 2020).

When unplanned downtime does not lead to immediate disturbance of society, the consequences can still be severe financially. One striking example comes from the semiconductor industry that heavily relies on lithography equipment; systems that are used for the production of chips that are built in virtually all electronic devices that we use in our everyday lives. Recent estimates indicate that unplanned downtime of such equipment can cost up to 72,000 euros per hour (Lamghari-Idrissi,

2021). Estimates for the global manufacturing industry in general, of which the semiconductor industry is a part, indicate that the cost of unplanned downtimes of systems and their aftermaths is roughly $50 billion annually, with system failures being the *sole* cause of almost half of this downtime (Wall Street Journal Custom Studios, 2017; Coleman et al., 2017). The implication of these numbers is twofold. First, unplanned downtime costs is a general problem faced by many industrial companies, stretching far beyond the examples put forward above. Second, and arguably more importantly, unplanned downtime costs can be reduced significantly through minimizing equipment failures. This latter implication also seems to be well-understood by executives in asset management. In a recent survey, they perceive unplanned failures as the most important risk to manage (Pacquin, 2014).

Motivated by the preceding examples and cost figures, asset managers responsible for these systems invest heavily in their maintenance operations to *prevent failures* and otherwise *mitigate their consequences*. The playing field in which these asset managers operate has been reshaped drastically over the years due to the increasing and improved availability of real-time data stemming from ubiquitous sensors installed in modern equipment, integrated in the so-called Internet-of-Things (IoT) (Coopers, 2014; Blackwell et al., 2017). These sensors continuously measure various conditions of components, thereby generating a real-time stream of data that offers ample opportunities. Asset managers can monitor and interpret this real-time data generated by each asset to assess their physical condition and intervene with maintenance when appropriate; this so-called *digital-physical marriage* holds the promise to revolutionize maintenance operations (Olsen and Tomlin, 2020).

Indeed, Manyika et al. (2015) estimated that, by 2025, such IoT-driven maintenance applications have the potential to (i) reduce equipment downtime by up to 50%, and (ii) decrease equipment capital investment by up to 5% through maximizing the utilization of the useful lifetime of equipment. The way real-time data is leveraged for decision making will ultimately determine the exact realized value. In order to maximize this realized value, a smart framework and smart mathematical models are needed as *foundation* to build such IoT-driven maintenance applications on.

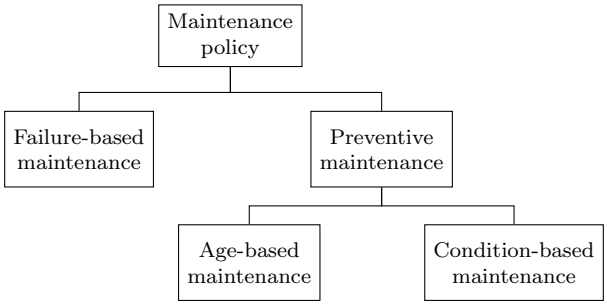
The overarching objective of this thesis is to address this need by developing mathematical models that leverage real-time data to improve decision making in maintenance. The remainder of this chapter is organized as follows. In Section 1.1, we give an overview of maintenance operations and discuss the most commonly applied maintenance policies. Section 1.2.1 discusses the conventional approach to maintenance optimization that is generally adopted by both practitioners and academics. This conventional approach, as we shall describe, does not allow for

incorporating real-time data, and we therefore discuss an alternative approach that does allow taking into account data in real-time (Section 1.2.2). This alternative approach serves as the foundation of our developed mathematical models. In Section 1.3, we position the work in this thesis in the literature, and in Section 1.4 we give an overview of our main contributions and provide an outline of this thesis.

1.1. Maintenance operations

Decision makers in asset management can resort to certain maintenance policies to ensure high system availability and low unplanned downtime. For the purpose of these policies, it is convenient to think of a system as a set of components. Figure 1.1 provides an overview of the maintenance policies that a decision maker can employ to decide for each of these components when to perform maintenance. In the remainder of this section we will describe these policies in more detail.

Figure 1.1 Classification of maintenance policies.



The first distinction is between failure-based maintenance (FBM) and preventive maintenance. The former policy is also called corrective maintenance and is usually adopted in the case where components do not wear, are cheap, or when their failures do not adversely affect the system's availability. When maintenance of components after failure is costly or such failures lead to very costly downtime (as is the case for advanced technical systems such as MRI systems, lithography systems, and wind turbines), a decision maker can choose to preventively maintain components *before* they fail.

Failures can thus be prevented by performing preventive maintenance regularly. However, early performance of preventive maintenance leads to high capital expenditures as the useful lifetime of components is cut short. The challenge faced by asset managers in preventive maintenance is therefore to optimize *two conflicting*

objectives: (i) minimize the risk of failure and unplanned downtime with all its adverse consequences and (ii) maximize the utilization of the useful lifetime of a component.

The traditional preventive maintenance policy is called the age-based maintenance (ABM) policy. Essentially, in an ABM policy (and all its variations), the decision to perform maintenance is based on the age of a component. That is, components are replaced at failure, or at a planned time after installation, whichever occurs first. We remark here that researchers sometimes make a further distinction between ABM and so-called usage-based maintenance (UBM), where a component is replaced based on the usage (e.g., number of kilometers or usage hours). However, since UBM works the same as ABM – the component is replaced whenever it reaches a certain amount of usage or failure, whichever occurs first – we only consider ABM. Although ABM can reduce unplanned downtime and failures considerably compared to FBM for components that do wear, the utilization of the useful lifetime is by default not maximized when a component is preventively replaced based on its age.

When asset managers have real-time data at their disposal that indicates the condition of a system, they can base their maintenance decision on that information. This policy, which is called condition-based maintenance (CBM), can maximize the utilization of the useful lifetime by ensuring that components are replaced early when necessary and late when possible; in the former the condition indicates that failure is near, while in the latter the opposite is the case. In CBM, an asset manager aims to predict, based on the condition, when the component will fail, and for this reason, CBM is also often called *predictive maintenance*. Due to the recent developments in sensor technology that enable real-time measurements of systems' conditions, as discussed in the previous section, CBM has gained momentum in recent years.

Maintenance optimization – the topic of this thesis – focuses on the development and analysis of mathematical models aimed at improving or optimizing maintenance policies for single components that fit in the policies described above. In this thesis, we mainly focus on CBM (Chapters 2, 3, 4, 5 and 7) and ABM (Chapter 6), though in Chapter 7 we also investigate policies that can smartly leverage opportunities that arise due to FBM. The latter policies are also sometimes referred to as *opportunistic* maintenance policies and they can be both condition-based and age-based.

1.2. Maintenance optimization

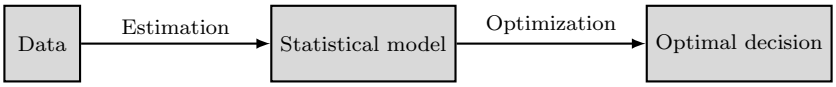
In this section we first describe the *conventional* approach to obtaining optimal maintenance policies for single components, and then contrast this approach with

the approach we take in this thesis.

1.2.1 The conventional approach

The conventional approach to maintenance optimization comprises two sequential steps, see Figure 1.2.

Figure 1.2 The conventional approach to maintenance optimization.



The first step is the *estimation* of a statistical model of either the deterioration process or time to failure of a system based on either condition or failure data, respectively. Conceivably, the nature of the needed statistical model depends on the policy that is adopted, where the former is usually needed for CBM while the latter is needed for ABM. The second step is the *optimization* of the decision when to perform maintenance based on the statistical model, thereby trading off costly preventive maintenance with costly corrective maintenance. For this second step, the general approach – since aging and the most common deterioration processes satisfy the Markov property – is to formulate the problem as a Markov decision process (MDP) to (i) get insights into the structure of the optimal policy, and/or (ii) compute optimal policies. Before we move to the next subsection, where we discuss MDPs in more detail, we already point out that there are two serious drawbacks with this conventional approach:

1. it assumes a homogeneous population where all components are statistically identical (i.e., they have the same statistical model) to any other component in the population;

2. it is based on the assumption that there is sufficient data available to accurately estimate the statistical model.

Note that these assumptions also enforce each other: Under the assumption of a homogeneous population (i.e., all components are statistically indistinguishable), all available data can be pooled together leading to sufficient data. Similarly, if the first assumption is violated and each component is statistically distinguishable, then there is immediately data scarcity on an individual component-level to estimate its statistical model.

In light of these drawbacks, this approach is not tenable when (i) a statistical model needs to be learned on-the-fly from real-time data, and/or when (ii) components stem from a heterogeneous population (i.e., all components are statistically distinguishable) and maintenance decisions need to be tailored individually. Consequently, this sequential approach cannot reap the benefits of all the opportunities that the advent of sensor technology offers for maintenance operations as described in the beginning of this chapter, and we need a different approach for that.

In the next section, we provide the necessary background information of the prototypical framework often used for the optimization step when a statistical model is readily available. It will then also become clear why such an approach indeed suffers from the drawbacks described above, and particularly, why it is not able to incorporate real-time data. In Section 1.2.2, we discuss how we can extend this prototypical framework such that it can (i) incorporate learning from real-time data and (ii) tailor decision making to each individual component.

Markov decision processes

When a statistical model is *available*, then the prototypical mathematical framework to compute and/or obtain insights into structural properties of optimal maintenance decisions is based on the theory of MDPs.

MDPs are models to analyze sequential decision making when uncertainty is involved – such as, but certainly not limited to, maintenance decision making. The sequentiality lies in that decisions are not made in isolation at a single point in time, but rather have both immediate and long-term effects, and are made repeatedly over a certain time horizon. We first informally discuss an MDP and then give a more rigorous introduction. This section serves merely as a short introduction to MDPs to the extent needed at this stage for this thesis; for a comprehensive treatment of MDPs we refer the reader to Puterman (2005), Bertsekas (2007), Guo and Hernández-Lerma (2009), Feinberg and Shwartz (2012), and Kallenberg (2020).

An MDP consists of decision epochs, states, actions, costs, and transition probabilities. Informally speaking, a decision maker has to take an action based on the current state (e.g., the system's deterioration or age in the context of maintenance) at a decision epoch, where such an action will (i) generate an immediate cost, and (ii) determine the state at the next decision epoch according to some transition probability function. This process repeats itself for either a finite or an infinite amount of time. For this introductory section it suffices to consider the setting in which the decision epochs are equidistant so that the resulting MDP is a so-called discrete-time MDP.

However, it is good to mention that the amount of time between decision epochs may also be a random variable, in which case the MDP is a so-called semi-MDP, or that decisions can even be made in continuous time, in which case the MDP is called a continuous-time MDP. In Chapter 7, where we consider semi-MDPs, and in Chapter 4, where we consider a continuous-time MDP, we shall discuss these in more detail.

A decision maker can employ a policy, which is essentially a prescription of which action to take for any possible state at every decision epoch. Decision makers seek policies that are optimal in some sense. In the context of maintenance, as we shall frequently see in this thesis, this is usually the minimization of either (i) the total expected (discounted) maintenance cost (both corrective and preventive), or (ii) the average maintenance cost per time unit. Here, we would like to remark that the optimal discounted policy need not be the same as the optimal average cost policy (in fact, the latter might not even exist), see e.g., Arapostathis et al. (1993) and Schäl (1993). Fortunately, in Chapter 3 we show that for archetypical maintenance problems, like the ones treated in this thesis, the two are usually equivalent.

We shall now mathematically describe the standard discrete-time MDP in which decision epochs are equidistant, with 1 time unit between epochs (w.l.o.g.), with the total expected discounted cost (over an infinite horizon) as optimality criterion in more detail. We do so because most maintenance optimization models, both in the literature and in this thesis, are based on this setting. An MDP is a tuple of elements $\langle \mathscr{S}, \mathscr{A}, \mathbf{C}, \mathbf{p}, \gamma \rangle$. Here, $\mathscr{S}$ is a set of states, called the state space, $\mathscr{A} = \big(\mathscr{A}_s\big)_{s\in\mathscr{S}}$ is the action space, where $\mathscr{A}_s$ is a bounded set of actions available in state $s \in \mathscr{S}$. $\mathbf{C} = \big(C(a,s)\big)_{s\in\mathscr{S}, a\in\mathscr{A}_s}$ is the collection of direct cost functions for taking action $a \in \mathscr{A}_s$ in state $s \in \mathscr{S}$, and $\mathbf{p} = \big(p(s'|s,a)\big)_{s,s'\in\mathscr{S}, a\in\mathscr{A}_s}$ is the collection of transition probabilities (i.e., $p(s'|s,a)$ is the probability of moving from s to s' when taking action a). One could also let $\mathbf{C}$ and $\mathbf{p}$ depend on time, or let $\mathbf{C}$ also depend on the future state s', but in order to keep the formulation concise we refrain from discussing those particular settings. Finally, $\gamma \in (0,1)$ is the discount factor.

Let $\tau \in \mathbb{N}_0$ ($\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$) be a decision epoch. A decision rule $\pi_\tau = \big(\pi_\tau(s)\big)_{s\in\mathscr{S}}$ indicates for all states $s \in \mathscr{S}$ which action to choose at decision epoch τ. Let Π be the set of all non-anticipatory policies. A policy $\pi \in \Pi$ is a sequence $\{\pi_\tau\}_{\tau\in\mathbb{N}_0}$ of decision rules for all decision epochs $\tau \in \mathbb{N}_0$. Let the random variable S_t be the state of the system at $t \in \mathbb{N}_0$. Given a policy $\pi \in \Pi$, the total expected γ-discounted cost given that the process starts in state $s \in \mathscr{S}$ at $\tau = 0$, denoted with $V_\pi(s)$, is given by

$$V_\pi(s) = \lim_{T\to\infty} \mathbb{E}_\pi \left[\sum_{\tau=1}^{T} \gamma^\tau C\big(\pi_\tau, S_\tau\big) |\ S_0 = s \right],$$

where the evolution of S_t depends on both π and S_0. The optimal total expected γ-discounted cost given that the process starts in state $s \in \mathscr{S}$ at $\tau = 0$, denoted with $V^*(s)$ is given by

$$V^*(s) = \inf_{\pi \in \Pi} V_\pi(s),$$

which throughout this thesis we refer to as the value function.

Under suitable conditions – satisfied by all discrete-time MDPs in this thesis, and in general for most archetypical maintenance problems – we know by Proposition 1.2.2 of Bertsekas (2007) that there exists an optimal Markov policy π^* depending only on the current state s independent of the decision epoch $\tau \in \mathbb{N}_0$, and this policy satisfies the following Bellman optimality equations:

$$V^*(s) = \min_{a \in \mathscr{A}_s} \left\{ C(a, s) + \gamma \sum_{s' \in \mathscr{S}} p(s'|s, a) \cdot V^*(s') \right\}. \tag{1.1}$$

If $\mathscr{S}$ is continuous, then the summation is replaced by an integral (in Chapter 3 we consider such a setting). The celebrated Bellman equation in (1.1) decomposes the value of a decision problem at a certain point in time in both the immediate cost $C(a, s)$ and the expected value of future actions that results from the current action a and continuing optimally according to the optimal policy (Bellman, 1952). This equation can be used for both computational purposes (i.e., to compute optimal policies via the value iteration algorithm or successive policy approximation) or analytical purposes (i.e., to establish structural properties of the optimal policy).

We now describe a canonical CBM example to make the previous theory more tangible from a maintenance optimization perspective.

Example 1.1. Consider a component whose condition deteriorates according to a Poisson process $\{X(t), t \geq 0\}$ with rate λ per time unit. At equidistant decision epochs (say with distance 1) we observe the condition $X(t)$ and decide whether we apply preventive maintenance at cost $c_p > 0$ or continue without performing maintenance. If we find that $X(t) \geq L$ (L is the failure threshold), the component has failed and we maintain correctively at cost $c_u > c_p$. When the component is maintained, which takes exactly 1 time unit, the deterioration level moves to 0 and the component will start deteriorating from that level again. We can model this problem as an MDP with $\mathscr{S} = \{0, 1, \ldots, L-1, \xi\}$ (ξ consists of all states greater than or equal to state L), $A_s = \{\text{maintain, continue}\}$ for $s < \xi$ and $A_\xi = \{\text{maintain}\}$, $C(a, s) = 0$ if $a = \text{continue}$, $C(a, s) = c_p$ if $s < \xi$ and $a = \text{maintain}$, and $C(a, s) = c_u$ if $s = \xi$

and $a =$ maintain. Since the deterioration increment between two decision epochs is distributed as a Poisson random variable, we have

$$p(s'|s,a) = \begin{cases} \frac{\lambda^{(s'-s)}e^{-\lambda}}{(s'-s)!}, & \text{for } s' \in \{s, s+1, \ldots, L-1\}, \\ 1 - \sum_{i=s}^{L-1} \frac{\lambda^{(i-s)}e^{-\lambda}}{(i-s)!}, & \text{for } s' = \xi, \end{cases}$$

for $a =$ continue, and $p(0|s,a) = 1$ if $a =$ maintain. ◇

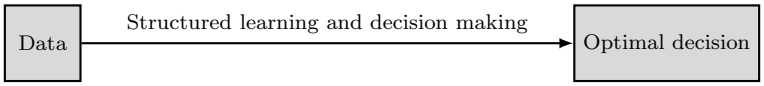
For the example above, it is optimal to perform maintenance if the deterioration is larger than or equal to a certain control limit, say $\delta \leq \xi$, at a decision epoch, and continue otherwise. This important result can be attributed to both Derman (1963) and Kolesar (1966), who, in their seminal papers on CBM, established the optimality of a control limit policy under mild conditions on the deterioration process. Starting with this important result, there is a long and rich history of establishing the optimality of control limit type CBM policies. The vast majority of this literature assumes that the dynamics are fully known with certainty. The reason for this assumption is also obvious at this stage; without the assumption it is much harder to tractably model the problem as an MDP to analyze the maintenance problem. That is, when the transition probabilities $\mathbf{p} = \big(p(s'|s,a)\big)_{s,s' \in \mathscr{S}, a \in \mathscr{A}_s}$ in the general MDP description, or the parameter λ in the example above, are *not* known.

One common way to overcome this, is by following the conventional approach that we described at the start of this section (see Figure 1.2). Using deterioration or aging data, one could estimate a statistical model that governs the transition probabilities and use that as an input to solve the corresponding MDP to obtain optimal maintenance policies. However, we already pointed out the shortcomings and limitations that this sequential approach suffers from. We therefore propose an integrated approach, which we discuss in the next section in more detail and that we mostly adopt throughout this thesis.

1.2.2 An integrated approach

At the start of this chapter we already mentioned that recently built systems usually have integrated sensor technology that allows deterioration data of individual components to be gathered and relayed in real-time to decision makers. Such an individual stream of real-time deterioration data can be used to integrate (i) learning the deterioration process on a component-level, and (ii) tailoring the maintenance decision to each component individually. In this section, we discuss the framework to do this on a conceptual level (see Figure 1.3 for a representation).

Figure 1.3 An integrated approach to maintenance optimization.



Bayesian learning

In order to integrate learning with decision making, we use a tractable approach: Bayesian inference, which enables us to sequentially learn unknown parameters with increasing accuracy. Bayesian inference is a statistical approach to parameter estimation based on Bayes' theorem, where knowledge about parameters is updated as more and more data is observed (Gelman et al., 1995; DeGroot, 2005).

We will now explain this process in more detail. Assume that in each period our data (e.g., deterioration, lifetime), denoted with X, is generated by a probability distribution with known probability density function $f(x|\theta)$ parametrized by an unknown parameter θ with $\theta \in \Theta$, where Θ is the set of all possible values that θ can take. In theory, Θ can be both discrete or continuous, and finite or infinite, but from a practical point of view, the nature of the situation we aim to model determines the choice of Θ. We will now clarify this as this is an *important* observation for the rest of this thesis.

Θ effectively models heterogeneity in the population of components. With *population heterogeneity*, we refer to the situation that components – belonging to a pool of components – are heterogeneous with respect to the (unknown) parameters of their underlying deterioration process or lifetime distribution. When Θ is a finite and discrete set, its elements represent possible values that the unknown parameter can take. In other words, there is a known, finite and discrete component population, where the relevant parameters of each component are known, and the decision maker is uncertain from which element of the component population a particular component stems. When Θ is continuous, there is no notion of a discrete set with known values that the unknown parameter can take. In this case, uncertainty in the parameter is modeled directly through the probability distribution of Θ. The former setting, which we refer to as *population uncertainty*, is applicable if a decision maker for instance knows that a component can be either a fast or slow deteriorating component but upon installment it is unknown which one of the two is installed. The latter setting, which we refer to as *parameter uncertainty*, is applicable when the decision maker has no knowledge regarding possible specific values an unknown parameter can take, but there is data available to fit a probability distribution that models the parameter uncertainty directly. In this thesis, the objective is to develop the general theory

for mathematical models for maintenance optimization characterized by parameter uncertainty. However, since the difference between the two settings is subtle and the used approaches similar, the population uncertainty setting is relevant – both in the context of ABM and CBM – and we shall therefore discuss works in this area in our literature overview in the next section. We now proceed with describing the general approach to Bayesian learning for the parameter uncertainty setting and we will remark when the procedure differs from the population uncertainty setting.

Given a prior belief with probability density function $\omega_n(\theta)$ (probability mass function in the population uncertainty setting) at decision epoch n and an observation $X_n = x_n$, the posterior density function $\omega_{n+1}(\theta|x_n)$ is given by Bayes' theorem:

$$\omega_{n+1}(\theta|x_n) = \frac{f(x_n|\theta)\omega_n(\theta)}{\int_\Theta f(x_n|\theta')\omega_n(\theta')\mathrm{d}\theta'}. \tag{1.2}$$

The recursive scheme in Equation (1.2) provides a sequential mechanism to learn an unknown parameter with increasing accuracy as data becomes available. This method does not require sufficient data to estimate a model but can rather be used on-the-fly. When the observation x_n in decision epoch n is observed, the prior belief $\omega_n(\theta)$ is updated to the posterior belief $\omega_{n+1}(\theta|x_n)$, which then becomes the prior belief for decision epoch $n+1$ (with slight abuse of notation, we drop the conditioning on x_n and write $\omega_{n+1}(\theta)$ for the prior at decision epoch $n+1$). This process is known as prior-to-posterior updating (Gelman et al., 1995; DeGroot, 2005).

Given a belief $\omega_n(\theta)$ of the uncertain parameter at decision epoch n (which is thus the result of the data accumulation up to decision epoch n), we can also make a prediction of the future data point (e.g., deterioration at the next decision epoch), encoded in the posterior predictive distribution of X_{n+1}. One obtains this posterior predictive by marginalizing the data point over its prior distribution:

$$f_{X_{n+1}}(x_{n+1}) = \int_\Theta f(x_{n+1}|\theta)\omega_n(\theta)\mathrm{d}\theta. \tag{1.3}$$

In general, the probability distributions characterizing the uncertainty of the unknown parameter – $\omega_n(\theta')$ in Equation (1.2) – are continuous densities defined over continuous domains. This means that performing the integrations (summations in the population uncertainty setting) prescribed by Bayes' theorem in Equation (1.2) can be a computationally intensive and arduous process. Fortunately, there are so-called conjugate families of distributions that considerably ease this process. A conjugate pair is a pair such that the posterior is in the same family of distributions as the prior, albeit with parameters updated based on some sufficient statistic of the observed data

(e.g., the number of data points, the size of the deterioration increment).

Remark 1.1. Prior-to-posterior updating needs to be initialized with an initial prior distribution at the first decision epoch. In practice, this initialization can be done using so-called empirical Bayes methods, which are procedures for statistical inference in which this prior distribution is estimated from *some* available data. In Chapter 2, we illustrate such an approach when we estimate the initial prior distribution from a real-life data set. Further, although conjugate priors are usually chosen because of their mathematical convenience, in practice, the prior distribution could also be interpreted as the distribution that *most likely* models the heterogeneity of the population. When some historical data is available from different components of a population, empirical Bayes methods can be used to statistically judge whether the conjugate prior distribution is indeed the correct distribution to model the population heterogeneity. We refer the interested reader for a comprehensive treatment of this topic to Carlin and Louis (2000) and Maritz and Lwin (2018). ◇

In this thesis, we intensively use such conjugate pairs to tractably learn the unknown parameters of the lifetime distribution (Chapter 6) or deterioration processes (Chapters 2, 3, 4 and 5). We will treat this in more detail and with more rigor in each individual chapter.

The next example is a continuation of Example 1.1. It illustrates a conjugate pair for the case that the rate λ of a Poisson random variable is unknown, in which case it is well-known that the Gamma distribution is a conjugate prior (Gelman et al., 1995). This example thus falls in the parameter uncertainty setting. An analog of this example for the population uncertainty setting would be if λ can only take two values, say λ_{high} and λ_{low}. In that case, the prior (and thus posterior) distribution models the belief that the component has λ_{high} (or λ_{low}) as true rate.

Example 1.2 (Continuation of Example 1.1)**.** Suppose that the rate λ is unknown. At decision epoch 0, upon installment of the component, we endow λ with a Gamma prior with shape parameter α_0 and rate parameter β_0. Suppose that at the next decision epoch 1, the component's deterioration is x_1. We can then update the parameters of the Gamma prior as $\alpha_1 = \alpha_0 + x_1$, and $\beta_1 = \beta_0 + 1$ to obtain the posterior, which is again a Gamma distribution. More generally, at decision epoch n if the deterioration is x_n, the Gamma posterior has parameters $\alpha_n = \alpha_0 + x_n$, and $\beta_n = \beta_0 + n$. In this case, the sufficient statistics are the decision epoch, n, and the deterioration level at that epoch, x_n. ◇

Bayesian Markov decision processes

The use of conjugate pairs has one additional advantage that is *imperative* for the integrated approach of learning and decision making. The sufficient statistics needed for the prior-to-posterior updating generally retain a Markovian structure (see Example 1.2 for an illustration). This implies that we can equip the standard state space description of the standard MDP (see Section 1.2.1) with the sufficient statistics needed for the learning, which results in a so-called Bayesian MDP. In this Bayesian MDP, the state-dependent transition probabilities are governed by the posterior predictive distribution that corresponds to each state. We illustrate a Bayesian MDP in the following example, which builds further on Example 1.2.

Example 1.3 (Continuation of Example 1.2)**.** The sufficient statistics are deterioration level x and decision epoch n (see Example 1.2). We thus need to extend the state space description from Example 1.1 to account for the decision epoch n, which can take values in $\mathbb{N}_0$. Hence we have for the Bayesian MDP, $\mathscr{S} = \{0, 1, \ldots, L-1, \xi\} \times \mathbb{N}_0$, $A_s = \{\text{maintain, continue}\}$ for $s \in \{(x,n) \in \mathscr{S} | x < \xi\}$ and $A_s = \{\text{maintain}\}$ for $s \in \{(x,n) \in \mathscr{S} | x = \xi\}$, $C(a,s) = 0$ if $a =$ continue for all $s \in \mathscr{S}$, $C(a,s) = c_p$ if $s \in \{(x,n) \in \mathscr{S} | x < \xi\}$ and $a =$ maintain, and $C(a,s) = c_u$ if $s \in \{(x,n) \in \mathscr{S} | x = \xi\}$ and $a =$ maintain. The deterioration increment between two decision epochs is now governed by the posterior predictive distribution, which in this case is a Poisson distribution with Gamma-distributed mean for which it is known that it is a Negative Binomial distribution. Specifically, when the current state is $(x,n) \in \mathscr{S}$ so that the current Gamma prior has parameters $\alpha = \alpha_0 + x$, and $\beta = \beta_0 + n$, then the posterior predictive distribution is a Negative Binomial distribution with success probability $p = \frac{\beta_0+n}{\beta_0+n+1}$ and number of successes $r = \alpha_0 + x$ (Gelman et al., 1995). The transition probabilities $p(s'|(x,n),a)$ for $s' \in \{(x', n+1) \in \mathscr{S} | x' \geq x\}$ can easily be derived for $a =$ continue, and $p((0,0)|s,a) = 1$ for all $s \in \mathscr{S}$ if $a =$ maintain. Observe that this latter probability implies that after maintenance, we start in state $(0,0)$, meaning that (i) we install a new component, and (ii) we reset the learning process and start again with our initial prior Gamma distribution, parametrized by α_0 and β_0. We remark here that if components are statistically indistinguishable; that is, they all have the same unknown parameter λ that needs to be learned on-the-fly, then one can model this by continuing with the current parameters instead of resetting the learning process. In this case, the sufficient statistic is the sum of all observed deterioration increments of each installed component (see also Chapter 5, were we study such a model). ◇

A Bayesian MDP, as opposed to the standard MDP, allows distribution updating and policy updating based on the real-time data that a component generates. For

the example above, it is again optimal to perform preventive maintenance if the deterioration is larger than or equal to a certain control limit, say $\delta(n) \leq \xi$, at a decision epoch, and continue otherwise. The difference with Example 1.1 is that this control limit now *depends* on the observed data through the sufficient statistic (i.e., n). As data is gathered and incorporated via prior-to-posterior updating, a decision maker can in real-time predict its future evolution through the posterior predictive distribution, and by incorporating this distribution in an MDP, she is able to make optimal decisions. In other words, a Bayesian MDP integrates learning and decision making and tailors these decisions to each system individually based on its own real-time data. The concept of Bayesian MDPs is not new; early applications can be found in the literature on the sequential design of experiments and the related theory on multi-armed bandits (cf. Bellman, 1956; Gittins and Glazebrook, 1977; Gittins and Jones, 1979). Recently, with the advent of information technology and the increasing availability of data, Bayesian MDPs have gained more popularity as a tool to integrate learning and decision making in many areas of operations management such as inventory management (e.g., Azoury, 1985; Lariviere and Porteus, 1999; Chen and Plambeck, 2008; Chen, 2010; Li and Ryan, 2011), and revenue management (Harrison et al., 2012; Afèche and Ata, 2013). In this thesis, we also use Bayesian MDPs as the main tool to integrate learning and decision making.

1.3. Positioning in the literature

In this section, we position each individual chapter in the literature. As the literature on maintenance optimization is very rich and extensive, we will mainly focus on the intersection with Bayesian optimization and we refer the interested reader to Scarf (1997), Wang (2002), and De Jonge and Scarf (2020) for comprehensive overviews of the complete area. We first discuss the field that the majority of this thesis is focused on: CBM. In Section 1.3.1, we position Chapters 2, 3, 7, and 5. We then continue with ABM models, in Section 1.3.2, where we position Chapter 6. Subsequently, we proceed with discussing two relevant areas in the literature that are strongly related with maintenance optimization – process control and condition-based production – and our contributions to these areas. Specifically, Chapter 3 and Chapter 4 will be positioned in Section 1.3.3 and Section 1.3.4, respectively. It is noteworthy to mention that a more extensive discussion of the literature related to the work in Chapter 7 – which does not relate to Bayesian optimization – is deferred to a focused literature review in that chapter (see Section 7.2).

1.3.1 Condition-based maintenance

Starting with the aforementioned seminal papers of Derman (1963) and Kolesar (1966), there is a long and rich history of establishing structural properties of optimal maintenance policies. Especially results about the optimality of control limit policies for the replacement of deteriorating components under many model variations have appeared in the literature; see e.g., Ross (1969), Kao (1973), Rosenfield (1976), Makis and Jardine (1992), Benyamini and Yechiali (1999), Makis and Jiang (2003), Maillart (2006), Kurt and Kharoufeh (2010), and Drent et al. (2019) (see also Chapter 7). A lot of research has been conducted in the area and most studies assume that the deterioration process is known a-priori. The distinguishing feature of our developed models is that we assume that the deterioration processes are a-priori unknown and need to be inferred from real-time data. In the previous section we described two different modeling approaches to use real-time data to learn an unknown parameter: the parameter and population uncertainty setting. Although our thesis centers around the parameter uncertainty setting, we will also discuss works adopting the second approach. Finally, we will also discuss works in a third stream, called the state uncertainty approach, which deals with settings in which the deterioration process itself is not (fully) observable but must be learned from data.

Parameter uncertainty

The parameter uncertainty setting is pioneered by Gebraeel et al. (2005), who developed a Bayesian framework to learn the a-priori unknown parameters of a Brownian motion from real-time sensory data. Using the Bayesian framework proposed in Gebraeel et al. (2005), Elwany et al. (2011) develop a Bayesian MDP and establish the optimality of a control limit policy that is non-decreasing in the age of the equipment. Similar to our research in Chapter 3, they model the deterioration process as a Brownian motion in which the drift parameter is initially unknown, but, unlike us, they assume this drift to be non-negative and their proofs rely explicitly on this assumption. In Chapter 3, we build further on this work and generalize the result of Elwany et al. (2011); their control limit policy can be viewed as the upper control limit of our established bandwidth policy. The intuition behind this is as follows. Since we do not impose a non-negativity assumption on the drift parameter of the Brownian motion, the Brownian motion can go negative, leading to a bandwidth policy that also prescribes what action to take whenever the Brownian motion takes on negative values. Incorporating the deviation on the negative side is *imperative* for modeling practical applications where it is the deviation from preferred operating

conditions that causes the failure; for instance, when the component's vibrations or internal temperature should be centered around a target value. In such settings, the Brownian motion models the deviation from the target value and then it is important to take into account the negative side. In Chapter 3, we are the first that establish structural results for CBM optimization characterized by parameter uncertainty in which it is the absolute deviation from the perfect operational state that causes a failure.

Since Elwany et al. (2011), some researchers have extended their idea and proved the optimality of control limit policies for other deterioration processes including the inverse Gaussian process (Chen et al., 2015) and the Gamma process (Zhang et al., 2016). In Chapter 2, we contribute to this area by considering a component in which the deterioration can be modeled by a compound Poisson process where there is *double parameter* uncertainty (i.e., of both the arrival process and compounding distribution). In this chapter, we consider the case where the deterioration is not only measured during planned downtime but continuously in real time. This allows us to learn not only the drift of the deterioration process (like the works mentioned above) but also *higher order properties*, in particular the volatility.

All the existing literature on CBM optimization with parameter uncertainty focuses exclusively on single components. In Chapter 5, we are the first to study the problem of optimally maintaining multiple systems with a common, unknown parameter over a finite horizon. When considering multiple systems, one can pool all data – stemming from all systems – together to jointly learn the unknown parameter on-the-fly as data becomes available. We show that this leads to substantial savings compared to not pooling any data for those systems and learning the unknown parameter independently from the other systems.

Population uncertainty

The previously mentioned works are focused on the parameter uncertainty setting and hence assume a certain parametric deterioration process of which the parameters are initially unknown but are inferred using Bayesian learning. This is also the approach that we take in this thesis, though there are also other approaches. A different modeling approach that researchers have used in the literature deals with the population uncertainty setting. Two recent examples include Van Oosterom et al. (2017) and Abdul-Malak et al. (2019). Van Oosterom et al. (2017) model deterioration as a Markov chain with an a-priori unknown transition matrix. Each installed system has a transition matrix that comes from a finite set of known transition matrices. By

endowing a prior distribution on this finite number of possible transition matrices they are able, using real-time data of the deterioration, to learn the transition matrix of the currently installed system. Using this Bayesian framework, the authors establish structural properties of the optimal policy. Abdul-Malak et al. (2019) take a similar approach and endow a prior distribution on a finite number of possible, known failure distributions. They too use a Bayesian framework to learn the current failure distribution and establish structural properties of the optimal policy. In this modeling approach, the uncertainty lies in not knowing which member of a finite population with known parameters is governing the deterioration behavior.

State uncertainty

Another stream that integrates Bayesian learning with decision making for maintenance optimization is focused on systems where the deterioration itself can only be *partially* observed. Using real-time observations that are stochastically related to the underlying deterioration state, one then needs to learn the *true* deterioration state. In this approach, the prior distribution then models the belief that the component's current deterioration state takes a certain value. Kim and Makis (2013), Kim (2016) and Van Staden and Boute (2021) all consider a continuous-time Markov chain for the underlying deterioration process, while Khaleghei and Kim (2021) consider semi-Markovian processes, and they all characterize the optimal policy using the theory of partially observable MDPs (POMDPs). The optimal policy in this setting is usually a control limit policy based on the posterior probability that the current state is close to the failure state (as the true deterioration state is not observable). Throughout this thesis we do not take such an approach; we always assume that data is perfect and hence that the deterioration state is fully observable, though in Chapter 2 we numerically assess the consequences if we relax this assumption.

1.3.2 Age-based maintenance

In the classical ABM problem, introduced in Barlow and Hunter (1960), a decision maker determines the optimal age threshold to preventively replace a single-component system subject to random failures to avoid high costs and/or low reliability associated with corrective replacements. The key assumption in this canonical ABM problem, and in many of its variations (we refer to De Jonge and Scarf (2020) for a recent, comprehensive overview of the area), is that the lifetime distribution is a-priori fully determined and known to the decision maker. When one relaxes this assumption, and the parameter of a lifetime distribution needs to be inferred, we

can again differentiate between the parameter uncertainty setting and the population uncertainty setting. As an age of an installed component is always observable, there is no analog in the ABM literature for the state uncertainty setting.

Parameter uncertainty

As described in the previous section, the concept of integrating Bayesian learning in optimal decision making has recently gained momentum in the literature of CBM. A crucial assumption in this literature, when incorporating Bayesian learning to learn the parameters of the deterioration process, is that all observations of the deterioration are uncensored. By contrast, in the study of ABM policies, the data accumulation process consists, by definition of an ABM policy, of both *censored* (i.e., preventive maintenance) and *uncensored* (i.e., corrective maintenance) observations of the underlying lifetime distribution.

We remark that these uncensored and censored observations should not be confused with full and partial observability, respectively, that researchers deal with in the state uncertainty setting in the CBM literature. In an ABM setting, a decision maker has *full knowledge* regarding the age of a currently installed component, yet when she replaces the component before failure, she only gets a censored observation of the true lifetime (the lifetime at failure). In a CBM setting, partial observability implies that a decision maker does not have full knowledge about the current deterioration of a component, but that she only gets observations that are stochastically related to the actual deterioration state. From a Bayesian perspective, the latter is much easier to analyze than the former. In fact, if the state-observation matrix is fully known, Bayesian learning can be done in a *tractable* way (see, e.g., Kim and Makis, 2013; Kim, 2016; Van Staden and Boute, 2021). By contrast, when data (partly) consists of censored observations, most lifetime distribution families immediately *lose* their conjugate property leading to tractability issues.

As a result, only very few papers focused on ABM have studied the integration of Bayesian learning in optimal decision making. Two important studies are Fox (1967) and Dayanik and Gürler (2002), who both consider a sequence of components subject to ABM with Weibull distributed lifetimes. In such a sequence, a component is repeatedly installed, and a decision maker needs to choose for each component an age threshold to preventively replace it. Fox (1967) assumes that only the scale parameter is unknown (i.e., scale parameter uncertainty), and formulates a Bayesian dynamic program (DP) to analyze the optimal policy of an infinite sequence of ABM problems. Due to the infinite sequence assumption, the author is not able to

investigate the so-called *exploration-exploitation trade-off* that is inherent to a finite sequence. In a finite sequence, a decision maker might explore with early components by deliberately using high age thresholds to acquire valuable, uncensored observations of the lifetime distribution, which can then be exploited as she nears the end of the sequence. This is commonly referred to as the exploration-exploitation trade-off. Nonetheless, Fox (1967) shows that the Bayesian DP converges to the corresponding DP of the setting in which the true scale parameter is known. The Bayesian DP is, however, computationally intractable and therefore difficult to implement in practice. Dayanik and Gürler (2002) therefore propose a myopic Bayesian policy that, at least numerically, performs close to the setting in which there is full knowledge of the unknown parameters. The authors do not, however, establish whether the learning and/or decision making of this myopic policy converge to the setting with full knowledge.

In Chapter 6, we focus on this gap in the literature. That is, we investigate the optimal policy for a finite sequence of components governed by an ABM policy, where there is an inherent exploration-exploitation trade-off. Additionally, we analyze asymptotic properties of a myopic policy in case of an infinite sequence of components. Throughout this chapter, we assume a particular class of distributions, the so-called Newsboy distributions, of which the Weibull distribution (assumed in Fox (1967) and Dayanik and Gürler (2002)) is a member, thereby making a direct link with the work in these two papers.

Population uncertainty

One other direction in ABM that integrates Bayesian learning with decision making can be found in De Jonge et al. (2015) and Dursun et al. (2022). They too consider an ABM policy to preventively replace components before failure, but in contrast to our work in Chapter 6 and the aforementioned papers, they adopt the population uncertainty approach to model uncertainty. They assume that components that are used for replacement come from either, what they call, a weak or a strong population. For both populations, the lifetime distributions and their parameters are known; both works assume a Weibull distribution for the two populations (with different, yet known parameters). However, the true population type is unknown to the decision maker and needs to be inferred from the observations. Dursun et al. (2022) build a POMDP and establish structural properties of the optimal policy, while De Jonge et al. (2015) consider a heuristic and numerically show that it performs quite well. Their approach can also be applied in CBM; see the discussion of Van Oosterom et al. (2017) and Abdul-Malak et al. (2019) in the previous section. As already discussed, in such an

approach, there is uncertainty with respect to the population, which contrasts with modeling uncertainty in the parameters of the underlying lifetime distribution.

1.3.3 Intersection with process control

Maintenance optimization models, like the ones discussed in the previous two sections, are developed to ensure *availability* of systems. When systems are up and running, they are usually responsible for production processes (for instance: lithography machines produce chips, MRI systems produce medical images) where it is of crucial importance for the *output quality* (e.g., quality of chips, medical image quality) that the system operates in close to *perfect* operating conditions. Process control (PC) of the production process is the most pivotal function to do so and has received considerable attention by researchers in the past.

The main objective in papers on PC (we refer the reader to Tagaras (1998) for an excellent overview) is to sequentially decide based on data samples of the system whether a production process has shifted from an in-control state to an out-of-control state associated with decreased quality, and if so, to inspect and potentially restore the production process. Duncan (1956) pioneered the area by proposing an $\bar{X}$-chart that uses the sample mean to detect such a shift in quality and subsequently restore the production process. Other charts have been proposed in the literature that use different measures, such as the fraction of defects in an *np*-chart (Chiu, 1976). Taylor (1965) and Taylor (1967) showed that such non-Bayesian approaches are in fact sub-optimal and that the decision should be based on the posterior probability that the process has shifted to an out-of-control state.

The focus in the area therefore shifted to the development of Bayesian control charts (see e.g., Tagaras, 1994, 1996; Calabrese, 1995; Porteus and Angelus, 1997; Makis, 2008). In Chapter 3, we also adopt a Bayesian approach, but instead of inferring the out-of-control probability, we directly learn the unknown drift of a system to move away from perfect operating conditions. Additionally, both Bayesian and non-Bayesian approaches typically assume only two operational states, whereas in Chapter 3, we assume a continuous set. Tagaras and Nikolaidis (2002) compare different Bayesian control charts and show that the sampling interval has the most positive impact on the economic performance, which is a rationale behind the continuous monitoring framework assumed in this chapter.

Despite the large amount of research in the area of both CBM and PC in isolation, there are relatively few studies focused on their joint optimization, and *none* of them study structural properties. This is rather surprising since PC and CBM are

conceivably related, especially if they utilize the same real-time data of the system to base their decision on (e.g., vibrations, temperature). The few papers that jointly consider PC and CBM all assume: (i) a discrete set of operational states, (ii) a predefined policy for when to perform maintenance based on an $\bar{X}$-chart for process control activities, and (iii) an $\bar{X}$-chart for inference of the operational state. By contrast, in Chapter 3, we assume a continuous set of operational states and like discussed before, we assume that there is parameter uncertainty surrounding the unknown drift and learn it in a Bayesian framework. Moreover, instead of limiting the analysis to a specific policy or a specific class of policies, we characterize the structural form of the optimal joint real-time data-driven PC and CBM policy.

Tagaras (1988) studies a piece of manufacturing equipment used for production processes with one failed state and several operational states that is subject to Markovian deterioration. An $\bar{X}$-chart is used to learn the operational state, while preventive maintenance is performed based on the age of the equipment; the age is thus the condition. Using reward-renewal theory, Tagaras (1988) computes the optimal parameters of the control chart and the optimal preventive maintenance age. Rahim and Banerjee (1993), Ben-Daya and Rahim (2000) Lee and Rahim (2001), and Linderman et al. (2005), all study a similar setting, but assume non-Markovian deterioration with two operational states, while the preventive maintenance policy is also based on the equipment's age.

Panagiotidou and Tagaras (2010) are the first, and to the best of our knowledge the only ones, who consider joint optimization of CBM and PC where the decision making is based on an actual condition – and not the age – of the equipment. Like us, they assume a single process characteristic for both PC and CBM purposes, but they propose a policy that relies on an $\bar{X}$-chart to initiate inspection. If the inspection reveals that the process is out-of-control, preventive maintenance takes place, otherwise the operation is continued. They provide numerical expressions for the average cost rate of the proposed policy, which are used to compute the optimal parameters of the $\bar{X}$-chart.

1.3.4 Intersection with condition-based production

Degradation behavior of components in advanced production systems is often affected by *adjustable settings* of the system itself (e.g., the production rate). The *natural* relationship between the two is that these settings have a direct impact on the deterioration rate, implying that the deterioration rate of the system can be controlled by adjusting the settings.

This novel idea was first proposed by Uit het Broek et al. (2020) who coined the term *condition-based production*. The authors introduce the use of condition monitoring to dynamically adjust the production rate over a finite planning horizon that ends with a maintenance moment. A higher production rate leads to more revenue, but also to a faster deterioration rate and possibly a corrective maintenance at the end of the planning horizon. The authors characterize the optimal condition-based production policy under the assumption that deterioration is *deterministic* and its parameters fully *known*. When deterioration is stochastic, the authors numerically explore the optimal policy and show that it exhibits similar behavior as in the deterministic case. In an extensive numerical study, they show that for such production-dependent deterioration, condition-based production rate decisions may result in significant revenue increases compared to policies that neglect this natural relationship. In Uit het Broek et al. (2021), the authors build further on the work of Uit het Broek et al. (2020) and combine the concept of condition-based production with CBM. The model considered in Uit het Broek et al. (2020) only allows maintenance at prespecified moments, and one can only influence the deterioration process in between by adjusting the production rate, while in Uit het Broek et al. (2021) it is allowed to perform CBM in between. The latter paper numerically evaluates a dynamic policy and shows that making both the production and maintenance schedule condition-based, can lead to high revenue increases compared to an isolated approach.

In Chapter 4, we generalize the model of Uit het Broek et al. (2020) considerably. We first relax the assumption that deterioration is deterministic by considering a *stochastic* deterioration process. For this model, when the parameter of the deterioration process is known, we establish structural properties of the optimal condition-based setting policy that are actually in line with the intuition built in the numerical exploration of Uit het Broek et al. (2020). We then relax the assumption that the parameter of the deterioration process is known. Specifically, we propose a Bayesian framework to tractably learn (under any production policy) the parameter of the deterioration process when it is a-priori *unknown*; that is, when there is parameter uncertainty.

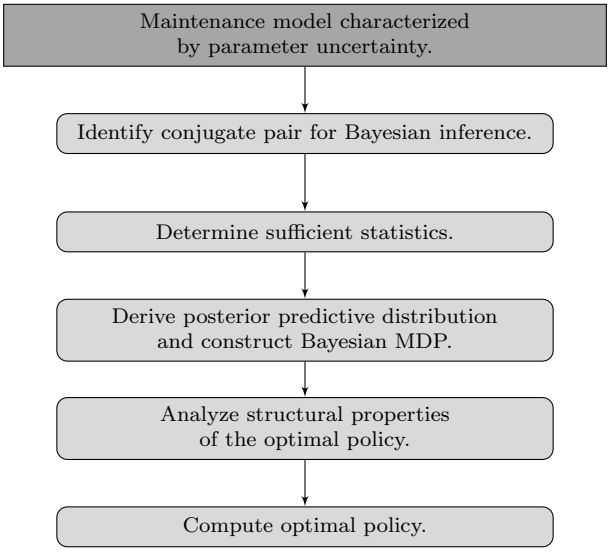
1.4. Main contributions and outline

The overarching topic of this thesis is the development of a general theory for a class of maintenance problems characterized by parameter uncertainty. For this class of problems, we develop mathematical models that integrate learning from real-time data with decision making for scenarios in which components' deterioration processes

or lifetime distributions are characterized by an a-priori unknown parameter. This approach, the structured learning and decision making approach, is able to leverage the opportunities that arise due to the ongoing developments in sensor technology and the IoT.

In each individual chapter, we consider stylized, yet representative models to analyze the essential maintenance trade-off: costly premature preventive maintenance versus costly tardy corrective maintenance in different scenarios that are prevalent in modern, data-driven manufacturing industries. Moreover, all scenarios (except for the model discussed in the last chapter) share one commonality: population heterogeneity through parameter uncertainty. Although some work has been done in that area (see previous section), we aim to provide a unified mathematical framework to analyze and tackle these problems based on our main methodology: Bayesian learning, MDPs, and their integration. Figure 1.4 provides a schematic representation of this structured learning and decision making framework.

Figure 1.4 Schematic representation of structured learning and decision making framework.



The unified framework provides a structured road map to develop mathematical models for the class of maintenance problems characterized by parameter uncertainty. The first step is to identify the conjugate pair to which the uncertain parameter belongs. This choice is largely determined by the underlying stochastic model assumed in the maintenance model. The next step is to determine the sufficient statistics required for the Bayesian inference of the uncertain parameter. This is usually a

function of the observed data, but it can also depend on certain actions that have impacted the observed data (e.g., when deterioration is controlled through production control). As a result, the sufficient statistics will also indicate what data is required and how often data should be sampled to effectively implement the Bayesian learning. Once you have determined the sufficient statistics, the next step is to derive the posterior predictive distribution and construct a Bayesian MDP encoded on (i) the sufficient statistics and (ii) the state(s) needed for maintenance purposes (e.g., the age or deterioration level) with the posterior predictive distribution governing the transition probabilities. In constructing this Bayesian MDP, one needs to take into account certain characteristics of the maintenance problem such as the length of the planning horizon (finite versus infinite) and the optimality criterion (discounted versus average cost). Using the developed Bayesian MDP, one is able to analyze its structural properties. For CBM, the optimal policy is usually a control limit that depends on the sufficient statistics. The nature of the dependence itself (e.g., monotonicity) can be proven by analyzing stochastic ordering properties of the posterior predictive distributions with respect to the sufficient statistics combined with an inductive argument on the value iteration algorithm. Insights into the structural properties of the optimal maintenance policy are not only interesting from a theoretical point of view, but they also aid the development of methods for computing the optimal policy, which is the last step of our framework. The resulting optimal policy indeed integrates learning and decision making and can thus utilize real-time data for better maintenance decision making.

In this thesis, we demonstrate the wide applicability of the proposed framework displayed in Figure 1.4 by adopting it for the analysis of scenarios that vary significantly in terms of modeling characteristics commonly used in the maintenance literature; see Table 1.1 for a classification of the model in each individual chapter based on these characteristics.

The first distinction is based on the maintenance policy under study in each chapter (see Section 1.1 for an explanation). Next, implementing such maintenance policies can be done by performing maintenance interventions at instances in time that are scheduled well in advance, or at unscheduled instances in time. The latter situation arises naturally in opportunistic maintenance, where if a component of a system fails and needs corrective maintenance, then this constitutes an unscheduled opportunity to do preventive maintenance on other components that have not yet failed. Most of the developed models in this thesis assume that when performing maintenance, the repair is perfect and the newly installed component is as-good-as-new, but we also model the situation that a repair does not restore the component to as-good-as-new. Next, when modeling the deterioration process, we distinguish three states

Table 1.1 Navigating the thesis by characteristics of developed model.

	Chapter					
	2	3	4	5	6	7
Maintenance policy						
Age-based					x	
Failure-based						x
Condition-based	x	x	x	x		x
Maintenance moments						
Scheduled	x	x	x	x	x	x
Unscheduled						x
Repair						
Perfect	x	x	x	x	x	
Imperfect						x
Deterioration state space (CBM)						
Discrete (3 states)						x
Discrete (>3 states)	x		x	x		
Continuous		x				
Time horizon						
Finite			x	x	x	
Infinite	x	x			x	x
Monitoring (CBM)						
Discrete		x		x		
Continuous	x		x			x
Interventions						
Discrete	x	x		x		
Continuous			x		x	x
Population						
Homogeneous			x			x
Heterogeneous	x	x	x	x	x	

(perfect, satisfactory and failed), a discrete state space with more than three states, and a continuous state space. The next classification is based on the length of the planning horizon, which can be based either on an infinite time horizon or finite time horizon. In obtaining condition data regarding the installed component, there are two types of monitoring that can be distinguished: monitoring by performing inspections at discrete time epochs, or continuously monitoring in real time. We further make a distinction between models in which a decision maker can only interfere with the system at equally spaced discrete decision epochs and models in which she can continuously interfere with the system. Finally, the majority of the developed models in this thesis are characterized by population heterogeneity in the components through parameter uncertainty, though we also look at homogeneous populations.

In studying these scenarios, we follow our structured approach and use various operations research techniques to tractably model and analyze each problem. The methodologies used to model each scenario in each individual chapter are summarized in Table 1.2. The ultimate goal, when analyzing each problem, is to characterize structural properties of the optimal policy. In establishing these structural properties, we use various proof techniques such as stochastic ordering, induction, contradiction, asymptotic convergence, and use various well-known theorems.

Table 1.2 Navigating the thesis by methodology.

	Chapter					
	2	3	4	5	6	7
Main methodology						
Discrete-time Markov decision process	x	x		x		
Semi-Markov decision process						x
Continuous-time Markov decision process			x			
Dynamic programming					x	
Bayesian inference	x	x	x	x	x	
Renewal theory		x				x
Simulation	x	x	x			

We now provide a brief overview of the contributions of each individual chapter in more detail, while we provide an outline of the thesis.

In **Chapter 2**, we consider a system whose condition deteriorates according to a compound Poisson degradation with a fairly general compounding distribution (viz., a member of the non-negative exponential family), where the parameters of both the Poisson process and the compounding distribution vary from one system to the next. We propose a Bayesian framework to learn these unknown parameters independently from each other and integrate this into an MDP. By doing so, we are the first that consider double parameter uncertainty: Uncertainty with respect to two different parameters which need to be inferred *simultaneously* from the *same* real-time data. We hereby fill a gap in the literature where all works are characterized by parameter uncertainty with respect to only a single parameter. By establishing new stochastic ordering properties, we establish the optimality of a control limit policy that depends on the entire deterioration path of the individual system. Another contribution in this chapter lies in a case study performed on real-life data from interventional X-ray machines to assess the value of integration of learning and decision making. We show that this approach leads to cost reductions of up to 11% relative to approaches that do not learn from real-time data and up to 4% relative to approaches that separate learning and decision making. This chapter is based on Drent et al. (2020a).

In **Chapter 3**, we address the *joint optimization* of PC and CBM. The primary

contributions of this chapter lie in (i) proposing and analyzing a novel real-time data-driven PC and CBM model, and in (ii) characterizing the structural form of the optimal policy under standard cost and operating assumptions. We are the first that rigorously look at the intersection of two related, yet often independently studied streams in the literature. We establish the optimality of an intuitive and easy-to-implement bandwidth policy that is monotone in the age of the production process, under both the discounted cost and the average cost criterion. With this bandwidth policy, the production process is continued if the process is within a bandwidth that is described by both an upper and a lower control limit. This bandwidth policy thus has characteristics both of control charts often used in the PC literature and of control limit policies typically established in the CBM literature. We again build a Bayesian MDP like in the previous chapter, yet this Bayesian MDP suffers from non-monotonic properties that render conventional proof techniques to establish structural properties not applicable. We overcome this challenge by translating the original Bayesian MDP into an alternative Bayesian MDP which *does* allow for establishing structural properties of the optimal policy. We first consider the discounting cost criterion and we then show that the optimal average cost criterion policy can be obtained as a limit of the discounted cost model when the discount factor approaches 1 from below. This is generally speaking highly nontrivial and often an onerous task that heavily depends on the context. Another important contribution of this chapter is that for archetypal replacement problems, we show that there is actually a *straightforward* road map to execute this onerous task. This chapter is based on Drent and Kapodistria (2021).

Chapter 4 centers around the novel idea of condition-based production to control deterioration when planned maintenance moments are fixed. In this chapter, we are the first to establish structural properties of the optimal condition-based setting policy (the production rate is essentially a setting) under *stochastic* deterioration. We thereby fill a substantial gap as previous works have only characterized structural properties for *deterministic* deterioration. Specifically, we consider a model in which a decision maker can continuously adjust settings which lead to an instantaneous change in both the revenue rate and the rate of a Poisson process. By employing such a condition-based setting policy, the decision maker can steer the deterioration as she nears the planned maintenance moment, at which a maintenance cost depending on the deterioration state is incurred. We model this problem as a continuous-time MDP and rigorously analyze both its Hamilton-Jacobi-Bellman equations and their discretized equivalents to characterize the monotonic behavior of the optimal policy. We also show that under optimal setting decisions, the length of the interval between planned maintenance moments can be easily optimized, thereby addressing the trade-

off that arises due to maintenance costs at such moments. We complement these theoretical results by an extensive numerical study in which we demonstrate that: (i) condition-based setting policies can lead to substantial profit increases relative to static setting policies, and (ii) integrating maintenance and setting policies can lead to significant profit increases compared to treating them sequentially (i.e., first deciding upon an interval length, and then implementing the setting policy). The intuitive behavior of the optimal condition-based setting policy – decrease the setting if close to the failure threshold and increase the setting if the planned maintenance moment is approaching – helps us to develop an *easy-to-implement* Bayesian heuristic for the case that there is parameter uncertainty regarding the base rate, that mimics this behavior. We provide two new technical results needed for this heuristic: (i) a Bayesian inference framework that is tractable under *any* setting policy, and (ii) new submodularity preservation properties that can also be useful in other application areas. In an extensive simulation study we show that this heuristic performs close to a clairvoyant Oracle policy. This chapter is based on Drent and Arts (2022).

In **Chapter 5**, we are the first that study optimal CBM policies for multiple systems, where data stemming from each individual system can be *pooled* to *jointly learn* an unknown parameter. We assume that each system has a critical component whose condition deteriorates according to a Poisson process with a common, yet unknown rate, and formulate this problem as a finite-horizon Bayesian MDP. The formulation suffers from the well-known *curse of dimensionality*: The cardinality of both the action and state space grow exponentially in the number of systems N. As a remedy, we prove a new decomposition result that establishes the equivalence between the original, high-dimensional MDP and N two-state MDPs with a binary action space, each focused on an individual system. We further show that the structure of the optimal policy of each individual system has a control limit structure. The decomposition result is *imperative* as it allows us to numerically analyze the benefits of data pooling for learning when N is relatively large. In a comprehensive numerical study, we investigate the savings that can be attained by pooling data while optimally maintaining the set of systems. We find that the savings can be significant, even for small values of N, but that the exact magnitude of these savings largely depends on the magnitude of the uncertainty in the parameter. This chapter is based on Drent and Van Houtum (2022).

In **Chapter 6**, we revisit the *canonical* ABM model proposed by Barlow and Hunter (1960), but assume that the systems lifetime distributions are parametrized by an a-priori unknown parameter that needs to be learned. The optimal policy for this relaxation has been an *open problem* since the introduction of its known counterpart in 1960. This is mainly because the Bayesian inference in this scenario is

complex: data is both *censored* (in case of a preventive maintenance) and *uncensored* (in case of a corrective maintenance). We overcome this challenge by adopting a certain family of lifetime distributions that is often used in inventory theory that *can* accommodate censored learning. We first analyze the optimal policy for a finite sequence of replacements. We formulate this problem as a Bayesian stochastic DP and by establishing new stochastic orders, we are able to establish some properties. This DP is unfortunately not tractable for practical purposes; hence we propose an *easy-to-implement* myopic Bayesian policy. Using well-known inequalities, the law of large numbers and epi-convergence results, we show that this myopic policy has *attractive* asymptotic properties: It almost surely learns the unknown parameter and *converges* to the optimal policy with *full knowledge* of the parameter. This chapter is based on Drent et al. (2020b).

The models considered in all previous chapters are characterized by parameter uncertainty. In **Chapter 7**, we model the deterioration behavior of systems as stochastic processes of which the parameters can be fully determined by historical data, and thus assume that the decision maker has no uncertainty regarding parameters. We analyze a scenario with two types of preventive maintenance: Planned maintenance at periodic, scheduled opportunities, and opportunistic maintenance at unscheduled opportunities. The latter type of maintenance arises in application areas with a network dimension (e.g., wind turbines in a wind farm), where if a system in the network fails (e.g., due to FBM), this constitutes an opportunity for preventive maintenance for the other systems. The structure of the optimal policy for this problem has been a long-standing *open question*, mainly because integrating the two types of maintenance results in decision epochs occurring at random points in time so that the theory of standard MDPs does not apply anymore. We overcome this by formulating the problem as a semi-MDP, which allows us to characterize the structure of the optimal policy under the discounted cost criterion. This policy, depending on the deterioration of the component and the remaining time until the next planned maintenance, indicates when it is optimal to perform preventive maintenance at both scheduled and unscheduled opportunities. We then show that the discounted cost policy is *also* optimal under the average cost criterion. The existing theory, and hence the usual approach to do this, fails to apply in our situation. We overcome this challenge by *extending* the theory of semi-MDPs such that it does apply to our scenario. This chapter is based on Drent et al. (2019).

Finally, in the last chapter, **Chapter 8**, we summarize our results by revisiting the applicability of the proposed framework (see Figure 1.4) in each chapter focused on parameter uncertainty. We also outline directions for future research.

Chapter 2

Condition-based maintenance with double parameter uncertainty

2.1. Introduction

In the introductory chapter, we already discussed that asset managers who are responsible for the up-time of advanced technical systems, continuously seek to *minimize* the risk of *failure* and *unplanned downtime*. Failures of these systems are usually caused by deterioration from components exceeding a certain critical level. In this chapter, we study a situation in which such deterioration is the result of random amounts of damage (e.g., wear, fatigue) that accumulate through shocks that occur randomly over time. For instance, certain metal and ceramic components in trains and aircraft only degrade at events at which they are subjected to shocks (e.g., propagation of cracks in a brake pad due to braking or in a propeller due to a heavy wind gust, respectively) rather than in a continuous fashion. For such degradation processes, it is natural to model them as so-called jump processes, i.e., stochastic processes that have discrete movements at random times (cf. Sobczyk, 1987; Singpurwalla, 1995).

This chapter is based on Drent et al. (2020a).

In this chapter, we study a condition-based maintenance (CBM) model for a single component whose condition deteriorates according to a jump process often assumed in the CBM literature (cf. Van Noortwijk, 2009). Specifically, we assume that the sequence of random shocks arrives as a Poisson process with a randomly varying shock size, that is, degradation is modeled as a compound Poisson process. The inter-arrival times between two consecutive shocks are thus Exponentially distributed, and we model the random amount of damage by a fairly general class of distribution which we introduce later (see Section 2.2.1).

Components are subject to compound Poisson degradation, but the individual parameters of both the Poisson process as well as the compounding distribution *vary* from one component to the next. That is, the population of components is heterogeneous with respect to two parameters resulting in *double parameter uncertainty*. These two parameters cannot be observed directly and they therefore need to be learned by observing the degradation signal that is relayed in real time through sensors on the component. Although we observe the degradation level of a component *continuously* through condition monitoring, we can only interfere with the system at scheduled moments occurring at equally spaced discrete epochs.

Learning parameters from real-time sensory data has been done before, but only in settings where measurements are only possible at planned downtimes; see Elwany et al. (2011), Kim and Makis (2013), Chen et al. (2015), and Van Oosterom et al. (2017). When the condition of a system can only be measured at planned downtimes, the amount of information that can be learned from the degradation level is limited compared to the situation with real-time data. Accordingly, attention in previous literature is restricted to population heterogeneity within a finite set of possibilities or heterogeneity in degradation drift only, with drift defined as the expected degradation increment per unit of time. By contrast, this chapter uses the entire degradation path of each component. This allows us to infer *higher-order properties* of the degradation behavior of the individual component, in particular the volatility, defined as the variance of a degradation increment per unit time.

The costs to replace a component after failure is much higher than before failure because they include the costs of unplanned downtime, and the decision maker is interested in minimizing the total expected discounted cost of corrective and preventive replacements over an infinite horizon. Learning the parameters from real-time data implies that the entire past degradation path of a component is relevant state information, which can lead to tractability issues. We circumvent these issues by using conjugate prior pairs to model the heterogeneity of the component population, allowing us to tractably build an infinite-horizon Bayesian MDP to study the decision

problem. We further collapse the state space by identifying structure in the prior to posterior updating procedure. This enables us to tractably compute optimal policies as well as prove structural results about optimal policies.

2.1.1 Contributions

This chapter makes the following contributions:

1. We tractably model the situation where components are heterogeneous in their degradation processes with respect to two parameters by using conjugate prior pairs. We collapse the high-dimensional state space to only 3 dimensions while retaining all relevant information. This collapse gives insight into how all relevant information in a real-time degradation signal can be parsimoniously represented. Furthermore, this collapse makes the model both tractable numerically and amenable to structural analysis.

2. We characterize the optimal replacement policy as a control limit policy where the control limit is non-decreasing in the age of a component and additionally depends on the volatility of the observed degradation signal.

3. In our first simulation study, we study (i) the benefits of explicitly modeling heterogeneity, (ii) the benefits of integrating learning with decision making, and (iii) the impact of the amount of available historical degradation data for estimation of the population heterogeneity on their performance. The results of this simulation study indicate that the integration of learning and decision making leads to excellent results with gaps of only 0.6% on average relative to an Oracle that knows the true population heterogeneity. By contrast, ignoring heterogeneity altogether leads to average gaps of 15% relative to an Oracle that knows the true population heterogeneity. Failing to integrate learning with decision making leads to average gaps of 7% relative to that same Oracle. Furthermore, we show that models that ignore population heterogeneity do *not* perform appreciably better when the amount of historical degradation data for model calibration increases.

4. In our second simulation study, we assess the performance of the optimal policy (under real-time, perfect data) when applied in a setting where (i) the degradation signal is not perfect, and (ii) the degradation signal is not relayed in real time but only at planned downtimes. The results indicate that having access to data in real time is valuable, while at the same time, this data need not be perfect to achieve excellent performance.

5. We demonstrate the efficacy of integrated learning and decision making on a real data set of X-ray tube degradation in an interventional X-ray machine. We find that integrated learning can save around 10% compared to approaches without learning and around 4% compared to an approach where learning is separated from decision making.

2.1.2 Organization

The remainder of this chapter is organized as follows. We first present the model formulation in Section 2.2. In Section 2.3, we characterize the optimal replacement policy. We then report on the results of a comprehensive simulation study in Section 2.4. We discuss the application of our approach to alternate settings in which the degradation signal is imperfect or not relayed in real time in Section 2.5. We establish the practical value of our model in Section 2.6, where we discuss a real life case study on the X-ray tube degradation in an interventional X-ray machine. Finally, we provide concluding remarks in Section 2.7.

2.2. Model formulation

This section describes the degradation model and the integrated learning problem of learning the degradation behavior of a component and deciding when to replace it.

2.2.1 Compound Poisson degradation

We consider a component that degrades as random shocks arrive. Shocks arrive as a Poisson process and the damage that accumulates during a shock is random variable, i.e., degradation is a compound Poisson process. The Poisson intensity of shock arrivals is denoted by λ. The compounding distribution is quite general; the only restriction we impose is that this distribution belongs to the one-parameter, non-negative exponential family. This family includes many well-known distributions such as the Geometric distribution and the Poisson distribution. We let the one parameter be denoted by $\phi \in \mathbb{R}_+$, where $\mathbb{R}_+$ denotes the non-negative real line. Hence, the probability density or mass function of this random amount can be expressed in the form

$$f(x|\phi) = h(x)e^{\phi T(x) - A(\phi)}, \tag{2.1}$$

where $T(x)$ is the sufficient statistic, and $h(x)$ and $A(\phi)$ are known functions. We assume that $T(x) \triangleq x$, which enables a state space collapse in our optimization problem (see Section 2.2.3). In the literature, this family of distributions is often referred to as the linear (due to the linear sufficient statistic) exponential family or natural exponential family and was first introduced by Morris (1982). This class encompasses many well-known distributions used in maintenance such as the Geometric distribution, the Poisson distribution, the Gamma distribution with known shape parameter, and the Binomial distribution with known number of trials (see Morris (1982) for a complete overview). The following examples illustrate how the Geometric distribution (with support $\mathbb{N}_0 \triangleq \{0, 1, \ldots\}$) with unknown success probability $p \in (0, 1)$ and the Poisson distribution with unknown mean $\mu > 0$ can be expressed in the canonical form of the natural exponential family.

Example 2.1 (Geometric distribution)**.** Let the damages be Geometrically distributed with (unknown) success probability $p \in (0, 1)$. The probability mass function of the random amount of damage, denoted by $f(x|p)$, then takes the form

$$f(x|p) = (1-p)^x p = e^{x \ln(1-p) - \ln(1/p)}. \tag{2.2}$$

Comparing Equation (2.2) with Equation (2.1), we find that $h(x) = 1$, $T(x) = x$, $\phi = \ln(1-p)$, and $A(\phi) = \ln(1/p) = \ln(1/(1-e^{\phi}))$ for the Geometric distribution. $\diamond$

Example 2.2 (Poisson distribution)**.** Let the damages be Poisson random variables with (unknown) mean $\mu > 0$. The probability mass function of the random amount of damage, denoted by $f(x|\mu)$, then takes the form

$$f(x|\mu) = \frac{\mu^x e^{-\mu}}{x!} = \frac{1}{x!} e^{\ln(\mu)x - \mu}. \tag{2.3}$$

Comparing Equation (2.3) with Equation (2.1), we find that $h(x) = \frac{1}{x!}$, $T(x) = x$, $\phi = \ln(\mu)$, and $A(\phi) = \mu = e^{\phi}$ for the Poisson distribution. $\diamond$

For simplicity and due to its practical appeal (see Section 2.6), throughout this chapter, we use the Geometric distribution with support $\mathbb{N}_0$ to illustrate further results – building further on Example 2.1 – but we emphasize that all structural results hold for *any* compounding distribution whose probability density or mass function can be expressed in the form displayed in Equation (2.1) with $T(x) = x$.

The degradation level is observed continuously, but it is only possible to interfere with the system at equally spaced decision epochs. These decision epochs correspond to planned downtimes. For convenience, we rescale time such that the time between two decision epochs equals 1. Furthermore, there exists a threshold $\xi \in \mathbb{N}_+$, where

$\mathbb{N}_+ \triangleq \{1, 2, \ldots\}$, such that a component has failed if its degradation is equal to or exceeds ξ.

Let $N_{(0,t]} \equiv N_t$ denote the total number of shocks received by a component from the start of its life (i.e., from the installation of the component) up to its age t. The number of shocks that arrive between age $t-1$ and t ($t \in \mathbb{N}_+$) is denoted by $K_{(t-1,t]} \triangleq N_t - N_{t-1}$. Observe that integer ages of components coincide with decision epochs. Moreover, let Y_i denote the damage incurred at the $i-$th shock since the installation of the component. Letting $X_0 = N_0 = 0$, the compound Poisson process at component age $t \in \mathbb{N}_+$ satisfies

$$X_t = \sum_{i=1}^{N_t} Y_i = X_{t-1} + \sum_{i=N_{t-1}+1}^{N_t} Y_i, \quad t \in \mathbb{N}_+,$$

with $\sum_{i=1}^{0} \cdot \equiv 0$. We also use the notation $\boldsymbol{Y}_t \triangleq \left(Y_{N_{t-1}+1}, Y_{N_{t-1}+2}, \ldots, Y_{N_t}\right)$ and $Z_{(t-1,t]} \triangleq \sum_{i=N_{t-1}+1}^{N_t} Y_i$ throughout this chapter.

2.2.2 Learning the degradation model

We assume that each component stems from a heterogeneous population of components in which each component has different degradation parameters λ and ϕ, which are unknown to the decision maker. Hence, the degradation parameters differ from one component to the next. This reflects the fact that the degradation process may be affected by the individual component's endogenous conditions. We treat the parameters λ and ϕ as random variables, denoted by Λ and Φ, which can be inferred with increasing accuracy by observing the degradation signal of the component in a Bayesian manner.

Λ has a Gamma distribution with shape α_0 and scale β_0 (i.e. $\Lambda \sim \text{Gamma}(\alpha_0, \beta_0)$) and Φ is distributed according to the general prior for a member of the exponential family (parametrized by a_0 and b_0), with $\alpha_0, \beta_0, a_0, b_0 > 0$, with prior density distribution denoted by $f_\Lambda(\lambda|\alpha_0, \beta_0)$ and $f_\Phi(\phi|a_0, b_0)$, respectively. We refer to α_0, β_0, a_0, and b_0 as the hyperparameters. Upon the installation of a new component, the parameters of the compound Poisson degradation process, λ and ϕ, are drawn from these distributions. Let k_t denote the observed number of shocks a component has sustained between ages $t-1$ and t, i.e., k_t is the realization of $K_{(t-1,t]}$. Furthermore let $\boldsymbol{y}_t \triangleq \left(y_t^1 \; y_t^2 \; \ldots \; y_t^{k_t}\right)$ be the array of the observed amounts of damage of the shocks sustained, i.e., $\boldsymbol{y}_t$ is the realization of $\boldsymbol{Y}_t$. Finally, let $z_t \triangleq \sum_{i=1}^{k_t} y_t^i$, be the sustained damage between ages $t-1$ and t, i.e., z_t is the realization of $Z_{(t-1,t]}$. The

tuple $\boldsymbol{\theta}_t \triangleq (k_t, \boldsymbol{y}_t)$ is then the observed degradation signal of a component between ages $t-1$ and t.

The sequential Bayesian updating procedure – also referred to as prior-to-posterior updating (Ghosh et al., 2007) – works as follows. When a new component is installed, there is no observed degradation signal accumulated yet, and hence Λ and Φ follow independent prior distributions, respectively. This joint prior density distribution, denoted by $f^0_{\Lambda,\Phi}(\lambda,\phi) \triangleq f_\Lambda(\lambda|\alpha_0,\beta_0)\cdot f_\Phi(\phi|a_0,b_0)$, may be obtained from historical or testing data. (In Section 2.B of the Appendix we discuss an appropriate estimation procedure.) At component age t, we use the observed degradation signal $\boldsymbol{\theta}_t$ and the joint posterior density distribution of Λ and Φ updated at component age $t-1$, say $f^{t-1}_{\Lambda,\Phi}(\lambda,\phi) \triangleq f_{\Lambda,\Phi}(\lambda,\phi|\boldsymbol{\theta}_0,\ldots,\boldsymbol{\theta}_{t-1})$, to derive the newly updated joint posterior distribution of Λ and Φ, denoted by $f^t_{\Lambda,\Phi}(\lambda,\phi) \triangleq f_{\Lambda,\Phi}(\lambda,\phi|\boldsymbol{\theta}_0,\ldots,\boldsymbol{\theta}_t)$.

For tractability purposes, so-called conjugate pairs, which have the appealing computational property that the posterior is in the same family as the prior, are often of interest in prior-to-posterior updating. It is well-known that the Gamma distribution is a conjugate prior distribution for the Poisson distribution and that a member of the exponential family has a conjugate prior whose density can be expressed in the form (cf. Ghosh et al., 2007)

$$f_\Phi(\phi|a_t,b_t) = H(a_t,b_t)e^{a_t\phi - b_t A(\phi)}.$$

However, since we infer the joint distribution of Λ and Φ using the same observed degradation signal, it is not evident which form the joint posterior distribution of Λ and Φ takes. Proposition 2.1 shows that this joint posterior distribution at component age t can be decomposed into two independent distributions of the same form with updated parameters that only depend on the information obtained in the last period ($\boldsymbol{\theta}_t$).

Proposition 2.1. *Given the last observed degradation signal at component age t, $\boldsymbol{\theta}_t = (k_t, \boldsymbol{y}_t)$, and the joint prior distribution $f^{t-1}_{\Lambda,\Phi}(\lambda,\phi) = f_\Lambda(\lambda|\alpha_{t-1},\beta_{t-1})\cdot f_\Phi(\phi|a_{t-1},b_{t-1})$, the joint posterior distribution, $f^t_{\Lambda,\Phi}(\lambda,\phi)$, is equal to $f_\Lambda(\lambda|\alpha_{t-1}+k_t,\beta_{t-1}+1)\cdot f_\Phi(\phi|a_{t-1}+z_t,b_{t-1}+k_t)$.*

Proof: The joint posterior distribution of Λ and Φ at component age t is proportional to the product of the joint likelihood function and the joint prior distributions on Λ and Φ at component age $t-1$. The joint likelihood of observing

$\boldsymbol{\theta}_t$ given (λ, ϕ), denoted by $\mathcal{L}(\boldsymbol{\theta}_t|\lambda,\phi)$, is equal to

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}_t|\lambda,\phi) &\triangleq \mathbb{P}\Big[K_{(t-1,t]} = k_t, \boldsymbol{Y}_t = \boldsymbol{y}_t|\Lambda=\lambda,\Phi=\phi\Big] \\ &= \frac{\lambda^{k_t}e^{-\lambda}}{k_t!}\prod_{i=1}^{k_t}\Big[h(y_t^i)e^{\phi y_t^i - A(\phi)}\Big] \\ &= \frac{\lambda^{k_t}e^{-\lambda}}{k_t!}e^{\phi\sum_{i=1}^{k_t}y_t^i - k_t\cdot A(\phi)}\prod_{i=1}^{k_t}h(y_t^i).\end{aligned}$$

where $\prod_{i=1}^{0}\cdot \equiv 1$.

This yields

$$\begin{aligned}&f_{\Lambda,\Phi}(\lambda,\phi|\boldsymbol{\theta}_0,\ldots,\boldsymbol{\theta}_t) \\ &\propto \mathcal{L}(\boldsymbol{\theta}_t|\lambda,\phi)\cdot f_\Lambda(\lambda|\alpha_{t-1},\beta_{t-1})\cdot f_\Phi(\phi|a_{t-1},b_{t-1}) \\ &= \frac{\lambda^{k_t}e^{-\lambda}}{k_t!}e^{\phi\sum_{i=1}^{k_t}y_t^i - k_t\cdot A(\phi)}\prod_{i=1}^{k_t}h(y_t^i)\cdot\frac{\beta_{t-1}^{\alpha_{t-1}}\lambda^{\alpha_{t-1}-1}e^{-\beta_{t-1}\lambda}}{\Gamma(\alpha_{t-1})}\times \\ &\qquad\qquad H(a_{t-1},b_{t-1})e^{a_{t-1}\phi - b_{t-1}A(\phi)} \\ &\propto \lambda^{\alpha_{t-1}+k_t-1}e^{-(\beta_{t-1}+1)\lambda}H(a_{t-1},b_{t-1})e^{\left(a_{t-1}+\sum_{i=1}^{k_t}y_t^i\right)\phi - \left(b_{t-1}+k_t\right)A(\phi)},\end{aligned}$$

which is after normalization (over hyperparameters) equal to

$$\begin{aligned}&\frac{(\beta_{t-1}+1)^{(\alpha_{t-1}+k_t)}\lambda^{\alpha_{t-1}+k_t-1}e^{-(\beta_{t-1}+1)\lambda}}{\Gamma(\alpha_{t-1}+k_t)}\times \\ &\qquad H(a_{t-1}+\sum_{i=1}^{k_t}y_t^i, b_{t-1}+k_t)e^{\left(a_{t-1}+\sum_{i=1}^{k_t}y_t^i\right)\phi - \left(b_{t-1}+k_t\right)A(\phi)}, \qquad (2.4)\end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function and $H(a_{t-1}+\sum_{i=1}^{k_t}y_t^i, b_{t-1}+k_t)$ is the new normalization factor with the updated hyperparameters. Observe that the joint posterior distribution of Λ and Φ in Equation (2.4) is equal to the product of a Gamma$(\alpha_{t-1}+k_t, \beta_{t-1}+1)$ distribution and the general prior with updated hyperparameters $(a_{t-1}+\sum_{i=1}^{k_t}y_t^i, b_{t-1}+k_t)$ for a member of the exponential family, which completes the proof. □

Proposition 2.1 induces a simple scheme to infer the true parameters of the degradation process of a component with increasing accuracy. The following example illustrates how the parameter of the compounding distribution can be inferred in the

case of Geometrically distributed damages.

Example 2.3 (Continuation of Example 2.1)**.** We endow a prior on the canonical parameter ϕ with density

$$f_\Phi(\phi|a_t,b_t) = H(a_t,b_t)e^{b_t\phi - a_t A(\phi)} = H(a_t,b_t)e^{\phi b_t}(1-e^\phi)^{a_t},$$

or equivalently, parametrized in terms of p using $p = 1-e^\phi$,

$$f_P(p|a_t,b_t) = H(a_t,b_t)p^{a_t}(1-p)^{b_t},$$

in which we recognize, after normalization, the Beta distribution with shape parameter $a_t - 1$ and scale parameter $b_t - 1$. Note also that $a_t = a_0 + \sum_{i=1}^t z_i$ and $b_t = b_0 + \sum_{i=1}^t k_i$. ◇

We now determine the posterior predictive distribution at component age t of the random variable $Z_{(t,t+1]}$ given the learned information contained in α_t, β_t, a_t, and b_t.

Lemma 2.1. *The posterior predictive distribution at component age t of the random variable $Z_{(t,t+1]}$ given the joint posterior distribution of Λ and Φ, $f_\Lambda(\lambda|\alpha_t,\beta_t)\cdot f_\Phi(\phi|a_t,b_t)$, is equal to:*

$$\mathbb{P}\Big[Z_{(t,t+1]} = z|\alpha_t,\beta_t,a_t,b_t\Big] = \sum_{k=0}^{\infty}\int_{-\infty}^{+\infty} f^{(k)}(z|\,\Phi=\phi)f_\Phi(\phi|a_t,b_t)\mathrm{d}\phi\binom{k+\alpha_t-1}{k}\left(\frac{1}{\beta_t+1}\right)^k\left(\frac{\beta_t}{\beta_t+1}\right)^{\alpha_t},$$

where $f^{(k)}(z|\,\Phi=\phi)$ denotes the k-fold convolution of the probability density (or mass) function of the random variable $\{Y|\,\Phi=\phi\}$.

Proof: We first consider the posterior predictive distribution conditioned on $K_{(t,t+1]}$,

$$\mathbb{P}\Big[Z_{(t,t+1]} = z\Big|K_{(t,t+1]} = k,\alpha_t,\beta_t,a_t,b_t\Big] = \int_0^\infty f^{(k)}(z|\,\Phi=\phi)f_\Phi(\phi|a_t,b_t)\mathrm{d}\phi, \quad (2.5)$$

where $f^{(k)}(z|\,\Phi=\phi)$ denotes the k-fold convolution of probability density (or mass) function of the random variable $\{Y|\,\Phi=\phi\}$. The distribution of $K_{(t,t+1]}$ is a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate follows a Gamma(α_t,β_t) distribution, which is known to be the Negative

Binomial distribution with $p = \frac{1}{\beta_t+1}$ and $r = \alpha_t$. Hence, we have

$$\mathbb{P}\Big[K_{(t,t+1]} = k \Big| \alpha_t, \beta_t\Big] = \binom{k+\alpha_t-1}{k} \left(\frac{1}{\beta_t+1}\right)^k \left(\frac{\beta_t}{\beta_t+1}\right)^{\alpha_t}. \tag{2.6}$$

Unconditioning Equation (2.5) using Equation (2.6) yields the desired result. □

Lemma 2.1 can be used to construct an updated posterior predictive distribution at each component's age t of the next observed damage increment in real time based on the observed degradation signal. Hence, the posterior distribution of the degradation parameters of the system is a Markov process whose evolution is induced by the degradation trajectory of the current component.

At first sight, the posterior predictive distribution in Lemma 2.1 seems rather intractable due to the convolution term involved. Fortunately, members of the natural exponential family with a linear sufficient statistic are closed under convolution with itself and hence possess a tractable form (Morris, 1982). Upon insertion of the expression for this convolution term and the corresponding conjugate prior, the posterior predictive distribution reduces to a closed-form expression that can be used for computational purposes. This is illustrated in the example below.

Example 2.4 (Continuation of Example 2.3)**.** In this example we use the parametrization in terms of unknown success probability p, which we treat as a random variable denoted by P. Due to the discrete nature of the Geometric distribution and as $p \in (0,1)$, we have by Lemma 2.1 that

$$\begin{aligned}
\mathbb{P}\Big[Z_{(t,t+1]} = z \Big| K_{(t,t+1]} =& k, \alpha_t, \beta_t, a_t, b_t\Big] \\
&= \int_0^1 \mathbb{P}\Big[\sum\nolimits_{i=1}^k Y_i = z \Big| P = p\Big] f_P(p|a_t, b_t)\mathrm{d}p \\
&= \frac{1}{\mathrm{B}(a_t,b_t)} \binom{z+k-1}{z} \int_0^1 p^k (1-p)^z p^{a_t-1}(1-p)^{b_t-1}\mathrm{d}p \\
&= \frac{\mathrm{B}(k+a_t, z+b_t)}{\mathrm{B}(a_t,b_t)} \binom{z+k-1}{z},
\end{aligned} \tag{2.7}$$

where $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t$ is the Beta function. Note that the distribution of $K_{(t,t+1]}$ is a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate follows a Gamma(α_t, β_t) distribution, which is known to be the Negative Binomial distribution with success probability $q = \frac{1}{\beta_t+1}$ and $r = \alpha_t$ number

of required successes. Hence, we have

$$\mathbb{P}\Big[K_{(t,t+1]} = k | \alpha_t, \beta_t\Big] = \binom{k+\alpha_t-1}{k} \left(\frac{1}{\beta_t+1}\right)^k \left(\frac{\beta_t}{\beta_t+1}\right)^{\alpha_t}. \tag{2.8}$$

Unconditioning Equation (2.7) using Equation (2.8) yields the closed form expression of the posterior predictive distribution:

$$\begin{aligned}&\mathbb{P}\Big[Z_{(t,t+1]} = z | \alpha_t, \beta_t, a_t, b_t\Big] \\ &\quad = \sum_{k=0}^{\infty} \frac{\mathrm{B}(k+a_t, z+b_t)}{\mathrm{B}(a_t, b_t)} \binom{z+k-1}{z} \binom{k+\alpha_t-1}{k} \left(\frac{1}{\beta_t+1}\right)^k \left(\frac{\beta_t}{\beta_t+1}\right)^{\alpha_t}.\end{aligned}$$

◇

2.2.3 Markov decision process formulation

Each component will incur a cost due to either corrective or preventive replacement. If the degradation level at a decision epoch is greater than or equal to the failure threshold ξ, then the failed component is replaced correctively at cost c_u. If the degradation level at a decision epoch does not exceed ξ, then we can either perform preventive replacement at cost c_p, or continue to the next decision epoch at no cost. We assume that replacements take negligible time and that $0 < c_p < c_u < \infty$ to avoid trivial cases.

Recall that each component stems from a heterogeneous population that includes components with different degradation parameters λ and ϕ. Note that these parameters cannot be observed; only the degradation signal $\boldsymbol{\theta}_i$ for $i = 1, \cdots, t$ is observable at component age t. We will therefore use a Bayesian MDP to model the integrated problem of learning the degradation parameters of a component and deciding when to replace.

We first observe that due to the results in the previous section, the information state of a component at age t can be represented by $(\alpha_t, \beta_t, a_t, b_t)$. Furthermore, the decision maker knows the current degradation level x. The state at decision epoch τ is therefore given by $(x, \alpha, \beta, a, b, \tau) \in \mathbb{N}_0^6$ where x denotes the degradation level of the component that is in service (note that when the compounding distribution is continuous, then $x \in \mathbb{R}_{\geq 0}$) and α, β, a, and b encode the most current degradation information of the component that is in service. This six dimensional state representation can be collapsed into an equivalent four dimensional state representation (x, n, t, τ) where n denotes the number of shocks that the current component has sustained and t denotes

its age. Indeed observe that by Proposition 2.1 we have $\alpha = \alpha_0 + n$, $\beta = \beta_0 + t$, $a = a_0 + x$, and $b = b_0 + n$. This representation is insightful: All the information in the degradation signal is encoded in the total degradation level, the number of shocks sustained, and the age of the component. The crucial assumption for this collapse is that the sufficient statistic $T(x)$ for the damage distribution is equal to x, see Equation (2.1). That is, all the relevant information for prior-to-posterior updating is contained in only the total sum of the accumulated damage and the number of load arrivals. If the sufficient statistic would not be linear in the damage, then the state space would need to include all individual damage arrivals.

We are interested in finding the optimal replacement policy $\pi^* = \{\pi_\tau\}_{\tau \in \mathbb{N}_0}$ that minimizes the total expected discounted cost of corrective and preventive replacements over an infinite horizon, where the costs are discounted by a factor $\gamma \in (0, 1)$. This policy is a sequence of decision rules that prescribe whether or not to perform preventive maintenance if $x < \xi$. Since (i) the action space is finite, (ii) the costs per decision epoch are uniformly bounded, and (iii) $\gamma \in (0, 1)$, we know by Proposition 1.2.3 of Bertsekas (2007), that there exists an optimal Markov policy depending only on state (x, n, t) independent of the decision epoch $\tau \in \mathbb{N}_0$. Let

$$V(\boldsymbol{s}) = \inf_{\pi \in \Pi} \lim_{T \to \infty} \mathbb{E}_{\boldsymbol{s}} \left[\sum_{\tau=1}^{T} \gamma^\tau C\left(\boldsymbol{S}_\tau, \pi(\boldsymbol{S}_\tau)\right) \right]$$

denote the expected total discounted cost given that the process starts in state $\boldsymbol{s} = (x, n, t) \in \mathbb{N}_0^3$ where $\boldsymbol{S}_\tau$ denotes the state of the component operating at decision epoch τ, Π denotes the set of Markov policies, $\mathbb{E}_{\boldsymbol{s}}$ is the conditional expectation given that the process starts in state $\boldsymbol{s} = (x, n, t)$, and $C(\boldsymbol{s}, \pi(\boldsymbol{s}))$ denotes the cost function defined as

$$C(\boldsymbol{s}, \pi(\boldsymbol{s})) = \begin{cases} c_p & \text{if } x < \xi \text{ and } \pi(\boldsymbol{s}) = \text{replace}, \\ 0 & \text{if } x < \xi \text{ and } \pi(\boldsymbol{s}) = \text{continue}, \\ c_u & \text{if } x \geq \xi. \end{cases}$$

Due to this state space collapse, the posterior distribution of $Z_{(t,t+1]}$ is a function of the state $\boldsymbol{s} = (x, n, t)$. Therefore we will use the shorthand notation $Z(\boldsymbol{s})$ to indicate the random variable $Z_{(t,t+1]}$ given the state $\boldsymbol{s} = (x, n, t)$. Similarly, we use the shorthand notation $K(\boldsymbol{s})$ and $\{Y(\boldsymbol{s})\}_{i \in \mathbb{N}}$ for the random variable $K_{(t,t+1]}$ and the random variables $\{Y_i\}_{i \in \mathbb{N}}$, given state $\boldsymbol{s} = (x, n, t)$. Note that $Z(\boldsymbol{s})$ and $K(\boldsymbol{s})$ are dependent random variables. This is intuitively clear as damage can only accumulate when the component sustains shocks. It is convenient to define the random vector $A(\boldsymbol{s}) = (Z(\boldsymbol{s}), K(\boldsymbol{s}), 1)$, where the probability distributions of each element of this random vector can be determined with Proposition 2.1 and Lemma 2.1. Because (i)

the action space is finite, (ii) the costs per decision epoch are uniformly bounded, and (iii) $\gamma \in (0,1)$, we know that the optimal replacement policy π^* satisfies the following Bellman optimality equations (Bertsekas, 2007, Proposition 1.2.2):

$$V(\boldsymbol{s}) = \begin{cases} c_u + \gamma \mathbb{E}\left[V(\boldsymbol{s_0} + A(\boldsymbol{s}_0))\right], & \text{if } x \geq \xi, \\ \min\Big\{c_p + \gamma \mathbb{E}\left[V(\boldsymbol{s}_0 + A(\boldsymbol{s}_0))\right]; \gamma \mathbb{E}_{\boldsymbol{s}}\left[V(\boldsymbol{s} + A(\boldsymbol{s}))\right]\Big\}, & \text{if } x < \xi, \end{cases} \tag{2.9}$$

where $\boldsymbol{s}_0 \triangleq (0,0,0)$.

The first case in Equation (2.9) follows because failed components must be replaced correctively at cost c_u. If the component's degradation level is less than ξ, we can either perform a preventive replacement, which costs c_p, or leave the component in operation until the next decision epoch at no cost. Upon preventive or corrective replacement, the parameters of the new component are unknown and need to be learned as the replacement component ages. In the next section, we analyze the value function defined in (2.9) and characterize the structure of the optimal replacement policy.

2.3. Optimal replacement policy

This section presents structural results on the optimal replacement policy. First we present two examples of optimal replacement policies that illustrate several structural properties. Motivated by practice (see Section 2.6), we use the Geometric compounding distribution (with support $\mathbb{N}_0$) in these illustrations.

2.3.1 Examples of optimal replacement policies

Figure 2.1 illustrates the optimal replacement policy for two different degradation sample paths. In these examples, the solid black line depicts the observed degradation path as the component ages, where a star denotes a shock arrival. The dashed line depicts the optimal control limit, where at each decision epoch, the optimal action is to carry out a preventive replacement if the degradation level is at or above the optimal control limit. The failure level ξ is depicted with the dot-dashed line, so that a corrective replacement has to be performed if the solid black line is at or above this dot-dashed line.

We can state three properties that we consistently observe, which are best explained by the illustrations in Figure 2.1.

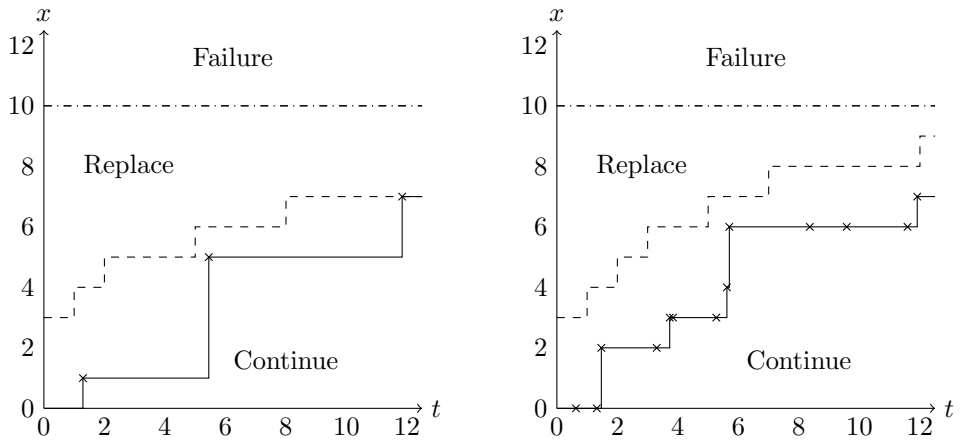


Figure 2.1 Two sample degradation paths and the optimal replacement policy, with $\alpha_0 = \beta_0 = a_0 = b_0 = 1$, $c_u = 4$, $c_p = 1$ and $\xi = 10$. For the compound Poisson processes that generate the sample paths, we have $\lambda = 0.25$ and $\phi = 0.2$ (left), and $\lambda = 1.5$ and $\phi = 0.8$ (right).

1. The optimal replacement policy is a control limit policy where the control limit depends on the past degradation signal.

2. The optimal control limit is non-decreasing in the age of the component, at least for a fixed number of shock arrivals n.

3. If we compare two components whose degradation is identical at age t, then the component with fewer shock arrivals has a lower control limit. This is illustrated in Figure 2.1, where both components have accrued the same cumulative degradation at age $t = 12$, yet the optimal policy prescribes to preventively replace the component on the left, but not the component on the right. Observe that the degradation path on the left may be described as more volatile.

We remark that the monotonic behavior with respect to the number of arrived shocks (property 3) is usually but not always observed. An example where this property is violated is shown in Figure 2.2.

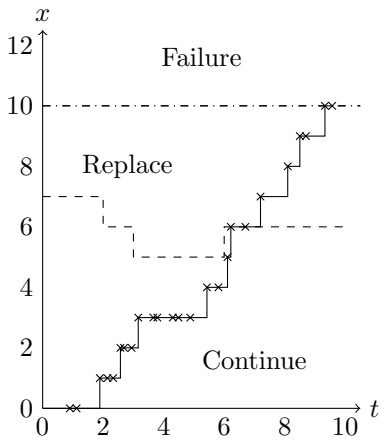


Figure 2.2 Sample degradation path and the optimal replacement policy, with $\alpha_0 = \beta_0 = 1$ and $a_0 = 60$ and $b_0 = 10$, $c_u = 4$, $c_p = 1$ and $\xi = 10$. For the compound Poisson process that generates the sample path, we have $\lambda = 1.75$ and $\phi = 0.9$.

2.3.2 Structural properties

In this section, we establish both the optimality of a control limit policy, as well as the monotonic structure of this control limit (see observations above). Although we illustrated the structural properties with the Geometric distribution as the compounding distribution in the previous section, the results in this section generalize to the complete class of distributions that we consider (see Section 2.2.1).

Lemma 2.1 shows how the decision maker can update the posterior predictive distribution of the upcoming degradation increment by utilizing the observed degradation signal $\boldsymbol{s}$. Proposition 2.2 presents two properties that provide insight into what the decision maker knows about the distribution of future degradation increments. In this result, we use the following definition quantifying the concept of one random variable being "larger" than another random variable:

Definition 2.1 (1.A.1 Definition, Shaked and Shanthikumar, 2007)**.** A random variable X is stochastically larger than a random variable Y in the usual stochastic order, denoted by $X \geq_{\text{st}} Y$, if and only if

$$\mathbb{P}[X \geq x] \geq \mathbb{P}[Y \geq x], \quad \text{for all } x \in \mathbb{R}.$$

Definition 2.1 implies that random variable Y is less likely than random variable X

to take on large values, where large means any value greater than x, and that this holds for all $x \in \mathbb{R}$.

Proposition 2.2 presents two stochastic ordering properties for the random variable $Z(\boldsymbol{s})$.

Proposition 2.2. *The random variable $Z(\boldsymbol{s})$ satisfies the following stochastic orders:*

(i) $Z(\boldsymbol{s})$ is stochastically decreasing in t in the usual stochastic order;

(ii) $Z(\boldsymbol{s})$ is stochastically increasing in x in the usual stochastic order.

PROOF: In proving the result, we use another stochastic order; the likelihood ratio order.

Definition 2.2 (1.C.1 Definition, Shaked and Shanthikumar, 2007)**.** Let X and Y be continuous (discrete) random variables with probability densities f and g, respectively, such that

$$\frac{g(t)}{f(t)} \text{ increases in } t \text{ over the union of the supports of } X \text{ and } Y$$

(here $a/0$ is taken to be equal to ∞ whenever $a > 0$), or, equivalently,

$$f(x)g(y) \geq f(y)g(x), \quad \text{for all } x \leq y.$$

Then X is said to be smaller than Y in the likelihood ratio order (denoted by $X \leq_{\text{lr}} Y$). Note that the likelihood ratio order is stronger than the usual stochastic order, as such, if $X \leq_{\text{lr}} Y$ then $X \leq_{\text{st}} Y$, cf. (Shaked and Shanthikumar, 2007, Theorem 1.C.1).

We now proceed with the proofs of parts (i) and (ii) of Proposition 2.2:

(i) We consider two different component ages, t^+ and t^- $(t^+ > t^-)$, with the same degradation level x and the same total number of shocks, n, received by the component. By Proposition 2.1, we have $\beta_{t^+} > \beta_{t^-}$, whereas all other hyperparameters (i.e., α_t, a_t and b_t) remain fixed when only t changes. Note that

$$\mathbb{P}[K_{(t,t+1]} = u \,|\, \Lambda = \lambda] = e^{-\lambda}\frac{\lambda^u}{u!} \quad \text{and} \quad \mathbb{P}[\Lambda = \lambda \,|\, \alpha, \beta] = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}.$$

We can then show that the random variables $\Lambda(\alpha, \beta) = \{\Lambda \,|\, \alpha, \beta\}$ are stochastically increasing in α and stochastically decreasing in β. This easily follows from the appropriate likelihood ratio.

We now show that the random variables $\mathcal{K}(\lambda) = \{K_{(t,t+1]} \,|\, \Lambda = \lambda\}$ are stochastically increasing in λ, i.e., $\lambda \leq \lambda'$ implies $\mathcal{K}(\lambda) \leq_{\text{st}} \mathcal{K}(\lambda')$. Consider the likelihood ratio

$$\frac{\mathbb{P}[\mathcal{K}(\lambda') = u]}{\mathbb{P}[\mathcal{K}(\lambda) = u]} = \frac{\mathbb{P}[K_{(t,t+1]} = u \,|\, \Lambda = \lambda']}{\mathbb{P}[K_{(t,t+1]} = u \,|\, \Lambda = \lambda]} = e^{-(\lambda'-\lambda)} \left(\frac{\lambda'}{\lambda}\right)^u.$$

It is immediately evident that for $\lambda \leq \lambda'$, the ratio is increasing in u. This yields the likelihood ratio order and as a consequence the usual stochastic order.

All in all, for $t^+ > t^-$, $\beta_{t^+} > \beta_{t^-}$, thus $\Lambda(\alpha, \beta_{t^+}) \leq_{\text{st}} \Lambda(\alpha, \beta_{t^-})$. From the stochastic monotonicity of $\mathcal{K}(\Lambda)$, the above yields $\mathcal{K}^{(t^+)} := \mathcal{K}(\Lambda(\alpha, \beta_{t^+})) \leq_{\text{st}} \mathcal{K}(\Lambda(\alpha, \beta_{t^-})) := \mathcal{K}^{(t^-)}$, cf. (Shaked and Shanthikumar, 2007, Theorem 1.A.2.). Since the parameters of the random variables Y_i remain fixed, we have

$$Z_{(t^+,t^++1]} = \sum_{i=1}^{\mathcal{K}^{(t^+)}} Y_i \leq_{\text{st}} \sum_{i=1}^{\mathcal{K}^{(t^-)}} Y_i = Z_{(t^-,t^-+1]},$$

cf. (Shaked and Shanthikumar, 2007, Theorem 1.A.4.).

(*ii*) We consider two degradation levels x^+ and x^- ($x^+ > x^-$) at the same component age t when we have observed the same total number of shocks n. By Proposition 2.1, we have $a_{x^+} > a_{x^-}$, whereas all other hyperparameters are equivalent as they remain fixed when only x changes.

It is well-known that a member of the one-parameter exponential family is likelihood-ratio ordered (and thus stochastically ordered in the usual stochastic order) according to its parameter (see e.g., Karlin and Rubin, 1956; Bapat and Kochar, 1994). Hence, we know that (i) the random variables $\mathcal{Y}(\phi) = \{Y \,|\, \Phi = \phi\}$ are stochastically non-decreasing in ϕ, i.e., $\phi \leq \phi'$ implies $\mathcal{Y}(\phi) \leq_{\text{st}} \mathcal{Y}(\phi')$, and (ii) the random variables $\Phi(a, b) = \{\Phi \,|\, a, b\}$ are stochastically non-decreasing in a, i.e., $a \leq a'$ implies $\Phi(a, b) \leq_{\text{st}} \Phi(a', b)$. Then, by Theorem 6 of Huang and Mi (2020), which relates the stochastic order of a posterior predictive distribution with the stochastic order of the posterior and the corresponding conditional distribution, we can conclude that $\mathcal{Y}^{(x^+)} := \mathcal{Y}(\Phi(a_{x^+}, b)) \geq_{\text{st}} \mathcal{Y}(\Phi(a_{x^-}, b)) =: \mathcal{Y}^{(x^-)}$. The parameters of the random variables $K_{(t,t+1]}$ remain fixed, hence

$$Z^{(x^+)}_{(t,t+1]} := \sum_{i=1}^{K_{(t,t+1]}} \mathcal{Y}_i^{(x^+)} \geq_{\text{st}} \sum_{i=1}^{K_{(t,t+1]}} \mathcal{Y}_i^{(x^-)} =: Z^{(x^+)}_{(t,t+1]}.$$

□

Part (i) of Proposition 2.2 shows that older components will accumulate – in expectation – less damage than younger components with the same deterioration level. As a component ages without failing, the decision maker infers that large increments are unlikely to happen for this component. The intuition behind Part (ii) of Proposition 2.2 is that when more damage has accumulated already, then we should expect more damage to accumulate in the future.

There is no monotone stochastic ordering of $Z(\boldsymbol{s})$ in n in general as shown in the following example.

Example 2.5 (Continuation of Example 2.4)**.** Consider again the Geometric compounding distribution with a Beta prior. Let $b_0-\alpha_0-1>0$, then the expectation of $Z(\boldsymbol{s})$.

$$\mathbb{E}[Z(\boldsymbol{s})] = \frac{a(\alpha_0+n)}{\beta(b_0+n-1)}$$

increases in n. However, the second moment

$$\begin{aligned}\mathbb{E}[Z(\boldsymbol{s})^2] =& \frac{a(\alpha_0+n)}{\beta^2(b_0+n-1)^2}\times \\ & \left(\beta^3\left(\frac{2(a+1)}{b_0+n-2}+\beta(b_0+n)+2b_0+a-\beta+2n\right)+a(\alpha_0+n)\right)\end{aligned}$$

is not monotonically increasing. (The derivation of these two moments is provided in Appendix 2.A). For example, choosing $b_0 = 2.62$, $\alpha_0 = \beta = 1$ and $a = 10$ yields that the second moment is increasing for all $n \leq 15$ and decreasing for all $n \geq 16$. We can therefore conclude that there is no monotone stochastic ordering of the random variables $Z(\boldsymbol{s})$ in n (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). ◇

The following two properties of the value function are essential to establish the structure of the optimal replacement policy. Proposition 2.2 is pivotal to establish these two properties. As such, here too there is no monotonic behavior of the value function in n in general.

Lemma 2.2. *The value function $V(\boldsymbol{s})$ is*

(i) non-increasing in t;

(ii) non-decreasing in x.

PROOF: We prove part (i) and omit the proof of part (ii) as its proof structure follows verbatim.

For $\boldsymbol{s} = (x, n, t) \in \mathbb{N}_0^3$, let $V^m(\boldsymbol{s})$ denote the value function at the m-th iteration of the value iteration algorithm, so that the value iteration algorithm produces the sequence $\{V^m(\boldsymbol{s})\}_{m\in\mathbb{N}_0}$. We use induction on the steps of the value iteration algorithm as a proof technique. Since (i) the action space is finite, (ii) the costs per decision epoch are uniformly bounded, and (iii) $\gamma \in (0, 1)$, we know that the value function is guaranteed to converge to the optimal value function that satisfies Equation (2.9) (i.e., $V^m(\boldsymbol{s}) \to V(\boldsymbol{s})$ for all $\boldsymbol{s} \in \mathbb{N}_0^3$ as $m \to \infty$) from any arbitrary starting position through the value iteration algorithm (cf. Bertsekas, 2007, Proposition 1.2.1).

For $\boldsymbol{s} = (x, n, t) \in \mathbb{N}_0^3$, we set $V^0(\boldsymbol{s}) = 0$. Note that $V^0(\boldsymbol{s})$ is non-increasing in t. We assume that the theorem holds for the m-th iteration, i.e., $V^m(\boldsymbol{s})$ is non-increasing in t. Then according to Equation (2.9), we have

$$V^{m+1}(\boldsymbol{s}) = \begin{cases} c_u + \gamma\mathbb{E}\left[V^m(\boldsymbol{s}_0 + A(\boldsymbol{s}_0))\right], & \text{if } x \geq \xi, \\ \min\Big\{c_p + \gamma\mathbb{E}\left[V^m(\boldsymbol{s}_0 + A(\boldsymbol{s}_0))\right]; \gamma\mathbb{E}_{\boldsymbol{s}}\left[V^m(\boldsymbol{s} + A(\boldsymbol{s}))\right]\Big\}, & \text{if } x < \xi. \end{cases} \tag{2.10}$$

Since $V^m(\boldsymbol{s})$ is non-increasing in t by the induction hypothesis and the random variable $A(\boldsymbol{s})$ is stochastically decreasing in t in the usual stochastic order (cf. Proposition 2.2 and the proof therein), the expectation $\mathbb{E}_{\boldsymbol{s}}\left[V^m(\boldsymbol{s} + A(\boldsymbol{s}))\right]$ has this property as well (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Because the terms of the right-hand side of Equation (2.10) are non-increasing in t, $V^{m+1}(\boldsymbol{s})$ is also non-increasing in t. Due to the convergence in which the structure of $V^m(\boldsymbol{s})$ is preserved (cf. Bertsekas, 2007, Proposition 1.2.1), we conclude that $V(\boldsymbol{s})$ is also non-increasing in t . □

We are now in the position to present the main result of this section.

Theorem 2.1. *At each component age $t \in \mathbb{N}_0$, for a given number of shock arrivals $n \in \mathbb{N}_0$, there exists a control limit $\delta^{(n,t)}$, $0 < \delta^{(n,t)} \leq \xi$, such that the optimal action is to carry out a preventive replacement if and only if $x \geq \delta^{(n,t)}$. The control limit $\delta^{(n,t)}$ is monotonically non-decreasing in t, for all n.*

PROOF: Preventive replacement is optimal when the following equation holds

$$c_p + \gamma\mathbb{E}\left[V(A(\boldsymbol{s}_0))\right] \leq \gamma\mathbb{E}_{\boldsymbol{s}}\left[V(\boldsymbol{s} + A(\boldsymbol{s}))\right]. \tag{2.11}$$

Since the expectation $\mathbb{E}\left[V(A(\boldsymbol{s}_0))\right]$ is constant with respect to x, the left-hand side of Inequality (2.11) is constant with respect to x. Based on part (ii) of Lemma 2.2, the right-hand side of Inequality (2.11) is non-decreasing in x. Hence, if the optimal decision is to carry out preventive replacement in state $(\delta^{(n,t)}, n, t)$, then the same decision is optimal for any state (x, n, t) with $x \geq \delta^{(n,t)}$, which implies the control limit policy.

Similarly, the right-hand side of Inequality (2.11) is non-increasing in t by part (i) of Lemma 2.2. Hence, if it is optimal to carry out a preventive replacement in state $(\delta^{(n,t)}, n, t)$, then it is optimal to carry out a preventive replacement in any state $(\delta^{(n,t)}, n, t')$ with $t' < t$, which implies that $\delta^{(n,t)}$ is monotonically non-decreasing in t. □

The optimal control limit is non-decreasing in t because as the component ages without failing, the decision maker is increasingly assured that the currently installed component degrades slowly relative to the general population of components.

The structural results are not only intuitive and convenient for the implementation of an optimal policy in practice. They can also be exploited to decrease the computational burden of finding the optimal policy by employing existing algorithms that rely on these monotonicity properties such as the monotone policy iteration algorithm (see Puterman, 2005, Section 6.11.2).

Remark 2.1. The structure of the optimal policy in Theorem 2.1 is characterized under the discounted cost criterion. In Chapter 3, we show that for this type of replacement problems, it can be shown that this structure is also optimal under the average cost criterion. ◇

2.4. Simulation study

This section reports the results of a comprehensive simulation study. Although the established structural results hold for any one-parameter member of the exponential family with positive support, we assume in this section, motivated by practice, that the damages are Geometrically distributed. This simulation study starts from the premise that the true hyperparameters of the degradation behavior are unknown to the decision maker; they only have access to historical degradation data for model calibration. This premise differs from previous contributions that use Bayesian techniques to model real-time learning, where hyperparameters for prior distributions are generally assumed to be given (see, e.g., Chen, 2010). This latter approach

is, however, arguably not the case in practice. Indeed, decision makers only have access to historical degradation data that they should leverage in order to estimate hyperparameters. The main objective of this simulation study is, therefore, twofold:

1. To examine the value of integrating learning and decision making, which takes into account explicitly that degradation of components is heterogeneous (value of integration).
2. To assess how the amount of available historical degradation data that is used for model calibration affects the performance (value of data).

To assess the value of integration and data, we define three heuristic approaches, all of which start from the same historical degradation data, but differ in how they calibrate their models, learn from the degradation of components, and integrate learning with decision making. We shall compare the performance of each approach with the performance of an Oracle who does know the true hyperparameters of the degradation behavior of different components. The Oracle thus follows the replacement policy that solves the Bellman optimality equations in (2.9), calibrated with the true hyperparameters. We denote the Oracle policy by $\pi_{\mathcal{O}}$.

For each instance of our simulation study, which we describe in detail later, the true hyperparameters of the Gamma and Beta prior distribution of Λ and Φ that model the population heterogeneity are denoted by $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{a}$, and $\tilde{b}$. These are known only to the Oracle. The historical degradation data that serves as starting point for the heuristic approaches of that same instance is obtained by simulating degradation paths of components whose degradation parameters λ and ϕ are drawn from a Gamma distribution with the true hyperparameters $\tilde{\alpha}$ and $\tilde{\beta}$, and a Beta distribution with $\tilde{a}$ and $\tilde{b}$, respectively. We now proceed with defining the three heuristic approaches, after which we describe the simulation set-up and discuss the results.

2.4.1 Offline approach

The first heuristic approach ignores both the population heterogeneity and the real-time degradation signal, which is the current state-of-the-art. This approach assumes that the degradation of each component upon installation follows a compound Poisson process with the same parameters λ and ϕ. Under this assumption, the decision maker faces the classical replacement problem for which we know that the optimal replacement policy is given by a stationary control limit (e.g., Derman, 1963; Kolesar, 1966). The optimal control limit can be readily found by solving the following one

dimensional Bellman optimality equations:

$$V(x) = \begin{cases} c_u + \gamma \mathbb{E}\left[V(\tilde{Z}(\lambda,\phi))\right], & \text{if } x \geq \xi, \\ \min\left\{c_p + \gamma \mathbb{E}\left[V(\tilde{Z}(\lambda,\phi))\right]; \gamma \mathbb{E}\left[V(x+\tilde{Z}(\lambda,\phi))\right]\right\}, & \text{if } x < \xi, \end{cases} \tag{2.12}$$

where $\tilde{Z}(\lambda,\phi) \triangleq \sum_{i=1}^{K(\lambda)} Y_i(\phi)$ denotes the degradation increment in between two consecutive decision epochs, $K(\lambda)$ is a Poisson distributed random variable with parameter λ and the $Y_i(\phi)$'s are Geometrically distributed with parameter ϕ.

Hence, under the first approach, the decision maker approximates the degradation parameters by the corresponding point estimates $\bar{\lambda}$ and $\bar{\phi}$, which are obtained using Maximum Likelihood Estimation (MLE) based on the available historical degradation data, and then solves the optimality equations (2.12) with those estimates to obtain a single control limit, that is used for all components. In the remainder of this section, we refer to this approach as the offline approach because it ignores both the population heterogeneity and the real-time degradation signal. We denote this approach by $\pi_{\mathcal{N}}$ since it is the most naive heuristic approach of all three approaches.

2.4.2 Myopic online approach

The second heuristic approach does utilize the real-time degradation signal to update the point estimates of the degradation parameters of an individual component. As such, this approach takes into account that the population of components is heterogeneous. This approach is calibrated as follows. Given the historical degradation paths, we estimate the initial hyperparameters α_0, β_0, a_0, and b_0, by maximizing the likelihood of those degradation paths being induced by components stemming from a population whose heterogeneity is modeled through a Gamma and Beta distribution with these hyperparameters. Thus, based on the available historical degradation data, we estimate the initial hyperparameters α_0, β_0, a_0, and b_0, using MLE (further details regarding this MLE procedure are relegated to Appendix 2.B).

After calibration, this heuristic approach works as follows: At component age t, the decision maker updates the information state encoded in $f_\Lambda(\lambda|\alpha_t,\beta_t)$ and $f_\Phi(\phi|a_t,b_t)$ in the same way as in the original Bayesian model (cf. Proposition 2.1). The decision maker then updates the point estimates of the degradation increment based on the minimum mean square error (MMSE) estimator. In a Bayesian setting, MMSE estimates correspond to posterior means. Hence, at component age t, the MMSE estimates for the degradation parameters are given by $\bar{\lambda}_t \triangleq \int_0^\infty u f_\Lambda(u|\alpha_t,\beta_t)\mathrm{d}u =$

α_t/β_t and $\bar{\phi}_t \triangleq \int_0^1 u f_\Phi(u|a_t, b_t)\mathrm{d}u = a_t/(a_t + b_t)$. The decision maker then computes a control limit by solving the optimality equations (2.12), where the parameters of the Poisson and of the Geometric distribution of the degradation increment $\tilde{Z}(\lambda, \phi)$ are now given by $\bar{\lambda}_t$ and $\bar{\phi}_t$, respectively. Although the second approach partly captures the Bayesian learning benefits, it does not integrate learning with optimization but rather solves a myopic optimization problem repeatedly with the latest point estimates of the degradation parameters. In the remainder of this section, we therefore refer to this approach as the myopic online approach, denoted by $\pi_{\mathcal{M}}$.

2.4.3 Integrated Bayesian approach

The third heuristic approach is the most sophisticated approach. It is similar to $\pi_{\mathcal{O}}$ in that it follows the replacement policy that solves the Bellman optimality equations in (2.9). However, as we have argued before, the true hyperparameters that model the population heterogeneity are unknown to the decision maker in practice. As such, this approach differs from $\pi_{\mathcal{O}}$ in that we calibrate this approach using MLE, in the same way as how we calibrate $\pi_{\mathcal{M}}$. That is, we estimate the initial hyperparameters α_0, β_0, a_0, and b_0, using MLE based on the available historical degradation data. These estimated hyperparameters are then used as input for finding the optimal replacement policy through solving the Bellman optimality equations in (2.9). We henceforth refer to this approach as the integrated Bayesian approach, denoted by $\pi_{\mathcal{I}}$. It is important to note that this approach is precisely how our model should be applied in practice, where the hyperparameters describing the heterogeneity of the component population should be estimated from historical degradation data.

2.4.4 Results

The main performance metric in this simulation study is the gap between the long run average cost rate induced by the Oracle policy and the offline approach, the myopic online approach, and the integrated Bayesian approach, where the latter three models are calibrated based on MLE. More formally, we are interested in $\%GAP_\pi = 100 \cdot (C_\pi - C_{\pi_{\mathcal{O}}})/C_{\pi_{\mathcal{O}}}$, where C_π is the long run average cost rate of approach $\pi \in \{\pi_{\mathcal{I}}, \pi_{\mathcal{M}}, \pi_{\mathcal{N}}\}$. Here we would like to point out that although we studied the structure of the integrated Bayesian policy under the discounted cost criterion in the previous section, this structure also holds for the average cost criterion; see Remark 2.1.

To achieve the two main objectives stated in the beginning of this section, we set up a large test bed consisting of instances obtained through all combinations of the

parameter values in Table 2.1, with $\xi = 20$ and $c_p = 1$. To vary the heterogeneity in the population of components, we naturally vary the coefficient of variation of the Gamma and Beta prior distributions of Λ and Φ, respectively, while keeping their respective means fixed at 1 and 0.5 for all instances.

Table 2.1 Input parameter values for simulation study.

	Input parameter	No. of choices	Values
1	Coefficient of variation of prior Gamma distribution, cv_Λ	2	0.3, 0.6
2	Coefficient of variation of prior Beta distribution, cv_Φ	2	0.01, 0.02
3	Corrective maintenance cost, c_u	2	5, 10
4	Number of simulated degradation paths, ν	2	10, 50

For each instance of the test bed, we determine the true hyperparameters $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{a}$, and $\tilde{b}$, based on the coefficient of variation of the Gamma and Beta distribution of that instance (and their fixed means). We use these hyperparameters to compute $\pi_{\mathcal{O}}$ and we also simulate a number of degradation paths, denoted ν, corresponding to that instance. We subsequently use these simulated degradation paths to calibrate approaches $\pi_{\mathcal{I}}$, $\pi_{\mathcal{M}}$ and $\pi_{\mathcal{N}}$. That is, we use MLE to estimate α_0, β_0, a_0, and b_0, in case of $\pi_{\mathcal{I}}$ and $\pi_{\mathcal{M}}$, and point estimates $\bar{\lambda}$ and $\bar{\phi}$, in case of $\pi_{\mathcal{N}}$. We then simulate $15 \cdot 10^3$ components, where upon installation of a new component, its degradation parameters are drawn from a Gamma and Beta distribution with the true hyperparameters $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{a}$ and $\tilde{b}$. For each simulated component, we keep track of the relevant cost rates and subsequently calculate $\%GAP_\pi$ for each approach $\pi \in \{\pi_{\mathcal{I}}, \pi_{\mathcal{M}}, \pi_{\mathcal{N}}\}$. We repeat this procedure 30 times for each instance of the test bed to ensure that the confidence intervals of the cost rates are sufficiently small. Throughout this simulation study, we use a discount factor of 0.99 in solving the corresponding optimality equations of each approach, and we truncate both state variables n and t at a sufficiently large value (i.e., 40) in computing the optimal policy under both $\pi_{\mathcal{I}}$ and $\pi_{\mathcal{O}}$.

The results of the simulation study are summarized in Table 2.2. In this table, we present the minimum, average, and maximum $\%GAP_\pi$ for each approach $\pi \in \{\pi_{\mathcal{I}}, \pi_{\mathcal{M}}, \pi_{\mathcal{N}}\}$. We first distinguish between subsets of instances with the same value for a specific input parameter of Table 2.1 (first eight rows) and then present the results for all instances (last row).

The following main observations can be drawn from Table 2.2: First, $\pi_{\mathcal{I}}$ yields excellent results with a gap of only 0.60% on average relative to the Oracle. Both heuristic approaches $\pi_{\mathcal{M}}$ and $\pi_{\mathcal{N}}$ perform poorly, with gaps of 7.08% and 15.02% on average relative to the Oracle. Ignoring both the degradation signal and the heterogeneity in the population can be quite detrimental, as gaps of up to 24.04%

Table 2.2 Results of simulation study.

		$\%GAP_\pi$								
		$\pi_{\mathcal{I}}$			$\pi_{\mathcal{M}}$			$\pi_{\mathcal{N}}$		
Input	Value	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
cv_Λ	0.3	0.05	0.58	1.26	3.52	6.14	10.05	7.91	13.97	22.93
	0.6	0.15	0.63	1.34	5.24	8.02	11.43	10.16	16.08	24.04
cv_Φ	0.01	0.05	0.46	1.02	3.52	5.54	8.11	7.91	11.31	15.25
	0.02	0.25	0.75	1.34	5.83	8.62	11.43	13.85	18.73	24.04
c_u	5	0.05	0.48	1.16	3.52	5.66	7.74	7.91	12.01	15.73
	10	0.24	0.72	1.34	5.21	8.50	11.43	10.83	18.03	24.04
ν	10	0.42	0.95	1.34	3.97	7.23	11.43	8.63	15.46	24.04
	50	0.05	0.26	0.49	3.52	6.93	11.36	7.91	14.58	23.00
Total		0.05	0.60	1.34	3.52	7.08	11.43	7.91	15.02	24.04

relative to the Oracle do occur under $\pi_{\mathcal{N}}$. Although learning the degradation signal is beneficial, failing to integrate this with decision making directly can still lead to gaps with the Oracle of up to 11.43%. All three approaches $\pi_{\mathcal{I}}$, $\pi_{\mathcal{M}}$, and $\pi_{\mathcal{N}}$ seem to perform worse when the heterogeneity in the population increases, and when the cost of performing corrective maintenance becomes higher.

Second, it is generally believed that increasing the amount of data available for model calibration leads to better decisions. However, when the underlying assumptions of the model are wrong, then this may not be true. Indeed, the performance of both heuristic approaches $\pi_{\mathcal{M}}$ and $\pi_{\mathcal{N}}$ does not increase considerably in the number of simulated degradation paths that serve as input to these models. The performance of $\pi_{\mathcal{I}}$ increases however significantly when this number increases. Furthermore, even when the amount of available data for estimating the population heterogeneity is limited, the integrated Bayesian approach, $\pi_{\mathcal{I}}$, still yields excellent results.

2.5. Alternate settings

We have so far assumed that the degradation signal is relayed in real time and that it provides a perfect observation of the actual degradation level. These assumptions are in line with what we have observed at our industrial partner who instigated this research (see also the case study in the next section). We note that there may be practical settings where such assumptions are not justified. It is however intractable to relax these assumptions and consequently compute optimal policies within our current modeling framework. In this section, we therefore study the performance of our integrated Bayesian approach when it is applied to settings where the degradation

signal is imperfect or relayed periodically. We do so in a simulation study that follows the same procedure as in the previous section.

2.5.1 Imperfect degradation signal

In line with most of the research on imperfect condition monitoring (e.g., Maillart, 2006; Kim and Makis, 2013), we model an imperfect degradation signal by constructing a state-observation matrix $\mathbf{Q}$ that captures the stochastic relationship between the actual degradation level X_t and the (imperfect) observation, denoted with $\bar{X}_t$, that the decision maker observes. More formally, let $\mathbf{Q} \triangleq \left(q_{ij}\right)_{\xi\times\xi}$, whose entry $q_{ij} \triangleq \mathbb{P}\left[\bar{X}_t = j \mid X_t = i\right]$ is equal to

$$\frac{\varphi_i(j|\sigma)}{\sum_{j=0}^{\xi-1}\varphi_i(j|\sigma)}, \tag{2.13}$$

where $\varphi_x(y|\sigma)$ is the density function of a normal random variable with mean x and standard deviation σ evaluated at y. Recall that in our simulation study, we assume that the damages are integer valued, so that we have ξ non-failed states, i.e. $0, 1, \ldots, \xi-1$, that are not observable, and one failed state that is observable. The matrix $\mathbf{Q}$ applies to the ξ non-failed states. Constructing this matrix using the parameterization in Equation (2.13) is an approach often used in literature (see, e.g., Maillart, 2006; Kim and Makis, 2013; Liu et al., 2021), where the standard deviation σ is a measure of the noise (or imperfectness) of the observations when the system has not failed yet. By varying the value of σ, we can thus investigate the robustness of our approach to the level of imperfectness of the observations. Note that by setting σ equal to zero, we are in the situation that we have a perfect observation of the actual degradation level.

2.5.2 Intermittent degradation signal

When the degradation signal is relayed periodically at decision epochs, the decision maker has only access to the degradation level of the current component x_t (and its age t). However, to apply the integrated Bayesian approach we require the number of sustained loading epochs n_t too. To resolve this, we rely on the following recursive formula to estimate n_t:

$$\bar{n}_t = \frac{x_t(a_0 + x_{t-1} - 1)}{b_0 + \bar{n}_{t-1}}, \qquad t > 0,$$

and $\bar{n}_0 = 0$. We round to the nearest integer in case $\bar{n}_t$ is not integral. This estimation procedure has intuitive appeal. To obtain an estimate for n_1 at $t = 1$, we divide the degradation level x_1 by the expected damage per loading epoch given the initial hyperparameters a_0 and b_0 that are obtained through the MLE procedure for model calibration, i.e. $\frac{b_0}{(a_0-1)}$. (Observe that this is the expectation of a Geometric random variable whose parameter is Beta distributed with parameters a_0 and b_0, see Appendix 2.A for the derivation.) We then apply the updating rules of Proposition 2.1 using this estimate, that is $a_1 = a_0 + x_1$ and $b_1 = b_0 + \bar{n}_1$, and consequently apply the same logic at $t = 2$ to obtain an estimate for n_2. This procedure repeats until the component is replaced.

2.5.3 Results

The long run average cost rate of the integrated Bayesian approach applied to the imperfect degradation signal setting and the intermittent degradation signal setting are denoted $C^{\text{imp}}_{\pi_{\mathcal{I}}}$ and $C^{\text{int}}_{\pi_{\mathcal{I}}}$, respectively. We are interested in how the integrated Bayesian approach performs in these alternate settings, and we therefore compare said costs rates with the cost rate of the integrated Bayesian approach that does have access to the true degradation signal in real time. That is, we compute $\%VAL_{\text{int}} = 100 \cdot (C^{\text{int}}_{\pi_{\mathcal{I}}} - C_{\pi_{\mathcal{I}}})/C_{\pi_{\mathcal{I}}}$ and $\%VAL_{\text{imp}} = 100 \cdot (C^{\text{imp}}_{\pi_{\mathcal{I}}} - C_{\pi_{\mathcal{I}}})/C_{\pi_{\mathcal{I}}}$ for each instance of the test bed. The instances of our test bed are identical to the test bed of the previous section, see Table 2.1. In the test bed of the imperfect degradation signal, we additionally vary σ over 5 different levels, i.e. $\sigma \in \{0.25, 0.5, 0.75, 1, 1.25\}$. Note that $C_{\pi_{\mathcal{I}}}$ is a lower bound on the cost rate of the *optimal* policy for both alternate settings.

The average $\%VAL_{\text{int}}$ and average $\%VAL_{\text{imp}}$ are presented in Table 2.3 and Table 2.4, respectively. As before, we first distinguish between subsets of instances with the same value for a specific input parameter and then present the results for all instances.

From Table 2.3 we see that when the degradation signal is not relayed in real time, our integrated Bayesian approach performs relatively poorly with an $\%VAL_{\text{int}}$ of almost 9 percent on average. This also implies that relaying a degradation signal in real time rather than only periodically has considerable value. Indeed, it allows the decision maker to not only learn the drift of a degradation signal (encoded in the degradation level and the age) but also the volatility of the degradation signal; the latter can only be inferred if one has access to the individual arrivals of loading epochs and their corresponding damages in between decision epochs.

By contrast, our integrated Bayesian approach performs quite well when the

Table 2.3 Results for intermittent degradation signals.

Input parameter	Value	$\%VAL_{\text{int}}$
Coefficient of variation of prior Gamma distribution, cv_Λ	0.3	6.51
	0.6	11.40
Coefficient of variation of prior Beta distribution, cv_Φ	0.01	10.37
	0.02	7.54
Number of simulated paths, ν	10	9.89
	50	8.02
Total		8.95

Table 2.4 Results for imperfect degradation signals.

Input parameter	Value	$\%VAL_{\text{imp}}$
Coefficient of variation of prior Gamma distribution, cv_Λ	0.3	3.29
	0.6	5.95
Coefficient of variation of prior Beta distribution, cv_Φ	0.01	4.59
	0.02	4.65
Number of simulated paths, ν	10	4.66
	50	4.58
Standard deviation of the noise, σ	0.25	0.01
	0.5	1.67
	0.75	3.50
	1	6.70
	1.25	11.23
Total		4.62

degradation signal is imperfect. Indeed, from Table 2.4 we see that $\%VAL_{\text{imp}}$ is less than 5 percent on average, and even remains below 3.5 percent for moderate levels of noisiness.

2.6. Case study

Interventional X-ray (IXR) systems are used by physicians for minimally-invasive image-guided procedures to diagnose and treat diseases in nearly every organ system. X-ray tubes (denoted by the rectangle in Figure 2.3) are the most expensive replacement components of an IXR system and therefore of major concern. Philips Healthcare produces the IXR system and does maintenance and service for many

hospitals that use the IXR system.

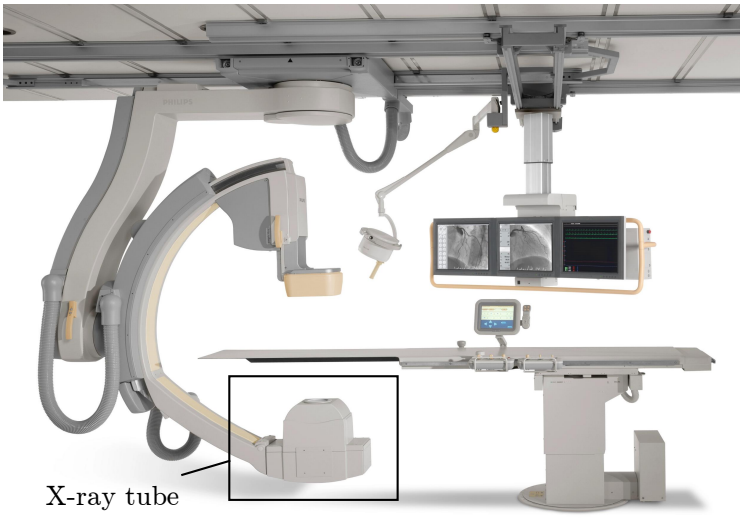


Figure 2.3 Example of IXR system with X-ray tube denoted by rectangle. (Philips, 2020)

Unexpected downtime incidents have a major impact, especially for the patients whose medical procedure is cut short or postponed. The cost of premature maintenance is also substantial. Medical imaging equipment have list-prices on the order of one million US dollars and the annual maintenance expenses of such equipment are around 10% of the list-price (ECRI, 2013). Since these medical imaging systems generally last up to 10 years, roughly half of the total cost of ownership of such a system (excluding downtime costs) consists of maintenance costs.

Philips healthcare faces the challenge of replacing the expensive X-ray tube before failures occur but also to maximize the utilization of their useful lifetime. Below we describe a case study on the most critical component of an IXR system: the X-ray tube.

This section is organized as follows: We describe the dominant failure mechanism of the X-ray tube (Section 2.6.1), give a description of the data set (Section 2.6.2), illustrate the operation of the optimal policy with real data (Section 2.6.3) and compare the three approaches outlined in Section 2.4 on real data with a bootstrapping study (Section 2.6.4).

2.6.1 IXR Filaments

A failure analysis performed by Philips Healthcare indicates that X-ray tube failures are predominantly caused by worn out filaments (Albano et al., 2019, Section 5.3.3.4). These tungsten filaments are heated to a high temperature by a voltage differential such that they emit electrons. These electrons are then accelerated by a high voltage potential differential towards the target so that they emit X-rays when they hit the target. The X-rays are then used to produce the desired image during image guided medical procedures. This process is depicted in Figure 2.4.

The tungsten evaporates slowly when the filament heats up. This filament usually develops a "hot-spot" at the thinnest location. The evaporation causes the hotspot to become thinner with every image taken. This continues until the tungsten melts at the hot-spot and the filament fails, see Covington (1973). The degradation

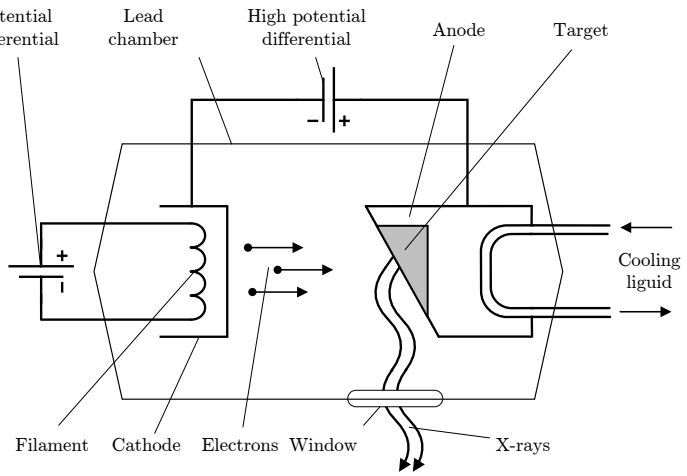


Figure 2.4 Simplified X-ray tube schematic.

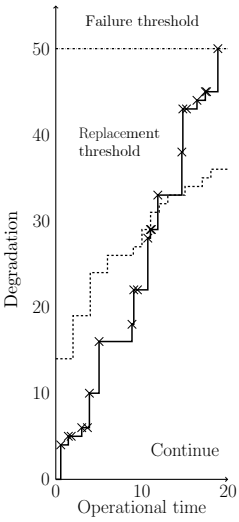
state of a filament can be inferred from the resistance of the filament. Philips Healthcare performed a statistical analysis to derive a single-dimensional health indicator that governs the degradation state over time of a filament (Albano et al., 2019, Section 7.9.7.12). This single-dimensional health indicator, the degradation, is recorded in a database each time that an IXR system is used for an image-guided procedure.

2.6.2 Degradation data of IXR X-ray tubes

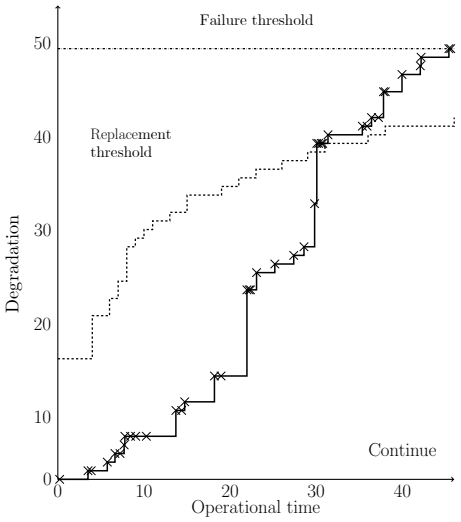
The data set of X-ray tube degradation consist of 52 time-series of degradation levels. Let $\mathcal{I}$ be the set of all X-ray tubes for which there is available data; $|\mathcal{I}| = 52$. The time-series of a single X-ray tube $i \in \mathcal{I}$ is denoted $\mathcal{J}_i$. Each datum $j \in \mathcal{J}_i$ in such a time-series is a tuple $(t_j, x_j)_i$ of the age of the X-ray tube t_j and the degradation level x_j at that age. Each tuple $(t_j, x_j)_i$ of X-ray tube i is generated when an IXR system is used and each time-series consists of 20.000-300.000 data points originating from a time period of 2-5 years. Due to confidentiality reasons we have left-truncated the data and normalized the data. That is, all time-series start with $x_0 = 0$ for $t_0 = 0$, and end at $x_{|\mathcal{J}_i|} = 50$ (i.e., $\xi = 50$) for all $i \in \mathcal{I}$. For each time-series, we computed the inter-arrival times (i.e., $t_j - t_{j-1}$) between succeeding data points and damage increments per data point (i.e., $x_j - x_{j-1}$). After removing outliers in the inter-arrival times due to either weekend or other prolonged non-operational periods and removing data points for which the image-guided procedure was considered too short to wear out the X-ray tube (i.e., shock arrival is regarded as non-critical), the assumption that shocks arrive as a Poisson process was not rejected based on the Kolmogorov-Smirnov test. Additionally, we normalized the time such that one unit of time corresponds to roughly the operational time that is considered as minimally achievable for performing maintenance practices from a practical perspective (e.g., sending a service engineer to the location of the hospital). Furthermore, based on the Akaike information criterion, the damage size distribution is within the class of the applicable probability distributions best represented by the Geometric distribution. Finally, pair-wise Kolmogorov-Smirnov tests in our data set show that the parameters of the distributions for the inter-arrival time and damage size, differed from one component to another (i.e., there is heterogeneity).

2.6.3 Illustration of optimal replacement policy

Figure 2.5 shows two examples of the optimal replacement policy, applied a-posteriori to two time-series of the filament data set. In these examples, and also throughout the rest of this section, the ratio $\frac{c_u}{c_p}$ is set to 5 based on discussions with Philips Healthcare. Furthermore, the prior values of the hyperparameters are estimated using the MLE method described in Appendix 2.B applied on the remaining time-series, which resulted in $\alpha_0 = 44.88$, $\beta_0 = 32.43$, $a_0 = 4.29$, and $b_0 = 4.76$ for this example. Figure 2.5 visually confirms that there is heterogeneity in both the shock arrival process and damage size distribution. In this example, the optimal replacement policy prescribes to preventively replace the X-ray tube at $x = 29$ and $n = 14$ at $t = 12$ (left)



(a) Replacement when $(x, n, t) = (29, 14, 12)$ with $\delta^{(14,12)} = 29$.



(b) Replacement when $(x, n, t) = (39, 28, 31)$ with $\delta^{(28,31)} = 39$.

Figure 2.5 Two examples of the optimal replacement policy applied to IXR filament degradation paths.

and $x = 39$ and $n = 28$ at $t = 31$ (right), respectively, such that the useful lifetime of the X-ray tube is best utilized. The examples illustrate how integration of learning and decision making allows for the X-ray tube to be replaced early when necessary and late when possible.

2.6.4 Bootstrapping study

The goal of this section is to illustrate the optimal replacement policy, and to assess the performance of the integrated Bayesian approach compared to both the offline and myopic online approach described in Section 2.4 on real-life data. The main performance metric is therefore the relative cost savings that can be attained by using the integrated Bayesian approach instead of the two heuristic approaches commonly used in practice and described in Section 2.4. More formally, we are interested in

$$\%SAV_\pi = 100 \cdot (C_\pi - C_{\pi_\mathcal{I}})/C_\pi,$$

where C_π is the average cost rate of approach $\pi \in \{\pi_{\mathcal{M}}, \pi_{\mathcal{N}}\}$. In addition, we are interested in the impact of the amount of available historical degradation data on the $\%SAV_\pi$ of the two heuristic approaches.

We evaluate the performance metric $\%SAV_\pi$ retroactively on the data set described in Section 2.6.2. To evaluate the impact of the amount of historical data available we bootstrap the amount of available data. That is, we first sample with replacement $s \in \{5, 10, 15\}$ time-series from the data set $\mathcal{I}$. We then use these time-series to estimate the required parameters for the integrated Bayesian approach as well as for the heuristic approaches using MLE (see Appendix 2.B for further details regarding the MLE procedure). We then implement and evaluate both the integrated Bayesian approach and the heuristic approaches on the remaining time-series that were not used for the estimation. This procedure is repeated 150 times per choice of s, such that the confidence intervals on the average cost rates are sufficiently small, resulting in 450 bootstrap instances. The resulting average values for $\%SAV_\pi$ are reported in Table 2.5. Observe that this bootstrapping study indicates the induced savings when the integrated Bayesian approach is implemented instead of the offline approach or the myopic online approach in a real-life setting.

Table 2.5 Results of bootstrapping study.

	$\%SAV_\pi$	
Number of time-series	$\pi_{\mathcal{M}}$	$\pi_{\mathcal{N}}$
5	4.68	10.46
10	4.08	10.35
15	4.07	10.70
Total	4.28	10.50

Table 2.5 shows that the integrated Bayesian approach reduces the average cost rate with 10.50% on average compared to the state-of-the-art approach. Moreover, cost savings of 4.28% can be attained by integrating the learning with the decision making instead of using a data-driven approach that does not integrate the two. Moreover, the attainable savings are not significantly influenced by the amount of available historical degradation data. Hence, the integrated Bayesian approach does not need a large amount of historical data to perform well. These results highlight the value of the proposed method to integrate learning and decision making in a real-life setting.

2.7. Conclusion

We have studied CBM of components that are subject to compound Poisson degradation, where the compounding distribution is a member of the one-parameter non-negative exponential family, with uncertainty in both degradation parameters of components. Using conjugate prior pairs to model this double parameter uncertainty, we have proposed a Bayesian MDP to analyze this decision problem. We have characterized the optimal replacement policy as a control limit policy, where the control limit is non-decreasing in the age of a component. Furthermore, we have shown that the control limit also depends on the volatility of the observed degradation signal of a component. This volatility can only be captured by observing the *entire* degradation signal in real time.

In an extensive numerical study, we have shown that the integration of learning and decision making performs close to an Oracle. By contrast, ignoring heterogeneity or failing to integrate learning with decision making performs much worse. Furthermore, we have shown that models that ignore population heterogeneity do *not* perform appreciably better when the amount of historical degradation data for model calibration increases. We have also assessed the performance of the optimal policy when applied in a setting where (i) the degradation signal is not perfect, and (ii) the degradation signal is not relayed in real time. The results indicated that having access to data in real time is valuable, while at the same time, this data need not be perfect to achieve excellent performance.

Finally, we have established the practical value of integrated learning and decision making based on a real-life data set of filament wear in an interventional X-ray machine. We find that integrated learning can save up to 10.50% compared to approaches without learning and up to 4.28% compared to an approach where learning is separated from decision making.

For further research, it would be interesting to study the following extension. In our modeling approach, we have assumed that the underlying degradation state can be observed with certainty. It would be interesting to study the case with partially observable degradation, in which case a partially observable MDP can be used to tractably model the problem and analyze its structural properties.

The cost structure assumed in this chapter only deals with the trade-off between costly failures and expensive preventive maintenance activities. One important factor that this cost structure neglects is that deterioration can also have an impact on the output of the production process that the system is used for. For instance, the quality of the medical images that the X-ray machine of our case study produces can become

worse as the filament wears, possibly to the point that the images cannot be used anymore. In the next chapter, we shall take this so-called deterioration dependent quality into account.

2.A. Deriving moments for the compound Poisson process

We first repeat some notation. Let $\boldsymbol{s} = (x, n, t) \in \mathbb{N}_0^3$ and let $Z(\boldsymbol{s}) = \sum_{i=1}^{K(\boldsymbol{s})} Y_i(\boldsymbol{s})$, where $K(\boldsymbol{s})$ is a Poisson random variable unknown rate λ, and $\{Y_i(\boldsymbol{s})\}_{i\in\mathbb{N}}$ is a sequence of independent and (identically) Geometrically distributed random variables with unknown success probability p, independent from $K(\boldsymbol{s})$. We endow p and λ with prior Beta and Gamma distributions, denoted by P and Λ, respectively. We can then find the first and second moment of $Z(\boldsymbol{s})$ by repeatedly using the tower property of expectations and the law of total variance.

We have

$$\begin{aligned}
\mathbb{E}[Y_i(\boldsymbol{s})] &= \mathbb{E}\big[\mathbb{E}[Y_i(\boldsymbol{s}) \mid P]\big] \\
&= \mathbb{E}\big[\frac{P}{1-P}\big] \\
&= \frac{\int_0^1 \frac{p}{1-p} p^{a_t-1}(1-p)^{b_t-1}\mathrm{d}p}{\mathrm{B}(a_t, b_t)} \\
&= \frac{\mathrm{B}(a_t+1, b_t-1)}{\mathrm{B}(a_t, b_t)} \\
&= \frac{a_t}{b_t-1},
\end{aligned} \tag{2.14}$$

where the fourth equality follows from manipulating the term inside the integral to the density of another Beta distribution.

Proceeding with $\mathbb{E}[K(\boldsymbol{s})]$, we have

$$\begin{aligned}
\mathbb{E}[K(\boldsymbol{s})] &= \mathbb{E}\big[\mathbb{E}[K(\boldsymbol{s}) \mid \Lambda]\big] \\
&= \mathbb{E}[\Lambda] \\
&= \frac{\alpha_t}{\beta_t}.
\end{aligned} \tag{2.15}$$

Combining (2.14) and (2.15) and using the updating rules of Proposition 2.1 for

$\boldsymbol{s} = (x, n, t)$, we have

$$\begin{aligned}\mathbb{E}[Z(\boldsymbol{s})] &= \mathbb{E}\big[\sum\nolimits_{i=1}^{K(\boldsymbol{s})} Y_i(\boldsymbol{s})\big] \\ &= \mathbb{E}[K(\boldsymbol{s})]\mathbb{E}[Y_i(\boldsymbol{s})] \\ &= \frac{a_0 + x}{b_0 + n - 1}\frac{\alpha_0 + n}{\beta_0 + t}.\end{aligned}$$

For the second moment, we first compute the variance. We have

$$\begin{aligned}\mathrm{Var}[Z(\boldsymbol{s})] &= \mathrm{Var}\big[\sum\nolimits_{i=1}^{K(\boldsymbol{s})} Y_i(\boldsymbol{s})\big] \\ &= \mathbb{E}\Big[\mathrm{Var}\Big[\sum\nolimits_{i=1}^{K(\boldsymbol{s})} Y_i(\boldsymbol{s}) \mid K(\boldsymbol{s})\Big]\Big] + \mathrm{Var}\Big[\mathbb{E}\sum\nolimits_{i=1}^{K(\boldsymbol{s})} Y_i(\boldsymbol{s}) \mid K(\boldsymbol{s})\Big]\Big] \\ &= \mathbb{E}\big[K(\boldsymbol{s})\big]\mathrm{Var}\big[Y_i(\boldsymbol{s})\big] + \mathrm{Var}\big[K(\boldsymbol{s})\big]\mathbb{E}\big[Y_i(\boldsymbol{s})\big] \\ &= \frac{\alpha_t}{\beta_t}\mathrm{Var}\big[Y_i(\boldsymbol{s})\big] + \frac{a_t}{b_t - 1}\mathrm{Var}\big[K(\boldsymbol{s})\big], \end{aligned} \tag{2.16}$$

where $\mathbb{E}\big[Y_i(\boldsymbol{s})\big]$ and $\mathbb{E}\big[K(\boldsymbol{s})\big]$ are computed in (2.14) and (2.15), respectively. We now proceed with the two variances in (2.16). For $\mathrm{Var}\big[Y_i(\boldsymbol{s})\big]$ we have

$$\begin{aligned}\mathrm{Var}\big[Y_i(\boldsymbol{s})\big] &= \mathbb{E}\Big[\mathrm{Var}\big[Y_i(\boldsymbol{s}) \mid P\big]\Big] + \mathrm{Var}\Big[\mathbb{E}\big[Y_i(\boldsymbol{s}) \mid P\big]\Big] \\ &= \mathbb{E}\Big[\frac{P}{(1-P)^2}\Big] + \mathrm{Var}\Big[\frac{P}{1-P}\Big]. \end{aligned} \tag{2.17}$$

For $\mathbb{E}\Big[\frac{P}{(1-P)^2}\Big]$, we have

$$\begin{aligned}\mathbb{E}\Big[\frac{P}{(1-P)^2}\Big] &= \frac{\int_0^1 \frac{p}{(1-p)^2} p^{a_t - 1}(1-p)^{b_t - 1}\mathrm{d}p}{\mathrm{B}(a_t, b_t)} \\ &= \frac{\mathrm{B}(a_t + 1, b_t - 2)}{\mathrm{B}(a_t, b_t)} \\ &= \frac{a_t(a_t + b_t - 1)}{(b_t - 2)(b_t - 1)}, \end{aligned} \tag{2.18}$$

where the second equality follows from manipulating the term inside the integral to

the density of another Beta distribution. Similarly, we can write

$$\begin{aligned}\mathrm{Var}\Big[\frac{P}{1-P}\Big] &= \mathbb{E}\Big[\frac{P^2}{(1-P)^2}\Big] - \mathbb{E}^2\Big[\frac{P}{(1-P)}\Big] \\ &= \frac{\mathrm{B}(a_t+2, b_t-2)}{\mathrm{B}(a_t, b_t)} - \Big(\frac{a_t}{b_t-1}\Big)^2 \\ &= \frac{a_t(a_t+1)}{(b_t-2)(b_t-1)} - \Big(\frac{a_t}{b_t-1}\Big)^2, \end{aligned} \tag{2.19}$$

which completes all the ingredients for $\mathrm{Var}\big[Y_i(\boldsymbol{s})\big]$. Using again the law of total variance, we can compute $\mathrm{Var}\big[K(\boldsymbol{s})\big]$ as follows

$$\begin{aligned}\mathrm{Var}\big[K(\boldsymbol{s})\big] &= \mathbb{E}\Big[\mathrm{Var}\big[K(\boldsymbol{s}) \mid \Lambda\big]\Big] + \mathrm{Var}\Big[\mathbb{E}\big[K(\boldsymbol{s}) \mid \Lambda\big]\Big] \\ &= \mathbb{E}\Big[\Lambda\Big] + \mathrm{Var}\Big[\Lambda\Big] \\ &= \frac{\alpha_t}{\beta_t} + \frac{\alpha_t}{\beta_t^2}. \end{aligned} \tag{2.20}$$

Plugging (2.18) and (2.19) in (2.17), and combining this with (2.20) in (2.16) gives us $\mathrm{Var}[Z(\boldsymbol{s})]$. We can then use the variance expansion $\mathrm{Var}[Z(\boldsymbol{s})] = \mathbb{E}[Z(\boldsymbol{s})^2] - \mathbb{E}^2[Z(\boldsymbol{s})]$ to obtain an explicit expression for the second moment in terms of the hyperparameters associated with state $\boldsymbol{s} = (x, n, t)$.

2.B. Maximum likelihood estimation

Let $\mathcal{I}$ denote the set containing all simulated sample paths (without maintenance). For each sample path $i \in \mathcal{I}$, we let x_i, n_i, and t_i, denote, respectively, the total damage accumulated by the component, the number of critical loading epochs sustained by the component, and the number of decision epochs that the component was operated until its failure. Thus the tuple $\boldsymbol{s}_i = (x_i, n_i, t_i)$ contains all relevant information for sample path i. Because the prior distributions of Φ and Λ are independent, we next look at their respective likelihoods given the set of simulated degradation paths $\mathcal{I}$ separately.

From Equation (2.5) we know that the likelihood of the degradation paths in $\mathcal{I}$ being induced by components stemming from a population whose degradation parameter Φ

follows a Beta prior distribution with parameters a_0 and b_0 is given by

$$\mathcal{L}_\Phi(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{I}|} \,|\, a_0, b_0) = \prod_{i=1}^{|\mathcal{I}|} \left[\frac{B(n_i + a_0, x_i + b_0)}{B(a_0, b_0)} \binom{x_i + n_i - 1}{x_i} \right].$$

Observe that the number of critical loading epochs sustained by the i-th component of age t_i has a Poisson distribution with parameter $t_i\Lambda$, where $\Lambda \sim \Gamma(\alpha_0, \beta_0)$. By the scaling property of the Gamma distribution, this number is therefore a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate follows a $\Gamma(\alpha_0, \beta_0/t_i)$ distribution, which is known to be a Negative Binomial (or Pascal) distribution with parameters $r = \alpha_0$ and $p = \frac{t_i}{\beta_0 + t_i}$. Hence, the likelihood of the degradation paths in $\mathcal{I}$ being induced by components stemming from a population whose degradation parameter Λ follows a Gamma prior distribution with parameters α_0 and β_0 is given by

$$\mathcal{L}_\Lambda(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{I}|} \,|\, \alpha_0, \beta_0) = \prod_{i=1}^{|\mathcal{I}|} \left[\binom{n_i + \alpha_0 - 1}{n_i} \left(\frac{t_i}{\beta_0 + t_i} \right)^{n_i} \left(\frac{\beta_0}{\beta_0 + t_i} \right)^{\alpha_0} \right].$$

We are interested in maximizing both likelihoods. For convenience, we maximize the log likelihoods. That is,

$$\underset{\alpha_0, \beta_0, a_0, b_0}{\arg\max} \; \ln \mathcal{L}_\Lambda(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{I}|} \,|\, \alpha_0, \beta_0) + \ln \mathcal{L}_\phi(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{|\mathcal{I}|} \,|\, a_0, b_0),$$

which is a nonlinear multidimensional maximization problem that can be solved using standard numerical methods (e.g., the Nelder-Mead method).

Chapter 3

Joint optimization of condition-based maintenance and process control with drift uncertainty

3.1. Introduction

In the previous chapter, we developed and analyzed a mathematical model to study the essential condition-based maintenance (CBM) trade-off between preventive maintenance versus corrective maintenance to improve the *availability* of technical systems. In practice, however, it is not only the availability, but also the *conditions* under which these systems operate when they are up and running that are of *crucial* importance. The latter is the case because these systems are often used for production processes, so that poor operating conditions can be detrimental for the output quality of the system. For instance, lithography machines in semiconductor fabrication plants can only produce microcircuits when up and running, while if they are running, even the slightest deviations in machine-related characteristics (e.g., internal pressure,

This chapter is based on Drent and Kapodistria (2021).

temperature) from good operating conditions can have a huge detrimental impact on the quality of the produced microcircuits (Senoner et al., 2021). The cost of poor quality can be significant. In the manufacturing industry, poor quality typically amounts to 10% to 30% of annual revenues (Bell, 2020). Motivated by such cost estimates, manufacturers continuously seek to achieve two goals: (i) improve system availability, and (ii) guarantee high production output quality. The first goal, as seen already in the previous chapters, can be achieved by using smart CBM models (such as the one developed in the previous chapter). To achieve the second goal, (statistical) *process control* (PC) of the production process is arguably the most pivotal function to resort to.

CBM and PC are conceivably related, especially if they utilize real-time data generated from the *same* process characteristic, yet most researchers have so far treated them independently. In practice they have also been treated predominantly in isolation, although the necessity of their joint treatment has been recognized as imperative for potential cost reductions by current best-in-class decision makers in asset management (Coleman et al., 2017; Blackwell et al., 2017). Consequently, there is an urgent need for methods that both manage CBM and PC in an integrated way and better incorporate real-time data. The objective of the present chapter is to address this need by developing a unified data-driven model that treats CBM and PC as a joint optimization problem and establishing the structural form of the optimal policy.

One of the tenets of PC is that there is a relationship between the production process' operating conditions and its output quality, which can be monitored through real-time data of a process characteristic (e.g., temperature, vibrations, pressure). CBM is more prescriptive in nature and prescribes maintenance actions based on data collected about a critical characteristic indicative of the equipment's degradation and hence proneness to imminent failure. In order to integrate the two, in this chapter we assume that there is a single characteristic suitable for both PC and CBM purposes, henceforth referred to as the *process characteristic*.

In this chapter, we model this process characteristic as a Brownian motion with constant drift and constant volatility. The Brownian motion falls into a different class of deterioration processes than the jump process considered in the previous chapter. For a jump process, damage accumulates at discrete epochs, while for the Brownian motion considered in this chapter, damage accumulates in a *continuous* fashion. A Brownian motion has often been used for degradation behavior modeling in the CBM literature (see e.g., Elwany et al., 2011; Liu et al., 2017; Si et al., 2018), but it is rather unconventional in PC literature. Traditionally, in PC, one or multiple samples of the

production output are sequentially taken to infer whether the production process has shifted from an in-control to an out-of-control state. The advent of technology – specifically the IoT that allows the equipment and the production process itself to be accurately monitored in real-time – makes such an approach outdated (Blackwell et al., 2017; Olsen and Tomlin, 2020). We therefore instead assume that the drift of the process characteristic is *unknown*, which we infer directly from real-time data. This drift term can be interpreted as the tendency to move away from the perfect operational state over time (in Section 3.2.1, we discuss this in more detail). We remark that since samples in traditional PC are usually modeled as Normal random variables (see, e.g., Tagaras and Nikolaidis, 2002; Panagiotidou and Tagaras, 2010), these conventional models can be seen as *time-invariant* cases of our model.

We model this problem as a sequential decision problem similar to the approach in the previous chapter. At scheduled, discrete epochs, the decision maker acquires data of the process characteristic and processes this data using Bayesian learning to obtain a posterior distribution over the future evolution of the process. Unlike the model in the previous chapter, the process characteristic need not be monitored continuously but only at these discrete epochs. A decision is then made to either perform preventive maintenance (and restore the production process) or to continue the production process – with (i) the risk of sudden failure of the manufacturing equipment, and (ii) possibly poor output quality – until the next decision epoch. We build an infinite-horizon Bayesian Markov decision problem (MDP) where the dynamics are governed by the posterior predictive distribution of the process characteristic, so that *learning* and *optimization* are integrated.

We assume the same cost structure as in the previous chapter: corrective maintenance upon failure is more expensive than performing preventive maintenance, but we additionally impose costs that relate to quality. To model the increased operational cost and/or decreased product quality, we assume that operational costs increase in the deviation of the process characteristic from the perfect operational state. By adopting this cost structure, we incentivize that the process remains close to good operating conditions. We further make the novel assumption that equipment fails if the process characteristic deviates more than a certain threshold from the perfect operational state. Virtually all research in the area of CBM assumes that equipment failure results from a gradual accumulation of damage that at some point exceeds a threshold (see Jardine et al. (2006) and Alaswad and Xiang (2017) for excellent overviews). However, in many practical applications, it is not per se the amount of accumulated damage itself, but rather the *absolute deviation* from normal operating conditions – of which an accumulation of damage is an example – that is instrumental in the failure. This is a subtle difference, yet this technicality renders conventional

proof techniques to establish structural properties not applicable. We overcome this challenge by translating the original MDP into an alternative MDP which does allow for establishing structural properties of the optimal policy.

3.1.1 Contributions

The main contributions of this chapter can be summarized as follows:

1. We propose and analyze a novel joint CBM and PC model and characterize the structural form of the optimal policy under standard cost and operating assumptions. We establish the optimality of an intuitive and easy-to-implement *bandwidth policy* that is monotone in the age of the production process, under both the *discounted cost* and *the average cost* criterion. With this bandwidth policy, the production process is continued as long as the process characteristic is within a bandwidth that is described by both an upper and a lower control limit. This bandwidth policy thus has characteristics both of control charts often used in the PC literature and of control limit policies typically established in the CBM literature.

2. We show that for archetypal replacement problems similar to the one treated in this chapter (and also in Chapter 2), there is a straightforward road map to show that the optimal average cost criterion policy can be obtained as a limit of the discounted cost model when the discount factor approaches 1 from below. This is generally speaking highly nontrivial and often an onerous task that heavily depends on the context. We believe that this proof can also be used to characterize structural properties under the average cost criterion in other replacement problems.

3. From a CBM perspective, we are the first to report structural results for CBM optimization when failures occur if the *absolute deviation* exceeds a threshold. From a PC perspective, we make two novel contributions. First, we adopt a continuous set of operational states instead of only an in- and out-of-control-state. Secondly, we model the process characteristic directly by a non-stationary stochastic process characterized by an unknown parameter – as opposed to computing latent variables (e.g., proportions, means, variances) based on samples of a stationary process – to guide decision making.

4. In an extensive simulation study, we numerically show that the optimal policy performs excellently compared to a clairvoyant policy that a-priori knows the true drift – with relative cost gaps of 2.1% on average. The results also suggest

that the data-driven policy is robust against misspecification of the initial (prior) parameters, which is appealing for practitioners who would like to implement the policy but do not (yet) have data to accurately estimate the initial parameters.

3.1.2 Organization

The remainder of this chapter is organized as follows. In Section 3.2, we present the problem formulation. We discuss the model in Section 3.2.1, the Bayesian framework in Section 3.2.2, and the optimization problem the decision maker faces in Section 3.2.3. Section 3.3 contains the main results. We present examples of the optimal joint PC and CBM policy in Section 3.3.1, and analytically establish the structural form in Sections 3.3.2 and 3.3.3 under the discounted cost criterion. In Section 3.3.4, we extend these results to the average cost setting. In Section 3.4, we benchmark the optimal policy when the drift is unknown with the performance of a clairvoyant policy that has access to the true value of the drift. Section 3.5 provides concluding remarks.

3.2. Problem formulation

In this section, we subsequently introduce the model, describe the process characteristics and discuss how its unknown parameter can be inferred from real-time data. We conclude with a formal description of the optimization problem that the decision maker wishes to solve.

3.2.1 Model description

Consider manufacturing equipment that is used for production processes and which may operate in different operational states. We observe real-time data of a process characteristic statistic (e.g., vibrations, temperature, pressure et cetera) at equally spaced discrete time epochs $t \in \{0, 1, \ldots\}$, where we assume, without loss of generality, 1 time unit between two consecutive observations. Each time the decision maker observes the signal, the process characteristic indicates the operational state of the system. We henceforth use the wording operational state and process characteristic interchangeably, as they refer to the same entity. We model the process characteristic

$\{X_t : t \geq 0\}$ as a Brownian motion with unknown drift α:

$$X_t = \alpha t + \bar{\sigma} W_t, \tag{3.1}$$

where $\bar{\sigma}$ is the known volatility of the process characteristic, and $\{W_t : t \geq 0\}$ is a standard Wiener process which can be interpreted as the randomness of the process characteristic. We thus use a continuous-time model for the process characteristic, while the decision maker can only interfere with the equipment at equally spaced decision epochs. We assume that the perfect operational state is 0, and that $X_0 = 0$, so that a new equipment starts in the perfect operational state. The unknown drift α thus models the mean tendency, per unit time, of the process characteristic to drift away from the perfect operational state over time. We shall henceforth call this process *deterioration* of the equipment and of the production process.

Let $\xi > 0$ denote the failure threshold, so that the equipment fails if the absolute value of the process characteristic is greater than or equal to ξ at a decision epoch (recall that the perfect operational state is 0 by convention). We assume that the set of operational states of the equipment is a bounded, open, and symmetric interval centered around the perfect operational state, viz. $(-\xi, \xi)$. At each decision epoch, the equipment can reside in one of the failed states, i.e. $(-\infty, -\xi]$ or $[\xi, \infty)$, or in an operational state. If the equipment is operational, which occurs when $|X_t| < \xi$, the decision maker has to decide whether to instantaneously perform preventive maintenance, or continue to the next epoch risking sudden failure of the equipment. If the equipment fails, which occurs when $|X_t| \geq \xi$, the system needs to be correctively maintained. Here, we make the common assumption (see, e.g., Elwany et al., 2011; Liu et al., 2017) that if the process characteristic crosses the failure threshold in between decision epochs and then returns to an operational state at the next decision epoch, that this does not constitute a failure. It is important to note that after a maintenance activity, either corrective or preventive, the newly installed system has a *new* drift term – which need *not* be the same as in the replaced system – that the decision maker needs to infer again, and that the stochastic process X_t is reset to $t = 0$. In the remainder, we shall therefore also refer to t as the age of a piece of equipment, which can thus also be interpreted as the time since the last maintenance activity.

Two kinds of costs are incurred: operating costs and maintenance costs. The per period operating cost and the preventive maintenance cost for a piece of equipment that is in operational state x at an epoch are given by the functions $l : (-\xi, \xi) \to \mathbb{R}_{\geq 0}$ and $c_p : (-\xi, \xi) \to \mathbb{R}_{\geq 0}$, respectively, and the corrective maintenance cost is $c_u < \infty$. We make three assumptions regarding these cost functions.

Assumption 3.1. $l(x)$ and $c_p(x)$ are even, bounded, and non-decreasing in $|x|$.

Assumption 3.2. $l(x) - c_p(x)$ is non-decreasing in $|x|$.

Assumption 3.3. $c_u > \max\left\{\sup_x\{c_p(x)\}, \sup_x\{l(x)\}\right\}$.

Assumption 3.1 ensures that the higher the deviation of the process is from the perfect operational state, the higher the operational costs are. This reflects for instance decreased product quality when the equipment is operating under deteriorated conditions. Assumption 3.1 ensures that a preventive maintenance activity (e.g., calibration, lubrication) costs more when the restoration is from a more deteriorated state. Assumption 3.2, often made in the literature (see e.g., Kawai et al., 2002; Van Oosterom et al., 2017), implies that an increase in the deviation from the perfect operational state leads to a more significant increase in operating costs than in preventive maintenance costs. Assumption 3.3 avoids trivial analysis. These assumptions are very general and allow a variety of settings, including the canonical CBM and PC cost structures (see Remark 3.1).

Remark 3.1. The cost structure addresses the trade-off between quality costs associated with deteriorated operational states and maintaining the equipment (objective of PC), *and* the trade-off between early preventive maintenance that potentially leads to high capital expenditures and the risk of failures (objective of CBM). The cost structure extends the canonical settings typically assumed in the CBM and PC literature. A single preventive maintenance cost, $c_p(x) \triangleq c_p < c_u$, and no operating costs, $l(x) \triangleq 0$, is often assumed in CBM models. Conventional PC models only assume two states, in-control and out-of-control, which can be obtained by constructing $l(x)$ accordingly. ◇

The objective is to find the policy that achieves the minimal expected cost over an infinite horizon. We treat both the discounted cost – where costs are discounted at rate $\gamma \in (0, 1)$ – and the average cost criterion.

3.2.2 Learning the unknown drift

We assume that the drift α is a-priori unknown to the decision maker and can only be inferred using data obtained through monitoring of the process characteristic. To this end, we take a Bayesian approach and treat α as a random variable denoted by Θ. For the initial prior distribution on this random variable Θ, we assume a (univariate) Normal distribution with mean $\mu_0 = 0$ and variance σ_0^2. This assumption is often made in the OM/OR literature when Bayesian inference involves a Gaussian random variable due to the conjugacy of the Normal distribution (see e.g., Azoury

and Miyaoka, 2009; Elwany et al., 2011; Russo, 2021). More importantly, it is also a very *natural* and *intuitive* assumption for the application at hand. The initial prior distribution models the decision maker's belief about the drift before any data is obtained after a new equipment is installed (either because of corrective or preventive maintenance). At that stage, the decision maker has no reason to believe that the likelihood of the drift being positive is larger than it being negative and vice versa, so that the Normal distribution with mean 0 is the *perfect candidate* to model this initial belief. Hence, upon installment of a new equipment, at $t = 0$, we have $\Theta_0 \sim \mathcal{N}(0, \sigma_0^2)$.

Recall that the process characteristic X_t satisfies Equation (3.1). Since the error term follows a standard Wiener process, it is convenient to work with the observed process characteristic increments which are independent and identically distributed (due to the Wiener process' properties). Let $\bar{\mathbf{X}}_t \triangleq [X_1, X_2 - X_1, \ldots, X_t - X_{t-1}]^\top$ be the observed process characteristic increments up to time t, with $\top$ denoting the transpose operator. We can then write the available data at time t in matrix form as follows:

$$\bar{\mathbf{X}}_t = \mathbf{A}_t \alpha + \boldsymbol{\epsilon}_t, \tag{3.2}$$

where $\boldsymbol{\epsilon}_t$ is a column vector with t independent Normally distributed random variables with mean 0 and variance $\bar{\sigma}^2$, and $\mathbf{A}_t$ is the $(t \times 1)$ design matrix at time t:

$$\mathbf{A}_t = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^\top. \tag{3.3}$$

We now state a result from Bayesian regression theory (see e.g., Bishop, 2006, Section 3.3). At time t, conditional on the observed process characteristic increments, we have $\Theta_t \sim \mathcal{N}\left(\mu_t\left(\bar{\mathbf{X}}_t\right), \sigma_t^2\left(\bar{\mathbf{X}}_t\right)\right)$ with:

$$\sigma_t^2\left(\bar{\mathbf{X}}_t\right) = \left(\frac{1}{\sigma_0^2} + \frac{1}{\bar{\sigma}^2}\mathbf{A}_t^\top \mathbf{A}_t\right)^{-1}, \quad \text{and} \quad \mu_t\left(\bar{\mathbf{X}}_t\right) = \sigma_t^2\left(\bar{\mathbf{X}}_t\right)\left(\frac{\mu_0}{\sigma_0^2} + \frac{1}{\bar{\sigma}^2}\mathbf{A}_t^\top \bar{\mathbf{X}}_t\right). \tag{3.4}$$

Combining (3.2) and (3.3), we have $\mathbf{A}_t^\top \mathbf{A}_t = t$ and $\mathbf{A}_t^\top \bar{\mathbf{X}}_t = X_t$, from which it is clear that the Bayesian inference only depends on the current process characteristic X_t and the equipment's age t. We use the notation $\mu_\Theta(x, t)$ and $\sigma_\Theta^2(t)$ to denote the updated hyperparameters of the density of Θ_t when $X_t = x$ is the last observed signal at time t since the installation of the equipment (time and age are equivalent).

Equation (3.4) reduces then to:

$$\sigma^2_{\Theta}(t) = \Big(\frac{1}{\sigma_0^2} + \frac{t}{\bar{\sigma}^2}\Big)^{-1}, \quad \text{and} \quad \mu_{\Theta}(x,t) = \sigma^2_{\Theta}(t)\frac{x}{\bar{\sigma}^2}, \tag{3.5}$$

where σ_0^2 is the initial prior parameter (recall that we assume that $\mu_0 = 0$), and $\bar{\sigma}^2$ is the known variance of the error process.

At each decision epoch, the decision maker wishes to *predict* the operational state of the process at the next decision epoch based on the observed data, so that one can utilize this to ascertain the *optimal* course of action. Specifically, we are interested in the posterior predictive distribution of X_{t+1} conditional on the current process characteristic $X_t = x$ at time t, denoted with $X_{t+1}(x,t)$. We then have the following recursive formula for the distribution of the next deterioration signal:

$$X_{t+1}(x,t) \overset{d}{=} x + \Theta_t + \epsilon, \tag{3.6}$$

where ϵ is a Normally distributed random variable with mean 0 and variance $\bar{\sigma}^2$ and $\overset{d}{=}$ denotes equality in distribution. It follows that $X_{t+1}(x,t)$ is also Normally distributed, that is:

$$X_{t+1}(x,t) \sim \mathcal{N}\big(\mu_{X_{t+1}}(x,t), \sigma^2_{X_{t+1}}(t)\big), \tag{3.7}$$

with

$$\mu_{X_{t+1}}(x,t) \triangleq x + \mu_{\Theta}(x,t) = x + \Big(\frac{1}{\sigma_0^2} + \frac{t}{\bar{\sigma}^2}\Big)^{-1}\frac{x}{\bar{\sigma}^2}, \tag{3.8}$$

$$\sigma^2_{X_{t+1}}(t) \triangleq \sigma^2_{\Theta}(t) + \bar{\sigma}^2 = \Big(\frac{1}{\sigma_0^2} + \frac{t}{\bar{\sigma}^2}\Big)^{-1} + \bar{\sigma}^2. \tag{3.9}$$

For notational brevity, we drop the conditioning on (x,t) when it is clear from the context, and we use the notation $\phi(y|x,t)$ to represent the *posterior predictive density function* of the next process characteristic when the current operational state is x at time t.

Equation (3.6) together with the updating schemes in (3.5) can be used to construct an updated posterior predictive distribution at each equipments age t of the next operational state in real-time based on the observed data. Since the posterior distribution of the signal process of the system is fully described by only the *current* operational state x and the equipment's age t, it is a Markov process. This allows us to formulate the optimization problem as an MDP encoded on state variables x and t, which is the objective of the next section.

3.2.3 Markov decision process formulation

In this section, we formulate the optimization problem for the discounted cost model, referred to as the γ-discounted cost model (in Section 3.3.4 we shall consider the average cost model).

At each decision epoch $\tau \in \mathbb{N}_0$ ($\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$), the decision maker can decide to do maintenance (which is preventive if the process is in a non-failed state and corrective otherwise) or continue if the system has not failed. Hence, if the operational state is x, then the action space $\mathcal{A}_x$ can be written as

$$\mathcal{A}_x = \begin{cases} \{\text{maintain, continue}\}, & \text{if } |x| < \xi, \\ \{\text{maintain}\}, & \text{if } |x| \geq \xi. \end{cases} \tag{3.10}$$

Let Π be the set of all non-anticipatory policies. A policy $\pi \in \Pi$ is a sequence $\{\pi_\tau(x,t)\}_{\tau \in \mathbb{N}_0}$ that prescribes which feasible action $\pi_\tau(x,t) \in \mathcal{A}_x$ (cf. Equation (3.10)) to take at all decision epochs $\tau \in \mathbb{N}_0$. If at decision epoch τ, the state is $(x_\tau, t_\tau) \in \mathbb{R} \times \mathbb{N}_0$, with x_τ the current operational state and t_τ the age of the equipment, and action $\pi_\tau(x_\tau, t_\tau)$ is followed, then the controlled dynamics of the discrete-time system $\{(X_\tau, T_\tau)\}_{\tau \in \mathbb{N}_0}$ can be described as

$$(X_{\tau+1}, T_{\tau+1}) = \begin{cases} \big(X_{t_\tau+1}(x_\tau, t_\tau), t_\tau + 1\big), & \text{if } |x_\tau| < \xi \text{ and } \pi_\tau(x_\tau, t_\tau) = \text{continue}, \\ (0,0), & \text{if } \pi_\tau(x_\tau, t_\tau) = \text{maintain}. \end{cases} \tag{3.11}$$

Note that from these dynamics, it is clear that if maintenance is performed – either preventive or corrective – then a new equipment is installed with age $t = 0$, starting in the perfect state $x = 0$.

Let

$$V_\gamma(x,t) = \inf_{\pi \in \Pi} \lim_{T \to \infty} \mathbb{E}\left[\sum_{\tau=1}^{T} \gamma^\tau C\big(\big(X_\tau, T_\tau\big), \pi(X_\tau, T_\tau) \big) \big| \ (X_0, T_0) = (x,t) \right] \tag{3.12}$$

denote the optimal total expected discounted cost given that the process starts in state $(x,t) \in \mathbb{R} \times \mathbb{N}_0$ at $\tau = 0$, where (X_τ, T_τ) denotes the state of the discrete-time

system at decision epoch τ, and $C(\cdot,\cdot)$ denotes the cost function defined as

$$C\big((x,t),\pi(x,t)\big) = \begin{cases} l(x), & \text{if } |x| < \xi \text{ and } \pi(x,t) = \text{continue}, \\ c_p(x), & \text{if } |x| < \xi \text{ and } \pi(x,t) = \text{maintain}, \\ c_u, & \text{if } |x| \geq \xi \text{ and } \pi(x,t) = \text{maintain}. \end{cases} \tag{3.13}$$

The cost function thus makes the distinction between a preventive and corrective maintenance action by assigning the corresponding cost to that specific action.

We are interested in finding the optimal policy $\pi^* = \{\pi_\tau\}_{\tau\in\mathbb{N}_0}$ that attains the infimum in (3.12). Since (i) the action space is finite, (ii) the costs per decision epoch are uniformly bounded, and (iii) $\gamma \in (0,1)$, we know by Proposition 1.2.3 of Bertsekas (2007), that there exists an optimal Markov policy π^* depending only on state (x,t) independent of the decision epoch $\tau \in \mathbb{N}_0$. Because of (i)-(iii), we know by Proposition 1.2.2 of Bertsekas (2007), that this optimal policy satisfies the following Bellman optimality equations:

$$V_\gamma(x,t) = \begin{cases} c_u + V_\gamma\big(0,0\big), & \text{if } |x| \geq \xi, \\ \min\Big\{c_p(x) + V_\gamma\big(0,0\big)\ ;\ l(x) + \gamma \displaystyle\int_{\mathbb{R}} V_\gamma\big(y,t+1\big)\phi(y|x,t)\mathrm{d}y\Big\}, & \text{if } |x| < \xi. \end{cases} \tag{3.14}$$

The first case in Equation (3.14) follows because equipment in the failed state must be maintained correctively at a cost of c_u. If the equipment's process characteristic is within the $(-\xi,\xi)$ bandwidth, and hence the system is in an operational state, we can either perform preventive maintenance, which costs $c_p(x)$, or leave the equipment in operation until the next decision epoch, which costs $l(x)$. The next operational state is then a Normal random variable with mean $x+\mu_\Theta(x,t)$ and variance $\sigma^2_\Theta(t)+\bar{\sigma}^2$, see Equation (3.6), which has density $\phi(y|x,t)$. Upon maintenance, the new equipment with age $t=0$ starts in the perfect operational state $x=0$, and the drift parameter of the stochastic process $\{X_t, t \geq 0\}$ is *unknown again* so that we start with the initial prior, encoded in $\phi(y|0,0)$ for $V_\gamma(0,0)$. Note that this evolution is in accordance with the described dynamics of Equation (3.11), and the costs are incurred according to Equation (3.13).

In the remainder, when we treat the discounted cost criterion, we drop the index γ in the value function and write $V(x,t)$ instead of $V_\gamma(x,t)$. In Section 3.3.4, where we show that the optimal policy under the average cost criterion can be obtained as a limit of the γ-discounted cost model when $\gamma \uparrow 1$, we shall re-use the notation $V_\gamma(x,t)$

to explicitly write the dependence on γ.

We refer to the setting described above as the *original MDP*. In the next section, we analyze the value function defined in Equation (3.14) and determine the structure of the optimal joint PC and CBM policy.

3.3. Optimal policy

This section presents structural results on the optimal joint PC and CBM policy. First, we present numerical examples of the optimal policy obtained via the value iteration algorithm that illustrate several structural properties. We then analytically establish these properties under both the discounted and average cost criterion.

3.3.1 Examples of the optimal policy

Figure 3.1 illustrates the optimal joint PC and CBM policy for two different sample paths of the process characteristic X_t. In these examples, the dashed line with circles depicts the observed process characteristic as the system ages (X_t) where the circles represent the observations at discrete epochs. The solid lines depict the optimal lower (LCL) and upper control limits (UCL), where at each decision epoch, the optimal action is to carry out a preventive maintenance if the observed process characteristic is at or below the LCL or at or above the UCL, otherwise the optimal action is to continue. In these examples, $\xi = 5$, $l(x) = \frac{|x|}{\xi}$ and $c_p(x) = \max(1, |x|)^{\frac{1}{\xi}}$, $c_u = 8$, $\gamma = 0.99$, $\bar{\sigma}^2 = 1$, and $\sigma_0^2 = 1$. The sample path on the left is simulated with drift term $\alpha = 0.25$, while the sample path on the right is simulated with drift term $\alpha = -0.25$. In the example on the left, we perform preventive maintenance at $t = 15$, while on the right we do so at $t = 14$ (both are marked with an arrow).

We can state two observations regarding the structure of the optimal policy that consistently arise, which are best explained by the illustrations in Figure 3.1.

1. The optimal joint PC and CBM policy is a bandwidth policy characterized by two control limits, referred to as LCL and UCL, such that we continue if the observed process characteristic is within the bandwidth, and perform preventive maintenance otherwise.

2. The LCL and the UCL are non-increasing and non-decreasing, respectively, in the age of the equipment.

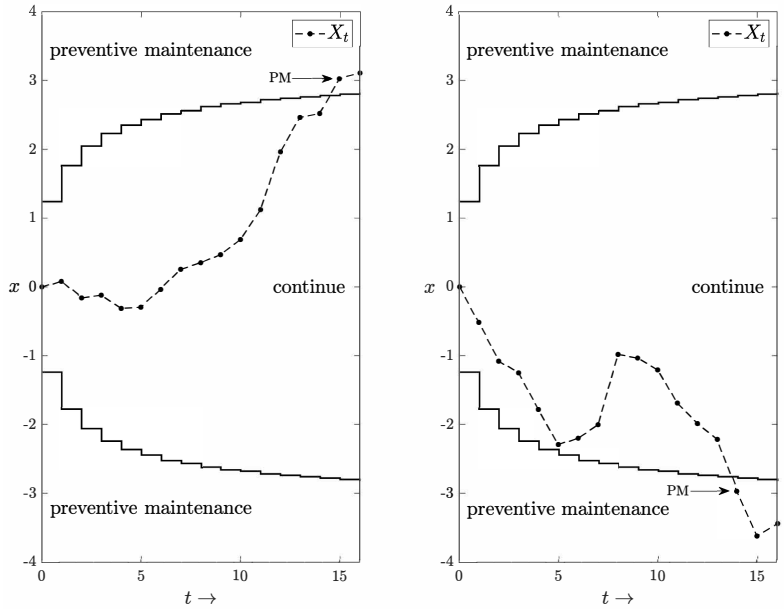


Figure 3.1 Two examples of the optimal joint PC and CBM policy.

3.3.2 Structural results for the optimal policy

In the two theorems that follow, we establish that observations 1 and 2 are in fact structural properties of the optimal data-driven joint PC and CBM policy. The proofs of these results rely on Propositions 3.1 and 3.3, and Theorem 3.3, which we shall formulate in Section 3.3.3.

The first theorem characterizes the monotonicity of the value function $V(x,t)$ in both x and t (see Figure 3.3 for an illustration of the result).

Theorem 3.1. *The value function $V(x,t)$ is:*

(i) non-increasing in t for all $x \in \mathbb{R}$;

(ii) even in x, and non-decreasing in $|x|$ for all $t \in \mathbb{N}_0$.

PROOF: The result follows directly from Proposition 3.3 and Proposition 3.1. □

The following theorem establishes the optimality of a monotonic bandwidth policy. Under a monotonic bandwidth policy, the equipment is kept operating until the observed process characteristic exits a certain bandwidth (see Figure 3.1 for illustrations

of the optimal policy). Since the optimal policy is even in x, we can denote this bandwidth with one parameter; $\delta(t)$. The equipment is thus kept operating until the observed process characteristic exceeds $\delta(t)$ or subceeds $-\delta(t)$. While control limit policies, where the equipment operates until the deterioration exceeds a control limit, have a longstanding and rich history in the CBM literature, this is the first result that establishes the optimality of a bandwidth policy.

Theorem 3.2. *At each equipment age t, there exists a bandwidth $\delta(t)$, $0 < \delta(t) \leq \xi$, such that the optimal action is to carry out preventive maintenance if and only if $|x| \geq \delta(t)$. The bandwidth $\delta(t)$ is monotonically non-decreasing in t.*

PROOF: The result follows directly from Theorem 3.3 and Proposition 3.1. □

Theorem 3.2 thus allows practitioners to monitor the process characteristic and interfere with the equipment when deemed necessary. The intuitive interpretation of the monotonicity in the age of the equipment is explained by the artificial example in Figure 3.2. In this example, we omit the sample paths as they obscure the message

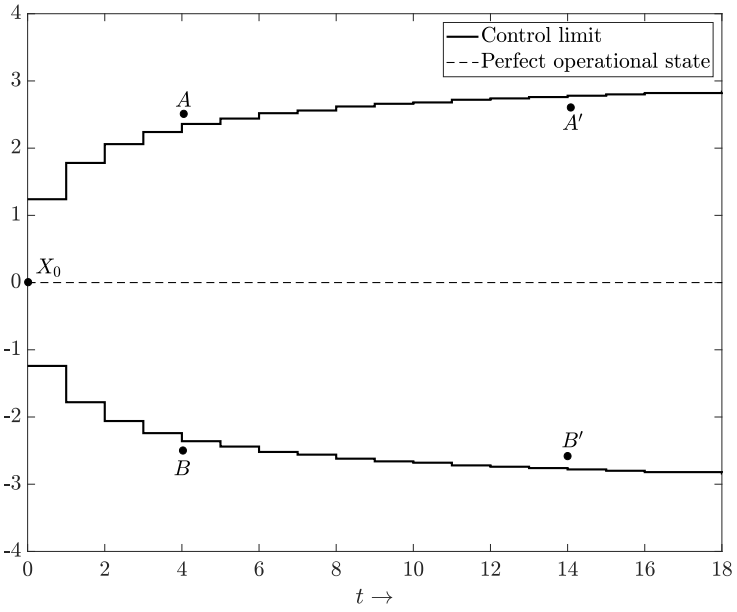


Figure 3.2 Illustration of the monotone behavior of the bandwidth policy in the equipment's age.

it needs to convey, and only focus on two pairs of process characteristic observations at two time epochs. If the process characteristic is at A (B) at time $t = 4$, then the

decision maker believes that the drift is much larger in absolute value than when she would observe the same value at time $t = 14$, i.e. A' (B'). Moreover, the prediction of the next process characteristic at $t = 14$ is not only smaller in expectation, it is also *more accurate* (i.e., less variable) compared to $t = 4$. This can be readily verified in Equation (3.5), which shows that the absolute value of the mean and variance of the posterior belief of the unknown drift decrease in time. Since the prediction is both smaller and less variable, the decision maker is conceivably less conservative in decision making – that is, the preventive maintenance threshold is closer to the failure threshold – at A' (B') than at A (B) where the prediction is both larger and more variable.

The policy thus ensures that the equipment is replaced early when necessary and late when possible, thereby maximizing the utilization of the useful lifetime. This bandwidth policy is not only intuitive and easy to implement in practice, it can also be exploited to decrease the computational time complexity of finding the optimal policy by employing existing algorithms that rely on such monotonic properties (e.g., the monotone policy iteration algorithm, see Puterman, 2005, Section 6.11.2).

3.3.3 Proofs of the structural results for the optimal policy

Establishing the structural properties via analyzing the original optimality equations as presented in Equation (3.14) is inherently complex due to the absolute value operator: it renders the common methodological approach via an inductive argument on the value iteration algorithm inapplicable. Such an argument relies on the monotonic behavior of actions and costs, and the stochastic monotonicity in the state variables, which all do not apply here.

Instead, we first introduce the so-called folded representation of the original MDP and establish its structural properties. We then discuss how these properties translate to the original MDP.

Chakravorty and Mahajan (2018) introduced the concept of the so-called folded MDPs (in discrete-time). We refer the interested reader to Chakravorty and Mahajan (2018) for a detailed discussion on this method and we discuss only the details that are relevant to the work in this chapter. To formalize this, we first need some definitions (taken from Chakravorty and Mahajan (2018) and adapted to our model).

Definition 3.1 (Even transition density)**.** For a given action $a \in \mathcal{A}_x$ at time t, we say that the controlled transition density $p_a(X_{t+1} = y|X_t = x) \triangleq p_a(y|x,t)$ on $\mathbb{R} \times \mathbb{R}$ is even if for all $x, y \in \mathbb{R}$, $p_a(y|x,t) = p_a(-y| - x, t)$.

Definition 3.2 (Even cost function)**.** For a given action $a \in \mathcal{A}_x$, we say that the cost function $c_a(x,t)$ is even if for all $(x,t) \in \mathbb{R} \times \mathbb{N}_0$, $c_a(x,t) = c_a(-x,t)$.

Definition 3.3 (Even MDP)**.** We say that an MDP is even if for all state-action pairs $(x,t,a) \in \mathbb{R} \times \mathbb{N}_0 \times \mathcal{A}_x$, the cost function $c_a(x)$ and transition density $p_a(y|x,t)$ are even.

Chakravorty and Mahajan (2018) show that if an MDP is even then one can construct a folded MDP defined only on the non-negative values of the state space. One can then analyze the structural properties of this folded MDP, which can then be *unfolded* to the original MDP. We shall proceed with formalizing this idea in the remainder of this subsection.

First of all, it is easy to see that the original MDP is even. To see this, first note that the cost function $C\big((x,t),\pi(x,t)\big)$ is even for all actions, see Equation (3.13) together with Assumption 3.1. Next, recall that in state (x,t), the process characteristic at the next epoch, $X_{t+1}(x,t)$, is Normal with mean $\mu_{X_{t+1}}(x,t) = x + \sigma^2_\Theta(t)\Big(\frac{x}{\bar{\sigma}^2}\Big)$ and variance $\sigma^2_{X_{t+1}}(t)$ (see (3.7)). Then, for all $t \in \mathbb{N}_0$, we have for all $x, y \in \mathbb{R}$ that

$$
\begin{aligned}
\phi(-y|-x,t) &= \frac{1}{\sqrt{2\pi\sigma^2_{X_{t+1}}(t)}} \exp\left(-\frac{\big(-y-\mu_{X_{t+1}}(-x,t)\big)^2}{2\sigma^2_{X_{t+1}}(t)}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2_{X_{t+1}}(t)}} \exp\left(-\frac{\Big(-y+x+\sigma^2_\Theta(t)\big(\frac{x}{\bar{\sigma}^2}\big)\Big)^2}{2\sigma^2_{X_{t+1}}(t)}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2_{X_{t+1}}(t)}} \exp\left(-\frac{\Big(y-\mu_{X_{t+1}}(x,t)\Big)^2}{2\sigma^2_{X_{t+1}}(t)}\right) = \phi(y|x,t).
\end{aligned}
$$

Hence, based on Definition 3.1, the controlled transition density is even. Informally speaking, this means that there is a symmetry in the posterior predictive density with respect to the current observation of the process characteristic. That is, if one observes a positive process characteristic x at some time t, then one believes that at the next observation $t+1$, the probability of observing some other positive value y is the same as the probability of observing $-y$ conditional on the observation $-x$ at time t. Observe that for this controlled density, we only need to consider the action 'continue', since it is the only action that involves stochasticity. Based on Definition

3.3, we can conclude that the original MDP is even and we shall now construct the *folded MDP*.

The folded MDP acts on $(x,t) \in \mathbb{R}_{\geq 0} \times \mathbb{N}_0$ with an adapted transition density, and with the same cost structure. The adapted transition density is given by: $\tilde{\phi}(y|x,t) \triangleq \phi(y|x,t) + \phi(-y|x,t)$ for all $t \in \mathbb{N}_0$ and $y \geq 0$. We denote the optimal expected discounted cost and the optimal policy of the folded MDP with $\tilde{V}(x,t)$ and $\tilde{\pi}^*(x,t)$, respectively. The folded MDP satisfies the following Bellman optimality equations:

$$\tilde{V}(x,t) = \begin{cases} c_u + \tilde{V}\big(0,0\big), & \text{if } x \geq \xi, \\ \min\Big\{c_p(x) + \tilde{V}\big(0,0\big) \;;\; l(x) + \gamma \displaystyle\int_{\mathbb{R}_{\geq 0}} \tilde{V}\big(y,t+1\big)\tilde{\phi}(y|x,t)\mathrm{d}y\Big\}, & \text{if } x < \xi. \end{cases} \tag{3.15}$$

Since the original MDP is even, we can present the first result. This result establishes the connection between the value function and the optimal policy of the original MDP and its folded counterpart. This connection is pivotal to the analysis in the rest of this section.

Proposition 3.1. *For all $t \in \mathbb{N}_0$, the value function $V(x,t)$ and the optimal policy $\pi^*(x,t)$ are even in x, and*

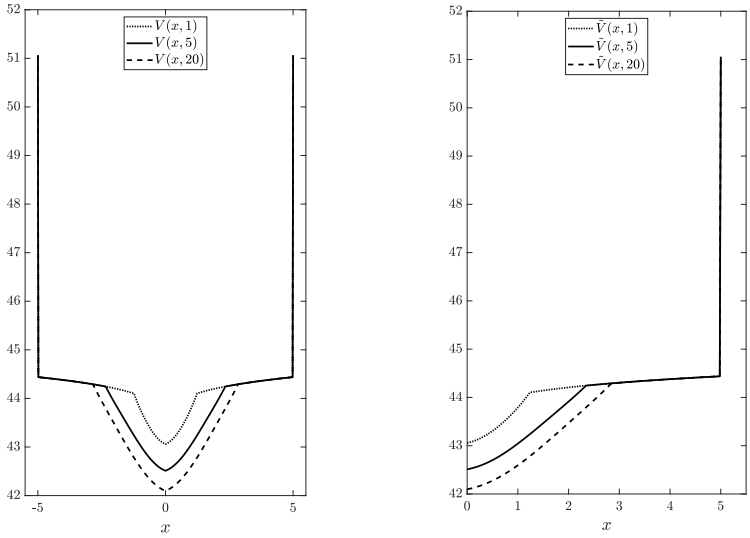
(i) $V(x,t) = V(-x,t) = \tilde{V}(|x|,t)$;

(ii) $\pi^*(x,t) = \pi^*(-x,t) = \tilde{\pi}^*(|x|,t)$.

Proof: Since the original MDP is even, the result follows directly by applying Proposition 2 of Chakravorty and Mahajan (2018) and Corollary 3 of Chakravorty and Mahajan (2018). □

Proposition 3.1 is illustrated in Figure 3.3 and Figure 3.4.

Figure 3.3 shows the value functions and Figure 3.4 shows the optimal policy of both the original MDP and the folded MDP. The same settings as in the example from Figure 3.1 are used, and we obtained these graphs by numerically computing the solution to (3.14) and (3.15), respectively, using the value iteration algorithm. We indeed observe that $V(x,t) = V(-x,t) = \tilde{V}(|x|,t)$, and that $\pi^*(x,t) = \pi^*(-x,t) = \tilde{\pi}^*(|x|,t)$. The discontinuities at $x = -5$ and $x = 5$ are due to the high expected cost associated with starting in the failed state while the kinks in the value functions correspond to where the optimal action changes from continue to perform preventive



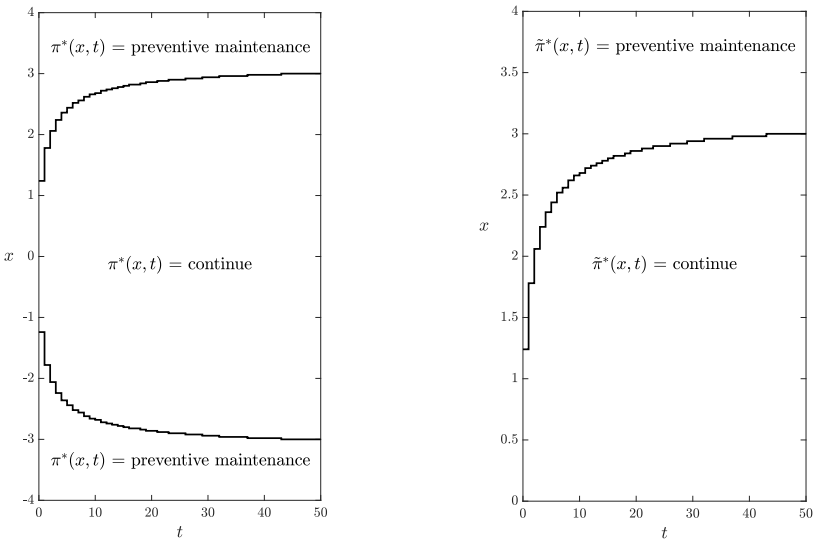
(a) Original MDP (b) Folded MDP

Figure 3.3 Graphs of the value function of the original MDP (a) and the folded MDP (b) as function of x for $t \in \{1, 5, 20\}$.

maintenance. Figure 3.3 further suggests that $\tilde{V}(x,t)$ is non-decreasing in x, and non-increasing in t, while Figure 3.4 suggests that the threshold induced by $\tilde{\pi}(x,t)$ is non-decreasing in t.

In the remainder of this subsection, we establish that these are in fact structural properties of $\tilde{V}(x,t)$ and $\tilde{\pi}(x,t)$. We can then, based on Proposition 3.1, characterize structural properties of the original MDP (both of the value function (3.14) and optimal policy). Let $\tilde{X}_{t+1}(x,t)$ denote the posterior predictive random variable when the state is (x,t) in the folded MDP. We first analyze the stochastic monotonicity of $\tilde{X}_{t+1}(x,t)$ in x and t, which are instrumental in characterizing the optimal policy of the folded MDP.

First, recall that in the original MDP, the posterior predictive random variable when the state is $(x,t) \in \mathbb{R} \times \mathbb{N}_0$, $X_{t+1}(x,t)$, is Normal with mean $\mu_{X_{t+1}}(x,t) = x + \mu_\Theta(x,t)$ and variance $\sigma^2_{X_{t+1}}(t) = \sigma^2_\Theta(t) + \bar{\sigma}^2$. The folded posterior predictive random variable $\tilde{X}_{t+1}(x,t)$ in state $(x,t) \in \mathbb{R}_{\geq 0} \times \mathbb{N}_0$ then has a Folded Normal distribution with location $\mu_{X_{t+1}}$ and scale $\sigma^2_{X_{t+1}}$ (i.e., the distribution of the absolute value of a Normally distributed random variable with mean $\mu_{X_{t+1}}$ and variance $\sigma^2_{X_{t+1}}$). To see



(a) Original MDP (b) Folded MDP

Figure 3.4 Graphs of the optimal policy of the original MDP (a) and the folded MPD (b).

this, note that:

$$\begin{aligned}\tilde{\phi}(y|x,t) &= \phi(y|x,t) + \phi(-y|x,t) \\ &= \frac{1}{\sqrt{2\pi\sigma^2_{X_{t+1}}}} \exp\left(-\frac{(y-\mu_{X_{t+1}})^2}{2\sigma^2_{X_{t+1}}}\right) + \frac{1}{\sqrt{2\pi\sigma^2_{X_{t+1}}}} \exp\left(-\frac{(y+\mu_{X_{t+1}})^2}{2\sigma^2_{X_{t+1}}}\right),\end{aligned}$$

which is the density of a Folded Normal random variable with location $\mu_{X_{t+1}}$ and scale $\sigma^2_{X_{t+1}}$. There is thus an intuitive connection between the posterior predictive random variable in the original MDP which acts on the reals and its folded counterpart that acts only on the non-negative part of the reals, namely $\tilde{X}_{t+1}(x,t) \overset{d}{=} |X_{t+1}(x,t)|$.

The following result, which is due to Wang and Wang (2011), presents stochastic ordering properties (see Definition 2.1 in Chapter 2 for the definition of $\leq_{\text{st}}$) of the Folded Normal distribution which are instrumental in establishing the stochastic monotonicity of $\tilde{X}_{t+1}(x,t)$ in x and t.

Lemma 3.1 (Theorem 1, Wang and Wang, 2011)**.** *Let X be a Normal random variable with mean μ_X, variance σ^2_X and probability density function $\phi_X(\cdot)$, and let Y be a Normal random variable with mean μ_Y, variance σ^2_Y and probability density*

function $\phi_Y(\cdot)$*. Then,*

$$|X| \leq_{st} |Y| \text{ if and only if } \phi_X(0) \geq \phi_Y(0) \text{ and } \sigma_X \leq \sigma_Y.$$

Lemma 3.1 is quite remarkable if one compares this result to the stochastic ordering properties of regular Normal random variables. It is well-known that Normal random variables are ordered in the usual stochastic order if and only if their variances are equal. Then, a smaller mean implies that it is smaller in the usual stochastic order than one with a larger mean (see, e.g., Müller, 2001). For the setting with unequal variances, there is no stochastic monotonicity in the usual stochastic order, even with additional conditions on the order of the means. Lemma 3.1 establishes that this is not the case if one looks at absolute Normal random variables, where the usual stochastic order is established under additional conditions on the ordering of their densities at 0.

The following result establishes the stochastic monotonicity of $\tilde{X}_{t+1}(x,t)$ in x and t, which will be used to establish the structural properties of the value function of the folded MDP.

Proposition 3.2. *The random variable* $\tilde{X}_{t+1}(x,t)$ *satisfies the following stochastic orders:*

(i) for $x^+ \geq x^- \geq 0$*, we have* $\tilde{X}_{t+1}(x^-,t) \leq_{st} \tilde{X}_{t+1}(x^+,t)$ *for all* $t \in \mathbb{N}_0$*;*

(ii) for $t^+ \geq t^- \geq 0$ *we have* $\tilde{X}_{t^-+1}(x,t^-) \geq_{st} \tilde{X}_{t^++1}(x,t^+)$ *for all* $x \in \mathbb{R}_{\geq 0}$*.*

Proof: We first prove Assertion (i) and then Assertion (ii) of Proposition 3.2:

(i) For fixed t and for $x^- \leq x^+$, only $\mu_{X_{t+1}}(x,t)$ changes, while $\sigma^2_{X_{t+1}}(t)$ does not change (cf. Equation (3.8) and Equation (3.9), respectively). Hence, in light of Lemma 3.1, we only need to show that $\phi(0|x,t)$ is non-increasing in x. For notational clarity, we write σ^2 instead of $\sigma^2_{X_{t+1}}(t)$ in the remainder of the proof of Assertion (i). We have for $\phi(0|x,t)$,

$$\phi(0|x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\mu_{X_{t+1}}(x,t)\right)^2}{2\sigma^2}\right). \tag{3.16}$$

Taking the derivative of (3.16) with respect to x yields

$$\frac{\partial\phi(0|x,t)}{\partial x} = \frac{-2\mu_{X_{t+1}}(x,t)\phi(0|x,t)}{2\sigma^2}\frac{\partial\mu_{X_{t+1}}(x,t)}{\partial x}.$$

Since $t \geq 0$, and $\sigma_0, \bar{\sigma}, \sigma > 0$, and $\mu_{X_{t+1}}(x,t) = x + \left(\frac{1}{\sigma_0^2} + \frac{t}{\bar{\sigma}^2}\right)^{-1} \frac{x}{\bar{\sigma}^2}$ (cf. Equation (3.8)), it is easy to see that $\mu_{X_{t+1}}(x,t) \geq 0$ and $\frac{\partial \mu_{X_{t+1}}(x,t)}{\partial x} \geq 0$ for all $x \geq 0$. As $\phi(0|x,t)$ is non-negative (property of a density function), we conclude that $\frac{\partial \phi(0|x,t)}{\partial x} \leq 0$ for all $x \geq 0$, so that $\phi(0|x,t)$ is non-increasing in x. This completes the proof of Assertion (i).

(ii) In light of Lemma 3.1, we need to show that, for fixed $x \geq 0$, $\phi(0|x,t)$ is non-decreasing in t and that, for $t^- \leq t^+$, $\sigma^2_{X_{t^+ +1}}(t^+) \leq \sigma^2_{X_{t^- +1}}(t^-)$. Starting with the latter, it is easy to see from Equation (3.9) that this inequality trivially holds as the variance is inversely related to t. We now proceed with showing that $\phi(0|x,t)$ is non-decreasing in t for $t \geq 0$.

For notational clarity, define, using Equations (3.8) and (3.9), the auxiliary function

$$f(x,t) \triangleq \frac{-1}{2}\left(\frac{\mu_{X_{t+1}}(x,t)}{\sigma_{X_{t+1}}(t)}\right)^2 = \frac{-x^2\left(\sigma_0^2(t+1) + \bar{\sigma}^2\right)}{2\bar{\sigma}^2\left(\sigma_0^2 t + \bar{\sigma}^2\right)}.$$

Taking the partial derivative of $f(x,t)$ with respect to t yields

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} &= -\frac{\sigma_0^2 x^2}{2\bar{\sigma}^2\left(\sigma_0^2 t + \bar{\sigma}^2\right)} + \frac{\sigma_0^2\left(\sigma_0^2(t+1) + \bar{\sigma}^2\right)x^2}{2\bar{\sigma}^2\left(\sigma_0^2 t + \bar{\sigma}^2\right)^2} \\ &= \frac{\sigma_0^4 x^2}{2\bar{\sigma}^2\left(\sigma_0^2 t + \bar{\sigma}^2\right)^2}. \end{aligned} \tag{3.17}$$

We can then write $\phi(0|x,t)$ as

$$\phi(0|x,t) = \frac{1}{\sqrt{2\pi}\sigma_{X_{t+1}}(t)}\, e^{f(x,t)}. \tag{3.18}$$

Taking the derivative of (3.18) with respect to t yields

$$\frac{\partial \phi(0|x,t)}{\partial t} = \frac{e^{f(x,t)}}{\sqrt{2\pi}}\left(\frac{1}{\sigma_{X_{t+1}}(t)}\frac{\partial f(x,t)}{\partial t} - \frac{1}{\sigma^2_{X_{t+1}}(t)}\frac{\partial \sigma_{X_{t+1}}(t)}{\partial t}\right). \tag{3.19}$$

For the partial derivative of $\sigma_{X_{t+1}}(t)$ with respect to t, we have

$$\begin{aligned}\frac{\partial \sigma_{X_{t+1}}(t)}{\partial t} &= \frac{\partial}{\partial t}\sqrt{\Big(\frac{1}{\sigma_0^2}+\frac{t}{\bar{\sigma}^2}\Big)^{-1}+\bar{\sigma}^2} \\ &= -\left(2\Big(\frac{1}{\sigma_0^2}+\frac{t}{\bar{\sigma}^2}\Big)^2\bar{\sigma}^2\sigma_{X_{t+1}}(t)\right)^{-1}. \end{aligned} \tag{3.20}$$

Observe that since $t \geq 0$, $x \geq 0$, and $\sigma_0, \bar{\sigma}, \sigma > 0$, the partial derivative in (3.17) is non-negative, while the partial derivative in (3.20) is non-positive. It is then easy to see that (3.19) is non-negative. Hence, $\frac{\partial \phi(0|x,t)}{\partial t} \geq 0$ for all $t \geq 0$, so that $\phi(0|x,t)$ is non-decreasing in t. This completes the proof of Assertion (ii).

□

The next result, building on the results of Proposition 3.2, shows that the value function of the folded MDP, $\tilde{V}(x,t)$, is monotone in both the age of the equipment and the current position of the process characteristic (see Figure 3.3 for an illustration). This result is instrumental in characterizing the optimal policy of the folded MDP.

Proposition 3.3. *The value function $\tilde{V}(x,t)$ is:*

(i) non-decreasing in x for all $t \in \mathbb{N}_0$;

(ii) non-increasing in t for all $x \in \mathbb{R}_{\geq 0}$.

PROOF: We prove these assertions using induction on the n-th step of the value iteration algorithm. Let $\tilde{V}^n(x,t)$ denote the value function at the n-th iteration of the value iteration algorithm. Since (i) the action space is finite, (ii) the costs per decision epoch are uniformly bounded, and (iii) $\gamma \in (0,1)$, we know that the value function is guaranteed to converge to the optimal value function that satisfies Equation (3.15) (i.e., $\tilde{V}^n(x,t) \to \tilde{V}(x,t)$ for all $x \in \mathbb{R}_{\geq 0}$ and $t \in \mathbb{N}_0$ as $n \to \infty$) from any arbitrary starting position through the value iteration algorithm (Bertsekas, 2007, Proposition 1.2.1). We set $\tilde{V}^0(x,t) = 0$ for all $x \in \mathbb{R}_{\geq 0}$ and $t \in \mathbb{N}_0$, so that $\tilde{V}^0(x,t)$ is non-decreasing in x and non-increasing in t. Then according to Equation (3.15) (in expectation form), we have

$$\tilde{V}^{n+1}(x,t) = \begin{cases} c_u + \tilde{V}^n(0,0), & \text{if } x \geq \xi, \\ \min\Big\{c_p(x) + \tilde{V}^n(0,0)\ ;\ l(x) + \gamma \mathbb{E}\left[\tilde{V}^n\big(\tilde{X}_{t+1}(x,t), t+1\big)\right]\Big\}, & \text{if } x < \xi. \end{cases} \tag{3.21}$$

We now proceed with the proofs of Assertion (i) and (ii) of Proposition 3.3:

(i) We inductively assume that $\tilde{V}^n(x,t)$ is non-decreasing in x for all $t \in \mathbb{N}_0$. The random variable $\tilde{X}_{t+1}(x,t)$ is stochastically non-decreasing in x in the usual stochastic order (cf. Assertion (i) of Proposition 3.2), which together with the induction hypothesis implies that the expectation $\mathbb{E}\left[\tilde{V}^n\big(\tilde{X}_{t+1}(x,t),t+1\big)\right]$ is non-decreasing in x for all $t \in \mathbb{N}_0$ (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Since $c_p(x)$ and $l(x)$ are non-decreasing in x (see Assumption 3.1), all terms of the right-hand side of Equation (3.21) are non-decreasing in x and hence $\tilde{V}^{n+1}(x,t)$ is also non-decreasing in x. Due to the convergence in which the structure of $\tilde{V}^n(x,t)$ is preserved (cf. Bertsekas, 2007, Proposition 1.2.1), we conclude that $\tilde{V}(x,t)$ is also non-decreasing in x for all $t \in \mathbb{N}_0$.

(ii) The proof is similar to the proof of Assertion (i). Let us assume that $\tilde{V}^n(x,t)$ (and hence $\tilde{V}^n(x,t+1)$) is non-increasing in t for all $x \in \mathbb{R}_{\geq 0}$. Since the random variable $\tilde{X}_{t+1}(x,t)$ is stochastically non-increasing in t in the usual stochastic order (cf. Assertion (ii) of Proposition 3.2), the expectation $\mathbb{E}\left[\tilde{V}^n\big(\tilde{X}_{t+1}(x,t),t+1\big)\right]$ is non-increasing in t for all $x \in \mathbb{R}_{\geq 0}$ (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Because the right-hand side of Equation (3.21) is non-increasing in t, $\tilde{V}^{n+1}(x,t)$ is also non-increasing in t. Due to the convergence in which the structure of $\tilde{V}^n(x,t)$ is preserved (cf. Bertsekas, 2007, Proposition 1.2.1), we conclude that $\tilde{V}(x,t)$ is also non-increasing in t for all $x \in \mathbb{R}_{\geq 0}$.

□

The next result, the main result for the folded MDP, characterizes the structural form of the optimal policy of the folded MDP (see Figure 3.4 for an illustration of the result).

Theorem 3.3. *Given the equipment's age t, the optimal policy is an age-dependent control limit policy with control limit $\tilde{\delta}(t)$, $0 < \tilde{\delta}(t) \leq \xi$: the optimal action is to carry out preventive maintenance if and only if $x \geq \tilde{\delta}(t)$. The control limit $\tilde{\delta}(t)$ is non-decreasing in t.*

Proof: Preventive maintenance is optimal when the following equation holds

$$c_p(x) + \tilde{V}(0,0) \leq l(x) + \gamma \mathbb{E}\left[\tilde{V}\big(\tilde{X}_{t+1}(x,t),t+1\big)\right],$$

or equivalently when,

$$\tilde{V}(0,0) \ \leq l(x) - c_p(x) + \gamma \mathbb{E}\left[\tilde{V}\big(\tilde{X}_{t+1}(x,t), t+1\big)\right]. \tag{3.22}$$

The left-hand side of Inequality (3.22) is constant with respect to x. Since the value function $\tilde{V}(x,t)$ is non-decreasing in x (cf. Assertion (i) of Proposition 3.3) and the random variable $\tilde{X}_{t+1}(x,t)$ is stochastically non-decreasing in x in the usual stochastic order (cf. Assertion (i) of Proposition 3.2), the expectation $\mathbb{E}\left[\tilde{V}\big(\tilde{X}_{t+1}(x,t), t+1\big)\right]$ is non-decreasing in x (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Hence, since $l(x) - c_p(x)$ is non-decreasing in x (see Assumption 3.2), we find that the right-hand side of Inequality (3.22) is non-decreasing in x. Hence, if the optimal decision is to carry out preventive maintenance in state $(\tilde{\delta}(t), t)$, then the same decision is optimal for any state (x,t) with $x \geq \tilde{\delta}(t)$, which implies the control limit policy.

Similarly, since the value function $\tilde{V}(x,t)$ is non-increasing in t (cf. Assertion (ii) of Proposition 3.3) and the random variable $\tilde{X}_{t+1}(x,t)$ is stochastically non-increasing in t in the usual stochastic order (cf. Assertion (ii) of Proposition 3.2), the expectation $\mathbb{E}\left[\tilde{V}\big(\tilde{X}_{t+1}(x,t), t+1\big)\right]$ is non-increasing in t (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Hence, the right-hand side of Inequality (3.22) is non-increasing in t. Thus, if it is optimal to carry out preventive maintenance in state $(\tilde{\delta}(t), t)$, then it is optimal to carry out preventive maintenance in any state $(\tilde{\delta}(t), t')$ with $t' < t$, which implies that $\tilde{\delta}(t)$ is monotonically non-decreasing in t. □

Theorem 3.3 establishes the optimality of a control limit policy, which is an archetypal result in the CBM literature. The monotonicity of the control limit in the age has also been established, albeit in different settings, by Elwany et al. (2011), Si et al. (2018), Liu et al. (2017), and in Drent et al. (2020a) which is based on Chapter 2. The intuition behind this monotonicity is that if the equipment ages, whilst everything else is kept fixed, the next process characteristic is stochastically smaller (see Proposition 3.2) thereby allowing a control limit that is closer to the failure threshold.

These results complete the analysis of the folded MDP. Using these results for the folded MDP, we can immediately establish the results in Theorems 3.1 and 3.2 for the characterization of the optimal policy of the original MDP.

3.3.4 Average cost criterion

In this subsection, we shall demonstrate that the bandwidth policy (see Theorem 3.2), which is optimal for the γ-discounted cost model, is also optimal under the expected

average cost criterion, and can be obtained as a limit of the γ-discounted cost model when $\gamma \uparrow 1$. In the literature, there exist several versions of sufficient conditions that guarantee this convergence for the average cost criterion (see Arapostathis et al. (1993) for an overview). In the present chapter, we follow the conditions of Schäl (1993), who considered general infinite-state MDPs with compact actions spaces. Before stating Schäls conditions in the context of our setting, we first need to introduce two cost quantities. Given a policy π and initial state (x,t), the expected total T-period cost is

$$J_\pi^T(x,t) = \mathbb{E}\left[\sum_{\tau=1}^{T} C\left((X_\tau, T_\tau), \pi(X_\tau, T_\tau)\right) \mid (X_0, T_0) = (x,t)\right],$$

and the expected average cost per unit time is

$$\Phi_\pi(x,t) = \limsup_{T\to\infty} \frac{J_\pi^T(x,t)}{T},$$

where we use the same notation as in Section 3.2.3.

Schäl (1993) established a set of three sufficient conditions, but since our action space is finite for each (x,t), i.e. continue or maintain, we only need to verify two of these. The first condition ensures that there exists a policy for which the average cost is finite.

Condition (G). *There exists a policy π and an initial state (x,t) such that*

$$\Phi_\pi(x,t) < \infty.$$

The next condition ensures that the optimal discounted cost function $V_\gamma(x,t)$ behaves well as $\gamma \uparrow 1$. It states that for any starting state (x,t) and any discount factor $\gamma \in (0,1)$, the extra cost for not starting from some reference state is upper bounded, and that this bound does not depend on the discount factor.

Condition (B). *For each (x,t), we have that for some reference state (x_0,t_0),*

$$\sup_{\gamma\in(0,1)} \big(V_\gamma(x,t) - V_\gamma(x_0,t_0)\big) < \infty.$$

The next result establishes that the optimal average cost policy can be obtained as a limiting case of the γ-discounted optimal policy, denoted with $\pi_\gamma(x,t)$, when $\gamma \uparrow 1$.

Theorem 3.4. *There exists a stationary policy π^* which is average cost optimal in*

the sense that, for all $(x,t) \in \mathbb{R} \times \mathbb{N}_0$,

$$\Phi_{\pi^*} = \inf_{(x,t)\in\mathbb{R}\times\mathbb{N}_0} \inf_{\pi\in\Pi} \Phi_\pi(x,t) =: g,$$

and π^* *is limit discount optimal in the sense that for any* $(x,t) \in \mathbb{R} \times \mathbb{N}_0$,

$$\pi^*(x,t) = \lim_{\gamma\uparrow 1} \pi_\gamma(x,t).$$

Furthermore,

$$g = \lim_{\gamma\uparrow 1}(1-\gamma) \inf_{(x,t)\in\mathbb{R}\times\mathbb{N}_0} V_\gamma(x,t),$$

is the optimal average cost.

PROOF: By Theorem 3.8 of Schäl (1993), we need to verify Condition (G) and (B).

We verify Condition (G) by constructing a policy for which the average cost per time unit is bounded from above. We consider the failure-based policy, denoted with π_{CM}, in which the decision maker only replaces the equipment upon failure and we compute the long-run average cost rate using renewal reward theory. The installments of new equipment constitute the renewal epochs, and we are interested in the expected cost per cycle, as well as the expected cycle length.

For $t \geq 1$, let $\bar{X}_t \triangleq X_t - X_{t-1}$ denote the random deterioration increment during interval $(t-1,t]$, so that $X_t = \sum_{i=1}^t \bar{X}_i$ with $X_0 = 0$. Let τ_ξ be the stopping time $\tau_\xi \triangleq \inf_{t\in\mathbb{N}_0}\{|X_t| \geq \xi\}$, which is the decision epoch at which the equipment fails and has to be replaced correctively. Since the $\bar{X}_i$'s are not identically zero and $\xi < \infty$, we know that $\mathbb{E}\left[\tau_\xi\right] < \infty$. Moreover, since $X_0 = 0$, we also know that $\mathbb{E}\left[\tau_\xi\right] \geq 1$. Since $\sup_x\{l(x)\} < c_u$, we further know that the operating cost per time unit is at most c_u, so that the expected operating costs per cycle are bounded from above by $\mathbb{E}\left[\tau_\xi\right] \cdot c_u$. Thus, we can conclude that

$$\Phi_{\pi_{CM}}(0,0) = \frac{\mathbb{E}[\text{expected operating costs per cycle}] + c_u}{\mathbb{E}[\text{expected cycle length}]} < \frac{\left(\mathbb{E}\left[\tau_\xi\right] + 1\right) \cdot c_u}{1} < \infty.$$

So Condition (G) holds.

We next verify Condition (B). Recall that the Bellman optimality equations are given by (3.14). Since $\sup_x\{c_p(x)\} < c_u$ (see Assumption 3.3), we can rewrite (3.14) and

bound the second branch using the inequality for the minimum operator as follows

$$V_\gamma(x,t) = c_u + V_\gamma\big(0,0\big), \text{ if } |x| \geq \xi, \text{ and}$$
$$V_\gamma(x,t) \leq c_p(x) + V_\gamma\big(0,0\big) \leq c_u + V_\gamma\big(0,0\big), \text{ if } |x| < \xi.$$

All in all, $V_\gamma(x,t) \leq c_u + V_\gamma\big(0,0\big)$ for all $(x,t) \in \mathbb{R} \times \mathbb{N}_0$. Now, set $(x_0,t_0) \triangleq (0,0)$ as the reference state, so that for all $\gamma \in (0,1)$, we have that, for all $(x,t) \in \mathbb{R} \times \mathbb{N}_0$,

$$V_\gamma(x,t) - V_\gamma\big(x_0,t_0\big) \leq c_u.$$

Since $c_u < \infty$, Condition (B) holds, and the proof is completed. □

Remark 3.2. Generally speaking, verifying Condition (G) and (B) is highly nontrivial and often an onerous task that heavily depends on the context. However, in archetypal replacement problems similar to the one in this chapter (where the action space is usually finite), Condition (G) will be satisfied by a failure-based policy, and Condition (B) can be verified since there is often a state that you can always enter by paying a cost that is bounded from above by the corrective maintenance cost c_u. We therefore believe that the proof of Theorem 3.4 can be used for characterizing the structural form of the optimal average cost policy in other replacement problems (including the problem studied in Chapter 2). ◇

3.4. Simulation study

In the previous section, we analytically established the structural properties of the optimal joint real-time data-driven PC and CBM policy under both the discounted, as well as the average cost criterion. This policy, henceforth denoted with $\pi_{\mathcal{J}}$, can be applied in practice when the drift is a-priori unknown and needs to be learned from real-time observations of the process characteristic. A natural question to ask then is: *How well does the optimal policy perform compared to the optimal policy in the case without uncertainty regarding the drift?* In this section, we numerically evaluate the answer to this question. Specifically, we report the results of a simulation study in which we benchmark the performance of policy $\pi_{\mathcal{J}}$ against the best achievable performance corresponding to a clairvoyant Oracle that knows the true value of the drift of each newly installed equipment.

We first discuss the Oracle approach, after which we describe the simulation set-up and discuss the results.

3.4.1 Oracle policy

If at the installment of a new equipment, the value of the drift α is known to the decision maker, then the optimal policy can be readily obtained by solving an MDP with only one state: the process characteristic. This MDP satisfies the following Bellman optimality equations:

$$V^{\alpha}(x) = \begin{cases} c_u + V^{\alpha}(0), & \text{if } |x| \geq \xi, \\ \min\Big\{ c_p(x) + V^{\alpha}(0) \ ; l(x) + \gamma \mathbb{E}\left[V^{\alpha}\big(x + \alpha + Z\big) \right] \Big\}, & \text{if } |x| < \xi, \end{cases} \tag{3.23}$$

where Z is Normally distributed with mean 0 and variance $\bar{\sigma}^2$, and $V^{\alpha}(x)$ is the minimal total expected discounted cost when the equipment with drift α starts with process characteristic x. In this case, the only randomness stems from the noise/error in the observations, modeled by the Wiener process. Since this policy is clairvoyant by having access to the true value of α, we refer to it as the *Oracle* policy and denote it with $\pi_{\mathcal{O}}$. From (3.23) it is easy to see that this policy is time independent and will only depend on the process characteristic x.

We observed numerically that $\pi_{\mathcal{O}}$ is also of a bandwidth type, but the UCL and LCL need not be of the same absolute value. The reason for this is that if $\alpha < 0$ (> 0), then $\pi_{\mathcal{O}}$ has a more conservative – closer to the perfect operational state – LCL (UCL) and a less conservative – further from the perfect operational state – UCL (LCL). Establishing this property analytically is complex since the MDP that satisfies (3.23) is not even (see Definition 3.3), which renders the framework used in the previous section not applicable. Nevertheless, we can still numerically obtain $\pi_{\mathcal{O}}$ using the value iteration algorithm and use the policy in our simulation study.

3.4.2 Simulation set-up

We perform a numerical experiment in which we vary five different input parameters (see Table 3.1 for a summary) and assess all possible combinations of these five input parameters (i.e. a full factorial numerical experiment). We use the cost functions $l(x) = \frac{|x|}{\beta}$ and $c_p(x) = \max(1, |x|)^{\frac{1}{\beta}}$, and let $\beta \in \{2, 4, 6\}$. We use three values for the known volatility of the error process $\bar{\sigma}^2 \in \{1, 1.5, 2\}$, three values for the corrective maintenance cost $c_u \in \{6, 8, 12\}$, and four values for the failure threshold $\xi \in \{8, 10, 12, 14\}$. Finally, in a simulation instance (we shortly describe the simulation procedure), the drifts are sampled from a Normal distribution with true mean μ^*_{Θ} and variance $\sigma_0^2 = 1$, and we use four values for $\mu^*_{\Theta} \in \{0, 0.5, 1, 1.5\}$. These choices lead to a total of 432 instances.

Table 3.1 Input parameter values for simulation study.

	Input parameter	No. of choices	Values
1	Volatility of process characteristic, $\bar{\sigma}^2$	3	1, 1.5, 2
2	True mean of Θ, μ^*_Θ	4	0, 0.5, 1, 1.5
3	Corrective maintenance cost, c_u	3	6, 8, 12
4	Failure threshold, ξ	4	8, 10, 12, 14
5	Cost parameter, β	3	2, 4, 6

The simulation procedure is as follows. For each instance,

Step 1: Compute policy $\pi_{\mathcal{J}}$ by numerically solving (3.14) using the value iteration algorithm;

Step 2: Sample a drift, say $\hat{\alpha}$, from $\mathcal{N}(\mu^*_\Theta, 1)$, and compute policy $\pi_{\mathcal{O}}$ based on $\hat{\alpha}$ by numerically solving $V^{\hat{\alpha}}(x)$ using the value iteration algorithm;

Step 3: Simulate N^{max} sample paths of the process characteristic $\{X_t, t \geq 0\}$ with drift $\hat{\alpha}$ from $t = 0$ to failure, or until both policy $\pi_{\mathcal{J}}$ and $\pi_{\mathcal{O}}$ have prescribed maintenance, and record realized costs and lifetimes for both policies;

Step 4a: Compute the average cost per cycle for policy $\pi_{\mathcal{J}}$ and $\pi_{\mathcal{O}}$ based on the results of Step 3;

Step 4b: Compute the average cycle length for policy $\pi_{\mathcal{J}}$ and $\pi_{\mathcal{O}}$ based on the results of Step 3;

Step 5: Repeat Steps 2-4 s^{max} times and compute the average of the cost rates obtained at Step 4 to estimate the long-run average cost rate of policy $\pi_{\mathcal{J}}$ and $\pi_{\mathcal{O}}$ as the ratio of the sample average of the average cost per cycle (step 4a) over the sample average of the average cycle length (step 4b).

We choose $N^{max} = 20 \cdot 10^3$ and $s^{max} = 3 \cdot 10^3$ so that the 95% confidence intervals of the average cost rates computed in Step 5 are sufficiently small (max $\leq \pm 1\%$). Note that in the simulation, the drift is sampled from a Normal distribution with mean μ^*_Θ and variance σ_0^2, while policy $\pi_{\mathcal{J}}$ initially assumes that the drift is Normally distributed with mean 0 and variance σ_0^2. Hence by varying the value of μ^*_Θ, we can evaluate how robust the relative regret is with respect to the decision maker's initial belief. Finally, we use a discount factor that is close to 1, i.e. $\gamma = 0.999$, so that the resulting policies are also optimal according to the average cost criterion (see Theorem 3.4).

3.4.3 Results

We benchmark the performance of policy $\pi_{\mathcal{J}}$ by the notion of relative regret, i.e. the percentage of expected additional cost relative to the clairvoyant Oracle policy $\pi_{\mathcal{O}}$. We compute this relative regret as $\mathcal{R} = 100 \cdot (C_{\pi_{\mathcal{J}}} - C_{\pi_{\mathcal{O}}})/C_{\pi_{\mathcal{O}}}$, where C_π is the long-run average cost rate (obtained via simulation) of policy $\pi \in \{\pi_{\mathcal{J}}, \pi_{\mathcal{O}}\}$. Note that as $\pi_{\mathcal{O}}$ is unattainable in practice, the only purpose of $\mathcal{R}$ is to serve as an upper bound on the relative regret with the best achievable performance.

The results of the simulation study are summarized in Table 3.2. In this table, we present the minimum, average, and maximum $\mathcal{R}$, where we first distinguish between subsets of instances with the same value for a specific input parameter of Table 3.1, and then present the results for all instances.

Table 3.2 Results of simulation study: Relative regret (in %) with Oracle performance.

		Relative regret $\mathcal{R}$		
Input parameter	Value	Min	Mean	Max
Volatility of process characteristic, $\bar{\sigma}^2$	1	1.42	2.86	5.96
	1.5	1.40	1.86	3.51
	2	0.94	1.50	2.20
True mean of Θ, μ^*_Θ	0	1.20	1.84	2.51
	0.5	1.23	1.90	2.81
	1	1.13	2.11	3.94
	1.5	0.94	2.45	5.96
Corrective maintenance cost, c_u	6	0.99	2.08	5.86
	8	0.94	2.07	5.96
	12	0.97	2.07	5.91
Failure threshold, ξ	8	0.94	1.94	4.94
	10	1.21	2.09	5.74
	12	1.22	2.13	5.96
	14	1.20	2.13	5.86
Cost parameter, β	2	1.40	1.90	2.51
	4	0.94	1.94	4.22
	6	0.98	2.37	5.96
Total		0.94	2.07	5.96

The following main observations can be drawn from Table 3.2. Firstly, policy $\pi_{\mathcal{J}}$ yields excellent results with a relative regret of only 2.07% on average, and a maximum relative regret over all instances of 5.96%. Secondly, the relative regret increases as the mean of the true generating distribution starts to deviate from what the decision maker initially believes. This is not surprising, as one can conceivably expect that the performance decreases when the initial hyperparameters of the prior distribution are misspecified by the decision maker (recall that $\mu_0 = 0$). Nevertheless, even in

this regime, when μ^*_Θ is large, the relative regret of $\pi_\mathcal{J}$ is still only 2.45% on average, suggesting that the policy corrects for the misspecification *on-the-go* and is thus robust in that sense. Thirdly, we see that the relative regret decreases as the volatility of the process characteristic increases. The intuition behind this is as follows. We observed that both policies become more conservative when the volatility increases – recall that both policies are computed with the same, known volatility $\bar{\sigma}$ as input – and in this case, the UCLs and LCLs of $\pi_\mathcal{J}$ and $\pi_\mathcal{O}$ are relatively close to each other, leading to a small relative regret. By contrast, in the low volatility case, the difference between the policies is relatively large, leading to a larger relative regret.

3.5. Conclusion

In this chapter, we have studied the joint optimization of CBM and PC of critical manufacturing equipment used for production processes. We have modeled this problem jointly by assuming a single process characteristic that is indicative of both the production output quality and the state of the manufacturing equipment. The evolution of this process characteristic is modeled as a Brownian motion with unknown drift that needs to be inferred from observations of the process characteristic. We have presented an MDP formulation encoded on the sufficient statistics needed for this inference, but due to its non-monotonicity properties, commonly used proof techniques to establish structural properties are not applicable. Instead, we have employed a novel approach and showed that the MDP can be translated into an alternative MDP which allowed us to establish structural properties. Using this approach, we showed that the optimal policy has an intuitive bandwidth structure with monotonic control limits, under both the average and discounted cost criterion. An extensive numerical study suggests that the Bayesian policy performs excellently compared to a clairvoyant policy that a-priori knows the true drift, and that it is robust against misspecification of the initial hyperparameters of the prior distribution.

As our work in this chapter is a first step in studying the joint optimization of CBM and PC, it has some limitations, and addressing these limitations are possible directions for future research.

First, although the operational and maintenance cost functions are quite general, we do assume that they are even, while in practice it could be the case that a positive deviation has a different cost structure than a negative deviation of the same magnitude. Similarly, we assume that equipment failure is due to a positive or negative deviation exceeding a certain threshold. In practical applications, however, it could be that the equipment is less prone to failure on either the positive or negative side,

meaning that there is not a single failure threshold but one for each side. Future research could investigate such generalizations, but they require different approaches as the resulting MDPs will likely not be even, deeming the framework used in this chapter not applicable.

Second, we assume that the volatility of the Brownian motion is known to the decision maker. One can relax this assumption by treating it as a random variable that needs to be inferred from the data in a Bayesian fashion (in addition to the unknown drift). The sample variance of the past observations is then additionally needed as sufficient statistic, and the resulting posterior predictive distribution will then be a non-central t-distribution (Gelman et al., 1995). As a non-central t-distribution is asymmetric (unless its non-centrality parameter is 0) there is little hope that the resulting MDP will be even.

In this chapter – and also in the previous chapter – we have assumed that the deterioration process itself behaves completely independently. Specifically, a decision maker can only reset the deterioration of a component by replacing it, but she cannot influence the evolution of the deterioration process. In practice, however, it might be reasonable that deterioration behavior of a component can be influenced by changing certain variables of the system. For instance, if a lithography machine deteriorates faster or produces worse quality in high environmental temperatures, then a simple solution would be to lower the environmental temperature. This gives rise to so-called controllable deterioration as another strategy to efficiently reduce system failures. In the next chapter, we shall study a CBM model that allows for such controllable deterioration.

Chapter 4

Optimal condition-based production policies with base rate (un)certainty

4.1. Introduction

In the two previous chapters, we studied mathematical models for condition-based maintenance (CBM) in which the degradation behavior is *not* influenced by other factors than by its own randomness. For many advanced technical systems, however, the random degradation behavior of components is often affected (to a certain degree) by *adjustable settings* of the system itself. Examples illustrating this relationship are manifold and can be found in various industries and application domains. For example, gearboxes and generators in a wind turbine deteriorate faster under higher rotational speeds (Feng et al., 2013; Macquart and Maheri, 2019) which can be adjusted by controlling the wind turbine's pitch (Spinato et al., 2009; Zhang et al., 2015). In the manufacturing industry, examples include cutting tools in high-speed cutting machinery that wear faster under higher speeds (Dolinšek et al., 2001), and conveyor belts in production lines that are increasingly more likely to fail when their

This chapter is based on Drent and Arts (2022).

rotational speeds are increased (Nourelfath and Yalaoui, 2012). The flip side of this coin is that settings that generally speaking lead to a lower deterioration rate, also lead to a system that generates less revenue. For instance (using some of the examples from above), wind turbines generate less wind energy when their rotational speeds are lower, and production lines and cutting tools have less output when conveyor belts and cutting speeds are slowed down, respectively.

As explained in the previous chapters, failures of these systems are usually the result of deterioration of components exceeding a critical level, and as such, they can be prevented before they occur by making instant maintenance decisions based on the component's deterioration level. However, in practice, maintenance planning has *limited flexibility* and cannot be done last minute because arranging the logistic support takes time, particularly for remotely located systems (e.g., wind farms). Moreover, it is inconvenient and costly to shut down an entire system each time an individual component needs maintenance. For these reasons, in practice, maintenance activities are usually performed at planned maintenance moments that are scheduled periodically and well in advance. At such planned maintenance moments, the entire system is deliberately shut down and components that require maintenance are maintained. In the previous two chapters, we analyzed the decision as to when maintenance should be performed at such a planned maintenance moment by allowing only interference with the system at equidistant epochs (i.e., planned maintenance moments).

When these planned maintenance moments are scheduled well in advance, another strategy that decision makers can employ to reduce maintenance costs and increase revenues, is to *control* the deterioration behavior in between such planned maintenance moments. Specifically, slowing down the deterioration rate (through the choice of corresponding settings) when a component is close to its failure threshold leads to less failures between consecutive maintenance moments. On the other side, if components degrade more slowly, decision makers can increase the revenue rate (through the choice of corresponding settings) without running the risk of system failure before the planned maintenance moment.

This operational strategy has become a viable option due to recent technological advances. Specifically, the advent of sensor technology integrated in the Internet of Things (IoT) enables both *remote accrual* of deterioration data and *remote control* of physical settings of components at near-zero costs and in real-time (Coopers, 2014; Manyika et al., 2015; Blackwell et al., 2017). The novel idea of controlling deterioration behavior was first proposed by Uit het Broek et al. (2020) who coined the term *condition-based production* for this operational strategy. The authors consider

condition-based production rate decisions for systems with an adjustable production rate (essentially a setting) that directly affects the deterioration rate. Inherent to this concept is a trade-off – as alluded to in the beginning of this section: a lower production rate has both a lower deterioration and revenue rate, while a higher production rate has both a higher deterioration and revenue rate. By formulating a continuous-time optimal control problem, they are able to study this trade-off, and establish structural insights and exact analytical solutions when deterioration is *deterministic*. Although they show that the optimal policy under stochastic deterioration exhibits similar properties as when deterioration is deterministic, they do not analytically study such a scenario.

In this chapter, we study such optimal *condition-based setting* policies (the production rate is essentially a setting), but in the context of *stochastically* deteriorating systems, and our goal is to characterize the structure of the optimal policy. To this end, we model the stochastic deterioration as a Poisson process whose *uncontrolled* intensity is equal to a certain *base rate*. This deterioration process is thus in the same class of deterioration process (i.e. jump processes) as the one considered in Chapter 2 and, as we shall later see, Chapter 5. The base rate represents the rate at which the system deteriorates under normal, uncontrolled operating conditions. A decision maker can continuously adjust settings which leads (via general functions) to an instantaneous change in the revenue rate and the rate of the Poisson process (proportional to the base rate). We shall discuss this in more detail in Section 4.2, but conceptually, we have that higher (lower) settings lead to higher (lower) revenue *and* deterioration rates. By employing such a condition-based setting policy, the decision maker can steer the deterioration as she nears the planned maintenance moment, at which a maintenance cost depending on the deterioration state is incurred.

We first consider the case that the base rate is a-priori *known* to the decision maker. We model this problem – essentially a Poisson intensity control problem – as a continuous-time MDP and rigorously analyze both its Hamilton-Jacobi-Bellman (HJB) equations and its discretized equivalent to characterize the monotonic behavior of the optimal policy. We then relax the assumption that the base rate is known a-priori and include *parameter uncertainty* regarding this base rate. For this case, we develop an *easy-to-implement* Bayesian heuristic, that mimics the structure of the optimal policy for the case that the base rate is known. In this scenario, a decision maker needs to learn the a-priori unknown base rate through the chosen settings and the deterioration she observes as a consequence of those chosen settings.

Next to the optimal condition-based setting policy, we also consider the trade-off that arises due to the maintenance cost at a planned maintenance moment. Specifically,

scheduling planned maintenance moments too frequently leads to unnecessary maintenance activities and associated costs, while a very long maintenance interval both increases the risk of a failed system and decreases the revenue accumulation as the decision maker is then forced to choose lower settings. In this chapter, we also study the optimization of the length of this interval given that an optimal condition-based setting policy is employed in between maintenance moments.

4.1.1 Contributions

In this chapter, we make the following contributions:

1. We are the first to analytically establish the structure of the optimal condition-based setting policy under stochastic deterioration. The structure of the optimal policy is remarkably intuitive: Decrease the setting if closer to the failure threshold and increase the setting if closer to the planned maintenance moment. In establishing these structural properties, we also provide new submodularity preservation properties that can be useful in other application areas.

2. We show that under these optimal setting decisions, the length of the interval between planned maintenance moments can be easily optimized. This allows for the optimal integration of setting policies on an operational level and maintenance decisions on a tactical level.

3. We complement these theoretical results by an extensive numerical study in which we demonstrate that: (i) condition-based setting policies can lead to profit increases of up to 50% on average compared to static setting policies, and (ii) integrating maintenance and setting policies can lead to profit increases of up to 21% on average compared to not integrating them.

4. For the case with parameter uncertainty, we propose a Bayesian framework that can tractably learn the unknown base rate under *any* setting policy, and use it to build a heuristic policy. In an extensive simulation study, we show that this heuristic performs close to a clairvoyant Oracle policy that knows the base rate a-priori.

4.1.2 Organization

The remainder of this chapter is organized as follows. We present the problem formulation in Section 4.2. Section 4.3 contains the main analytical results for the

case that the base rate is known. In Section 4.4, we report on an extensive numerical study that highlights the practical value of our theoretical results. Section 4.5 is devoted to the case in which there is uncertainty regarding the base rate. Finally, Section 4.6 provides concluding remarks.

4.2. Problem formulation

We consider a single-unit production system whose deterioration can be represented by a single variable. The deterioration process of the system is continuously monitored and described by a Poisson process $Y = (Y_t)_{t\geq 0}$, with $Y_0 = 0$. Deterioration level 0 indicates that the system is as-good-as-new, and failure occurs when the deterioration level is equal to a fixed failure level ξ. We denote the set of all deterioration levels with $\mathscr{X} \triangleq \{0, 1, \ldots, \xi\}$. We assume that we can adjust a univariate setting of the system (e.g., the production rate), that will both impact the deterioration process and the rate at which the production system generates revenue. That is, if the decision maker selects setting $s \in \mathscr{S} \triangleq [0, s^{max}]$ (where $0 < s^{max} < \infty$ is the maximum setting), then the rate of the Poisson process is instantaneously governed by the base rate λ multiplied by the *deterioration function* $f : \mathscr{S} \to \mathbb{R}_+$ ($\mathbb{R}_+ \triangleq [0, \infty)$), and the revenue rate is instantaneously governed by the *revenue function* $r : \mathscr{S} \to \mathbb{R}_+$. In practice, one can view the base rate as the rate at which the system deteriorates under normal operating conditions which can be adjusted by settings via the function f. For now we shall assume that the base rate λ is fully known to the decision maker and that each system upon replacement has the same rate; that is, the systems stem from a homogeneous population where all systems are statistically identical. The formulation in this section and the analysis in the next section – focused on a homogeneous population – are novel and insightful in their own right. They also help us to develop and benchmark a heuristic for the setting in which newly installed systems stem from a heterogeneous population, where each system has an a-priori *unknown* base rate (we discuss this setting in Section 4.5).

In what follows (and in the remainder of this chapter), we use increasing and decreasing in the weak sense. We assume that $f(s)$ and $r(s)$ are continuous and increasing in s, and that $f(0) = 0$ and $r(0) = 0$. Note that due to this assumption and the fact that $\mathscr{S}$ is compact, we are also guaranteed that f and r are bounded. The increasing assumption is reasonable in practice as a higher production rate generally leads to both a higher revenue rate and a faster deterioration rate. We remark that we do not use the assumption that $r(s)$ is increasing in our analysis, and in fact, all results also hold true if $r(s)$ is decreasing. However, if $r(s)$ is decreasing then this will result

in a practically meaningless policy in which the system is always turned off (turning on the machine will lead to a negative revenue since $r(s) = 0$). Maintenance is only performed at scheduled maintenance moments, and the next maintenance action is scheduled at time T (e.g., each half year a maintenance team is sent to an offshore wind farm to perform maintenance activities). The maintenance cost at the scheduled maintenance moment is governed by an increasing convex maintenance cost function $c_m : \mathscr{X} \to \mathbb{R}_+$, which maps the deterioration level at the maintenance moment to the cost of performing maintenance. Observe that the canonical cost structure with a single corrective maintenance cost c_u and a single preventive maintenance cost c_p $(0 < c_p < c_u)$ typically assumed in the maintenance literature is a special case of this assumption.

The decision maker's problem is to dynamically adjust the settings with the goal of maximizing the expected profit: the expected cumulative revenue minus the expected maintenance cost. We formulate this optimization problem using the deterioration level when the time until the next maintenance moment is t. To this end, let $X_t = \{Y_{\tilde{t}} \mid 0 \leq \tilde{t} \leq T\}$ be the deterioration level at time $t = T - \tilde{t}$ before the next maintenance moment. Since $Y_{\tilde{t}}$ is the number of degradation increments up to time $\tilde{t}$, $\mathrm{d}Y_{\tilde{t}} = 1$ if a degradation increment is realized at time $\tilde{t}$. Let $\mathscr{U}$ be the set of all admissible non-anticipating *setting policies* satisfying $\int_{k=0}^{t} \mathrm{d}Y_k \leq \xi - X_t$ (almost surely). This constraint ensures that whenever the system fails (i.e. when $Y_t = \xi$), the system is turned off. The existence of setting 0 with $\lambda(0) = 0$ in the set $\mathscr{S}$ guarantees that it can always be satisfied. For a given setting policy $u \in \mathscr{U}$, let $s^u_\lambda(x, k)$ denote the induced setting when the deterioration level is x, the time to planned maintenance moment is $k \in [t, 0)$ and the base rate is λ. The expected profit, starting from state (x, t) under a policy $u \in \mathscr{U}$ is then denoted by

$$J^u_\lambda(x, t) \triangleq \mathbb{E}_u \left[\int_0^t r\big(s^u_\lambda(X_k, k)\big) \mathrm{d}k \right],$$

where $\mathbb{E}_u$ denotes that the expectation is taken with respect to policy u, and with $J^u_\lambda(x, 0) \triangleq -c_m(x)$ to incorporate the maintenance costs at the scheduled maintenance moment. We make the dependence of the total expected profit on the base rate λ, and hence the correspondence to a homogeneous population, explicit by using the subscript. The goal of the decision maker is to find the optimal setting policy u^*_λ that maximizes the expected profit generated over $[t, 0]$, denoted with

$$J^*_\lambda(x, t) \triangleq \sup_{u \in \mathscr{U}} J^u_\lambda(x, t).$$

In our first result, we present the Hamilton-Jacobi-Bellman (HJB) equations for the optimal expected profit $J^*_\lambda(x,t)$. In the proof, we use a standard argument in which we consider the dynamics of the deterioration under a chosen setting – essentially an inhomogeneous Poisson process – over a very small time interval.

Lemma 4.1. *$J^*_\lambda(x,t)$ is the solution of*

$$\frac{\partial J^*_\lambda(x,t)}{\partial t} = \begin{cases} \max_{s\in\mathscr{S}} \Big[r(s) - \lambda f(s)\Delta_\lambda(x,t)\Big], & \text{if } x \in \mathscr{X}\setminus\xi, \\ 0, & \text{if } x = \xi, \end{cases} \tag{4.1}$$

with $\Delta_\lambda(x,t) \triangleq J^*_\lambda(x,t) - J^*_\lambda(x+1,t)$, and boundary conditions $J^*_\lambda(x,0) = -c_m(x)$ for all $x \in \mathscr{X}$.

Proof: Consider the dynamics over a small interval δt (that is, when the time until the next maintenance moment changes from t to $t-\delta t$) when $t > 0$. Let $x < \xi$. If the decision maker selects setting $s \in \mathscr{S}$, then with probability $\delta t\lambda f(s) + o(\delta t)$, x changes to $x+1$, and with probability $1-\delta t\lambda f(s)+o(\delta t)$, x stays the same. Secondly, if $x = \xi$, then the only admissible action is $s = 0$, so that with probability 1, x stays ξ. In all cases, t changes to $t-\delta t$. By the principle of optimality, for $t > 0$, we have

$$J^*_\lambda(x,t) = \max_{s\in\mathscr{S}} \Big[r(s)\delta t + \delta t\lambda f(s) J^*_\lambda(x+1,t-\delta t) + \big(1-\delta t\lambda f(s)\big) J^*_\lambda(x,t-\delta t) + o(\delta t)\Big], \tag{4.2}$$

when $x < \xi$, and,

$$J^*_\lambda(\xi,t) = J^*_\lambda(\xi,t-\delta t) + o(\delta t), \tag{4.3}$$

when $x = \xi$. Rearranging Equations (4.2) and (4.3), dividing by δt, and taking the limit $\delta t \to 0$, leads to the desired result. The boundary conditions follow from the maintenance cost that is incurred when the system has deterioration level x at the scheduled maintenance moment. □

The formulation in Lemma 4.1 falls under the scope of Markovian intensity control problems, and we refer the reader to Bremaud (1980) for an excellent treatment of the general theory on this subject. Since f and r are continuous and bounded (by assumption), and $\mathscr{S}$ is a compact subset of $\mathbb{R}$, we are guaranteed by Theorem III of Bremaud (1980) that there exists a solution to (4.1), which we denote by $s^*_\lambda(x,t)$; that is, $s^*_\lambda(x,t)$ is the optimal setting for state (x,t) when the base rate is λ.

From a methodological point of view, it is noteworthy to mention that the problem formulation in Lemma 4.1 shares similarities with the canonical pricing problem in revenue management, where a decision maker seeks a revenue-maximizing pricing strategy when one is selling a finite number of products during a limited planning horizon. In this problem, first introduced and characterized in the seminal paper by Gallego and Van Ryzin (1994), customers arrive according to a Poisson process whose intensity depends on the chosen price. Gallego and Van Ryzin (1994) formulate this problem as an intensity control problem, also leading to a set of HJB equations. Their work initiated a large stream of literature in which various properties and extensions are studied, such as multiple products (Gallego and Van Ryzin, 1997), unknown market responses (Araman and Caldentey, 2009; Farias and Van Roy, 2010), multiple offered prices (Aydin and Ziya, 2009), discounted criterion with time-varying demand (Cao et al., 2012), with discounts and rebates on products (Aydin and Ziya, 2008; Hu et al., 2017), and sales target constraints (Du et al., 2021). Since demand for products is modeled as a controlled Poisson process in the works in this stream of literature, they all model their problem as Markovian intensity control problems and study the corresponding HJB equations (or the equivalent finite difference equation) to establish structural properties of the optimal policy. There are two major differences between our formulation and the formulations in that field. First, a chosen price impacts both the revenue accumulation and demand process in a different way than a chosen setting impacts the revenue accumulation and deterioration process in our problem. In the pricing problem, revenue only accumulates at discrete demand epochs (i.e., at the Poisson arrivals) while in our setting, revenue is accumulated continuously. In the former, choosing a higher price leads to a *lower* demand rate but *potentially* a higher revenue rate if demand occurs, while in our problem, choosing a higher setting leads to a *higher* deterioration rate and *instantaneously* to a higher revenue rate. Second, in our problem, there is a boundary condition – the cost function $c_m(x)$ – that models the preventive maintenance cost incurred at the end of the planning horizon. This boundary condition needs to be taken into account when adopting a setting policy throughout the planning horizon to prevent high maintenance costs at the end, while in the literature on revenue management there are no such boundary conditions.

In the next section, we study the HJB equation in (4.1) to characterize structural properties of both the expected profit and the optimal setting policy.

4.3. Structural properties

In this section, we shed light on the optimal setting policy when the decision maker knows the base rate λ, or equivalently, when the population is homogeneous. We first prove that the optimal expected profit function $J^*_\lambda(x,t)$ behaves monotonically in its state variables, which we then use to prove that the optimal setting also has certain monotonicity properties. We highlight these theoretical results with numerical examples.

The lemma below establishes first order monotonic properties of $J^*_\lambda(x,t)$ with respect to x, t, and λ. These properties are not only very intuitive but also imperative for subsequent analysis in this section.

Lemma 4.2. *The optimal expected profit $J^*_\lambda(x,t)$ is decreasing in the deterioration level x, increasing in the time to the next maintenance moment t, and decreasing in the base rate λ.*

PROOF: We use sample path arguments to prove the result. For the monotonicity in x, consider a sample path where the system with deterioration level x uses the optimal policy for the system with deterioration level $x+1$ until either the system with deterioration level $x+1$ breaks down, or until the scheduled maintenance moment, whichever happens first. As they use the same policy, both systems face the same sample path. If the system with deterioration level $x+1$ breaks down at time τ before the scheduled maintenance moment, then the system with deterioration level x can still generate revenue over $(T-\tau, 0]$, possibly even with less maintenance costs if the system does not break down. On the other hand, if the scheduled maintenance moment arrives first, the two systems will have generated the same revenue and will have the same maintenance cost. Consequently, the system with deterioration level x generates at least as much revenue and incurs at most the same maintenance costs. Hence, for their expected profits, we have that $J^*_\lambda(x,t) \geq J^*_\lambda(x+1,t)$. Using similar arguments one can also prove the monotonicity in t and λ. □

Everything else fixed, Lemma 4.2 states that for a more deteriorated system, the expected profit will be lower than for a system that is less deteriorated. Similarly, everything else fixed, the expected profit is higher when the time to planned maintenance is higher, because there is more time to generate revenue. Finally, a system whose base rate is higher, deteriorates faster under any setting which leads to a lower expected profit.

We now turn our attention to higher order monotonic properties of the expected

profit with respect to its parameters, which are needed to prove structural properties of the optimal policy. We first introduce some additional notation. Let $\Delta^2_\lambda(x,t) \triangleq \Delta_\lambda(x,t) - \Delta_\lambda(x+1,t)$ for $x \leq \xi - 2$. The next result shows that the expected profit $J^*_\lambda(x,t)$ is concave in x for all $t \geq 0$ and each base rate, i.e. that $\Delta^2_\lambda(x,t) \leq 0$ for all $x \leq \xi - 2$ and $t, \lambda \geq 0$.

Proposition 4.1. *The expected profit $J^*_\lambda(x,t)$ is concave in x for all $t \geq 0$ and $\lambda \geq 0$.*

Proof: For $t = 0$, it is easy to verify that $\Delta^2_\lambda(x,0) \leq 0$ for all $x \leq \xi - 2$ and $\lambda \geq 0$, because $-c_m(x)$ is concave in x by assumption.

We now consider $t > 0$. We first construct some useful inequalities that we will use in an inductive proof to establish the result. Let $x \leq \xi - 2$, using (4.1), we have

$$\frac{\partial J^*_\lambda(x,t)}{\partial t} = r\big(s^*_\lambda(x,t)\big) - \lambda f\big(s^*_\lambda(x,t)\big)\Delta_\lambda(x,t), \tag{4.4}$$

and since $s^*_\lambda(x,t)$ is also admissible, but not necessarily optimal, for state $(x+1,t)$ with base rate λ, we have

$$\frac{\partial J^*_\lambda(x+1,t)}{\partial t} \geq r\big(s^*_\lambda(x,t)\big) - \lambda f\big(s^*_\lambda(x,t)\big)\Delta_\lambda(x+1,t). \tag{4.5}$$

Then, subtracting (4.5) from (4.4) leads to

$$\frac{\partial \Delta_\lambda(x,t)}{\partial t} \leq -\lambda f\big(s^*_\lambda(x,t)\big)\Delta^2_\lambda(x,t). \tag{4.6}$$

Similarly, we have

$$\frac{\partial J^*_\lambda(x+2,t)}{\partial t} = r\big(s^*_\lambda(x+2,t)\big) - \lambda f\big(s^*_\lambda(x+2,t)\big)\Delta_\lambda(x+2,t), \tag{4.7}$$

and since $s^*_\lambda(x+2,t)$ is also admissible, but not necessarily optimal, for state $(x+1,t)$ with base rate λ, we have

$$\frac{\partial J^*_\lambda(x+1,t)}{\partial t} \geq r\big(s^*_\lambda(x+2,t)\big) - \lambda f\big(s^*_\lambda(x+2,t)\big)\Delta_\lambda(x+1,t). \tag{4.8}$$

Then, subtracting (4.7) from (4.8) gives

$$\frac{\partial \Delta_\lambda(x+1,t)}{\partial t} \geq -\lambda f\big(s^*_\lambda(x+2,t)\big)\Delta^2_\lambda(x+1,t). \tag{4.9}$$

Next, subtracting (4.9) from (4.6) leads to

$$\frac{\partial \Delta_\lambda^2(x,t)}{\partial t} \leq -\lambda f\big(s_\lambda^*(x,t)\big)\Delta_\lambda^2(x,t) + \lambda f\big(s_\lambda^*(x+2,t)\big)\Delta_\lambda^2(x+1,t). \tag{4.10}$$

We now proceed with our inductive proof. We first prove the base case, i.e. $\Delta_\lambda^2(\xi - 2,t) \leq 0$ for all $t > 0$ and $\lambda \geq 0$. By (4.6) we have that

$$\frac{\partial \Delta_\lambda(\xi-2,t)}{\partial t} \leq -\lambda f\big(s_\lambda^*(\xi-2,t)\big)\Delta_\lambda^2(\xi-2,t). \tag{4.11}$$

Note that due to the linearity of the derivative, we have

$$\frac{\partial \Delta_\lambda(\xi-1,t)}{\partial t} = \frac{\partial J_\lambda^*(\xi-1,t)}{\partial t} - \frac{\partial J_\lambda^*(\xi,t)}{\partial t} = \frac{\partial J_\lambda^*(\xi-1,t)}{\partial t} \geq 0, \tag{4.12}$$

where the second equality follows from (4.1) for $x = \xi$ and the inequality follows from Lemma 4.2. Subtracting $\frac{\partial \Delta_\lambda(\xi-1,t)}{\partial t}$ from both sides of (4.11) and combining this with Inequality (4.12) leads to

$$\frac{\partial \Delta_\lambda^2(\xi-2,t)}{\partial t} \leq -\lambda f\big(s_\lambda^*(\xi-2,t)\big)\Delta_\lambda^2(\xi-2,t). \tag{4.13}$$

Applying Grönwall's Lemma to Inequality (4.13) leads to

$$\Delta_\lambda^2(\xi-2,t) \leq \Delta_\lambda^2(\xi-2,0) \cdot \exp\left(\int_0^t -\lambda f(s_\lambda^*(\xi-2,u))\mathrm{d}u\right).$$

Since $\Delta_\lambda^2(\xi-2,0) \leq 0$, (recall that $-c_m(x)$ is concave in x by assumption), and since $\exp\left(\int_0^t -\lambda f(s_\lambda^*(\xi-2,u))\mathrm{d}u\right) > 0$ for all $t > 0$ and $\lambda \geq 0$, we have that $\Delta_\lambda^2(\xi-2,t) \leq 0$, which proves the base case.

Assume now inductively that $\Delta_\lambda^2(x,t) \leq 0$ for an $x \leq \xi - 2$. We will show that this implies that $\Delta_\lambda^2(x-1,t) \leq 0$ for all $t > 0$ and $\lambda \geq 0$.

For $(x-1,t)$, we have

$$\begin{aligned}\frac{\partial \Delta_\lambda^2(x-1,t)}{\partial t} &\leq -\lambda f\big(s_\lambda^*(x-1,t)\big)\Delta_\lambda^2(x-1,t) + \lambda f\big(s_\lambda^*(x+1,t)\big)\Delta_\lambda^2(x,t) \\ &\leq -\lambda f\big(s_\lambda^*(x-1,t)\big)\Delta_\lambda^2(x-1,t). \end{aligned} \tag{4.14}$$

The first inequality is Equation (4.10), and the second inequality is due to the

induction hypothesis. Applying Grönwall's Lemma to Inequality (4.14) leads to

$$\Delta_\lambda^2(x-1,t) \leq \Delta_\lambda^2(x-1,0) \cdot \exp\left(\int_0^t -\lambda f(s_\lambda^*(x-1,u))\mathrm{d}u\right).$$

Since $\Delta_\lambda^2(x-1,0) \leq 0$ for all $x \leq \xi - 2$ (again by the concavity assumption of $-c_m(x)$ in x) and since $\exp\left(\int_0^t -\lambda f(s_\lambda^*(x-1,u))\mathrm{d}u\right) > 0$ for all $t > 0$ and $\lambda \geq 0$, we have that $\Delta_\lambda^2(x-1,t) \leq 0$ for all $t > 0$ and $\lambda \geq 0$. □

The concavity of the optimal expected profit in the deterioration level implies that the expected profit drops ever steeper as system deterioration is closer to the failure threshold for any fixed time. This is intuitive as for a system that is approaching the failure threshold, there is less and less room (in terms of remaining useful condition) to control the deterioration, and hence for generating revenue.

The next result establishes that the marginal value of less deterioration, increases in the remaining time until the next planned maintenance. In the proof, and also in other proofs in this chapter, we employ a commonly used transformation of the continuous model into a discrete model (see e.g., Bitran and Mondschein, 1997; Aydin and Ziya, 2008; Hu et al., 2017). We are allowed to do so because it is known that this discrete model is guaranteed to converge uniformly to the continuous model (cf. Kleywegt and Papastavrou, 2001).

Proposition 4.2. *$\Delta_\lambda(x,t)$ is increasing in t.*

Proof: We divide the entire planning horizon into $\hat{T} = \frac{T}{\delta t}$ periods, each of which is short enough that the probability of more than one degradation increment during an interval of length δt is negligible. Indeed, note that since the function f is bounded, we have that for each choice of s, the probability that more than 1 arrival will occur in an interval of δt is of $o(\delta t)$. Let $J_\lambda^*(x,k)$ denote the maximum expected revenue when the deterioration level is x and there are k periods to go until the planned maintenance moment, and the base rate is λ. We then need to show that

$$J_\lambda^*(x-1,k) - J_\lambda^*(x,k) \geq J_\lambda^*(x-1,k-1) - J_\lambda^*(x,k-1), \text{ for all } k \geq 1. \tag{4.15}$$

The dynamic program for the revenue maximization problem over the $\hat{T}$-period horizon is given by

$$J_\lambda^*(x,k) = \max_{s \in \mathscr{S}} \Big[r(s)\delta t + \lambda f(s)\delta t J_\lambda^*(x+1,k-1) + \big(1 - \lambda f(s)\delta t\big) J_\lambda^*(x,k-1) \Big],$$

with boundary conditions $J^*_\lambda(x,0) = -c_m(x)$ for all $x \in \mathscr{X}$.

Let $s^*_\lambda(x,k)$ denote the optimal setting for state (x,k) when the base rate is λ. We then have

$$\begin{aligned} J^*_\lambda(x,k) =& r\big(s^*_\lambda(x,k)\big)\delta t + \lambda f\big(s^*_\lambda(x,k)\big)\delta t J^*_\lambda(x+1,k-1) \\ &+ \big(1-\lambda f\big(s^*_\lambda(x,k)\big)\delta t\big)J^*_\lambda(x,k-1). \end{aligned} \tag{4.16}$$

Subtracting $J^*_\lambda(x,k-1)$ from both sides of (4.16) leads to

$$\begin{aligned} J^*_\lambda(x,k) - J^*_\lambda(x,k-1) =& r\big(s^*_\lambda(x,k)\big)\delta t \\ &+ \lambda f\big(s^*_\lambda(x,k)\big)\delta t\big(J^*_\lambda(x+1,k-1) - J^*_\lambda(x,k-1)\big). \end{aligned} \tag{4.17}$$

Since $s^*_\lambda(x,k)$ is admissible in state $(x-1,k)$ when the base rate is λ but not necessarily optimal, we have

$$\begin{aligned} J^*_\lambda(x-1,k) \geq& r\big(s^*_\lambda(x,k)\big)\delta t + \lambda f\big(s^*_\lambda(x,k)\big)\delta t J^*_\lambda(x,k-1) \\ &+ \big(1-\lambda f\big(s^*_\lambda(x,k)\big)\delta t\big)J^*_\lambda(x-1,k-1). \end{aligned} \tag{4.18}$$

Subtracting $J^*_\lambda(x-1,k-1)$ from both sides of (4.18) and then using the concavity of $J^*_\lambda(x,k)$ in x (see Proposition 4.1) yields

$$\begin{aligned} & J^*_\lambda(x-1,k) - J^*_\lambda(x-1,k-1) \\ & \qquad \geq \\ & r\big(s^*_\lambda(x,k)\big)\delta t + \lambda f\big(s^*_\lambda(x,k)\big)\delta t\big(J^*_\lambda(x,k-1) - J^*_\lambda(x-1,k-1)\big) \\ & \qquad \geq \\ & r\big(s^*_\lambda(x,k)\big)\delta t + \lambda f\big(s^*_\lambda(x,k)\big)\delta t\big(J^*_\lambda(x+1,k-1) - J^*_\lambda(x,k-1)\big). \end{aligned} \tag{4.19}$$

Observe that the right hand side of (4.19) can be replaced by (4.17), which yields the following inequality

$$J^*_\lambda(x-1,k) - J^*_\lambda(x-1,k-1) \geq J^*_\lambda(x,k) - J^*_\lambda(x,k-1),$$

which, after rearranging terms, equals (4.15). □

The intuition behind Proposition 4.2 is as follows. Consider a system with a better condition (i.e., with one deterioration level less) than another system. The former generates more revenue than the latter, see Lemma 4.2. Now, if the scheduled maintenance moment is further away, then for the former system, there is a longer

planning horizon to exploit this better condition – for generating revenue – than when there is less time to do so. This leads to a higher additional revenue.

Using a similar intuition one can also reason that the optimal expected profit must be concave in the time to the next scheduled maintenance moment for all deterioration levels and each base rate. This is indeed true and we present this result in the next proposition.

Proposition 4.3. *The optimal expected profit $J^*_\lambda(x,t)$ is concave in t for all $x \in \mathscr{X}$ and $\lambda \geq 0$.*

PROOF: We need to show that $\frac{\partial J^*_\lambda(x,t)}{\partial t}$ is decreasing in t. Let $u > t$. We have

$$\frac{\partial J^*_\lambda(x,u)}{\partial u} = r\big(s^*_\lambda(x,u)\big) - \lambda f\big(s^*_\lambda(x,u)\big)\Delta_\lambda(x,u), \tag{4.20}$$

and since $s^*_\lambda(x,u)$ is also admissible, but not necessarily optimal, for state (x,t) when the base rate is λ, we have

$$\frac{\partial J^*_\lambda(x,t)}{\partial t} \geq r\big(s^*_\lambda(x,u)\big) - \lambda f\big(s^*_\lambda(x,u)\big)\Delta_\lambda(x,t). \tag{4.21}$$

Subtracting (4.21) from (4.20) yields

$$\frac{\partial J^*_\lambda(x,t)}{\partial t} - \frac{\partial J^*_\lambda(x,u)}{\partial u} \geq \lambda f\big(s^*_\lambda(x,u)\big)\Big(\Delta_\lambda(x,u) - \Delta_\lambda(x,t)\Big) \geq 0,$$

where the last inequality follows from $\Delta_\lambda(x,u) \geq \Delta_\lambda(x,t)$ for $u > t$. (see Proposition 4.2). □

We now proceed with proving that the optimal expected profit is submodular in (x,λ), which enables us to establish the monotonic behavior of the optimal policy in the base rate λ. We first provide a formal definition of submodularity in the definition below.

Definition 4.1 (Submodularity)**.** A bivariate function, $g(x,y)$, is submodular if for all $x_1 \geq x_2$ and $y_1 \geq y_2$, we have

$$\big[g(x_1,y_2) + g(x_2,y_1)\big] - \big[g(x_1,y_1) + g(x_2,y_2)\big] \geq 0.$$

We now state two technical results that relate to preservation of submodularity.

Lemma 4.3. *Suppose $f(x,\theta)$ is submodular in (x,θ) and concave in x and the non-negative random variable $Z(\theta)$ is stochastically increasing in θ, then $\mathbb{E}[f(x+Z(\theta),\theta)]$ is submodular in (x,θ) and concave in x.*

PROOF: Let $x_1 \leq x_2$ and $\theta_1 \leq \theta_2$. Since $Z(\theta_1) \leq_{\text{st}} Z(\theta_2)$, there exist two random variables, $\tilde{Z}(\theta_1)$ and $\tilde{Z}(\theta_2)$, on the same probability space, such that $\tilde{Z}(\theta_1) =_{\text{st}} Z(\theta_1)$, $\tilde{Z}(\theta_2) =_{\text{st}} Z(\theta_2)$, and $\tilde{Z}(\theta_1) \leq \tilde{Z}(\theta_2)$ almost surely (here, $=_{\text{st}}$ denotes equality in law) (see, e.g., Shaked and Shanthikumar, 2007, Theorem 1.A.1.). We therefore have:

$$\begin{aligned}
\mathbb{E}[f(x_1 + Z(\theta_1), \theta_1)] - \mathbb{E}[f(x_2 + Z(\theta_1), \theta_1)] & \\
&= \mathbb{E}[f(x_1 + \tilde{Z}(\theta_1), \theta_1) - f(x_2 + \tilde{Z}(\theta_1), \theta_1)] \\
&\leq \mathbb{E}[f(x_1 + \tilde{Z}(\theta_1), \theta_2) - f(x_2 + \tilde{Z}(\theta_1), \theta_2)] \\
&\leq \mathbb{E}[f(x_1 + \tilde{Z}(\theta_2), \theta_2) - f(x_2 + \tilde{Z}(\theta_2), \theta_2)] \\
&= \mathbb{E}[f(x_1 + Z(\theta_2), \theta_2)] - \mathbb{E}[f(x_2 + Z(\theta_2), \theta_2)],
\end{aligned}$$

where the first inequality is from the submodularity of $f(x, \theta)$, and the second inequality is from the concavity in x of $f(x, \theta)$. Hence, $\mathbb{E}[f(x + Z(\theta), \theta)]$ is submodular in (x, θ). The preservation of concavity under expectation is a well-known result. For completeness, we provide a short proof here. Let $x_1, x_2 \in \mathbb{R}$ and $\alpha \in [0, 1]$. We then have

$$\begin{aligned}
\mathbb{E}[f(\alpha x_1 + (1-\alpha)x_2 + Z(\theta), \theta)] &= \mathbb{E}[f(\alpha x_1 + (1-\alpha)x_2 + \alpha Z(\theta) + (1-\alpha)Z(\theta), \theta)] \\
&= \mathbb{E}[f(\alpha\big(x_1 + Z(\theta)\big) + (1-\alpha)\big(x_2 + Z(\theta)\big), \theta)] \\
&\geq \mathbb{E}[\alpha f\big(x_1 + Z(\theta), \theta\big) + (1-\alpha) f\big(x_2 + Z(\theta), \theta\big)] \\
&= \alpha \mathbb{E}[f\big(x_1 + Z(\theta), \theta\big)] + (1-\alpha)\mathbb{E}[f\big(x_2 + Z(\theta), \theta\big)],
\end{aligned}$$

where the inequality follows from the concavity of $f(x, \theta)$ in x. □

Lemma 4.3 complements other preservation results – usually needed for establishing structural properties in optimization problems – that can be found in the literature, for instance a closely related result by Chao et al. (2009) that focuses on $\mathbb{E}[f\left(x, Z(\theta)\right)]$ (in our notation).

Lemma 4.4. *Suppose $g(y)$ is non-negative, and increasing in y, and $f(x, y)$ is non-negative, submodular in (x, y), and decreasing in x, then $g(y)f(x, y)$ is submodular in (x, y).*

PROOF: Let $x_1 \leq x_2$ and $y_1 \leq y_2$, we then have

$$\begin{aligned}
&g(y_1)f(x_1,y_1) - g(y_2)f(x_1,y_2) - g(y_1)f(x_2,y_1) + g(y_2)f(x_2,y_2) \\
&= g(y_1)\big(f(x_1,y_1) - f(x_2,y_1)\big) - g(y_2)\big(f(x_1,y_2) - f(x_2,y_2)\big) \\
&\leq g(y_2)\big(f(x_1,y_1) - f(x_2,y_1) - f(x_1,y_2) + f(x_2,y_2)\big) \\
&\leq 0.
\end{aligned}$$

The first inequality holds because $g(y_1) \leq g(y_2)$ and $f(x_1,y_1) \geq f(x_2,y_1)$, and the second inequality follows from the submodularity of $f(x,y)$ in (x,y). □

Proposition 4.4. *For all $t \geq 0$, $J_\lambda(x,t)$ is submodular in (x,λ).*

PROOF: Again, we use the dynamic program for the revenue maximization problem over the $\hat{T}$-period horizon and rewrite it in expectation form:

$$J^*_\lambda(x,k) = \max_{s\in\mathscr{S}} \Big[r(s)\delta t + \mathbb{E}[J^*_\lambda(x + Z(\lambda,s), k-1)] \Big],$$

with boundary conditions $J^*_\lambda(x,0) = -c_m(x)$ for all $x \in \mathscr{X}$, and where $Z(\lambda,s)$ is a Bernoulli random variable with success probability $p = \delta t \lambda f(s)$.

We need to show that for all $k \geq 0$, $J^*_\lambda(x,k)$ is submodular in (x,λ). We prove this by induction on k. Note that for the base case $k = 0$, $J^*_\lambda(x,0) = -c_m(x)$ is evidently submodular in (x,λ). Assume that for some $k-1$, $J_\lambda(x,k-1)$ is submodular in (x,λ), we will now show that $J_\lambda(x,k)$ is submodular in (x,λ).

Let $g_\lambda(s,x,k) \triangleq r(s)\delta t + \mathbb{E}[J^*_\lambda(x + Z(\lambda,s), k-1)]$. Since $Z(\lambda,s)$ is stochastically increasing in λ, the induction hypothesis and the fact that $J^*_\lambda(x,k-1)$ is concave in x (see Proposition 4.1), we can employ Lemma 4.3 to conclude that $g_\lambda(s,x,k)$ is submodular in (x,λ) for all s.

Now, let $s^*_\lambda(x,k) \in \arg\max_{s\in\mathscr{S}}\{g_\lambda(s,x,k)\}$, and let $x_1 < x_2$, and $\lambda_1 < \lambda_2$, we then have:

$$\begin{aligned}
J^*_{\lambda_1}(x_1,k) + J^*_{\lambda_2}(x_2,k) &= g_{\lambda_1}\big(s^*_{\lambda_1}(x_1,k), x_1, k\big) + g_{\lambda_2}\big(s^*_{\lambda_2}(x_2,k), x_2, k\big) \\
&\leq g_{\lambda_2}\big(s^*_{\lambda_1}(x_1,k), x_1, k\big) + g_{\lambda_1}\big(s^*_{\lambda_2}(x_2,k), x_2, k\big) \\
&\leq g_{\lambda_2}\big(s^*_{\lambda_2}(x_1,k), x_1, k\big) + g_{\lambda_1}\big(s^*_{\lambda_1}(x_2,k), x_2, k\big) \\
&= J^*_{\lambda_2}(x_1,k) + J^*_{\lambda_1}(x_2,k),
\end{aligned}$$

where the first inequality holds because of the submodularity of $g_\lambda(s,x,k)$ in (x,λ) for all s, and the second inequality follows from the suboptimality of $s^*_{\lambda_1}(x_1,k)$ for

(x_1, λ_2) and $s^*_{\lambda_2}(x_2, k)$ for (x_2, λ_1), respectively. Hence $J^*_\lambda(x, k)$ is submodular in (x, λ). □

The next result, the main result of this section, establishes the monotonicity of the optimal setting in the deterioration level x, the time to maintenance t, and the base rate λ.

Theorem 4.1. *The optimal setting when the deterioration level is x, the time until planned maintenance is t, and the base rate is λ, denoted with $s^*_\lambda(x, t)$, is decreasing in x, t, and λ.*

Proof: The proof is divided into three parts, focused on the monotonicity in x, t, and λ, respectively. In each part we use the notation $g_\lambda(s, x, t) = r(s) - \lambda f(s)\Delta_\lambda(x, t)$, that we previously introduced.

Proof of monotonicity in x: For notational clarity we drop the subscript λ. Let $\bar{s} \triangleq s^*(x, t)$ and consider $s > \bar{s}$, we have that

$$\begin{aligned}
&g(\bar{s}, x+1, t) - g(\bar{s}, x, t) + g(s, x, t) - g(s, x+1, t) \\
&= -\lambda f(\bar{s})\Delta(x+1, t) + \lambda f(\bar{s})\Delta(x, t) - \lambda f(s)\Delta(x, t) + \lambda f(s)\Delta(x+1, t) \\
&= \lambda f(\bar{s})\Delta^2(x, t) - \lambda f(s)\Delta^2(x, t) \\
&= \lambda\big(f(\bar{s}) - f(s)\big)\Delta^2(x, t) \\
&\geq 0. \qquad (4.22)
\end{aligned}$$

The inequality follows from the increasing property of $f(s)$ with $s > \bar{s}$, and since $\Delta^2(x, t) \leq 0$ (see Proposition 4.1). Rearranging Inequality (4.22) leads to

$$g(\bar{s}, x+1, t) - g(s, x+1, t) \geq g(\bar{s}, x, t) - g(s, x, t) \geq 0, \qquad (4.23)$$

where the second inequality holds because $\bar{s} \triangleq s^*(x, t)$. We can therefore conclude from (4.23), that any $s > \bar{s}$ cannot be optimal for $(x+1, t)$; hence $s^*(x+1, t) \leq s^*(x, t)$.

Proof of monotonicity in t: For notational clarity we again drop the subscript λ.

Let $u < t$ and $\bar{s} \triangleq s^*(x,u)$, and consider $s > \bar{s}$. We have that

$$\begin{aligned} & g(\bar{s},x,t) - g(\bar{s},x,u) + g(s,x,u) - g(s,x,t) \\ &= -\lambda f(\bar{s})\Delta(x,t) + \lambda f(\bar{s})\Delta(x,u) - \lambda f(s)\Delta(x,u) + \lambda f(s)\Delta(x,t) \\ &= \lambda\big(f(\bar{s}) - f(s)\big)\big(\Delta(x,u) - \Delta(x,t)\big) \\ &\geq 0. \end{aligned} \tag{4.24}$$

The inequality follows from the increasing property of $f(s)$ and $s > \bar{s}$, and since $\Delta(x,u) \leq \Delta(x,t)$ for $u < t$ (see Proposition 4.2). Rearranging Inequality (4.24) leads to

$$g(\bar{s},x,t) - g(s,x,t) \geq g(\bar{s},x,u) - g(s,x,u) \geq 0, \tag{4.25}$$

where the second inequality holds because $\bar{s} \triangleq s^*(x,u)$. We can therefore conclude from (4.25), that any $s > \bar{s}$ cannot be optimal for (x,t); hence $s^*(x,t) \leq s^*(x,u)$ for $u < t$.

Proof of monotonicity in λ: In this part, we use the analog discrete dynamic program. In light of the well-known Topkis' Theorem (see, for instance, Topkis, 1998, Theorem 2.8.1), we need to show that $g_\lambda(s,x,k) = r(s)\delta t + \lambda f(s)\delta t J^*_\lambda(x+1,k-1) + (1-\lambda f(s)\delta t)J^*_\lambda(x,k-1)$ is submodular in (s,λ) to conclude that $s^*_\lambda(x,k)$ is decreasing in λ.

Let $s_1 \leq s_2$ and $\lambda_1 \leq \lambda_2$. We then need to show that

$$g_{\lambda_1}(s_1,x,k) - g_{\lambda_2}(s_1,x,k) - g_{\lambda_1}(s_2,x,k) + g_{\lambda_2}(s_2,x,k) \leq 0.$$

We have (after some algebraic manipulations):

$$\begin{aligned} & g_{\lambda_1}(s_1,x,k) - g_{\lambda_2}(s_1,x,k) - g_{\lambda_1}(s_2,x,k) + g_{\lambda_2}(s_2,x,k) \\ &= \lambda_1 f(s_1)J^*_{\lambda_1}(x+1,k-1) - \lambda_1 f(s_1)J^*_{\lambda_1}(x,k-1) \\ &\quad - \lambda_2 f(s_1)J^*_{\lambda_2}(x+1,k-1) + \lambda_2 f(s_1)J^*_{\lambda_2}(x,k-1) \\ &\quad - \lambda_1 f(s_2)J^*_{\lambda_1}(x+1,k-1) + \lambda_1 f(s_2)J^*_{\lambda_1}(x,k-1) \\ &\quad + \lambda_2 f(s_2)J^*_{\lambda_2}(x+1,k-1) - \lambda_2 f(s_2)J^*_{\lambda_2}(x,k-1) \\ &= \big(f(s_1) - f(s_2)\big)\big(\lambda_1 J^*_{\lambda_1}(x+1,k-1) - \lambda_1 J^*_{\lambda_1}(x,k-1) \\ &\quad - \lambda_2 J^*_{\lambda_2}(x+1,k-1) + \lambda_2 J^*_{\lambda_2}(x,k-1)\big). \end{aligned}$$

Since $f(s_1) - f(s_2) \leq 0$ (f is increasing and $s_2 \geq s_1$), $g_\lambda(s,x,k)$ is submodular

in (s, λ) if

$$\lambda_1 J^*_{\lambda_1}(x+1, k-1) - \lambda_1 J^*_{\lambda_1}(x, k-1) - \lambda_2 J^*_{\lambda_2}(x+1, k-1) + \lambda_2 J^*_{\lambda_2}(x, k-1) \geq 0,$$

that is, if $\lambda J^*_\lambda(x, k-1)$ is submodular in (x, λ).

Note that $J^*_\lambda(x, k-1)$ is submodular in (x, λ) (see Proposition 4.4) and decreasing in x (see Lemma 4.2), hence by applying Lemma 4.4 with $g(\lambda) \triangleq \lambda$ and $f(x, \lambda) \triangleq J^*_\lambda(x, k-1)$, we conclude that $\lambda J^*_\lambda(x, k-1)$ is submodular in (x, λ).

□

The intuition behind Theorem 4.1 is as follows. For a fixed deterioration level, a decision maker would decrease (increase) the setting if the remaining time to the scheduled maintenance moment increases (decreases). Likewise, for a fixed remaining time, a decision maker would decrease (increase) the setting if the degradation level increases (decreases). This ensures that the system deteriorates more slowly (faster), thereby reducing the expected maintenance cost at the scheduled maintenance moment (increasing the accumulated revenue). If a system deteriorates faster under any setting (i.e. the base rate is higher), then we would naturally use lower settings for all deterioration levels and remaining times than for a system that deteriorates more slowly.

Remark 4.1. The established structure of the optimal policy is inherently different from that of the optimal policy under deterministic degradation. For the latter setting, Uit het Broek et al. (2020) establish that the optimal policy may prescribe to break down the system on purpose when a failure is unavoidable, breaking down any monotonic properties of the optimal policy. This is in sharp contrast with our result, where a decision maker would gradually slow down the deterioration when a potential failure is nearing. However, Theorem 4.1 does formally confirm their reported intuitions – based on numerical explorations – on the structure of the optimal policy for the stochastic setting (see Uit het Broek et al., 2020, Section 5.3). ◇

The monotonic behavior of the optimal setting policy is illustrated by Figure 4.1 and Figure 4.2. In both figures, we use $\xi = 10$ (failure threshold), $c_m(\xi) = 5$ (corrective maintenance cost), and $c_m(\cdot) = 1$ for all non-failed states (preventive maintenance cost). The planning horizon T is equal to 15 and the setting space is equal to the unit interval, i.e. $\mathscr{S} = [0, 1]$. For the revenue and degradation functions, we use $r(s) = s^{1/2}$ and $f(s) = s^2$. We obtained the optimal policy in these illustrations,

and elsewhere in this chapter, using the standard approach for solving continuous time dynamic programs. In this approach, we discretize the planning horizon $[T, 0]$ into $N \triangleq T/\Delta t$ time intervals of length Δt and use a finite difference equation to approximate the optimality equation in (4.1) (see, e.g., Kushner and Dupuis, 1992). Specifically, we replace the partial derivative in Equation (4.1) by the finite difference $\big(J^*_\lambda(x, t + \Delta t) - J^*_\lambda(x, t)\big)/\Delta t$, after which we obtain the finite difference equation:

$$J^*_\lambda(x, t + \Delta t) =$$
$$\begin{cases} J^*_\lambda(x,t) + \Delta t \cdot \max_{s \in \mathscr{S}} \Big[r(s) - \lambda f(s) \big(J^*_\lambda(x,t) - J^*_\lambda(x+1,t) \big) \Big], & \text{if } x \in \mathscr{X} \setminus \xi, \\ J^*_\lambda(x,t), & \text{if } x = \xi, \end{cases}$$

which can be solved by a backward recursion on the discrete time set $n\Delta t$, with $n \in \{0, 1, \ldots, N\}$, starting from the boundary condition $J^*_\lambda(x, 0) = -c_m(x)$ for all $x \in \mathscr{X}$. In all our numerical experiments and illustrations we use a sufficiently small Δt, such that the approximation is sufficiently accurate.

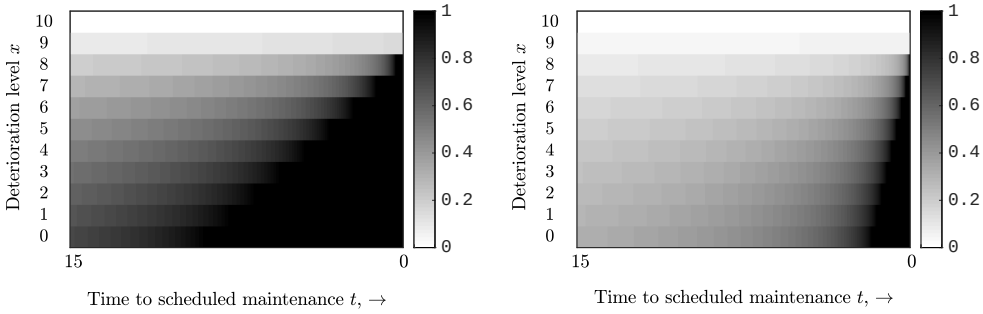


Figure 4.1 Illustration of monotonic behavior of the optimal policy with base rate $\lambda = 1$ (left) and base rate $\lambda = 4$ (right). The y-axis and x-axis represent the deterioration level and the time to scheduled maintenance, respectively, while the color-bar indicates the optimal setting.

In the left subfigure of Figure 4.1 the base rate is 1, while in the right subfigure, the base rate is 4. In each subfigure, we clearly observe that the optimal setting increases as the time to scheduled maintenance decreases for a fixed deterioration level (moving right horizontally), and that the optimal setting decreases as the condition deteriorates (moving up vertically). If we compare the left subfigure (base rate is 1) with the right subfigure (base rate is 4), we clearly see that lower settings are prescribed for all deterioration levels at all times in the situation displayed in the right subfigure.

Figure 4.2 depicts sample path illustrations of both the optimal setting policy

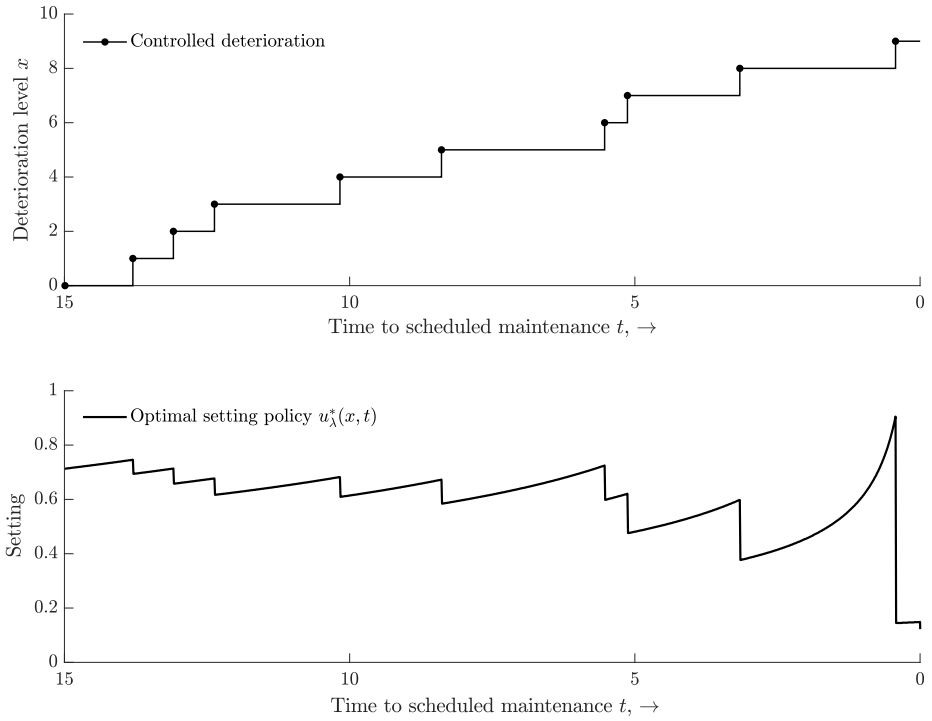


Figure 4.2 Sample path illustrations of both the optimal policy and the controlled deterioration process, when the base rate λ is 1.

(bottom) and the corresponding controlled deterioration (top). Since the downward jumps in the settings (bottom subfigure) correspond one-to-one to deterioration increments (top subfigure), we can state two interesting observations that demonstrate Theorem 4.1:

1. When the system's condition deteriorates (i.e., at the jump epochs of the controlled deterioration), the setting is immediately lowered, which highlights the monotonicity of the optimal setting in the deterioration level when the time to scheduled maintenance is kept fixed.

2. The upward creeping nature of the chosen setting between deterioration increments corresponds to the monotonic behavior of the optimal setting in the time to scheduled maintenance when the deterioration level is kept constant.

Theorem 4.1 characterizes the structure of the optimal policy for dealing with the

operational trade-off between revenue accumulation and deterioration. In practice, a decision maker also needs to decide on the planned maintenance interval T in which she deals with another trade-off that is more on a tactical level. Specifically, doing planned maintenance too often (a small T) leads to unnecessary maintenance activities and associated costs, while a very long maintenance interval (a large T) both increases the risk of a failed system and decreases the revenue accumulation as the decision maker is then forced to choose lower settings. In deciding on the length of this maintenance interval, a decision maker seeks to maximize the expected profit, by incorporating both the operational and tactical trade-off. The next result establishes that if an optimal setting policy is followed on an operational level, and T^* is a local maximizer of the average revenue per time unit, then we are guaranteed that this T^* is the global maximizer.

Theorem 4.2. *Let $g(T) \triangleq \frac{1}{T} J^*_\lambda(0,T)$ be the expected profit per time unit if the planned maintenance interval is set to T. If T^* is a local maximizer of $g(T)$, then T^* is the global maximizer.*

PROOF: We use the shorthand notation $J(T) \triangleq J^*_\lambda(0,T)$. Taking the first and second derivative of $g(T)$ with respect to T yields

$$g'(T) = \frac{TJ'(T) - J(T)}{T^2}, \tag{4.26}$$

and, after some algebraic simplifications,

$$g''(T) = \frac{T^2 J''(T) - 2 \cdot \big(TJ'(T) - J(T)\big)}{T^3}. \tag{4.27}$$

Now observe that for any t with $g'(t) = 0$, we have that $tJ'(t) - J(t) = 0$ (see Equation (4.26)), and hence that $g''(t) = \frac{t^2 J''(t) - 2 \cdot 0}{t^3} = \frac{J''(t)}{t}$ (see Equation (4.27)). We further know that $J(t)$ is concave in t, i.e. $J''(t) \leq 0$ for all $t \geq 0$ (see Proposition 4.3), so that the following condition holds

$$g''(t) \leq 0, \text{ for all } t \geq 0 \text{ such that } g'(t) = 0. \tag{4.28}$$

We now proceed with the proof. Let T^* be a local maximizer of $g(T)$. Let us assume the contrary of the result, i.e. that T^* is not a global maximizer. We will show that this leads to a contradiction with Condition (4.28) and conclude that T^* must be a global maximizer of $g(T)$.

Let T' denote the global maximizer and consider the case that $T' > T^*$ (the proof follows verbatim for the case that $T' < T^*$). By the mean value theorem, we know

that there exists a $t_0 \in (T^*, T']$ with

$$g'(t_0) = \frac{g(T') - g(T^*)}{T' - T^*} > 0, \tag{4.29}$$

where the inequality holds because $g(T') > g(T^*)$ since T' is the global maximizer and $T' > T^*$. As T^* is a local maximizer, we know that for small $\epsilon > 0$, we have $g'(T^* + \epsilon) < 0$. Hence, by Darboux's theorem (the mean value theorem for derivatives), we know that since $g'(T^* + \epsilon) < 0$ and $g'(t_0) > 0$ with $T^* + \epsilon < t_0$ (see Equation (4.29)), that there exists a $t_1 \in [T^* + \epsilon, t_0]$ with $g'(t_1) = 0$, $g'(t_1 - \epsilon) < 0$, and $g'(t_1 + \epsilon) > 0$ and hence that $g''(t_1) > 0$. This contradicts with Condition (4.28), and we can therefore conclude that T^* must be a global maximizer. □

Setting choices on an operational level and maintenance decisions on a tactical level should be made in conjunction rather than in isolation, and Theorem 4.2 provides us with a way to do so. Specifically, readily available methods that make use of the property established in Theorem 4.2 (e.g., the well-known golden-section search proposed by Kiefer (1953)) can be used to efficiently find the optimal length of the maintenance interval.

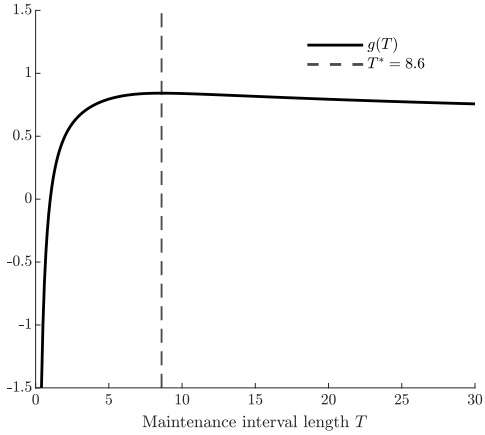


Figure 4.3 Expected profit per time unit as function of the maintenance interval length when the base rate is 1. The optimal maintenance interval is depicted with the dashed line.

Figure 4.3 provides an illustration of this. In this example, where we use the same settings as in Figures 4.1 and 4.2 with the base rate equal to 1, the optimal maintenance interval length is equal to 8.6 with an expected profit rate of 0.84.

4.4. Numerical study

In this section, we summarize the findings of two comprehensive numerical studies that highlight the practical value of our theoretical results. Specifically, the goal is twofold. Firstly, to assess the value of dynamically adjusting the settings based on the system's deterioration and time to scheduled maintenance (i.e. the optimal condition-based setting policy) by comparing it to the best static policy. Secondly, to investigate the value of integrating maintenance and setting decisions as opposed to treating the two problems independently. In both numerical studies, we choose the canonical maintenance cost structure with a fixed preventive maintenance cost $c_m(\cdot) \triangleq c_p$ for all non-failed states, and a corrective maintenance cost $c_m(\xi) \triangleq c_u$ for the failed state.

4.4.1 Value of condition-based setting policies

We assess the value of employing the optimal condition-based setting policy by comparing it to a static policy. In this static policy, the decision maker needs to decide on a fixed setting, which is employed throughout the whole planning horizon. Specifically, the system runs on the fixed setting until either failure or the moment of planned maintenance, whichever comes first. We call this static policy the fixed-setting policy, denoted by π_{FS}, and it can be found by solving the following maximization problem:

$$\pi_{FS} \in \arg\max_{s \in \mathscr{S}} \Big[\mathbb{E}\big[r(s) \cdot \min\{T_\xi, T\}\big] - c_p \Big],$$

where T_ξ is an Erlang random variable with shape ξ and rate $\lambda f(s)$, and T is the length of the maintenance interval. The first part of the objective function is the expected accumulated revenue, while the second part is the maintenance cost at time T. The maximization problem can be easily solved numerically to obtain π_{FS} using readily available software packages.

As indicated, we would like to assess the value of using the optimal condition-based setting policy, denoted by π_{DS}, instead of the static fixed-setting policy π_{FS}. We do so by using the relative profit increase, denoted with $\mathcal{R}$, which we compute as $\mathcal{R} = 100 \cdot (P_{\pi_{DS}} - P_{\pi_{FS}})/P_{\pi_{FS}}$, where P_π is the expected profit (over the planning horizon) of policy $\pi \in \{\pi_{DS}, \pi_{FS}\}$.

We perform an extensive numerical study with the following parameter settings. We use base rates $\lambda \in \{0.5, 0.75, 1, 1.25, 1.5\}$, failure thresholds $\xi \in \{10, 12, 14, 18, 20\}$,

and planning horizons $T \in \{10, 20\}$. We use the functions $f = s^{\gamma}$ and $r = s^{\nu}$, for the degradation and revenue function, respectively, and use the interval $\mathscr{S} = [0, 2]$ for the settings to choose from. By using the values $\{0.5, 0.75, 1, 1.33, 2\}$ for γ and ν, we model concave, linear, and convex functions, respectively. We vary the preventive maintenance cost c_p between $\{1, 2, 3\}$, and choose $c_m(\xi) = 10$ as corrective maintenance cost for the failed state. The complete set of parameter values is summarized in Table 4.1. We perform a full factorial numerical study, which results in 3750 instances.

Table 4.1 Input parameter values for numerical study.

	Input parameter	No. of choices	Values
1	Base rate, λ	5	0.5, 0.75, 1, 1.25, 1.5
2	Preventive maintenance cost, c_p	3	1, 2, 3
3	Failure threshold, ξ	5	10, 12, 14, 18, 20
4	Length of planning horizon, T	2	10, 20
5	Parameter of degradation function, γ	5	0.5, 0.75, 1, 1.33, 2
6	Parameter of revenue function, ν	5	0.5, 0.75, 1, 1.33, 2

The results of the numerical study are summarized in Table 4.2. In this table, we present the average and standard deviation of $\mathcal{R}$, and the $P_{\pi_{DS}}$, where we first distinguish between subsets of instances with the same value for a specific input parameter of Table 4.1, and then present the results for all instances.

From the results in Table 4.2, we can state the following main observation. Condition-based settings based on the system's deterioration and time to maintenance, instead of using the best fixed-setting policy, can lead to significant profit increases, with an average increase of 50.42%. The huge financial advantage is mainly because of two mechanisms. First, the condition-based setting policy is able to prevent failures by slowing down the deterioration when the failure threshold is nearing, thereby saving on high maintenance costs. This is in sharp contrast with the fixed-setting policy, which would just continue operating on the fixed setting until failure. Second, when the scheduled maintenance moment is nearing and the system is still in good condition, the condition-based setting policy would exploit the remaining useful condition for accumulating revenue by setting a higher setting. Such an opportunity is completely ignored by the fixed-setting policy. Based on the results in Table 4.2, we can conclude that these two mechanisms are especially important for systems that have a low expected time to failure, either through a high base rate or a low failure threshold.

Table 4.2 Relative profit increase (in %) of condition-based compared to static setting policies.

		$\mathcal{R}$		
Input parameter	Value	Mean	SD	$P_{\pi_{DS}}$
Base rate, λ	0.5	15.18	11.46	33.74
	0.75	26.82	18.26	24.89
	1	41.58	34.97	19.39
	1.25	72.31	133.78	15.73
	1.5	96.23	425.59	13.16
Preventive maintenance cost, c_p	1	55.94	262.30	22.38
	2	46.11	188.32	21.38
	3	49.23	136.83	20.39
Failure threshold, ξ	10	92.22	427.09	14.65
	12	66.07	129.02	17.70
	14	42.56	44.98	20.59
	18	27.65	21.93	25.81
	20	23.63	17.87	28.16
Length of planning horizon, T	10	47.94	151.18	19.47
	20	52.91	243.17	23.29
Parameter of degradation function, γ	0.5	67.08	376.91	31.59
	0.75	61.19	173.84	25.54
	1	48.59	168.19	20.75
	1.33	41.02	45.08	16.42
	2	34.25	42.68	12.60
Parameter of revenue function, ν	0.5	28.95	48.01	12.14
	0.75	32.31	43.44	13.35
	1	53.79	169.41	15.83
	1.33	67.49	175.60	21.49
	2	69.58	374.14	44.09
Total		50.42	202.46	21.38

4.4.2 Value of integrating maintenance and setting policies

In Theorem 4.2 we show that one can optimize the length of the maintenance interval when the optimal condition-based setting policy is followed in between those maintenance moments. This allows decision makers to jointly decide on the two policies (i.e., maintenance and setting policies) in an integrated manner. In practice however, these two decisions are often made sequentially; that is, first the tactical decision of choosing a maintenance interval is made and then, given that maintenance interval, operational choices pertaining to settings of the system are made. We now describe how this approach is usually carried out.

First, the problem of selecting the optimal maintenance interval without taking into account the setting policy and using only the knowledge of the base rate falls in the class of the classical age-based maintenance problem (Barlow and Hunter, 1960). Note

that the lifetime distribution under the base rate is an Erlang random variable with shape ξ and rate λ. Hence, if a maintenance interval of length t is selected, then this results in an average maintenance cost rate, denoted with $g_m(t)$, of

$$g_m(t) = \frac{c_p + (c_u - c_p) \cdot \mathbb{P}\big[T_\xi \leq t\big]}{\mathbb{E}\big[\min\{T_\xi, t\}\big]}, \tag{4.30}$$

where the numerator is the expected maintenance cost, and the denominator is the expected time to maintenance (either preventive or corrective). It is well-known that in this setting, the function $g_m(t)$ has a unique minimizer that we can easily compute (Barlow and Hunter, 1960).

Next, given this optimal maintenance interval length, we can then use the optimal setting policy to control the system's deterioration in between those maintenance moments, which yields an expected profit rate per time unit. We call this approach the sequential approach, denoted with π_S, and denote its resulting average profit rate with $\hat{P}_{\pi_S}$. Note that if t^* is the minimizer of Equation (4.30), then $\hat{P}_{\pi_S}$ is equal to $J^*_\lambda(0, t^*)/t^*$.

Alternatively we can use Theorem 4.2 and determine the optimal length of the maintenance interval when taking into account that the optimal condition-based setting policy is followed in between those maintenance moments. We call this approach the integrated approach, denoted with π_I, and denote its resulting average profit rate with $\hat{P}_{\pi_I}$.

Since we would like to assess the value of integrating maintenance and setting policies, we report on its relative profit rate increase, denoted with $\hat{\mathcal{R}}$, which we compute as $\hat{\mathcal{R}} = 100 \cdot (\hat{P}_{\pi_I} - \hat{P}_{\pi_S})/\hat{P}_{\pi_S}$, where $\hat{P}_\pi$ is the expected profit rate per time unit of approach $\pi \in \{\pi_S, \pi_I\}$. We perform a full factorial numerical study using the same parameters of the previous numerical study as displayed in Table 4.1 (excluding the parameter T for obvious reasons), which results in 1875 instances. We summarize the results of this numerical study in Table 4.3. Again, we present the average and standard deviation of $\hat{\mathcal{R}}$, and the $\hat{P}_{\pi_I}$, where we first distinguish between subsets of instances with the same value for a specific input parameter of Table 4.1, and then present the results for all instances.

The following main observations can be drawn from the results in Table 4.3. First, integrating maintenance and setting decisions leads to significant profit rate increases as opposed to treating them sequentially, with an average increase of 21.39%. We emphasize that this increase is solely due to this integrated approach since the optimal condition-based setting policy is applied on an operational level in both approaches.

Table 4.3 Relative profit rate increase (in %) of integrating maintenance and setting decisions compared to treating them sequentially.

		$\hat{\mathcal{R}}$		
Input parameter	Value	Mean	SD	$\hat{P}_{\pi_I}$
Base rate, λ	0.5	26.18	30.00	2.04
	0.75	23.22	28.99	1.92
	1	20.80	28.09	1.82
	1.25	19.04	27.29	1.73
	1.5	17.70	26.59	1.64
Preventive maintenance cost, c_p	1	18.74	24.84	2.04
	2	21.44	28.38	1.82
	3	23.99	31.25	1.63
Failure threshold, ξ	10	15.95	23.72	1.63
	12	18.39	25.87	1.75
	14	20.80	27.75	1.83
	18	25.05	30.62	1.95
	20	26.76	31.70	1.99
Parameter of degradation function, γ	0.5	2.73	1.98	1.98
	0.75	8.46	4.56	1.92
	1	17.40	8.99	1.87
	1.33	30.52	18.88	1.78
	2	47.84	47.40	1.60
Parameter of revenue function, ν	0.5	5.70	6.25	1.00
	0.75	11.25	8.79	1.22
	1	18.23	14.24	1.50
	1.33	27.73	23.56	1.99
	2	44.04	47.30	3.43
Total		21.39	28.35	1.83

Secondly, an integrated approach is especially more profitable for systems that have a large expected time to failure under normal deterioration, either through a larger failure threshold or a lower base rate. This is because for such systems, the revenue is relatively a higher contributor to the profit than for systems with low lifetimes. In an integrated approach, setting the maintenance interval takes into account that more revenue can be realized relative to maintenance costs, while in a sequential approach, this accumulation of revenue is not taken into account when choosing the maintenance interval. Thirdly, the form of both the degradation and revenue function has a huge impact on the profit increase; concave functions lead to small profit increases, while convex functions lead to high profit increases.

4.5. Heterogeneous systems: Learning the base rate

The previous sections focused on the situation in which the base rate is fully known to the decision maker and where each newly installed system has this base rate. In this section, our attention is focused on the situation in which the base rate is not a-priori known to the decision maker and where each newly installed system has a different base rate. For this case, where systems stem from a heterogeneous population with parameter uncertainty, a decision maker needs to learn the individual base rate based on both the system's deterioration path and the chosen settings. We first propose a Bayesian framework that is able to do so, which we then use to develop a heuristic policy. This heuristic policy is based on the optimal policy for the known case that we extensively analyzed in the previous section. We conclude this section with a simulation study in which we compare the performance of the heuristic policy with a clairvoyant policy that knows the true base rate upon installation.

4.5.1 Bayesian inference under setting policies

We assume that the heterogeneity of the systems' base rates, denoted by Λ, can be modeled by a Gamma distribution with known shape α_0 and rate β_0; that is, $\Lambda \sim \Gamma(\alpha_0, \beta_0)$. Each newly installed system has a realization of Λ as base rate which is unknown to the decision maker. Hence at $t = 0$ (which is T time units before the planned maintenance), upon installation of a new system, the decision maker's knowledge regarding the unknown base rate can be modeled by a prior distribution with density function:

$$\pi_0(\lambda|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0} \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda}}{\Gamma(\alpha_0)},$$

where the subscript 0 indicates that this represents the decision maker's knowledge at $t = 0$. Recall that the deterioration level at $t = 0$ is denoted with $Y_0 = 0$. Suppose now that the decision maker adopts a setting policy in which she chooses setting $s \in \mathscr{S}$ for the period $[0, u]$ and we observe deterioration level $Y_u = y$ at $t = u$. We can then obtain the posterior distribution at $t = u$ using Bayesian statistics (i.e. that the posterior distribution is proportional to the likelihood function times the prior

distribution) as follows:

$$\begin{aligned}\pi_u(\lambda|s,y,\alpha_0,\beta_0) &\propto \mathbb{P}[Y_u = y|\lambda,s]\cdot\pi_0(\lambda|\alpha_0,\beta_0)\\ &= \frac{\big(u\lambda f(s)\big)^y e^{-u\lambda f(s)}}{y!}\frac{\beta_0^{\alpha_0}\lambda^{\alpha_0-1}e^{-\beta_0\lambda}}{\Gamma(\alpha_0)}\\ &= \frac{\big(uf(s)\big)^y}{y!}\frac{\beta_0^{\alpha_0}\lambda^{y+\alpha_0-1}e^{-\lambda\left(\beta_0+uf(s)\right)}}{\Gamma(\alpha_0)}\\ &\propto \lambda^{\alpha_0+y-1}e^{-\lambda\left(\beta_0+uf(s)\right)}.\end{aligned}$$

Note that this is again a Gamma distribution with parameters α_0+y and $\beta_0+uf(s)$. It is well-known that the Gamma distribution is a conjugate for the Poisson distribution (see, e.g., Gelman et al., 1995), and we have now shown that this also holds under setting policies. This also implies that for the inference of the unknown base rate, we only need to keep track of the settings used and the current deterioration. That is, let s_i be the setting used in period i lasting from $[t_{i-1}, t_{i-1}+u]$, with $t_0 = 0$, and let $\hat{s}_t \triangleq u\cdot\sum_{i=1}^{\frac{t}{u}} f(s_i)$ be the aggregate sum of settings used until time t, and suppose that the system has deterioration level y, then the decision maker's knowledge regarding the unknown base rate is represented by the (updated) Gamma distribution $\Lambda\sim\Gamma(\alpha_0+y,\beta_0+\hat{s}_t)$. The mean of this Gamma distribution is equal to $\mathbb{E}[\Lambda|y,\hat{s}_t] = \frac{\alpha_0+y}{\beta_0+\hat{s}_t}$, which is intuitive for the following reason. If a system has been controlled by a policy with a higher aggregate sum of settings (recall that $f(s)$ is increasing in s) than another, but both systems have the same deterioration level, then we expect the base rate of the former to be smaller. Similarly, if a policy has been employed with the same aggregate sum of settings, but one system has a higher deterioration level than another, then we expect the system with the higher deterioration level to have a larger base rate.

4.5.2 A certainty-equivalent policy

We now propose an intuitive heuristic policy called the certainty-equivalent policy to deal with the situation of heterogeneous systems that makes use of the Bayesian framework described in the previous section. Under this policy, the decision maker estimates the base rate based on her current knowledge and assumes that the true base rate is equal to her estimation with certainty. The certainty-equivalent policy is parametrized by a positive integer $N^{\text{re-opt}}\in\mathbb{Z}_{>0}$, which denotes how often the decision maker re-optimizes the policy based on the accrued knowledge. In the certainty-equivalent policy, we divide the planning horizon in $N^{\text{re-opt}}+1$ phases, each of equal

length.

At the start of the planning horizon, upon installation of the system, the decision maker's estimate for the base rate, denoted with $\hat{\lambda}_0$, is equal to the mean of the initial prior distribution: $\hat{\lambda}_0 = \frac{\alpha_0}{\beta_0}$. Using this estimate, we obtain the optimal policy for the known case with $\hat{\lambda}_0$ as base rate, and implement this policy until the first re-optimization epoch at $t_1 = \frac{T}{N^{\text{re-opt}}+1}$. At this point, the decision maker updates the estimate of the base rate based on the current deterioration level, denoted with x_1, and the aggregate sum of settings used, denoted with $\hat{s}_{t_1}$, to $\hat{\lambda}_1 = \frac{\alpha_0+x_1}{\beta_0+\hat{s}_t}$ and computes the optimal policy as if $\hat{\lambda}_1$ is the true base rate. This policy is then implemented until the next re-optimization epoch, after which the procedure is repeated until the scheduled maintenance moment. A schematic illustration of the certainty-equivalent policy with $N^{\text{re-opt}} = 1$ is provided in Figure 4.4. Recall that u^*_λ denotes the optimal policy for a system with known base rate λ.

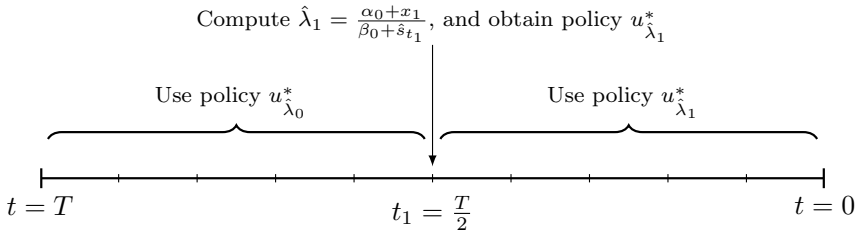


Figure 4.4 Schematic illustration of certainty-equivalent policy with $N^{\text{re-opt}} = 1$.

The certainty-equivalent policy is easy to implement and, based on the structural properties that we established in Theorem 4.1, we have a good understanding of how this policy behaves. Specifically, if the decision maker learns that a system has a high base rate, she will impose lower settings for all times and deterioration levels, while if she learns that the system has a low base rate, she will impose higher settings. The policy thus learns when it is possible to impose higher settings to accumulate more revenue, but also when lower settings are needed to prevent failures and high associated maintenance costs.

4.5.3 Simulation study

In the numerical studies in Section 4.4 we computed the performance measures needed for the comparisons in those studies. The performance of the certainty-equivalent policy, however, can only be evaluated through simulation. In this section, we report on the results of a comprehensive simulation study in which we compare the

performance of the certainty-equivalent policy with a clairvoyant policy that knows the true base rate at installation.

We use a full factorial design in which we vary five different input parameters (see Table 4.4 for a summary). We use failure thresholds $\xi \in \{10, 15, 20\}$ and planning horizons $T \in \{10, 15, 20\}$. Again, we use a single corrective maintenance cost equal to 10 and a single preventive maintenance cost for all non-failed states, denoted with c_p, where we use either $c_p = 1$ or $c_p = 3$. We use the functions $f = s^\gamma$ and $r = s^\nu$, for the degradation and revenue function, respectively, and use the same interval for the settings to choose from as in the previous numerical studies, i.e. $\mathscr{S} = [0, 2]$. By using the values $\{0.5, 1, 2\}$ for γ and ν, we model concave, linear, and convex functions, respectively. The certainty-equivalent policy is characterized by the parameter $N^{\text{re-opt}}$, which we vary between 5 values. In a simulation instance (we shortly describe the simulation procedure), the base rates are sampled from a Gamma distribution in which we fix the mean to 1, and vary its coefficient of variation, denoted with cv_Λ. This coefficient of variation is thus a measure for the population heterogeneity of the systems. These choices lead to a total of 4050 instances.

Table 4.4 Input parameter values for simulation study.

	Input parameter	No. of choices	Values
1	Coefficient of variation of prior Gamma distribution, cv_Λ	3	0.71, 1, 1.41
2	Preventive maintenance cost, c_p	2	1, 3
3	Length of planning horizon, T	3	10, 15, 20
4	Failure threshold, ξ	5	10, 12, 15, 18, 20
5	Parameter of degradation function, γ	3	0.5, 1, 2
6	Parameter of revenue function, ν	3	0.5, 1, 2
7	Number of re-optimizations, $N^{\text{re-opt}}$	5	1, 2, 4, 6, 8

We benchmark the performance of the certainty-equivalent policy, denoted with π_{CEP}, by the notion of relative regret, i.e. the percentage of expected profit loss (over a planning horizon) relative to the clairvoyant Oracle policy π_O. We compute this relative regret as $\bar{\mathcal{R}} = -100 \cdot (\bar{P}_{\pi_{CEP}} - \bar{P}_{\pi_O})/\bar{P}_{\pi_O}$, where $\bar{P}_\pi$ is the average profit of policy $\pi \in \{\pi_{CEP}, \pi_O\}$. Note that as π_O is unattainable in practice, $\bar{\mathcal{R}}$ serves as an upper bound on the relative regret with the best achievable performance.

The simulation procedure is as follows. For each instance, we first sample a base rate, say λ^*, from the Gamma distribution $\Lambda \sim \Gamma(\alpha_0, \beta_0)$. Then we compute the expected optimal profit of the Oracle policy using base rate λ^*: $J^*_{\lambda^*}(0, T)$. We then simulate a sample path of an inhomogeneous Poisson process with base rate λ^* that is controlled by the certainty-equivalent policy using the thinning algorithm proposed by Lewis and Shedler (1979) and obtain the resulting profit. We repeat this S^{max}

times to obtain estimates for both $\bar{P}_{\pi_O}$ and $\bar{P}_{\pi_{CEP}}$, respectively. Here, we choose S^{max} sufficiently large such that the confidence intervals of the estimated $\bar{P}_{\pi_O}$ and $\bar{P}_{\pi_{CEP}}$ are sufficiently small. We then use $\bar{P}_{\pi_O}$ and $\bar{P}_{\pi_{CEP}}$ to compute $\bar{\mathcal{R}}$.

The results of the simulation study are summarized in Table 4.5. In this table, we present the average and maximum $\bar{\mathcal{R}}$, and the $\bar{P}_{\pi_O}$, where we first distinguish between subsets of instances with the same value for a specific input parameter of Table 4.4, and then present the results for all instances.

Table 4.5 Results of simulation study: Relative profit loss (in %) compared with Oracle performance.

		$\bar{\mathcal{R}}$		
Input parameter	Value	Mean	Max	$\bar{P}_{\pi_O}$
Coefficient of variation of prior Gamma distribution, cv_Λ	0.71	0.59	7.29	10.32
	1	0.72	9.75	10.33
	1.41	0.84	11.63	10.33
Preventive maintenance cost, c_p	1	0.65	10.06	11.33
	3	0.78	11.63	9.33
Length of planning horizon, T	10	0.38	3.10	7.03
	15	0.70	7.16	10.51
	20	1.08	11.63	13.46
Failure threshold, ξ	10	1.37	11.63	9.03
	12	0.94	7.98	9.75
	15	0.64	4.98	10.32
	18	0.34	2.33	11.14
	20	0.28	2.09	11.43
Parameter of degradation function, γ	0.5	0.01	0.07	9.93
	1	0.78	11.63	10.24
	2	1.39	11.63	10.83
Parameter of revenue function, ν	0.5	1.39	11.63	10.83
	1	0.77	11.63	10.24
	2	0.01	0.05	9.93
Number of re-optimizations, $N^{\text{re-opt}}$	1	0.91	11.63	10.32
	2	0.79	10.21	10.33
	4	0.68	8.64	10.33
	6	0.62	7.90	10.33
	8	0.59	7.55	10.33
Total		0.72	11.63	10.33

The results in Table 4.5 indicate that the certainty-equivalent policy performs excellently with average profit losses of only 0.72% compared to the Oracle policy that knows the true base rate. Additionally, with a maximum profit loss of only 11.63% over all 4050 instances, the certainty-equivalent policy is also robust in providing good performance under varying conditions. We find that this profit loss is inversely related to the number of re-optimizations and the failure threshold. The decrease

in the number of re-optimizations is intuitive; the certainty-equivalent policy adapts more frequently to the accrued knowledge of the current system's base rate yielding a lower profit loss with respect to the optimal policy for that base rate. When the failure threshold increases, there is less need for controlled deterioration as the system can, on average, be set to its maximum setting more often. This holds for both the Oracle policy and the certainty-equivalent policy resulting in comparable performance. Finally, we see that the form of both the degradation and revenue function has a large impact on the relative profit loss; a concave degradation function and/or a convex revenue function lead to small profit losses.

We would like to emphasize that we have assumed that the decision maker has access to the true hyperparameters of the Gamma distributions from which the base rates are sampled. Hence, the certainty-equivalent policy already starts with a good understanding (in expectation) of the underlying heterogeneity, which might be a contributing factor to the excellent results. Nonetheless, in practice, decision makers often have access to data that can be used to provide a good estimate of these initial prior parameters, which makes this assumption reasonable.

4.6. Conclusion

In this chapter, we have studied optimal condition-based setting policies for stochastically deteriorating production systems when planned maintenance moments are given. First, we considered the case that the base rate is a-priori known to the decision maker. For this case, we modeled the problem as a continuous-time MDP and characterized the monotonic behavior of the optimal policy. We have shown that this structure is remarkably intuitive: decrease the setting if closer to the failure threshold and increase the setting if closer to the planned maintenance moment. We have further shown that under these optimal setting decisions, the length of the interval between planned maintenance moments can be easily optimized, thereby addressing the trade-off that arises due to maintenance costs at such moments. We have complemented these theoretical results by an extensive numerical study in which we demonstrate that: (i) condition-based setting policies can lead to substantial profit increases relative to static setting policies, and (ii) integrating maintenance and setting policies can lead to significant profit increases compared to treating them sequentially (i.e. first deciding on an interval length, and then implementing the setting policy, see Section 4.4.2).

We then relaxed the assumption that the base rate is known a-priori by assuming that there is parameter uncertainty regarding this base rate. For this case, we have developed an *easy-to-implement* Bayesian heuristic, that mimics the structure

of the optimal policy for the case that the base rate is known. In an extensive simulation study, we have shown that this heuristic performs excellently compared to a clairvoyant Oracle policy that knows the base rate a-priori.

Two immediate directions for further research are (i) extending the analysis to stochastic processes other than the Poisson process, and (ii) integrating Bayesian learning in the optimization problem directly instead of myopically like we proposed in this chapter. For the former, it is perhaps most promising to consider other point processes so that the resulting optimization problem still remains in the class of intensity control problems (see, e.g., Bremaud, 1980, Chapter VII). For the latter, one could embed the optimization problem with (in addition to state variables) a belief variable needed for the Bayesian learning. In a discrete-time setting, this gives rise to a tractable Bayesian MDP (see Chapter 1), but in a continuous-time setting – like the one discussed in this chapter – the evolution process of this belief variable is usually a martingale satisfying a certain stochastic differential equation (see, e.g., Araman and Caldentey, 2009; Kwon and Lippman, 2011; Kwon et al., 2012), which is likely much harder to tackle.

This chapter is the final of three chapters on CBM for single-component systems with parameter uncertainty. In the next chapter, we shall look at a set of multiple systems, where each system degrades according to the deterioration process considered in this chapter: a Poisson process. We shall study the case where the rate of the Poisson process is unknown a-priori, similar to the setting studied in Section 4.5.1 in this chapter. When studying a set of multiple systems, where each system is generating a real-time stream of data, a natural question to ask is whether combining these multiple streams of data can lead to savings compared to not combining them. Answering this question will be the main objective of the next chapter.

Chapter 5

Pooling data for joint learning in condition-based maintenance with rate uncertainty

5.1. Introduction

The condition-based maintenance (CBM) models that we studied in the three previous chapters all have two assumptions in common. The first assumption deals with modeling the parameter uncertainty, where we assume that each time upon replacement of a component a new – again unknown – parameter is drawn from some distribution describing the uncertainty. This new realization is *different* from the parameter of the replaced component, so that all components are statistically *distinguishable*. This is reflected in the corresponding optimizations problems in each chapter, where learning is reset (through resetting the sufficient statistic) once a component is replaced. The second assumption deals with the number of systems. In each model we focused on a single-component system in isolation, which is a *logical* consequence of the first assumption. Indeed, there is no incentive to look at multiple systems if the individual systems are in fact statistically distinguishable. In this

This chapter is based on Drent and Van Houtum (2022).

chapter, which is the last chapter on parameter uncertainty in CBM, we depart from both assumptions.

Specifically, we consider a set of systems where each system has a critical component whose condition deteriorates according to a Poisson process (similar to the model considered in Chapter 4) with an a-priori unknown rate. In contrast to the previous chapters, we assume that this unknown rate is *identical* for all systems and remains the same whenever a component of a system is replaced. Although all systems have the same rate, all other characteristics (preventive and corrective maintenance costs, and failure thresholds) differ for each system. When treating parameter uncertainty in this way, where all systems are now statistically *indistinguishable*, there is an incentive to look at all systems together. That is, one could *pool* all data – stemming from all systems – together to *jointly learn* the unknown parameter on-the-fly as data becomes available. In this chapter we study this data pooling and investigate the benefits of shared learning in CBM.

The benefit of pooling resources has been extensively studied in many application domains, yet almost exclusively related to pooling of physical resources. For instance, inventory pooling, where multiple inventory locations are consolidated into a single one, leads consistently to lower costs than having a set of distributed inventory locations (see, e.g., Eppen, 1979; Benjaafar et al., 2005; Berman et al., 2011). Pooling of server capacity in queuing networks – frequently used to model applications in service industries – is another example where pooling of physical resources can lead to lower costs or higher efficiency when compared to its unpooled equivalent (see, e.g., Smith and Whitt, 1981; Calabrese, 1992; Mandelbaum and Reiman, 1998).

Due to recent developments in information technology, many applications are characterized by (i) multiple sources of data, but at the same time, (ii) limited data per source because each of them has not yet generated a sufficient amount of data to accurately estimate values of relevant parameters needed for decision making. To alleviate the insufficiency of data per source, some researchers have recently started to investigate the possibility of pooling data from many sources to improve decision making that uses that data as input, mainly motivated by the proven benefits of pooling physical resources.

Bastani et al. (2022) consider a set of dynamic pricing problems where demand in each problem is parametrized by a common yet unknown parameter. They propose a meta dynamic pricing algorithm that learns the unknown parameter using Thompson sampling, and they show that pooling data speeds up learning significantly. Gupta and Kallus (2022) propose a sample average approximation algorithm for a set of stochastic optimization problems – each with limited data – and prove that combining

data across these problems can outperform the unpooled equivalents. Finally, Deprez et al. (2022) investigate the benefits of combining data from a set of heterogeneous systems in the context of preventive maintenance. In their work, Deprez et al. (2022) assume that a system's failure time follows a Cox proportional hazards model, where the intensity function can be decomposed in both a factor that is identical for all systems and a system specific factor. The authors propose a method in which limited data stemming from the statistically distinguishable (due to the system specific factor) systems can be aggregated and adapted such that it can be utilized for better maintenance decisions on each individual system. In this chapter, we contribute to this novel stream of literature by investigating the benefits of data pooling in the context of CBM.

5.1.1 Contributions

The main contributions of this chapter are as follows:

1. We formulate the problem of optimally maintaining N systems with a common, unknown deterioration rate over a finite lifespan T as a finite-horizon Bayesian Markov decision process (MDP) in which data is pooled for joint learning. This formulation suffers from the well-known curse of dimensionality: The cardinality of both the action and state space grow exponentially in N. As a remedy, we provide a new decomposition result that establishes the equivalence between the original MDP and N two-state MDPs with a binary action space, each focused on an individual system.

2. We show that the structure of the optimal policy of each individual system has a control limit structure.

3. In a comprehensive numerical study, we investigate the savings that can be attained by pooling data to learn the a-priori unknown deterioration rate, while optimally maintaining the systems. We find that the savings can be significant, even for small values of N, and that the exact magnitude of these savings largely depends on the magnitude of the uncertainty in the parameter. When there is high uncertainty, huge savings of close to 57% can be realized on average, while these savings become almost negligible when uncertainty decreases.

5.1.2 Organization

The rest of this chapter is organized as follows. We describe the model in more detail in Section 5.2. In Section 5.3, we formulate the problem as an MDP and we show that it can be decomposed into N alternative MDPs. We present some structural properties of both the expected cost and the optimal policy of the alternative MDP in Section 5.4. In Section 5.5, we report on an extensive numerical study that highlights the benefit of pooling data. Finally, Section 5.6 provides concluding remarks.

5.2. Model description

We consider a set of $N \geq 1$ systems subject to damage accumulation due to random shocks that arrive over time. We remark that although pooling has only value when $N > 1$, the analysis in this chapter also holds if $N = 1$. The set of all systems is denoted by $\mathscr{N}$, i.e. $\mathscr{N} = \{1, \ldots, N\}$. We assume that each system has a critical component such that the system breaks down whenever this component fails. The deterioration processes of these components are modeled as independent Poisson processes with the same rate λ, denoted by $\{X_i(t), t \geq 0\}$, with $X_i(0) = 0$, for $i \in \mathscr{N}$. A component of system $i \in \mathscr{N}$ deteriorates until its deterioration level reaches or crosses a deterministic failure threshold, denoted with $\xi_i \in \mathbb{N}_+$, where $\mathbb{N}_+ \triangleq \{1, 2, \ldots\}$, after which the component is considered to have failed.

The deterioration levels are monitored at equally spaced decision epochs, though failure moments can happen at any point in time (i.e. not only at decision epochs). These decision epochs also correspond to planned maintenance moments at which the decision maker can interfere with the components, which implies that if a component fails, corrective maintenance is executed at the subsequent decision epoch. In practice though, a failed component may be replaced immediately upon failure. If we assume that the time between two subsequent decision epochs is small compared to their deterioration rate, then replacing only at decision epochs is a reasonable assumption. For convenience, we rescale time such that the time between two decision epochs equals 1. If at a decision epoch, a component of system $i \in \mathscr{N}$ has failed, it needs to be replaced correctively which takes a negligible amount of time and costs $c_u^i > 0$. Such a failure can be prevented by performing an instantaneous preventive replacement at a decision epoch, which costs $c_p^i > 0$, with $c_p^i < c_u^i$ for all $i \in \mathscr{N}$. Corrective maintenance is more expensive because it includes costs caused by a component failure in addition to the costs related to the replacement. Both replacements lead to a newly installed component that is as-good-as-new which starts deteriorating again

from level 0 according to a Poisson process with rate λ, that is, $\{X_i(t), t \geq 0\}$ is reset to $X_i(0) = 0$.

The systems have a common, finite lifespan, with length T ($< \infty$) time units, which represents the time from the beginning of operating the systems until they are taken out of service. We let this lifespan consist of T discrete time steps corresponding to the intervals between consecutive decision epochs. That is, the systems start operating at $t = 0$ which coincides with the first decision epoch, while the last decision epoch is at $t = T$ which coincides with the end of the lifespan.

The decision maker, responsible for maintaining the set of N systems, seeks to minimize the total expected maintenance costs – due to both corrective and preventive replacements of components – over their lifespan. In dealing with this optimization problem, the decision maker faces another layer of uncertainty in addition to the random shock arrivals. That is, the components used for all replacements always have the same rate but this rate is a-priori unknown and needs to be inferred based on the observations of the deterioration processes throughout their lifespan. Since all components have the same rate, the decision maker can pool and utilize all accumulated data together when inferring this unknown rate.

To this end, we adopt a Bayesian approach and treat the unknown rate λ as a random variable denoted with Λ. Upon the start of operating all systems, at $t = 0$, Λ has a Gamma distribution with shape parameter α_0 and rate parameter β_0. The subscript notation reflects that this corresponds to $t = 0$; we adopt this notation in the remainder of this chapter. Thus, at $t = 0$, the density function of Λ is equal to

$$f_\Lambda(\lambda; \alpha_0, \beta_0) = \frac{\lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda} \beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \quad \text{for } \lambda > 0, \quad \alpha_0, \beta_0 > 0,$$

where $\Gamma(\cdot)$ denotes the Gamma function. Suppose that at decision epoch $t \in \mathbb{N}_+$, we observed a cumulative amount of k deterioration increments from all installed components. It is well-known that the Gamma distribution is a conjugate prior for the Poisson distribution. We can therefore readily obtain the new posterior distribution describing our belief of Λ, which is again a Gamma distribution but with updated parameters (see, e.g., Gelman et al., 1995, Chapter 2):

$$\alpha_t = \alpha_0 + k \quad \text{and} \quad \beta_t = \beta_0 + N \cdot t. \tag{5.1}$$

Observe that from the updating scheme in Equation (5.1), it is immediately clear that the data stemming from all N systems is pooled for learning the unknown rate λ that the systems have in common. At each decision epoch, based on her current belief of

Λ, the decision maker wishes to predict the future evolution of the deterioration of each component so that she can decide on potential replacements. This prediction is encoded in the posterior predictive distribution. For this Gamma-Poisson model, it is well-known that the posterior predictive distribution is a Negative Binomial distribution (see, e.g., Gelman et al., 1995, Chapter 2). Specifically, given parameters α_t and β_t, the deterioration increment of a component at system i at the next decision epoch, denoted with X_i, is Negative Binomially distributed with parameters

$$r = \alpha_t \quad \text{and} \quad p = \frac{\beta_t}{\beta_t + 1}, \tag{5.2}$$

where r is the number of successes and p is the success probability, so that X_i can be interpreted as the number of failures until the r^{th} success. In the remainder we use the notation $X \sim NB(r, p)$ to denote that X is a Negative Binomially distributed random variable with parameters r and p.

Equation (5.2) together with the updating scheme in (5.1) can be used to construct an updated posterior predictive distribution at each decision epoch of the next deterioration increments in real-time based on the observed data. Since the posterior predictive distributions of the deterioration increments of each system are fully described by only the current decision epoch t and cumulative amount of deterioration increments k, it is a Markov process. This allows us to formulate the optimization problem as a finite-horizon (with length T) MDP equipped with the state variable k for Bayesian inference of the unknown rate, which is the objective of the next section.

5.3. Markov decision process formulation

In this section, we formulate the problem described in the previous section as an MDP.

The state space of the MDP is the set $\mathscr{S} \triangleq \mathbb{N}_0^{N+1}$ where $\mathbb{N}_0 \triangleq \mathbb{N}_+ \cup \{0\}$. For a given state $(\boldsymbol{x}, k) \in \mathscr{S}$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ represents the vector of all deterioration levels, and k denotes the sum of all deterioration increments. For a given state $(\boldsymbol{x}, k) \in \mathscr{S}$, let $\mathscr{A}(\boldsymbol{x})$ denote the action space, with $\mathscr{A}(\boldsymbol{x}) = \{(\boldsymbol{a}_i)_{i \in \mathscr{N}}\}$ with

$$\boldsymbol{a}_i = \begin{cases} \{0, 1\} & \text{if } x_i < \xi_i, \\ \{1\} & \text{if } x_i \geq \xi_i, \end{cases}$$

where $a_i = 0$ corresponds to taking no action and $a_i = 1$ corresponds to performing

maintenance on the component of system i, respectively. This implies that if the critical component of system i has failed (i.e. $x_i \geq \xi_i$), then the decision maker must (correctively) replace it. For all components that have not failed, the decision maker can choose to either preventively replace it, or do nothing and continue to the next decision epoch.

Given the state $(\boldsymbol{x}, k) \in \mathscr{S}$ and an action $a \in \mathscr{A}(\boldsymbol{x})$, the decision maker incurs a direct cost, denoted by $C(\boldsymbol{x}, \boldsymbol{a})$, equal to

$$C(\boldsymbol{x}, \boldsymbol{a}) \triangleq \sum_{i \in \mathscr{N}} \Big(a_i \big(1 - \mathbb{I}_i(\boldsymbol{x})\big) c_p^i + \mathbb{I}_i(\boldsymbol{x}) c_u^i \Big), \tag{5.3}$$

where $\mathbb{I}_i(\boldsymbol{x})$ is an indicator function that indicates whether the component of system i has failed in the deterioration vector $\boldsymbol{x}$; that is,

$$\mathbb{I}_i(\boldsymbol{x}) = \begin{cases} 0 & \text{if } x_i < \xi_i, \\ 1 & \text{if } x_i \geq \xi_i. \end{cases}$$

Let $V_t^N(\boldsymbol{x}, k)$ denote the optimal expected total cost over decision epochs $t, t{+}1, \ldots, T$, starting from state $(\boldsymbol{x}, k) \in \mathscr{S}$, and let the terminal cost, $V_T^N(\boldsymbol{x}, k)$, be equal to the function $C(\boldsymbol{x}) \triangleq \sum_{i \in \mathscr{N}} \mathbb{I}_i(\boldsymbol{x}) c_u^i$ for all k. Then, for all $t \in \{0, 1, \ldots, T-1\}$, $V_t^N(\boldsymbol{x}, k)$ satisfies the following recursive Bellman optimality equations

$$V_t^N(\boldsymbol{x}, k) = \min_{\boldsymbol{a} \in \mathscr{A}(\boldsymbol{x})} \left\{ C(\boldsymbol{x}, \boldsymbol{a}) + \mathbb{E}_{\boldsymbol{a}} \left[V_{t+1}^N \Big(\boldsymbol{x'} + \boldsymbol{X}, k + \sum_{i \in \mathscr{N}} X_i \Big) \right] \right\}, \text{ for all } (\boldsymbol{x}, k) \in \mathscr{S}, \tag{5.4}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ is an N-dimensional random vector with $X_i \sim NB\Big(\alpha_0 + k, \frac{\beta_0 + N \cdot t}{\beta_0 + N \cdot t + 1}\Big)$ (all X_i's are independent and identically distributed), and $\boldsymbol{x'} = (x'_1, x'_2, \ldots, x'_N)$ with

$$x'_i = \begin{cases} x_i & \text{if } a_i = 0, \\ 0 & \text{if } a_i = 1. \end{cases} \tag{5.5}$$

We also refer to $V_t^N(\boldsymbol{x}, k)$ as the value function. The first part between the brackets is the direct costs while the second part is the expected future costs of taking action $\boldsymbol{a}$ in state $(\boldsymbol{x}, k)$. Specifically, each component's deterioration accumulates further according to the posterior predictive distribution that corresponds to state $(\boldsymbol{x}, k)$ that we described in the previous section, while k increases with the sum of all those

increments. Systems that are maintained start with an as-good-as-new component, which is governed by the auxiliary vector $\boldsymbol{x'}$ which ensures that $x'_i = 0$ when $a_i = 1$, see (5.5). The formulation in (5.4) shows that the learning process about the unknown rate λ is pooled through the evolution of the common state variable k, while the future evolution of all individual deterioration processes depends on all pooled information and the parameter N.

Because (i) the action space is finite, and (ii) the state space is countable, the existence of an optimal policy for this finite-horizon MDP is guaranteed, see e.g., Proposition 4.4.3 of Puterman (2005). Observe that over the complete lifespan of length T, the minimum total expected cost for N systems is given by $V_0^N(\mathbf{0}, 0)$ ($\mathbf{0}$ denotes the N-dimensional zero vector) which can be found by solving Equation (5.4) via backward induction. It is however clear from the formulation in (5.4), that as the number of systems grows, the problem will increasingly suffer from the curse of dimensionality: The cardinality of both the action and state space grow exponentially in N.

Instead of solving (5.4) (referred to as the original MDP) directly, we will now construct an alternative MDP and show that the original MDP can be decomposed into N of these alternative MDPs: One for each system $i \in \mathscr{N}$. This decomposition is imperative as it allows us to analyze the benefits of pooling of learning when N is relatively large without suffering from the curse of dimensionality.

To this end, for each $i \in \mathscr{N}$, let $\tilde{V}_t^{N,i}(x, k)$ denote the optimal expected total cost over decision epochs $t, t+1, \ldots, T$, starting from state $(x, k) \in \mathbb{N}_0^2$, and let the terminal cost, $\tilde{V}_T^{N,i}(x, k)$, be equal to the function $C_i(x) \triangleq \mathbb{I}_i(x) c_u^i$ for all k. Then, for all $t \in \{0, 1, \ldots, T-1\}$ and all $(x, k) \in \mathbb{N}_0^2$, $\tilde{V}_t^{N,i}(x, k)$ satisfies the following recursive Bellman optimality equations

$$\tilde{V}_t^{N,i}(x,k) = \min_{a \in \mathscr{A}(x)} \left\{ C_i(x,a) + \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x \cdot (1-a) + X, k + X + K \Big) \right] \right\}, \tag{5.6}$$

where $X \sim NB\Big(\alpha_0 + k, \frac{\beta_0 + N \cdot t}{\beta_0 + N \cdot t + 1}\Big)$, $K \sim NB\Big((N-1) \cdot (\alpha_0 + k), \frac{\beta_0 + N \cdot t}{\beta_0 + N \cdot t + 1}\Big)$ (i.e., the probability distribution of K is the $(N-1)$-fold convolution of the probability distribution of X), and

$$C_i(x,a) \triangleq a\big(1 - \mathbb{I}_i(x)\big) c_p^i + \mathbb{I}_i(x) c_u^i. \tag{5.7}$$

The indicator functions and actions (spaces) are as defined before. It is noteworthy to mention that the formulation in (5.6) in fact resembles a single component optimization problem in isolation, where the transition probabilities depend on both

parameter N and the state variable k. The evolution of the state variable k depends on the random deterioration increment of the component (X) but it also accounts for the evolution of the other components through the random variable K. Below we present the decomposition result, which establishes that the value function of the original MDP is the sum of all N value functions of the alternative MDPs.

Proposition 5.1. *For each $t \in \{0, 1, \ldots, T\}$, we have:*

$$V_t^N(\boldsymbol{x}, k) = \sum_{i \in \mathscr{N}} \tilde{V}_t^{N,i}(x_i, k).$$

Proof: We prove the statement using induction on t. Note that the terminal values can be decomposed:

$$V_T^N(\boldsymbol{x}, k) = C(\boldsymbol{x}) = \sum_{i \in \mathscr{N}} \mathbb{I}_i(\boldsymbol{x}) c_u^i = \sum_{i \in \mathscr{N}} \mathbb{I}_i(x_i) c_u^i = \sum_{i \in \mathscr{N}} \tilde{V}_T^{N,i}(x_i, k),$$

so that the statement trivially holds for the base case T. Assume that the statement holds for some $t+1, 0 < t+1 \leq T$, we will show that the statement then also holds for t.

We have, by Equation (5.4),

$$\begin{aligned}
V_t^N(\boldsymbol{x}, k) &= \min_{\boldsymbol{a} \in \mathscr{A}(\boldsymbol{x})} \left\{ C(\boldsymbol{x}, \boldsymbol{a}) + \mathbb{E}_{\boldsymbol{a}} \left[V_{t+1}^N \Big(\boldsymbol{x'} + \boldsymbol{X}, k + \sum_{j \in \mathscr{N}} X_j \Big) \right] \right\} \\
&\overset{(a)}{=} \min_{\boldsymbol{a} \in \mathscr{A}(\boldsymbol{x})} \left\{ \sum_{i \in \mathscr{N}} C_i(x_i, a_i) + \mathbb{E}_{\boldsymbol{a}} \left[V_{t+1}^N \Big(\boldsymbol{x'} + \boldsymbol{X}, k + \sum_{j \in \mathscr{N}} X_j \Big) \right] \right\} \\
&\overset{(b)}{=} \min_{\boldsymbol{a} \in \mathscr{A}(\boldsymbol{x})} \left\{ \sum_{i \in \mathscr{N}} C_i(x_i, a_i) + \mathbb{E}_{\boldsymbol{a}} \left[\sum_{i \in \mathscr{N}} \tilde{V}_{t+1}^{N,i} \Big(x_i' + X_i, k + \sum_{j \in \mathscr{N}} X_j \Big) \right] \right\} \\
&\overset{(c)}{=} \min_{\boldsymbol{a} \in \mathscr{A}(\boldsymbol{x})} \left\{ \sum_{i \in \mathscr{N}} C_i(x_i, a_i) + \sum_{i \in \mathscr{N}} \mathbb{E}_{\boldsymbol{a}} \left[\tilde{V}_{t+1}^{N,i} \Big(x_i' + X_i, k + X_i + \sum_{j \in \mathscr{N} \setminus i} X_j \Big) \right] \right\} \\
&\overset{(d)}{=} \sum_{i \in \mathscr{N}} \min_{a_i \in \mathscr{A}(x_i)} \left\{ C_i(x_i, a_i) + \mathbb{E} \left[\tilde{V}_{t+1}^{N,i} \Big(x_i(1 - a_i) + X_i, k + X_i + K \Big) \right] \right\} \\
&= \sum_{i \in \mathscr{N}} \tilde{V}_t^{N,i}(x_i, k),
\end{aligned}$$

where (a) holds because the direct costs can be decomposed (see (5.3) and (5.7)), (b) holds due to the induction hypothesis, (c) holds because of the linearity of an expectation and extracting X_i from the summation, (d) holds because the sum of

$N-1$ independent Negative Binomially distributed random variables with $r = \alpha_0 + k$ and $p = \frac{\beta_0 + N \cdot t}{\beta_0 + N \cdot t + 1}$ is again Negative Binomially distributed with the same p but with $r = (N-1) \cdot (\alpha_0 + k)$ (see, e.g., DasGupta, 2010, Chapter 6), and the last equality follows from using Equation (5.6). □

The decomposition in Proposition 5.1 reduces the computational burden of solving (5.4) significantly. It collapses the original, high-dimensional MDP into N 2-dimensional MDPs with a binary action space, each with their own cost structure and failure threshold, while still taking into account pooled learning across the N systems. As already pointed out, the alternative MDP is in fact a single component optimization problem with an adapted state space and adapted transition probabilities. This also eases the process of establishing some of its structural properties, which is the topic of the next section.

5.4. Structural properties

In this section, we establish some structural properties of the alternative MDP that can be exploited to decrease the computational complexity of solving the MDP.

We first rewrite (5.6) into the conventional formulation for single component optimization problems:

$$\tilde{V}_t^{N,i}(x,k) = \tag{5.8}$$
$$\begin{cases} c_u^i + \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(X, k+X+K\Big)\right], & \text{if } x \geq \xi_i, \\ \min\left\{c_p^i + \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(X, k+X+K\Big)\right]; \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x+X, k+X+K\Big)\right]\right\}, & \text{if } x < \xi_i. \end{cases}$$

The first case in Equation (5.8) holds because failed components must be replaced correctively at cost c_u^i. If the components deterioration level is less than ξ_i, we can either perform a preventive replacement, which costs c_p^i, or leave the component in operation until the next decision epoch at no cost. The terminal costs are as introduced before. The next result establishes the monotonicity of the value function $\tilde{V}_t^{N,i}(x,k)$ in x.

Proposition 5.2. *For each $t \in \{0,1,\ldots,T\}$ and $k \in \mathbb{N}_0$, the value function $\tilde{V}_t^{N,i}(x,k)$ is non-decreasing in x.*

Proof: We prove the statement using induction on t. Note that the terminal costs

are non-decreasing in x since $\tilde{V}_T^{N,i}(x,k) = 0$ when $x < \xi_i$ and c_u^i when $x \geq \xi_i$, and $c_u^i > 0$, so that the statement holds for the base case T. Assume that the statement holds for some $0 < t+1 \leq T$, we will show that the statement then also holds for t. Consider $\tilde{V}_t^{N,i}(x,k)$. Since all terms, except the action of leaving the component in operation, on the right of (5.8) are constant with respect to x, we only need to consider $\mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x+X, k+X+K\Big)\right]$. Observe that, since the random variable X is constant with respect to x, the following stochastic order holds for $x^+ \geq x^- \geq 0$ (see Definition 2.1 in Chapter 2 for the definition of $\geq_{\text{st}}$):

$$x^+ + X \geq_{\text{st}} x^- + X.$$

Hence, together with the induction hypothesis, we conclude that the expectation $\mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x+X, k+X+K\Big)\right]$ is non-decreasing in x (cf. Shaked and Shanthikumar, 2007, Theorem 1.A.3). Since the other terms are constant with respect to x, we may conclude that $\tilde{V}_t^{N,i}(x,k)$ is non-decreasing in x. □

Proposition 5.2 implies that if deterioration states increase, we expect to incur higher costs. This is quite intuitive as a higher level of deterioration also increases (i) the probability of a costly failure and/or (ii) the need for a preventive maintenance to replace the deteriorated component with an as-good-as-new one.

By Proposition 5.1, we may also conclude that the value function $V_t^N(\boldsymbol{x}, k)$ is non-decreasing in the standard component-wise order. That is, for any deterioration vectors $\boldsymbol{x}$ and $\boldsymbol{x'}$ such that $x_i \leq x_i'$ for all $i \in \mathscr{N}$, we have that $V_t^N(\boldsymbol{x}, k) \leq V_t^N(\boldsymbol{x'}, k)$. The intuition behind this is similar to the intuition behind Proposition 5.2: Deterioration vectors with higher deterioration levels lead to higher expected costs than deterioration vectors with lower deterioration levels.

The next result, which builds further on Proposition 5.2, establishes the optimality of a control limit policy for the alternative MDP.

Proposition 5.3. *For each $t \in \{0, 1, \ldots, T-1\}$ and $k \in \mathbb{N}_0$, there exists a control limit $\delta_i^{(k,t)}$, $0 < \delta_i^{(k,t)} \leq \xi_i$, such that the optimal action is to carry out a preventive replacement if and only if $x \geq \delta_i^{(k,t)}$.*

Proof: Preventive maintenance at decision epoch $t \in \{0, 1, \ldots, T-1\}$ is optimal when the following equation holds:

$$c_p^i + \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(X, k+X+K\Big)\right] \leq \mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x+X, k+X+K\Big)\right]. \tag{5.9}$$

The left-hand side of Inequality (5.9) is constant with respect to x. Since $\mathbb{E}\left[\tilde{V}_{t+1}^{N,i}\Big(x+X,k+X+K\Big)\right]$ is non-decreasing in x (cf. proof of Proposition 5.2), we find that the right-hand side of Inequality (5.9) is non-decreasing in x. Hence, if the optimal decision is to carry out preventive maintenance in state $(\delta_i^{(k,t)},k)$ at decision epoch $t\in\{1,2,\ldots,T-1\}$, then the same decision is optimal for any state (x,k) at decision epoch t with $x\geq\delta_i^{(k,t)}$, which implies the control limit policy. □

The optimality of a control limit policy is not only intuitive and convenient for the implementation of this optimal policy in practice, it can also be exploited to further decrease the computational burden of solving the original MDP. That is, existing algorithms that rely on these monotonicity properties such as Monotone Backward Induction (see Puterman, 2005, Section 4.7.6) can be used to efficiently solve the alternative MDP, and hence the original MDP.

5.5. Numerical study

This section reports the results of a comprehensive numerical study in which we assess the benefits of pooling data to learn an a-priori unknown parameter, while optimally maintaining N systems over a finite lifespan T. Although the results in the previous sections hold for asymmetric – in terms of costs and failure thresholds – systems, we shall focus on symmetric systems in this numerical study. By doing so, the value function $\tilde{V}_0^N(0,0)$ (we drop the index i as we consider symmetric systems) gives us the cost per system over its lifespan when the data of N systems is pooled. We can use this cost per system to compute a suitable performance metric that allows us to assess the value of pooling learning as a function of N compared to not pooling. To this end, we define the following measure:

$$\%\Delta = 100\left[1-\frac{\tilde{V}_0^N(0,0)}{\tilde{V}_0^1(0,0)}\right],$$

which is the percentage savings per system over the lifespan when the learning of N systems is pooled compared to not pooling any data for those systems and learning the unknown rate independently from the other systems.

We first perform an extensive numerical study with the following parameter settings. We use failure thresholds $\xi\in\{7,10\}$, lifespan lengths $T\in\{50,70,90\}$, and preventive maintenance costs $c_p\in\{0.5,1,1.5\}$ and set the corrective maintenance cost c_u equal

to 10. We remark that the choices for the cost ratio $\frac{c_u}{c_p}$ are of the same order as is usually assumed in the maintenance literature (see, e.g., Van Oosterom et al., 2017; Dursun et al., 2022). Recall that the initial parameter uncertainty is modeled by the random variable Λ which has a Gamma distribution with shape α_0 and rate β_0. We can thus fix the mean of Λ to a certain value and vary its coefficient of variation to increase or decrease the uncertainty. We do so by solving the following set of equations for α_0 and β_0:

$$\mathbb{E}[\Lambda] = \frac{\alpha_0}{\beta_0}, \text{ and } cv_\Lambda = \frac{1}{\sqrt{\alpha_0}},$$

where cv_Λ is the coefficient of variation of Λ. This allows us to explicitly study the impact of the uncertainty (in terms of its mean and coefficient of variation) on the pooling effects. We use means $\mathbb{E}[\Lambda] \in \{0.5, 0.75, 1\}$ and vary the coefficient of variation between 6 values, that is $cv_\Lambda \in \{0.1, 0.25, 0.5, 1, 2, 4\}$. We further consider 7 values for N, $N \in \{1, 2, 4, 6, 8, 10, 20\}$. The choices for these means and coefficient of variations model a wide variety of mean-variability combinations in the parameter uncertainty of the deterioration rate. We further remark that the chosen values for the coefficient of variations are in line with previous work where rate uncertainty of a Poisson process is modeled by a Gamma distribution (see, e.g., Baker, 2001; Van Wingerden et al., 2017). The complete set of parameter values is summarized in Table 5.1. We perform a full factorial numerical study, which results in 324 instances per value of N and for each instance we compute the relative savings $\%\Delta$.

Table 5.1 Input parameter values for numerical study.

	Input parameter	No. of choices	Values
1	Number of systems, N	7	1, 2, 4, 6, 8, 10, 20
2	Failure threshold, ξ	2	7, 10
3	Length of lifespan, T	3	50, 70, 90
4	Preventive maintenance cost, c_p	3	0.5, 1, 1.5
5	Mean of Λ, $\mathbb{E}[\Lambda]$	3	0.5, 0.75, 1
6	Coefficient of variation of Λ, cv_Λ	6	0.1, 0.25, 0.5, 1, 2, 4

The results of the numerical study are summarized in Table 5.2. In this table, we present the average and maximum relative savings $\%\Delta$. For each value of N, we first present the average relative savings for subsets of instances with the same value for a specific input parameter of Table 5.1 (row wise), and then present the average results for all instances with that fixed value of N (bottom row), where each average value is accompanied with the maximum value in brackets.

Based on the results in Table 5.2, we can state the following main observations:

Table 5.2 Relative savings (%Δ) due to pooled learning.

Input parameter	Value	N = 2	4	6	8	10	20
ξ	7	2.7 (16.4)	9.8 (37.2)	14.8 (59.2)	18.0 (71.2)	19.7 (77.2)	24.5 (88.5)
	10	3.4 (19.7)	9.9 (36.8)	14.5 (59.4)	17.1 (71.8)	19.1 (79.7)	22.7 (89.2)
T	50	3.0 (19.7)	9.7 (36.3)	14.5 (57.3)	17.4 (70.2)	19.3 (78.4)	23.6 (88.4)
	70	3.1 (19.7)	9.8 (36.6)	14.7 (58.6)	17.6 (71.2)	19.4 (79.2)	23.6 (88.9)
	90	3.1 (19.7)	10.0 (37.2)	14.8 (59.4)	17.6 (71.8)	19.5 (79.7)	23.6 (89.2)
c_p	0.5	4.4 (19.7)	12.7 (37.2)	18.3 (59.4)	21.6 (71.8)	23.7 (79.7)	28.0 (89.2)
	1	2.9 (12.7)	9.5 (33.5)	14.2 (53.9)	17.1 (65.1)	18.9 (72.5)	22.9 (82.8)
	1.5	1.9 (8.7)	7.4 (29.9)	11.5 (48.9)	14.0 (59.4)	15.6 (65.8)	19.9 (77.6)
$\mathbb{E}[\Lambda]$	0.5	3.0 (19.7)	8.3 (33.5)	12.1 (47.5)	14.6 (60.1)	16.2 (69.1)	19.6 (83.6)
	0.75	3.0 (18.7)	10.0 (36.8)	14.9 (54.6)	17.8 (67.6)	19.7 (76.3)	23.8 (87.5)
	1	3.1 (14.3)	11.2 (37.2)	16.9 (59.4)	20.3 (71.8)	22.4 (79.7)	27.3 (89.2)
cv_Λ	0.1	0.0 (0.1)	0.0 (0.2)	0.1 (0.2)	0.1 (0.3)	0.1 (0.3)	0.2 (0.4)
	0.25	0.2 (0.5)	0.4 (0.8)	0.5 (1.0)	0.5 (1.1)	0.6 (1.2)	0.8 (1.4)
	0.5	0.6 (1.3)	1.3 (2.7)	1.8 (3.5)	2.1 (4.1)	2.4 (5.3)	3.1 (6.4)
	1	4.1 (12.0)	7.8 (22.5)	9.4 (26.3)	10.4 (28.3)	11.0 (31.1)	12.4 (35.7)
	2	10.3 (19.7)	22.7 (36.8)	29.5 (44.8)	33.8 (51.5)	36.7 (57.3)	46.0 (73.9)
	4	13.0 (24.6)	27.2 (37.2)	35.3 (59.4)	40.8 (71.8)	44.9 (79.7)	56.9 (89.2)
Total		3.6 (24.6)	9.8 (37.2)	14.0 (59.4)	16.5 (71.8)	18.2 (79.7)	22.3 (89.2)

1. Pooling of data for learning a common, unknown parameter can lead to significant savings compared to not pooling data and learning it independently.

2. The magnitude of the savings seems to be inextricably linked with the magnitude of uncertainty in the parameter measured by its coefficient of variation. When the coefficient of variation is high, savings of up to 56.9% on average (over all instances with $cv_\Lambda = 4$ and $N = 20$) can be achieved, while if the coefficient of variation is low, savings become almost negligible ($\leq 0.2\%$ on average). This can be explained as follows. When there is high uncertainty in the unknown parameter, pooling data allows the decision maker to faster learn the unknown parameter compared to learning it from data generated by a single system. When there is little uncertainty in the unknown parameter, the benefit of this vanishes; a decision maker already has an accurate belief of the unknown parameter that needs little updating.

3. When comparing the average savings for increasing values of N, we find that pooling has already a significant impact for small values of N, and that the marginal savings gradually decrease when N increases.

4. The savings for each value of N tend to increase as the mean of Λ increases. When $\mathbb{E}[\Lambda]$ increases and N is fixed, the expected deterioration increment between two consecutive decision epochs is larger and, as a result, the optimal

control limit will be more conservative (i.e., further from the failure threshold). The results suggest that in that regime, the choice of the control limit has a higher impact on the resulting costs than when $\mathbb{E}[\Lambda]$ is low and a less conservative control limit is chosen. By pooled learning, one is able to better choose this control limit, and as a result, the relative savings of pooled learning also increase when $\mathbb{E}[\Lambda]$ increases.

5. The savings for each value of N tend to decrease as the ratio $\frac{c_u}{c_p}$ decreases (recall that we keep c_u fixed and vary c_p). When this ratio decreases and N is fixed, maintenance decisions have less impact on the resulting costs – simply because their cost difference decreases – than when $\frac{c_u}{c_p}$ is larger. Consequently, the benefits of utilizing pooled learning in such maintenance decisions also decrease when $\frac{c_u}{c_p}$ decreases.

6. The relative savings per value of N are not so much affected by the length of the lifespan. This suggests that the absolute savings increase close to linearly in the lifespan T, so that the relative savings remain largely unaffected.

Observations 1-4 are also illustrated by Figure 5.1. In this figure we plot the relative savings ($\%\Delta$) as a function of N for various values of cv_Λ when $\mathbb{E}[\Lambda]$ is 0.75 (top) and 1 (bottom). In both figures, we used $\xi = 10$, $c_p = 0.5$, $c_u = 10$, and $T = 90$.

The plots show that for a given level of parameter uncertainty, pooling data across a larger number of systems increases the relative savings. The rate at which the savings increase in N increases significantly in the coefficient of variation. This confirms that pooling data can lead to significant cost reductions, especially when the uncertainty surrounding an unknown parameter is high. We further clearly see the decreasing marginal savings, since the marginal savings due to adding an extra system to the pooled systems decreases as N increases. Finally, when comparing the top figure with the bottom figure, we see that, in general, the savings for each value of N and cv_Λ tend to be larger in the bottom figure, where the mean of Λ is larger.

5.6. Conclusion

In this chapter, we have studied the benefits of data pooling when an unknown deterioration rate that multiple systems have in common needs to be learned over a finite lifespan. We formulated this problem as a finite-horizon Bayesian MDP in which learning is pooled. This formulation suffers from the well-known curse of dimensionality, even for small problem instances. As a remedy to this curse, we

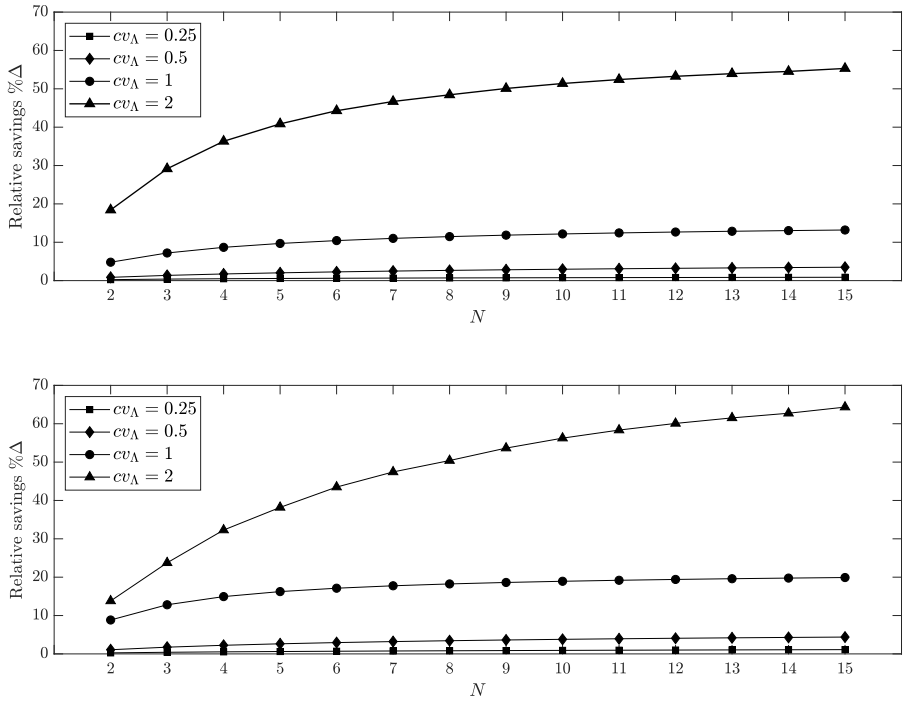


Figure 5.1 Relative savings ($\%\Delta$) as function of N for various values of cv_Λ for $\mathbb{E}[\Lambda] = 0.75$ (top) and $\mathbb{E}[\Lambda] = 1$ (bottom), respectively.

have proven a new decomposition result that establishes the equivalence between the original, highly dimensional MDP and multiple low-dimensional MDPs. For these smaller MDPs, we showed that the structure of the optimal policy has a control limit structure.

The results of a comprehensive numerical study indicated that significant savings can be attained by pooling data to learn the a-priori unknown parameter. We have shown that the exact magnitude of these savings largely depends on the magnitude of the uncertainty in the parameter. When there is high uncertainty, huge savings of close to 57% on average can be realized, while savings become almost negligible when uncertainty decreases.

For further research, it would be interesting to investigate the applicability of our decomposition result in other application areas, and its generalizability to other deterioration processes. For the former, one area that seems particularly promising is a set of inventory stock points; each with Poisson demand with a common but

unknown rate and each stock point with their own cost structures. Jointly learning the unknown demand rate by pooling sales data shares many similarities with the problem treated in this chapter, and our decomposition result might therefore be useful in analyzing that situation. For the latter, we believe that investigating which posterior predictive distributions are closed under convolutions – a crucial step in our proof – is an important direction.

This chapter is the final chapter on CBM with parameter uncertainty. In the next chapter, the attention will shift towards a different maintenance policy: the age-based maintenance policy. Similar to the models on condition-based maintenance, we shall focus on an age-based maintenance problem characterized by parameter uncertainty.

Chapter 6

Optimal age-based maintenance policies with censored learning

6.1. Introduction

In the previous chapters, we looked at condition-based maintenance (CBM) models characterized by parameter uncertainty. In this chapter, we depart from the stream of condition-based maintenance and look at an age-based maintenance (ABM) problem. In the classical ABM problem, introduced in Barlow and Hunter (1960), a decision maker determines the optimal age threshold to preventively replace a single-component system subject to random failures to avoid high costs and/or low reliability associated with corrective replacements. The key assumption in this canonical ABM problem, and in many of its variations (we refer to De Jonge and Scarf (2020) for a recent, comprehensive overview of the area), is that the lifetime distribution is a-priori fully determined and known to the decision maker. However, in many real-life applications, especially when a component has not yet generated (sufficient) data to estimate the lifetime distribution, this assumption is unfounded and necessitates an ABM policy that integrates learning and decision making.

In this chapter, we relax this assumption and revisit the classical ABM problem taking

This chapter is based on Drent et al. (2020b).

into account parameter uncertainty. The main difficulty in analyzing this model is that the data accumulation process consists of both censored and uncensored observations of the underlying lifetime distribution. That is, when a decision maker replaces the component before failure, she only gets a censored observation of the true lifetime (the lifetime if the component would have been kept in operation until failure). This difficulty is typical for ABM models characterized by parameter uncertainty and was not present in the CBM models that we studied in the previous chapters. We overcome this challenge by adopting a parametric distribution that allows for Bayesian learning when data consists of such censoring characteristics. Using this Bayesian framework, we build a finite horizon Bayesian dynamic program (DP) to investigate the optimal policy for a finite sequence of components. Since this optimal policy is analytically intractable, we propose a computationally appealing Bayesian policy and establish its asymptotic optimality.

6.1.1 Contributions

The main contributions of this chapter can be summarized as follows:

1. We build a Bayesian DP to analyze a sequence of ABM problems with parameter uncertainty for a fairly general lifetime distribution, and establish some of its structural properties.
2. We propose a myopic Bayesian policy that allows for learning from censored observations, and that is easy to implement.
3. We show that this myopic policy has asymptotic properties that are desirable. Specifically, we establish that (i) it almost surely learns the unknown parameter, and (ii) it converges to the optimal decision one would have taken with full knowledge of the unknown parameter.

6.1.2 Organization

The remainder of this chapter is organized as follows: We introduce our notation and the problem formulation in Section 6.2. In Section 6.3, we provide some background on censored learning in the parametric Bayesian framework for Newsboy distributions, and establish a new stochastic order result for this class of distributions. We investigate the optimal policy when the sequence of components is finite in Section 6.4. In Section 6.5, we propose a myopic Bayesian policy and establish its asymptotic properties. Finally, Section 6.6 contains concluding remarks.

6.2. Problem formulation

We consider a sequence of N components, where the components are indexed by $n = 1, 2, \ldots, N$. Each component is controlled by the classical ABM policy. That is, component n is either replaced preventively when it has been in operation for a fixed amount of time τ_n (at cost $c_p > 0$), or it is replaced correctively if it fails before this time (at cost $c_u > c_p$). We refer to τ_n as the *age-replacement threshold*.

We assume that the lifetimes of the components are independent and identically distributed, belonging to a family of distributions parametrized by an *unknown* parameter $\theta > 0$ with true value θ_0. Given parameter θ, the lifetime distribution has a probability density function denoted by $f_X(x \,|\, \theta)$ and a cumulative distribution function denoted by $F_X(x \,|\, \theta)$. To preserve the conjugate property under censored learning, we assume that the underlying lifetime distribution is from the class of Newsboy distributions, so that $F_X(x \,|\, \theta)$ has the form

$$F_X(x \,|\, \theta) = 1 - e^{-\theta \ell(x)}, \tag{6.1}$$

where $\ell(x) : [0, \infty) \to [0, \infty)$ is a differentiable, non-decreasing and unbounded function with $\ell(0) = 0$ that is known to the decision maker (Braden and Freimer, 1991). Note for instance that the Weibull distribution with known shape parameter $\beta > 0$ and unknown scale parameter θ can be expressed as a Newsboy distribution by setting $\ell(x) \triangleq x^\beta$. The Weibull distribution has been extensively used in modeling lifetimes due to its ability to model various aging classes of lifetime distributions (see, e.g., Ahmad and Kamaruddin, 2012). The following remark discusses the connection between the form of Newsboy distributions displayed in (6.1) and the hazard rate function of a lifetime distribution.

Remark 6.1. It is well-known that the distribution function of a positive continuous random variable X, denoted with $F_X(x)$, can be specified by its hazard rate function, denoted with $h_X(x)$, as follows (see, e.g., Ross, 2020, Chapter 14):

$$F_X(x) = 1 - \exp\left(- \int_0^x h_X(t) \mathrm{d}t \right). \tag{6.2}$$

Comparing Equation (6.2) with Equation (6.1), we find that for a Newsboy distribution, it holds that its hazard rate function can be decomposed in the product of (a function of) the unknown parameter, say $f(\theta)$, and a function independent of

the unknown parameter, say $\hat{h}(t)$. We can then namely write

$$-\int_0^x h_X(t)\mathrm{d}t = -\int_0^x f(\theta)h'(t)\mathrm{d}t = -f(\theta)\int_0^x \hat{h}(t)\mathrm{d}t,$$

such that $\int_0^x \hat{h}(t)\mathrm{d}t \triangleq \ell(x)$. As the hazard rate function is often used in maintenance and reliability applications (see, e.g., Boland et al., 1994), this characterization of Newsboy distributions might be useful as well. ◇

Newsboy distributions, introduced by Braden and Freimer (1991), received much attention from the inventory research community as they are used in the modeling of excess demand of the inventory level that is lost and thus unobserved. Lariviere and Porteus (1999), Ding et al. (2002), Chen and Plambeck (2008), Lu et al. (2008), Bensoussan et al. (2009), Chen (2010), Bisi et al. (2011), and Mersereau (2015) all assume a Newsboy distribution with unknown parameter, which permits an exact analysis of the optimal policy under different variations of this lost sales inventory control problem with censored demand learning. Despite the similarities between these inventory control problems and maintenance under lifetime censoring, the parametric Bayesian framework for censored learning assumed in these papers has not found its way to the maintenance community. Specifically, the inventory level and the age-replacement threshold influence the information accumulation process of the unknown demand and lifetime distribution, respectively, in the same way. That is, if all inventory is sold during a period, then the decision maker obtains a censored observation of the true demand in that period (it is at least as large as the inventory level), while if a decision maker does not sell out all inventory, she obtains an uncensored observation of the demand.

The decision maker can only observe censored observations, rather than lifetime realizations. For component n, the censored observation is given by $X_n \wedge \tau_n \triangleq \min\{X_n, \tau_n\}$, where X_n is the realized lifetime and τ_n the imposed age-replacement threshold. Let $\mathcal{F}_n$ be the filtration generated by this censored lifetime process. Hence, for $n \geq 1$, we have the σ-algebra

$$\mathcal{F}_n \triangleq \sigma\big(X_1 \wedge \tau_1, \tau_1, X_2 \wedge \tau_2, \tau_2, \ldots, X_n \wedge \tau_n, \tau_n\big),$$

and let $\mathcal{F}_0$ be the trivial σ-algebra. It is thus evident that the accumulated information about the lifetime up to the n^{th} component is impacted by all past replacements decisions.

The decision maker wishes to minimize the total expected discounted cost due to both corrective and preventive replacements, where costs are discounted with rate

$\alpha \in (0,1)$, over an N component horizon. This optimality criterion is often employed in finite-horizon problems in the ABM literature (De Jonge and Scarf, 2020). We are interested in finding the optimal non-anticipatory policy (i.e. the decision τ_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$) that attains this minimum.

Observe that when deciding on the age-replacement threshold for the n^{th} component, there is an inherent trade-off between the direct expected cost of the n^{th} component and the impact the decision has on future costs through the information accumulation process. Specifically, exploration (i.e., a higher value of the age-replacement threshold) increases the probability of a corrective replacement, but at the same time leads to accumulating more, valuable information. This phenomenon is often referred to as the *exploration-exploitation trade-off*.

6.3. Censored lifetime learning

In this section, we describe how the unknown parameter θ of the lifetime distribution of components can be inferred with increasing accuracy as information is accumulated. The approach is based on Braden and Freimer (1991), who show that the Gamma distribution is the conjugate prior for all Newsboy distributions.

Following a Bayesian approach, we treat the unknown parameter as a random variable, denoted with Θ, and assume that the decision maker has a prior density for the unknown parameter θ, denoted by $p_\Theta(\theta)$. This density captures the information about the unknown parameter of the lifetime distribution.

Let m and k denote the shape and rate parameter, respectively, of the Gamma distribution. The prior density is then given by

$$p_\Theta(\theta \,|\, m, k) \triangleq \frac{k^m \theta^{m-1} e^{-k\theta}}{\Gamma(m)}, \quad \text{for all } \theta > 0, \tag{6.3}$$

where $\Gamma(\cdot)$ denotes the gamma function. Using (cf. Equation (6.1))

$$f_X(x \,|\, \theta) = \frac{d}{dx} F_X(x \,|\, \theta) = \theta \ell'(x) e^{-\theta \ell(x)},$$

and unconditioning on θ using Equation (6.3), we obtain the posterior predictive

lifetime density and distribution function, respectively;

$$f_X(x|m,k) = \frac{mk^m \ell'(x)}{[k+\ell(x)]^{m+1}}, \text{ and,}$$
$$F_X(x|m,k) = 1 - \left[\frac{k}{k+\ell(x)}\right]^m.$$

We use the shorthand notation m_n and k_n to denote the updated shape and rate parameter conditional on $\mathcal{F}_n$. Here, we omit the dependence on $\mathcal{F}_n$ as there is a mapping between $\mathcal{F}_n$ and (m_n, k_n), which we explain henceforth. Let (m_0, k_0) denote the parameters before the installation of the first component. Then for $n \geq 1$, conditional on $\mathcal{F}_n$, the prior hyperparameters are computed as (see Section 5 of Braden and Freimer (1991)):

$$m_n = m_0 + \sum_{i=1}^{n} \mathbb{1}_{\{X_i < \tau_i\}} = m_{n-1} + \mathbb{1}_{\{X_n < \tau_n\}}, \text{ and,}$$
$$k_n = k_0 + \sum_{i=1}^{n} \ell(X_i \wedge \tau_i) = k_{n-1} + \ell(X_n \wedge \tau_n), \tag{6.4}$$

where $\mathbb{1}_{\{a\}}$ denotes the indicator function taking value 1 if event a occurs and value 0 otherwise. Here, we assume that $m_0 \cdot k_0 > 1$, and thus $m_n \cdot k_n > 1$ for all $n \geq 1$ through the update rules, so that the posterior predictive lifetime distribution of each component has a finite expectation. Observe that the rate parameter k_n is an aggregate of all observations. The shape parameter m_n counts the number of uncensored observations (corresponding to corrective replacements), and, as the coefficient of variation for the Gamma prior is equal to $\sqrt{1/m_n}$, it is also a measure for the precision of the accrued information on the unknown parameter θ.

Equation (6.4) induces a simple, Markovian scheme for sequentially inferring the lifetime distribution conditional on $\mathcal{F}_n$. In what follows, for notational simplicity and in order to enhance the readability of the chapter, we write, depending on our objective, either $\mathcal{F}_n$ or (m_n, k_n).

In order to derive structural properties of the conditional posterior predictive lifetime random variable given (m_n, k_n), denoted with $X(m_n, k_n) \triangleq \{X \,|\, m_n, k_n\}$, and in order to make comparisons between different conditional posterior predictive lifetime random variables, we use the hazard rate ordering:

Definition 6.1 (1.B.1 Definition, Shaked and Shanthikumar, 2007)**.** Let Y and Z be two non-negative random variables with absolutely continuous distribution functions and with hazard rate functions $h_Y(x)$ and $h_Z(x)$, respectively, such that $h_Y(x) \geq$

$h_Z(x)$ for all $x > 0$. Then Y is said to be smaller than Z in the hazard rate order.

We now present an important proposition that indicates how the accrued information, encoded in (m, k), affects the stochastic ordering of the conditional posterior predictive lifetime distribution.

Proposition 6.1. *The conditional posterior predictive lifetime random variable* $X(m, k)$ *is:*

(i) stochastically increasing in the hazard rate order in the rate parameter k*, and*

(ii) stochastically decreasing in the hazard rate order in the shape parameter m*.*

PROOF: Let $h_X(x \,|\, m, k)$ denote the hazard rate function of the conditional posterior predictive lifetime when the shape and rate parameters are m and k, respectively. We then have

$$h_X(x \,|\, m, k) \triangleq \frac{f_X(x \,|\, m, k)}{1 - F_X(x \,|\, m, k)} = \frac{\frac{mk^m \ell'(x)}{[k+\ell(x)]^{m+1}}}{\left[\frac{k}{k+\ell(x)}\right]^m} = \frac{m\ell'(x)}{k + \ell(x)}.$$

Observe that since $\ell(x)$ is non-negative and non-decreasing by assumption, we have the following. If $M > m > 0$ then $h_X(x \,|\, M, k) \geq h_X(x \,|\, m, k)$ for all $x > 0$, which establishes the hazard rate order in m. Finally, if $K > k > 0$ then $h_X(x \,|\, m, K) \leq h_X(x \,|\, m, k)$ for all $x > 0$, which establishes the hazard rate order in k. □

Assertion (i) establishes the monotonic increase in the expected conditional posterior predictive lifetime when the aggregate of observations increases for a fixed number of uncensored observations. This implies that if the aggregate of all observations is high, then past components have had a relatively long lifetime on average. Hence, the decision maker predicts that the next component will have a longer lifetime than in the case where the accrued information has a lower aggregate of observations.

Assertion (ii) establishes the stochastic-ordering property of lifetime distributions that are updated using censored observations and uncensored observations, respectively. It states that a censored lifetime observation results in a lifetime distribution that is stochastically greater than that from an uncensored observation. An intuitive explanation is as follows. With a censored observation, the true lifetime is at least as large as the censored observation, as opposed to the uncensored observation, where the true lifetime is equal to the uncensored observation.

As the usual stochastic order is implied by the hazard rate order (see e.g., Theorem 1.B.1. of Shaked and Shanthikumar (2007)), Proposition 6.2 and 6.3 of Braden and Freimer (1991), which state the usual stochastic order of the conditional posterior predictive lifetime, directly follow from Proposition 6.1. The hazard rate order is particularly useful in maintenance and reliability theory due to the importance of the hazard rate function in these areas (cf. Boland et al., 1994).

6.4. Optimal policy for a finite sequence

In this section, we explore both analytically and numerically the structure of the optimal policy when N is finite. In a finite sequence of components, the exploration-exploitation trade-off urges the decision maker to explicitly recognize the impact that current decisions have on both the direct expected costs and the future expected costs through the information accumulation process. This interdependence is made explicit by formulating the optimization problem as a DP. As we will see, in this DP, it is more convenient to work with the hyperparameters (m, k) to capture the dynamics of the information accumulation process, rather than directly with $\{\mathcal{F}_n\}_{1\leq n\leq N}$ itself.

To this end, let the value function $V_n(m, k)$ denote the minimum total expected discounted cost over components $n, n+1, \ldots, N$, starting with component n, when the updated hyperparameters are (m, k), respectively. We assume the terminal cost to be zero, hence $V_{N+1}(m, k) \triangleq 0$ for all (m, k). The optimality equations, for $n = 1, 2, \ldots, N$, are:

$$V_n(m,k) = \min_{\tau_n \geq 0} \left\{ C_n(\tau_n \,|\, m,k) + \int_0^\infty G_n(\tau_n, x \,|\, m,k) f_X(x \,|\, m,k) \mathrm{d}x \right\}, \tag{6.5}$$

where

$$C_n(\tau_n \,|\, m,k) = c_u \int_0^{\tau_n} e^{-\alpha x} f_X(x \,|\, m,k) \mathrm{d}x + c_p e^{-\alpha \tau_n} \big(1 - F_X(\tau_n \,|\, m,k)\big), \tag{6.6}$$

denotes the direct expected discounted cost function of component n when the age-replacement threshold is τ_n (the first part is due to corrective replacement and the second part is due to preventive replacement) and

$$G_n(\tau_n, x \,|\, m,k) \triangleq \begin{cases} e^{-\alpha x} V_{n+1}(m+1, k+\ell(x)), & \text{if } x < \tau_n, \\ e^{-\alpha \tau_n} V_{n+1}(m, k+\ell(\tau_n)), & \text{if } x \geq \tau_n, \end{cases}$$

denotes the discounted cost function over the remaining components $n+1, \ldots, N$ when

the age-replacement threshold of the n^{th} component is τ_n and the lifetime realization equals x.

Since the costs c_p and c_u are uniformly bounded, the existence of an optimal policy in this setting is guaranteed and this policy satisfies the optimality equations displayed in (6.5) (see, e.g., Bertsekas, 2012, Proposition 1.3.1). Observe that over an N component horizon, the minimum total expected discounted cost is given by $V_1(m_0, k_0)$ which can be found by solving Equation (6.5) via backward induction. As the age-replacement threshold affects not only the direct expected discounted costs, but also the posterior distribution of the future components it is difficult to analytically establish structural properties of the optimal policy $\{\tau_n\}_{n\in\{1,\ldots,N\}}$. This is a complexity that this type of Bayesian DP generally suffers from, see e.g., Ding et al. (2002); Chen and Plambeck (2008); Chen (2010). In the remainder of this section, we establish some structural properties of the value function and numerically investigate $\{\tau_n\}_{n\in\{1,\ldots,N\}}$.

We first state two properties regarding the direct expected discounted cost function that are instrumental in characterizing the behavior of $V_n(m,k)$ with respect to its parameters.

Lemma 6.1. *For all $\tau_n \geq 0$ and $n \in \{1, 2, \ldots, N\}$, $C_n(\tau_n \,|\, m, k)$ is*

(i) non-increasing in k, and

(ii) non-decreasing in m.

Proof: Note that Equation (6.6) can be rewritten as

$$\begin{aligned} C_n(\tau_n \,|\, m,k) =& - c_u \int_0^{\tau_n} e^{-\alpha x} \mathrm{d}\Big((1 - F_X(x \,|\, m,k)\Big) + c_p e^{-\alpha\tau_n}\Big(1 - F_X(\tau_n \,|\, m,k)\Big) \\ =& (c_p - c_u) e^{-\alpha\tau_n}\Big(1 - F_X(\tau_n \,|\, m,k)\Big) + c_u \\ & - c_u\alpha \int_0^{\tau_n} e^{-\alpha x}\Big(1 - F_X(\tau_n \,|\, m,k)\Big)\mathrm{d}x. \end{aligned} \tag{6.7}$$

Since $c_p < c_u$, $\alpha \in (0,1)$, and because of Assertion (i) (Assertion (ii)) of Proposition 6.1, both terms involving $F_X(\tau_n \,|\, m,k)$ in the last two lines of (6.7) are non-increasing (non-decreasing) in k (m), which establishes the result. □

We now proceed with two properties regarding the value function.

Theorem 6.1. *For all $n \in \{1, 2, \ldots, N+1\}$, $V_n(m,k)$ is*

(i) non-increasing in k, and

(ii) non-decreasing in m.

PROOF: We first prove Assertion (i) by backward induction. The base case, i.e. $V_{N+1}(m,k)$, holds trivially as terminal costs $V_{N+1}(m,k) = 0$ for all (m,k). Let $K > k$ and assume inductively that $V_{n+1}(m,K) \leq V_{n+1}(m,k)$, and let $\tau^n_{m,j}$ denote the optimal age-replacement threshold for the n^{th} component, when the shape and rate parameters are m and j, respectively. We have

$$
\begin{aligned}
&V_n(m,K) - V_n(m,k) \\
&= C_n(\tau^n_{m,K} \,|\, m,K) + \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,K) f_X(x \,|\, m,K) \mathrm{d}x \\
&\quad - C_n(\tau^n_{m,k} \,|\, m,k) - \int_0^\infty G_n(\tau^n_{m,k}, x \,|\, m,k) f_X(x \,|\, m,k) \mathrm{d}x \\
&\leq C_n(\tau^n_{m,K} \,|\, m,K) + \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,K) f_X(x \,|\, m,K) \mathrm{d}x \\
&\quad - C_n(\tau^n_{m,K} \,|\, m,k) - \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,k) f_X(x \,|\, m,k) \mathrm{d}x \\
&\leq \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,K) f_X(x \,|\, m,K) \mathrm{d}x - \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,k) f_X(x \,|\, m,k) \mathrm{d}x \\
&\leq \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,k) f_X(x \,|\, m,K) \mathrm{d}x - \int_0^\infty G_n(\tau^n_{m,K}, x \,|\, m,k) f_X(x \,|\, m,k) \mathrm{d}x \\
&= \mathbb{E}\big[G_n(\tau^n_{m,K}, X \,|\, m,k) \,|\, m,K\big] - \mathbb{E}\big[G_n(\tau^n_{m,K}, X \,|\, m,k) \,|\, m,k\big] \leq 0.
\end{aligned}
$$

The first inequality holds because $\tau^n_{m,K}$ is a feasible policy for m and k but not necessarily optimal. The second inequality holds by Assertion (i) of Lemma 6.1. The third inequality follows from the induction hypothesis. That is, $V_{n+1}(m,K) \leq V_{n+1}(m,k)$ implies that

$$
\begin{aligned}
&G_n(\tau^n_{m,K}, x \,|\, m,K) - G_n(\tau^n_{m,K}, x \,|\, m,k) \\
&\quad = \begin{cases} e^{-\alpha x}\Big(V_{n+1}(m+1, K+\ell(x)) - V_{n+1}(m+1, k+\ell(x))\Big) \leq 0, \\ e^{-\alpha \tau^n_{m,K}}\Big(V_{n+1}(m, K+\ell(\tau^n_{m,K})) - V_{n+1}(m, k+\ell(\tau^n_{m,K}))\Big) \leq 0, \end{cases}
\end{aligned}
$$

where the first branch corresponds to $x < \tau^n_{m,K}$ and the second branch to $x \geq \tau^n_{m,K}$. Hence, $G_n(\tau^n_{m,K}, x \,|\, m,K) \leq G_n(\tau^n_{m,K}, x \,|\, m,k)$ for all $x \geq 0$, which implies the third inequality. Then, following a similar reasoning, $G_n(\tau^n_{m,K}, x \,|\, m,k)$ is non-increasing in x by the induction hypothesis and since $e^{-\alpha x}$ is decreasing in x for $\alpha > 0$. We then have the expectation of a decreasing function $G_n(\tau^n_{m,K}, x \,|\, m,k)$, so that the

last inequality is implied by the stochastic order of Assertion (i) of Proposition 6.1 between $f_X(x \mid m, K)$ and $f_X(x \mid m, k)$ (see e.g., Proposition 9.1.2 of Ross (1996)).

The proof of Assertion (ii) follows verbatim the proof of Assertion (i), starting with $M > m$, looking at the difference $V_n(m,k) - V_n(M,k)$, and assuming inductively that $V_{n+1}(m,k) \leq V_{n+1}(M,k)$ with $M > m$. □

Theorem 6.1 establishes the monotonicity of the value function in both the aggregate of all observations k and the number of exact observations m. The intuition behind both parts (and their individual counterparts in Lemma 6.1) is as follows: If the aggregate of all observations increases and everything else is held fixed, it means that on average, each component has had a longer lifetime and is thus discounted at a higher rate. This leads to a lower total expected discounted cost. A similar reasoning holds if the number of exact observations increases and the aggregate of all observations is held fixed. Both preventive and corrective replacement costs are then, on average, discounted at an equal rate, but there are more exact observations so that c_u is incurred more often. This leads to a higher total expected discounted cost.

In extensive numerical experiments (see Appendix 6.A more results), with $N = 20$ components with Weibull distributed lifetimes with various shape values β and unknown scale parameter θ, for various cost ratios of $\frac{c_u}{c_p}$, we consistently observed three features of the optimal decisions that we state in the following conjecture.

Conjecture 6.1. *The optimal age-replacement threshold $\tau^n_{m,k}$ is:*

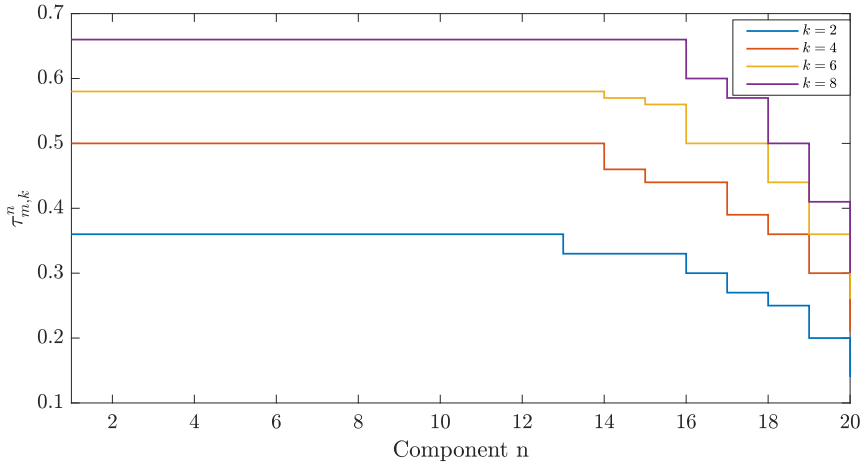
(i) non-decreasing in k,

(ii) non-increasing in m, and

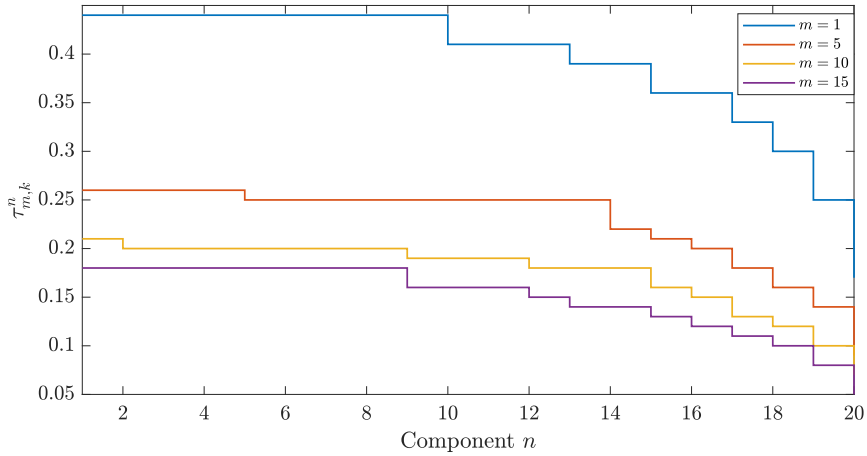
(iii) non-increasing in n.

These observations are illustrated by Figures 6.1a and 6.1b. In Figure 6.1a we observe that $\tau^n_{m,k}$ is non-increasing in n and non-decreasing in k (if we compare the different curves). In Figure 6.1b we observe that $\tau^n_{m,k}$ is non-increasing in n and, non-increasing in m (if we compare the different curves).

The intuition behind Assertions (i) and (ii) is the following. Everything else fixed, a higher k (lower m) (see Proposition 6.1) implies a higher (lower) component's lifetime on average, so that it is better to impose a higher (lower) age-replacement threshold. The intuition behind Assertion (iii) is precisely the exploration-exploitation trade-off. Indeed, if the information state (encoded in m and k) is equal, the exploratory



(a) $\tau^n_{m,k}$ as function of n for various values of k, with $m = 5$.



(b) $\tau^n_{m,k}$ as function of n for various values of m, with $k = 1$.

Figure 6.1 Illustration of observed behavior of optimal age-replacement threshold $\tau^n_{m,k}$, with $c_p = 1$, $c_u = 4$, $\ell(x) = x^3$ and $\alpha = 0.9$.

benefits of setting a higher age-replacement threshold are greater at the beginning of the sequence. There are then namely more remaining components that benefit from the accumulated information than when imposing the same age-replacement threshold later in the sequence.

6.5. An asymptotically optimal policy

In the previous section we investigated both analytically and numerically the structure of the optimal policy for a finite sequence of components (i.e. $N < \infty$). However, obtaining this optimal policy via the proposed DP is analytically intractable and computationally challenging for large instances. Therefore, in this section, we investigate asymptotic properties (as $N \to \infty$) of the information accumulation process and of a computationally tractable myopic policy.

We proceed in two steps. First we show that under any reasonable policy, the learning converges in the Bayesian sense. We then propose a myopic policy and prove its asymptotic optimality. It turns out that for this analysis, it is more convenient to work directly with $\mathcal{F}_n$ rather than the corresponding hyperparameters (m_n, k_n).

6.5.1 Convergence of learning

Recall that $\{\Theta \,|\, \mathcal{F}_n\}_{n=1,2,\ldots,N}$ denotes our posterior belief regarding the unknown parameter θ and that the hyperparameters are updated according to Equation (6.4). In the next result, we show that this posterior expectation, denoted with $\mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big]$ converges (a.s.) to the true value θ_0, as $N \to \infty$, and that the variance, denoted with $\mathbb{V}ar\big[\Theta \,|\, \mathcal{F}_N\big]$, converges (a.s.) to 0, as $N \to \infty$. This convergence is guaranteed under any policy in which the limiting age-replacement threshold is strictly positive. We believe that this assumption is justified as it is not natural to replace a component immediately after installation.

Theorem 6.2. *Under any policy for which* $\lim_{N\to\infty} \tau_N = \tau > 0$*, we have*

$$\mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big] \xrightarrow{a.s.} \theta_0 \text{ and } \mathbb{V}ar\big[\Theta \,|\, \mathcal{F}_N\big] \xrightarrow{a.s.} 0 \text{ as } N \to \infty.$$

Proof: By the updating scheme of the posterior density, see Equation (6.4), $\mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big]$ can be written as (cf. Equation (6.3)):

$$\begin{aligned}
\mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big] &= \frac{m_N}{k_N} \\
&= \frac{m_0 + \sum_{n=1}^{N} \mathbb{1}_{\{X_n < \tau_n\}}}{k_0 + \sum_{n=1}^{N} \ell(X_n \wedge \tau_n)} \\
&= \frac{\frac{m_0}{N} + \frac{\sum_{n=1}^{N} \mathbb{1}_{\{X_n < \tau_n\}}}{N}}{\frac{k_0}{N} + \frac{\sum_{n=1}^{N} \ell(X_n \wedge \tau_n)}{N}}.
\end{aligned}$$

Then, when $N \to \infty$, given $\lim_{N\to\infty} \tau_N = \tau$, we have by the strong law of large numbers, almost surely,

$$\lim_{N\to\infty} \mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big] = \lim_{N\to\infty} \frac{\frac{m_0}{N} + \frac{\sum_{n=1}^N \mathbb{1}_{\{X_n<\tau_n\}}}{N}}{\frac{k_0}{N} + \frac{\sum_{n=1}^N \ell(X_n\wedge\tau_n)}{N}} = \frac{F_X(\tau \,|\, \theta_0)}{\mathbb{E}[\ell(X\wedge\tau)]}.$$

Using straightforward calculus yields, for $\tau > 0$,

$$\begin{aligned}\frac{F_X(\tau \,|\, \theta_0)}{\mathbb{E}[\ell(X\wedge\tau)]} &= \frac{F_X(\tau \,|\, \theta_0)}{\int_0^\infty \mathbb{P}\Big(\ell(X\wedge\tau) > y\Big)\mathrm{d}y} \\ &= \frac{F_X(\tau \,|\, \theta_0)}{\int_0^{\ell(\tau)} \mathbb{P}\Big(X > \ell^{-1}(y)\Big)\mathrm{d}y} \\ &= \frac{F_X(\tau \,|\, \theta_0)}{\int_0^{\ell(\tau)} e^{-\theta_0 y}\mathrm{d}y} \\ &= \frac{F_X(\tau \,|\, \theta_0)}{\frac{1}{\theta_0}F_X(\tau \,|\, \theta_0)} = \theta_0.\end{aligned}$$

The second equality follows from the non-decreasing property of the function $\ell(\cdot)$. The third and fourth equality follow using the cumulative distribution function of a Newsboy distribution, see Equation (6.1).

For the second part, note that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\big[\Theta \,|\, \mathcal{F}_N\big] &= \frac{m_0 + \sum_{n=1}^N \mathbb{1}_{\{X_n<\tau_n\}}}{\Big(k_0 + \sum_{n=1}^N \ell(X_n\wedge\tau_n)\Big)^2} \\ &= \mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big] \cdot \frac{1}{k_0 + \sum_{n=1}^N \ell(X_n\wedge\tau_n)}.\end{aligned}$$

Using

$$\mathbb{E}\big[\Theta \,|\, \mathcal{F}_N\big] \xrightarrow{a.s.} \theta_0 \text{ and } \frac{1}{k_0 + \sum_{n=1}^N \ell(X_n\wedge\tau_n)} \xrightarrow{a.s.} 0,$$

when $N \to \infty$ leads to the desired result. □

Observe that the crucial part of the proof, i.e. the equality $\frac{F_X(\tau \,|\, \theta_0)}{\mathbb{E}[\ell(X\wedge\tau)]} = \theta_0$ for all $\tau > 0$, relies explicitly on the form of the cumulative distribution function of a Newsboy distribution. As such, this is a distinctive feature of this class of distributions.

Theorem 6.2 establishes the Bayesian consistency of the posterior distribution $\{\Theta \,|\, \mathcal{F}_N\}$ at the true value θ_0 (DeGroot, 2005). This implies that the true value will be learned with certainty as information is accumulated.

6.5.2 Convergence of myopic policy

Given full knowledge of the true value θ_0, the optimal age-replacement threshold for each component can be computed by minimizing the direct expected discounted cost function, that is,

$$\begin{aligned}\tau^*(\theta_0) &\triangleq \operatorname{argmin}_{\tau\geq 0} C(\tau \,|\, \theta_0) \\ &= \operatorname{argmin}_{\tau\geq 0} \left\{ c_u \int_0^\tau e^{-\alpha x} f_X(x \,|\, \theta_0) \mathrm{d}x + c_p e^{-\alpha\tau}\big(1 - F_X(\tau \,|\, \theta_0)\big) \right\},\end{aligned} \tag{6.8}$$

where we use the notation $C(\tau \,|\, \theta_0)$ to denote the direct expected discounted cost function when θ_0 is known. We refer to $\tau^*(\theta_0)$ as the *Oracle* as this decision requires full knowledge about the unknown parameter and is hence not attainable in practice. The following remark relates the uniqueness and finiteness of $\tau^*(\theta_0)$ to properties of $\ell(x)$.

Remark 6.2. It has been shown in Fox (1966) that $\tau^*(\theta_0)$ is unique and finite if and only if X has a strictly increasing hazard rate. An increasing hazard rate implies that the component degrades over time so that there is an incentive to perform preventive maintenance. Given full knowledge about θ_0, the hazard rate, denoted with $h_X(x \,|\, \theta_0)$, is equal to

$$h_X(x \,|\, \theta_0) = \frac{\frac{d}{dx} F_X(x \,|\, \theta_0)}{1 - F_X(x \,|\, \theta_0)} = \frac{\theta_0 \ell'(x) e^{-\theta_0 \ell(x)}}{e^{-\theta_0 \ell(x)}} = \theta_0 \ell'(x).$$

It is then obvious (recall that $\theta_0 > 0$) that $h(x \,|\, \theta_0)$ is strictly increasing if and only if $\ell'(x)$ is strictly increasing. Hence, $\tau^*(\theta_0)$ is unique and finite if and only if $\ell''(x) > 0$. This statement can thus be verified a-priori without any knowledge about the unknown parameter θ_0. ◇

A policy that is attainable in the absence of knowledge on the true value θ_0 is to, upon installation of the n^{th} component, implement the age-replacement threshold that only minimizes the direct expected discounted costs given the accumulated information. Recall that we used this function also in the DP formulation, where (m, k) captured the accumulated information, see Equation (6.6). In other words, this myopic policy does not integrate learning with decision making, and solely focuses on exploitation.

For $n = 1, 2, \ldots, N$, we denote the optimal age-replacement threshold of this myopic Bayesian policy with $\tau^{mb}(\mathcal{F}_n)$, so that

$$\begin{aligned}\tau^{mb}(\mathcal{F}_n) &\triangleq \operatorname{argmin}_{\tau\geq 0}\left\{c_u\int_0^\tau e^{-\alpha x} f_X(x\,|\,\mathcal{F}_n)\mathrm{d}x + c_p e^{-\alpha\tau}\big(1-F_X(\tau\,|\,\mathcal{F}_n)\big)\right\}\\ &= \operatorname{argmin}_{\tau\geq 0} C_n(\tau\,|\,\mathcal{F}_n).\end{aligned}$$

The following result establishes the asymptotic optimality of this myopic Bayesian policy in the sense that the induced decision converges to the Oracle. It relies on the condition that the Oracle is unique and finite, which, as stated before, can be easily verified before any information is accrued, see Remark 6.2.

Theorem 6.3. *Suppose $\tau^*(\theta_0)$ is unique and finite. Then the myopic Bayesian policy is asymptotically optimal; that is,*

$$\lim_{N\to\infty}\tau^{mb}(\mathcal{F}_N) = \tau^*(\theta_0).$$

Proof: We first prove that $C_N(\tau\,|\,\mathcal{F}_N)$ converges uniformly to $C(\tau\,|\,\theta_0)$ when $N\to\infty$. We then show that this uniform convergence together with properties of $C_N(\tau\,|\,\mathcal{F}_N)$ implies a stronger notion of convergence, namely epi-convergence, which leads directly to the desired result. We have

$$\begin{aligned}&\sup_{\tau\geq 0}\Big|C_N(\tau\,|\,\mathcal{F}_N) - C(\tau\,|\,\theta_0)\Big|\\ &= \sup_{\tau\geq 0}\Big|c_u\int_0^\tau e^{-\alpha x} f_X(x\,|\,\mathcal{F}_N)\mathrm{d}x + c_p e^{-\alpha\tau}\big(1-F_X(\tau\,|\,\mathcal{F}_N)\big)\\ &\quad - c_u\int_0^\tau e^{-\alpha x} f_X(x\,|\,\theta_0)\mathrm{d}x - c_p e^{-\alpha\tau}\big(1-F_X(\tau\,|\,\theta_0)\big)\Big|\\ &= \sup_{\tau\geq 0}\Big|c_u\int_0^\tau e^{-\alpha x}\big(f_X(x\,|\,\mathcal{F}_N) - f_X(x\,|\,\theta_0)\big)\mathrm{d}x\\ &\quad + c_p e^{-\alpha\tau}\int_\tau^\infty \big(f_X(x\,|\,\mathcal{F}_N) - f_X(x\,|\,\theta_0)\big)\mathrm{d}x\Big|\\ &\leq \sup_{\tau\geq 0}\left(c_u\int_0^\tau e^{-\alpha x}\Big|f_X(x\,|\,\mathcal{F}_N) - f_X(x\,|\,\theta_0)\Big|\mathrm{d}x\right. \qquad (6.9)\\ &\quad \left. + c_p e^{-\alpha\tau}\int_\tau^\infty \Big|f_X(x\,|\,\mathcal{F}_N) - f_X(x\,|\,\theta_0)\Big|\mathrm{d}x\right)\\ &\leq \sup_{\tau\geq 0}\left(c_u\int_0^\tau e^{-\alpha x}\Big|f_X(x\,|\,\mathcal{F}_N) - f_X(x\,|\,\theta_0)\Big|\mathrm{d}x\right) \qquad (6.10)\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{\tau \geq 0} \left(c_p e^{-\alpha\tau} \int_\tau^\infty \Big| f_X(x \,|\, \mathcal{F}_N) - f_X(x \,|\, \theta_0) \Big| \mathrm{d}x \right) \\
&= c_u \int_0^\infty e^{-\alpha x} \Big| f_X(x \,|\, \mathcal{F}_N) - f_X(x \,|\, \theta_0) \Big| \mathrm{d}x \\
&\quad + c_p \int_0^\infty \Big| f_X(x \,|\, \mathcal{F}_N) - f_X(x \,|\, \theta_0) \Big| \mathrm{d}x \Big). \qquad (6.11)
\end{aligned}
$$

Inequality (6.9) follows from the Triangle inequality and Hölder's inequality, and finally, Inequality (6.10) is a Triangle-like inequality for the supremum operator.

Note that

$$
f_X(x \,|\, \mathcal{F}_N) = \int_0^\infty f_X(x \,|\, \theta) p_\Theta(\theta \,|\, \mathcal{F}_N) \mathrm{d}\theta.
$$

Since $f_X(x \,|\, \theta)$ is a bounded, continuous function, we have by Theorem 6.2 and the weak convergence of measures that

$$
f_X(x \,|\, \mathcal{F}_N) \to f_X(x \,|\, \theta_0), \quad \text{when } N \to \infty.
$$

Using this weak convergence of measures, we have by Scheffé's Theorem (see e.g., Theorem 16.11 of Billingsley (1995)) that

$$
\lim_{N\to\infty} \int_0^\infty \Big| f_X(x \,|\, \mathcal{F}_N) - f_X(x \,|\, \theta_0) \Big| \mathrm{d}x = 0, \qquad (6.12)
$$

$$
\lim_{N\to\infty} \int_0^\infty e^{-\alpha x} \Big| f_X(x \,|\, \mathcal{F}_N) - f_X(x \,|\, \theta_0) \Big| \mathrm{d}x = 0. \qquad (6.13)
$$

Using the established bound in (6.11) in combination with (6.12) and (6.13), we have

$$
\lim_{N\to\infty} \sup_{\tau \geq 0} \Big| C_N(\tau \,|\, \mathcal{F}_N) - C(\tau \,|\, \theta_0) \Big| = 0,
$$

which establishes the uniform convergence of the direct expected cost functions.

Since $C_N(\tau \,|\, \mathcal{F}_N)$ converges uniformly to $C(\tau \,|\, \theta_0)$, and $C_N(\tau \,|\, \mathcal{F}_N)$ is finite (i.e. $0 < C_N(\tau \,|\, \mathcal{F}_N) \leq c_u$) and continuous for all $\tau \geq 0$ and $N \geq 0$, we have by Proposition 7.15 of Rockafellar and Wets (2009) that $C_N(\tau \,|\, \mathcal{F}_N)$ epi-converges to $C(\tau \,|\, \theta_0)$ when $N \to \infty$.

Note that the sequence $C_N(\tau \,|\, \mathcal{F}_N)$ is eventually level-bounded since we assume that $\tau^*(\theta_0)$ is unique and finite (hence the argmin set will eventually be bounded and nonempty). Then, as $C(\tau \,|\, \theta_0)$ is a proper and left semi-continuous function, we have

by Theorem 7.33 of Rockafellar and Wets (2009) that

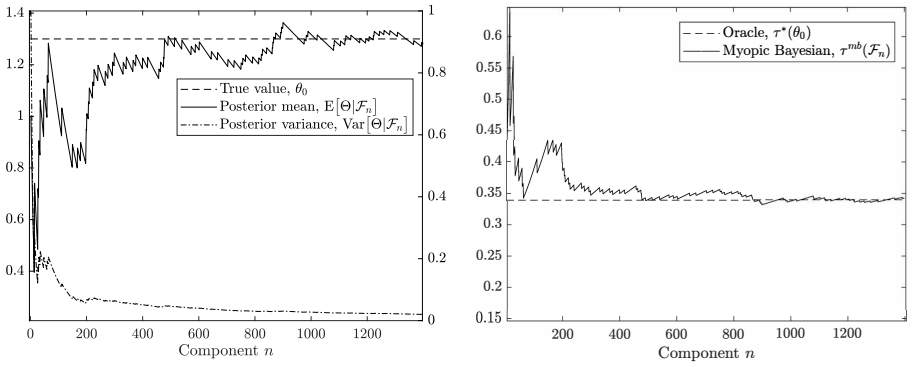
$$\lim_{N\to\infty} \operatorname{argmin}_{\tau\geq 0} C_N(\tau \mid \mathcal{F}_N) = \operatorname{argmin}_{\tau\geq 0} C(\tau \mid \theta_0),$$

which establishes the result. □

As noted before, computing the optimal policy via the proposed DP is analytically intractable and computationally challenging. As such, decision makers can resort to the myopic Bayesian policy. This policy is not only computationally appealing, but also, as is established in Theorem 6.3, asymptotically optimal.

6.5.3 Illustrative example

Figure 6.2 provides an illustrative example of the established asymptotic properties for the case when the lifetimes are Weibull distributed random variables with known shape $\beta = 3$ and unknown scale. Since $\beta > 1$, the lifetime distribution has an increasing hazard rate so that the assumption of a unique and finite $\tau^*(\theta_0)$ is justified. The value of $\tau^*(\theta_0)$ is computed using Equation (6.8) with $\alpha = 0.9$.



(a) Sequence of posterior mean (left scale) versus the true value θ_0, and sequence of posterior variance (right scale).

(b) Sequence of decisions induced by myopic Bayesian policy versus the decision induced by the Oracle $\tau^*(\theta_0)$.

Figure 6.2 Illustration of asymptotic properties for the consistency (a) and the decision making (b) when $c_p = 1$, $c_u = 4$, $\ell(x) = x^3$, $\theta_0 = 1.3$ and $\tau^*(\theta_0) \approx 0.34$. The data points in both sub-figures are obtained from the same sample path.

These two subfigures are generated using the same sample path, so that there is a one-to-one correspondence between the paths displayed in the sub-figures. Subfigure (a) shows that the posterior mean converges to the true value θ_0, while the posterior

variance converges to 0 at the same time. Subfigure (b) shows the corresponding sequence of decisions induced by the myopic Bayesian policy, and its convergence to $\tau^*(\theta_0)$. The posterior mean in Subfigure (a) and the myopic Bayesian policy in Subfigure (b) appear to be closely linked. The coupling between the evolution of the posterior mean and the myopic Bayesian policy can be explained intuitively. For instance, the jump in the myopic Bayesian policy starting at $n \approx 80$ can be explained as follows. If the posterior mean of Θ decreases based on the accumulated information, the decision maker expects that the posterior predictive lifetime is becoming larger in expectation (see Proposition 6.1), hence, a higher age-replacement threshold is imposed. The reverse holds true as well, as is nicely illustrated when the posterior mean starts to increase again after $n \approx 200$.

6.6. Conclusion

In this chapter, we have considered a sequence of ABM problems with a general lifetime distribution parametrized by an a-priori unknown parameter. By adopting a parametric Bayesian framework often used in the inventory research community, we have been able to investigate the exploration-exploitation trade-off that naturally arises when age-based decisions are integrated with learning from both censored and uncensored observations.

A new stochastic order for this parametric Bayesian framework is established that is particularly useful in maintenance problems. We have investigated the optimal policy for a finite sequence of components, and established some structural properties. For the infinite case, we have proposed a computationally appealing myopic policy and established its asymptotic optimality.

Three immediate directions for future research are (i) to investigate the – practically important – rate of convergence of the asymptotic properties, (ii) to analytically prove or disprove our posed conjecture, and (iii) to study how the myopic decision relates to the optimal decision in the finite sequence when the information state is the same. The latter direction is a special case of the third assertion of our posed conjecture as the decision for the age-replacement threshold of the last component in every sequence only exploits the current information in the same way as the proposed myopic Bayesian policy.

This chapter is the last chapter in which we considered a model facing parameter uncertainty. In the next chapter, we shall study a model in which we assume that there is full knowledge regarding the relevant parameters.

6.A. Numerical exploration of optimal age-replacement thresholds

In this section we report on our extensive numerical experiments in which we investigate properties of the optimal age-replacement threshold. Recall that $\tau^n_{m,k}$ denotes the optimal age-replacement threshold for component n when the shape parameter is m and the rate parameter is k. Due to the computational difficulty of solving Equation (4) for large instances, we consider only a small instance, that is, we consider a sequence of $N = 20$ components and we set $\alpha = 0.9$. We consider three cost ratios, namely $\frac{c_u}{c_p} \in \{2, 4, 6\}$. Due to its application in maintenance, we consider Weibull distributed lifetimes with known shape β and unknown scale, i.e. $\ell(x) = x^\beta$, for two values of $\beta \in \{3, 5\}$.

In the remainder of this section, we first describe the observed behavior of $\tau^n_{m,k}$ in n for (m, k) fixed. Here we can observe how the optimal age-replacement threshold behaves for the same information state, but for different positions in the sequence. We then describe the observed behavior of $\tau^n_{m,k}$ in the hyperparameters m and k respectively, while keeping the other two parameters fixed. Here we can observe how $\tau^n_{m,k}$ changes when part of the information state changes. It is noteworthy to mention that the results displayed in this document are by no means a complete enumeration of all the results, but since we consistently observed the described observations throughout our numerical experiments (i.e. for all parameter settings), we only discuss parts of these results in this document to convey the message.

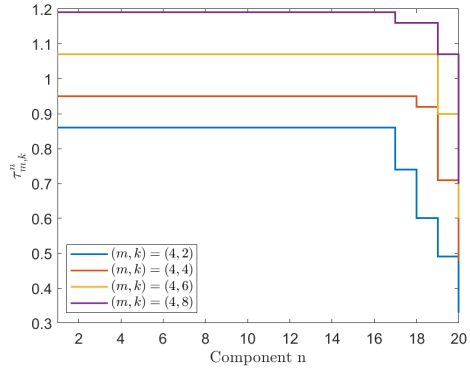
Note that in some figures, plots overlap where optimal policies coincide. Here, to prevent any confusion, we only show one colored line in the figure and one can then deduce that the colored line that is not plotted has the same value.

6.A.1 Monotonic behavior

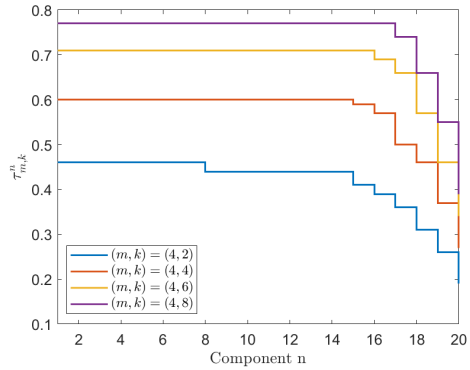
The subfigures in Figures 6.3 and 6.4 clearly suggest that $\tau^n_{m,k}$ is non-increasing in n, if we keep m and k fixed.

The subfigures in Figures 6.5 and 6.6 clearly suggest that $\tau^n_{m,k}$ is non-increasing in m, if we keep n and k fixed.

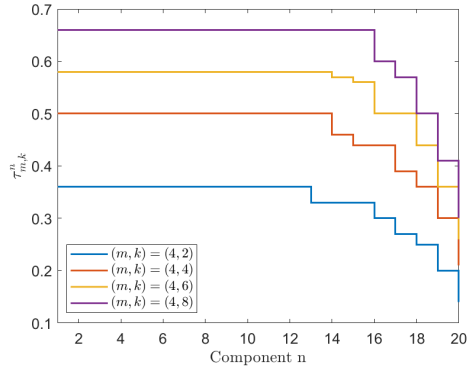
The subfigures in Figures 6.7 and 6.8 clearly suggest that $\tau^n_{m,k}$ is non-decreasing in k, if we keep n and m fixed.



(a) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 2$, and $\beta = 3$.

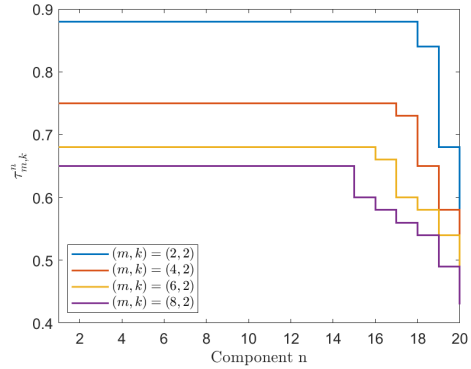


(b) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 4$, and $\beta = 3$.

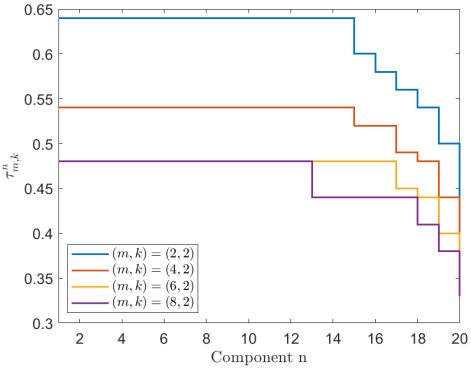


(c) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 6$, and $\beta = 3$.

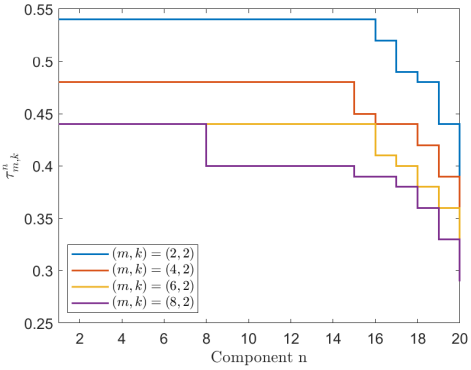
Figure 6.3 Illustration of behavior of $\tau^n_{m,k}$ in n for $\beta = 3$.



(a) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 2$, and $\beta = 5$.

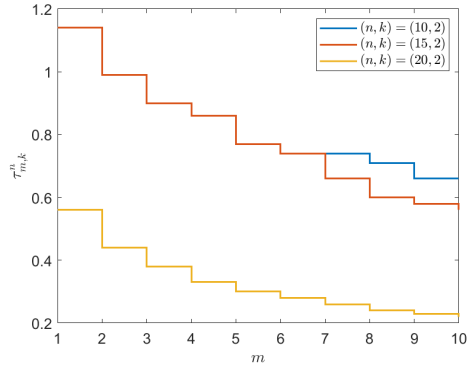


(b) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 4$, and $\beta = 5$.

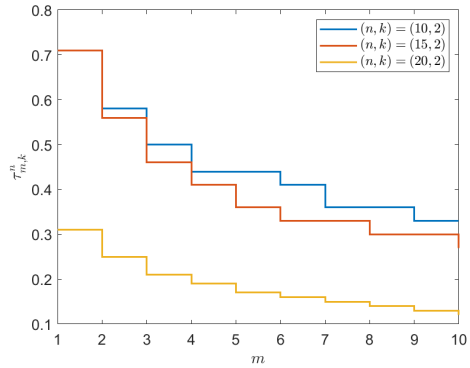


(c) $\tau^n_{m,k}$ as function of n for various values of (m,k) for $\frac{c_u}{c_p} = 6$, and $\beta = 5$.

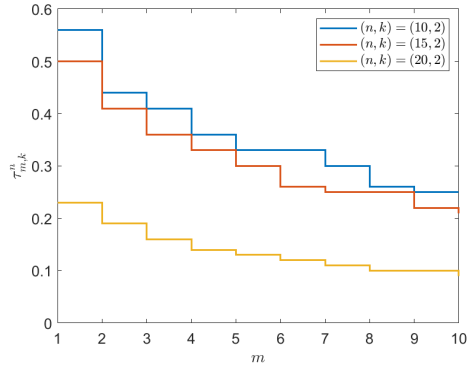
Figure 6.4 Illustration of behavior of $\tau^n_{m,k}$ in n for $\beta = 5$.



(a) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 2$, and $\beta = 3$.

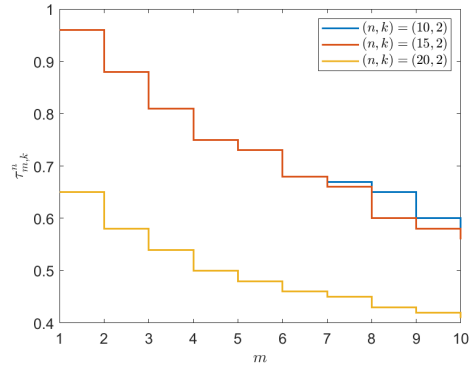


(b) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 4$, and $\beta = 3$.

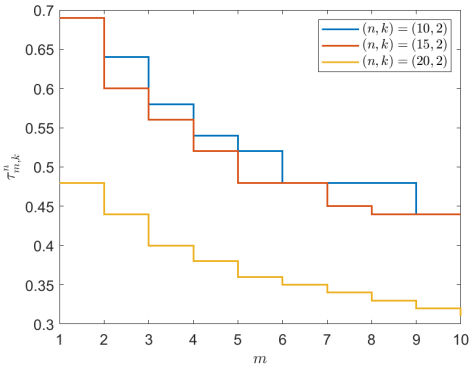


(c) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 6$, and $\beta = 3$.

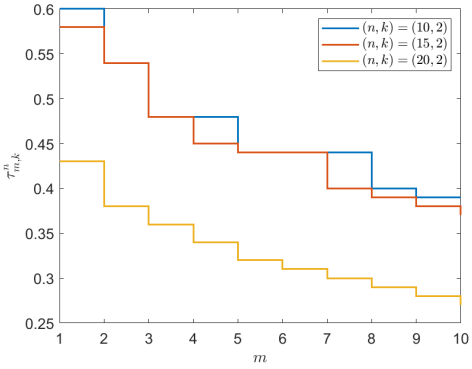
Figure 6.5 Illustration of behavior of $\tau^n_{m,k}$ in m for $\beta = 3$.



(a) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 2$, and $\beta = 5$.

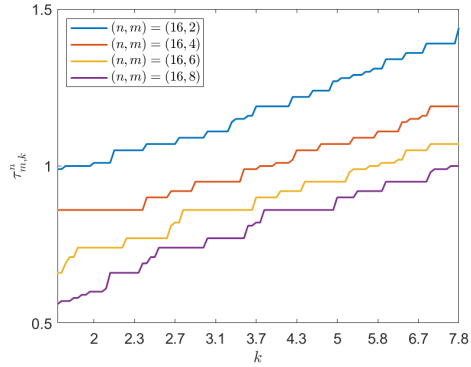


(b) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 4$, and $\beta = 5$.

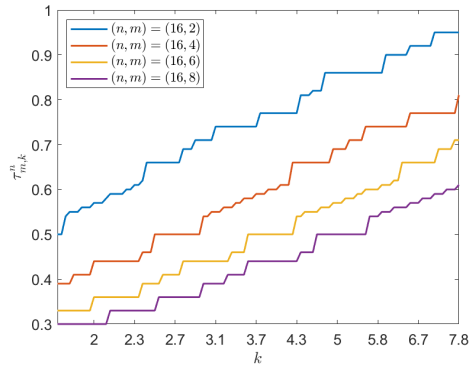


(c) $\tau^n_{m,k}$ as function of m for various values of n for $k = 2$, $\frac{c_u}{c_p} = 6$, and $\beta = 5$.

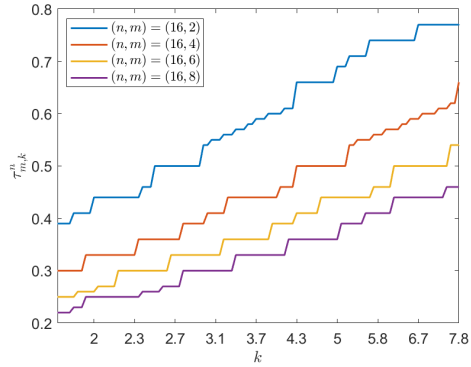
Figure 6.6 Illustration of behavior of $\tau^n_{m,k}$ in m for $\beta = 5$.



(a) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 2$, and $\beta = 3$.

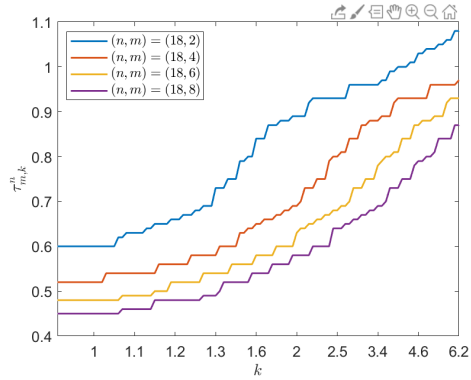


(b) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 4$, and $\beta = 3$.

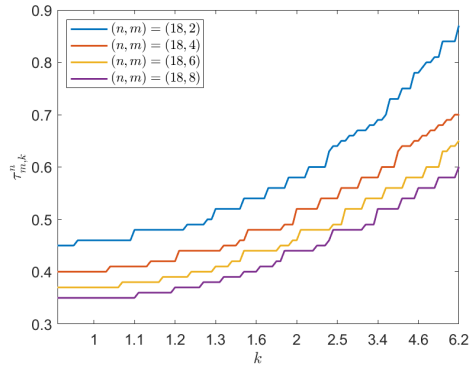


(c) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 6$, and $\beta = 3$.

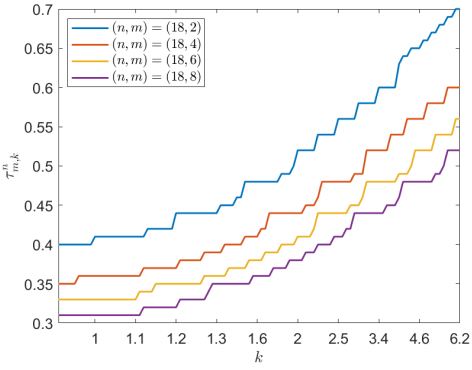
Figure 6.7 Illustration of behavior of $\tau^n_{m,k}$ in k for $\beta = 3$.



(a) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 2$, and $\beta = 5$.



(b) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 4$, and $\beta = 5$.



(c) $\tau^n_{m,k}$ as function of k for various values of m for $n = 16$, $\frac{c_u}{c_p} = 6$, and $\beta = 5$.

Figure 6.8 Illustration of behavior of $\tau^n_{m,k}$ in k for $\beta = 5$.

Chapter 7

Optimal opportunistic condition-based maintenance policies with parameter certainty and imperfect repairs

7.1. Introduction

Up to this point, we have studied mathematical models for condition-based maintenance (CBM) and age-based maintenance (ABM) characterized by parameter uncertainty, either in the deterioration process or in the lifetime distribution. In this last chapter, we depart from this modeling choice and focus on a setting in which we have full knowledge of all relevant parameters; that is, there is parameter certainty.

In the previous chapters, we generally assumed that maintenance is so-called *planned maintenance*. That is, in most models that we studied, we assumed that a decision maker can only replace a component at discrete epochs, i.e. at *scheduled opportunities* that are planned at equidistant moments in time, say at instances $\tau, 2\tau, 3\tau, \ldots$, (e.g., τ

This chapter is based on Drent et al. (2019).

= 6 months). In the context of a network of assets, such as a wind park or a network of hospitals in close geographic proximity (from the viewpoint of the service provider), there is a second type (in addition to the above scheduled instances) of opportunities to perform preventive maintenance. In the event that a failure occurs, its *corrective maintenance* instance can be viewed as an unscheduled opportunity for preventive maintenance for the other assets in the network. In these instances, *opportunistic maintenance* can take place, with the respective instances constituting the *unscheduled opportunities* of preventive maintenance. This form of network dependency can be viewed on two levels: (i) the economic dependency between the various systems of a network, and (ii) the structural degradation and failure dependencies.

Incorporating opportunistic maintenance may also affect the scheduling of planned maintenance, as it might be beneficial to defer the planned maintenance opportunity to take place after a period of length τ after the occurrence of an opportunistic maintenance. This decision of *deferring or not the scheduling of planned maintenance* after the occurrence of opportunistic maintenance may have a positive or negative effect on the total costs.

Up to this point, we have consistently assumed that a maintenance activity is perfect, i.e. it restores the system to a state of 'as good as new'. However, this assumption may not be true in practice. For instance, a misidentification of the root cause of the (imminent) failure can lead to an erroneous repair not resolving the actual issue, or some minor repair activity (such as exchange of parts, changes or adjustment of the settings, software update, lubrication or cleaning, see e.g., Spinato et al. (2009)) may not restore the system to a state of 'as good as new'. In that case, it is more reasonable to assume that the system is restored to a state in between 'as bad as old' and 'as good as new'. This concept will be referred to as *imperfect maintenance*. Evidently, this assumption impacts the resulting cost. Hence, knowledge regarding the degree of how successful a maintenance activity is should not be ignored in the maintenance planning.

In conclusion, asset owners are oftentimes faced with the following questions:

(i) What is the advantage of incorporating planned maintenance in comparison to exercising only corrective maintenance?

(ii) What is the benefit of sharing resources in the network (in the form of incorporating opportunistic maintenance in addition to the planned maintenance)?

(iii) What is the influence of deferring the planned maintenance after the occurrence of opportunistic maintenance?

(iv) What is the influence of imperfect maintenance on the maintenance planning and on the costs (long-run rate of cost)?

(v) When should preventive maintenance be performed (so as to minimize the long-run rate of cost)?

7.1.1 Contributions

The contributions of this chapter are threefold:

1. We consider a stylized, yet representative model that incorporates planned and opportunistic maintenance, as well as imperfect maintenance. We build a semi-Markov decision process (semi-MDP), which we use to (i) prove the existence of an optimal maintenance policy, and (ii) establish its structural properties. The optimal policy has an intuitive control limit structure where the control limit depends on the time until the next planned maintenance opportunity. Moreover, using this approach, we are able to derive a closed-form expression for this control limit.

2. Considering the class of control limit policies (depending on the remaining time until the next planned maintenance), we derive, using the theory of regenerative processes, an explicit expression for the long-run rate of cost.

3. We consider data from the wind energy industry and provide, based on these values, concrete answers to questions (i)–(v) mentioned above. More specifically, we analyze the benefit of using planned and opportunistic maintenance compared to only corrective maintenance. We also analyze the influence of deferring planned maintenance after the occurrence of opportunistic maintenance. Finally, we also highlight the cost savings that can be attained by reducing the probability of an imperfect maintenance.

7.1.2 Organization

The remainder of this chapter is structured as follows. In Section 7.2, we review the related literature. In Section 7.3, we describe in detail the model at hand, which captures the condition of the asset and which incorporates imperfect maintenance at scheduled and unscheduled maintenance opportunities. Subsequently, in Section 7.4, we characterize the structure of the optimal policy for CBM using the average cost criterion, see Section 7.4.1, and we compute the long-run rate of cost for any

policy with the same structure as the optimal policy (i.e. the class of control limit policies depending on the remaining time until the next planned maintenance), see Section 7.4.2. In Section 7.5, we permit the deferral of planned maintenance after the occurrence of opportunistic maintenance, and we compute the long-run rate of cost. A numerical illustration is provided in Section 7.6, where, based on data from the wind energy industry, we compare the long-run rate of cost for various policies, we show the effect of imperfect maintenance, and the effect of deferring planned maintenance. Finally, Section 7.7 contains concluding remarks.

7.2. Literature review

Maintenance optimization models have been extensively studied in the literature. Optimal maintenance policies aim to provide optimal system reliability/availability and safety performance at lowest possible maintenance costs (Pham and Wang, 1996). Due to the fast development of sensing techniques in recent years, the state of a capital asset can be monitored or inspected at a much lower cost and in a continuous fashion, which facilitates CBM. CBM recommends maintenance actions based on information collected through online monitoring of the capital asset and it can significantly reduce maintenance costs by decreasing the number of unnecessary maintenance operations (see e.g., Jardine et al., 2006; Peng et al., 2010; Lam and Banjevic, 2015). The CBM model that we propose builds on the delay time model proposed by Christer (1982) and Christer and Waller (1984). We refer the reader to Baker and Christer (1994), Christer (1999), Wang (2008), and Wang (2012) for excellent overviews of delay time models, and their use in maintenance modeling and optimization.

Not only are delay time models well-known in literature, but they are also very frequently appearing in practice. Practice-based research with real diagnostic data, such as data related to the spectrometry of oil (see, e.g., Makis et al., 2006; Kim et al., 2011) and data related to vibrations (see, e.g., Yang and Makis, 2010), showed that it is usually sufficient, and even preferable from a modeling and decision making perspective, to consider only two operational states. The first state is the perfect state, in which the system lasts from newly installed to the point that a hidden defect has been identified. After the occurrence of a hidden defect in the system until the occurrence of a failure (which is typically referred to as the delay time), the system resides in the second state also referred to as the satisfactory state. Such a classification of the operational states has the property that maintenance actions are initiated only when the system is degraded to the state that can actually lead to a direct failure, i.e. the satisfactory state, but not when the system is functioning

perfectly, i.e. the perfect state. The vast majority of the literature on delay time models is restricted to numerical methods or approximations to solve the models at hand, due to their underlying complexity. Few recent exceptions are Maillart and Pollock (2002), Kim and Makis (2013) and Van Oosterom et al. (2014), who study two-state systems under periodic inspection, partial observability, and postponed replacement, respectively, and provide analytical results regarding the structure of the optimal policy. However, all of them do not consider the option of resource sharing in the network (in the form of opportunistic maintenance), nor do they incorporate the notion of imperfect repair.

Most delay time model analyses assume that the system after a maintenance action is restored to a state of 'as good as new'. Contrary to this assumption, in imperfect maintenance it is assumed that upon preventive maintenance, the system lies in a state somewhere between 'as good as new' and 'as bad as old'. This is first introduced by Nakagawa (1979a) and Nakagawa (1979b) and is called the (p, q)-rule. Under the (p, q)-rule, the system is returned to an 'as good as new' state (perfect preventive maintenance) with probability p and it is returned to the 'as bad as old' state (minimal preventive maintenance) with probability $q = 1 - p$ after preventive maintenance. Clearly, the case $p = 0$ corresponds to having no preventive maintenance. Also, from a practical point of view, imperfect maintenance can describe a large set of realistic maintenance actions (Pham and Wang, 1996).

When planning CBM strategies, see, e.g., Jardine et al. (2006), Jardine and Tsang (2005) and Prajapati et al. (2012), a typical assumption in the literature is that the system at hand is monitored continuously and one can intervene and maintain the system at any given moment. However, due to accessibility reasons (e.g., in the case of off-shore wind parks) or for cost reduction purposes, it is cost optimal and more practical to allow only for discrete time opportunities. The simplest amongst the discrete time opportunities are the periodic planned maintenance instances (also referred to as scheduled downs), with period say τ, that serve as a scheduled opportunity to do maintenance for a network of systems. Furthermore, unplanned maintenance instances (due to opportunistic maintenance) can be modeled as discrete instances occurring according to a multi-dimensional counting process.

For recent works related to opportunistic maintenance, the interested reader is referred to Zhu et al. (2016, 2017), Kalosi et al. (2016), and Arts and Basten (2018). In Zhu et al. (2016, 2017), the authors consider a single-unit system and account for both scheduled and unscheduled opportunities. In these analyzes, the authors model the age and the condition, respectively, of the system and derive, based on approximations, the long-run rate of cost under a given policy. In both papers,

the arrivals of unscheduled opportunities are modeled according to a homogeneous Poisson process. This approximation is justified by the Palm-Khintchine theorem (Khinchin, 1956), which states that even if the failure times of some systems do not follow Exponential distributions, the superposition of a sufficiently large number of independent renewal processes behaves asymptotically like a Poisson process. Arts and Basten (2018) build further on Zhu et al. (2016, 2017), but they only consider scheduled maintenance opportunities (excluding unscheduled opportunities). Furthermore, Arts and Basten (2018) assume that at a scheduled opportunity, the system is restored to a perfect condition (i.e. $p = 1$), while at a failure they assume that the system is restored to a state which is stochastically identical to the state just prior to the system's failure. In a recent conference paper, Kalosi et al. (2016) looked at a model with both planned and unplanned maintenance opportunities, at which the system is restored to a perfect condition, showing some preliminary results that a control limit policy (depending on the remaining time until the next planned maintenance) is optimal.

In contrast to Arts and Basten (2018) and to Zhu et al. (2016, 2017), in which the long-run rate of cost is computed for a given policy, we first characterize the structure of the optimal policy explicitly and thereafter, for the optimal policy class, we compute the long-run rate of cost. Furthermore, we include both scheduled and unscheduled maintenance opportunities. In contrast to Kalosi et al. (2016), we extend the model by incorporating the (p, q)-rule, making it more generic and realistic. Moreover, we are the first to analyze the influence of deferring planned maintenance and we illustrate the financial effects of the maintenance policy in a realistic context using data stemming from the wind industry.

7.3. Model description

We consider a single unit system (equivalently, a component or asset) that is monitored continuously and whose condition is fully observable. We assume that the condition of the system degrades over time and that it can be modeled according to a delay time model. That is, the states are classified as *perfect*, *satisfactory* and *failed*. We shall refer to the state of perfect condition as state 2, the state of satisfactory condition as state 1 and the failure state as state 0. Furthermore, we assume that as soon as a system failure occurs, the system is instantaneously replaced by an 'as good as new' system. So, in the mathematical formulation of the model, we may assume, due to the instantaneous replacement at failure, that the model evolves between only states 1 and 2. The system spends an exponential amount of time with rate μ_i in state

i, $i \in \{1, 2\}$. The above model formulation implies that initially the system starts in state 2 (perfect state), then after an exponential amount of time with rate μ_2, the system deteriorates and the condition of the system goes to state 1 (satisfactory state). The system spends an exponential amount of time with rate μ_1 in state 1, after which a failure occurs. At a failure the system is instantaneously replaced by an 'as good as new' system and the condition is restored to 2 (perfect state). A schematic evolution of the condition of the component and the corresponding times of transitions are depicted in Figure 7.1.

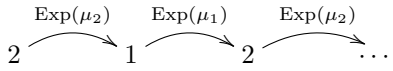


Figure 7.1 Schematic evolution of the condition of the component and the corresponding times of transitions.

We assume that we have two types of opportunities in which we can perform preventive maintenance (PM) before failure: the scheduled and the unscheduled opportunities. The scheduled opportunities correspond to pre-arranged opportunities occurring according to a fixed schedule. These opportunities can be attributed to either service/maintenance agreements or to regulation imposition checks. We assume that the scheduled opportunities occur at epochs $\tau, 2\tau, 3\tau, \ldots$, with $\tau > 0$. This is also in accordance with what happens in practice as maintenance actions once planned are typically not rescheduled. The unscheduled opportunities correspond to random opportunities triggered by failures of other systems in close proximity. We assume that these unscheduled opportunities occur according to a Poisson process at rate λ.

The unscheduled and scheduled opportunities, abbreviated by USO and SO, respectively, serve as opportunities to perform preventive maintenance. Such a preventive maintenance is assumed to cost less than a corrective maintenance (CM) upon failure, which costs c_{cm}. Moreover, incorporating a planning perspective, we may assume that the preventive maintenance cost at an SO, $c_{\text{pm}}^{\text{so}}$, is less than or equal to the corresponding cost at a USO, say $c_{\text{pm}}^{\text{uso}}$, that is $0 < c_{\text{pm}}^{\text{so}} \leq c_{\text{pm}}^{\text{uso}} < c_{\text{cm}}$ (however, we also extend our analysis to the case $c_{\text{pm}}^{\text{so}} > c_{\text{pm}}^{\text{uso}}$). Following the (p, q)-rule of Nakagawa (1979a) and Nakagawa (1979b), we assume that after preventive maintenance a system is returned to the 'as good as new' state with probability $p \in (0, 1]$ and returned to the 'as bad as old' state (i.e. the amount of time left until the failure has not altered) with probability $q = 1 - p$.

Our aim is to determine a policy when to perform preventive maintenance on the system based on its condition and the opportunity type, i.e. scheduled or unscheduled.

More explicitly, we will need to formally define the state space, which refers to the condition of the system, the action space and the decision epochs. The state space is governed by the process depicting the condition of the system, i.e. the Markov chain evolving between the states $\{1, 2\}$. The action space consists of only two actions: perform preventive maintenance or do nothing. Lastly, the decision epochs are the SO and USO epochs. In Figure 7.2, we depict the SO epochs by $(*)$ and the USO epochs by (o).

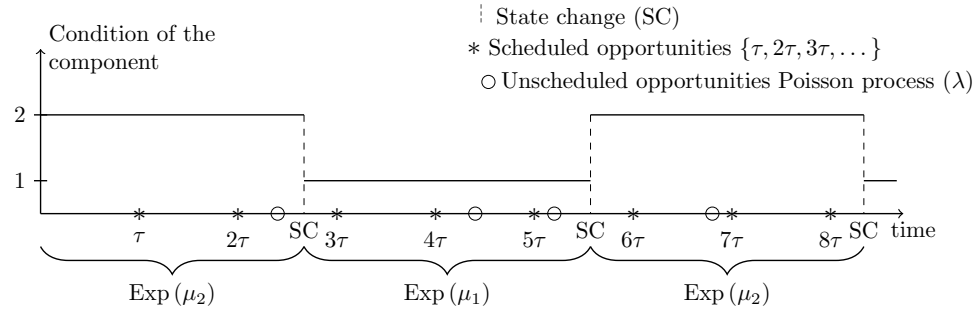


Figure 7.2 A sample path of the model.

7.4. Optimal policy

The goal of this section is twofold: We first characterize the structure of the optimal average cost CBM policy. We then derive an explicit form for the long-run rate of cost per time unit for any given policy that has the same structure as the optimal policy.

7.4.1 Average cost criterion

This section is devoted to the derivation of the optimal policy on when to perform preventive maintenance for the system at hand using the average cost criterion. To this purpose, we set up our problem as a (controlled) semi-MDP. Due to the stochastic nature of the problem, it does not suffice to know the type of the decision epoch (SO or USO), but it is also required to keep track of the remaining time till the next SO. That time may impact our decision, i.e. the optimal policy may depend on the residual time till the next SO. Thus, for the full description of the condition (state)

of the system, we use a triplet descriptor

$$\mathcal{S} = \big\{(i,j,t):\ i \in \{1,2\},\ j \in \{\text{SC},\text{USO}\},\ t \in (0,\tau)\big\} \cup \big\{(i,\text{SO},0):\ i \in \{1,2\}\big\},$$

where i indicates the condition of the system. If $j = \text{SC}$, then this means that the condition of the system is about to change and there is no decision associated with this epoch, while if $j = \text{SO}$ or $j = \text{USO}$, this means that this is a decision moment at either a scheduled (SO) or unscheduled opportunity (USO), respectively. Finally, the third element indicates the remaining time until the SO. Note that if $j = \text{SO}$ then $t = 0$. The introduction of the remaining time until the upcoming SO in the full description of the condition of the system renders the model inhomogeneous, and for this reason we use techniques that stem from semi-MDPs. Note here that the inclusion of the remaining time until the upcoming SO in the state, although it complicates the analysis, permits us to prove that there is an optimal policy in the class of deterministic stationary policies, cf. Propositions 7.1 and 7.3. At each decision epoch (depending on the values of $(i,j,t) \in \mathcal{S}$), we can choose to perform preventive maintenance or do nothing or in case of a failure to do corrective maintenance (CM), that is $\mathcal{A} = \{\text{perform PM, do nothing, perform CM}\}$, where $\mathcal{A}$ represents the overall action space.

Proposition 7.1. *For the model at hand, the deterministic stationary policy is optimal for the average cost criterion.*

A formal version of the above proposition, cf. Proposition 7.3, and its proof can be found in Appendix 7.A, together with a full formal definition of the model in the context of semi-MDPs. In addition to the theoretical validation that the above proposition offers on the existence and nature of the optimal maintenance policy, in the following theorem we compute the optimal policy.

Theorem 7.1. *Under the assumption that $c_{pm}^{so} < c_{pm}^{uso}$ and given the imperfect preventive maintenance probability $1-p \in (0,1]$, the optimal policy under the average cost criterion is: For state 2 to do nothing. For state 1 to perform preventive maintenance at scheduled opportunities, if $\mu_1 c_{cm} > (\mu_1+\mu_2)\frac{c_{pm}^{so}}{p}$, and to do nothing otherwise, and to perform preventive maintenance at unscheduled opportunities for which the residual time until the next scheduled opportunity is in $[\hat{t},\tau)$, if $\mu_1 c_{cm} > \left(\frac{c_{pm}^{uso}}{p} - \frac{c_{pm}^{uso}-c_{pm}^{so}}{e^{(\mu_1+\mu_2)\tau}-1}\right)(\mu_1+\mu_2)$, and to do nothing otherwise. Where,*

$\hat{t} = \min\{\tau, \max\{0, t^*\}\}$, *with* t^* *satisfying*

$$\frac{c_{pm}^{uso}}{p} = \frac{\mu_1 c_{cm} + \lambda c_{pm}^{uso}}{\mu_1 + \mu_2 + \lambda p} + \left(\frac{-c_{pm}^{so} + \frac{\mu_1 c_{cm}}{\mu_1+\mu_2} + \left(\frac{c_{pm}^{uso}}{p} - \frac{\mu_1 c_{cm}}{\mu_1+\mu_2} \right) e^{(\mu_1+\mu_2)t^*}}{1-p} - \frac{\mu_1 c_{cm} + \lambda c_{pm}^{uso}}{\mu_1 + \mu_2 + \lambda p} \right) e^{(\mu_1+\mu_2+\lambda p)(\tau - t^*)}. \tag{7.1}$$

PROOF: See Appendices 7.B and 7.C. □

For USOs, Theorem 7.1 establishes a control limit policy depending on the remaining time until the next SO: if the residual time until the next SO is smaller than $\hat{t}$, then it is optimal to not take the opportunity to perform preventive maintenance in state 1. This is intuitive in the sense that the urgency for preventive maintenance in state 1 at a USO should decrease as the cheaper opportunity at an SO is approaching.

Note that in the special case when preventive maintenance costs at SOs and USOs are equal, the optimal policy reduces to a stationary control limit policy, which is shown in Proposition 7.2.

Proposition 7.2. *Under the assumption that* $c_{pm}^{so} = c_{pm}^{uso} = c_{pm} > 0$ *and given the imperfect preventive maintenance probability* $1-p \in (0, 1]$*, the optimal policy under the average cost criterion is: For state* 2 *to do nothing. For state* 1 *to perform preventive maintenance at both SOs and USOs, if* $\mu_1 c_{cm} > (\mu_1 + \mu_2)\frac{c_{pm}}{p}$*, and to do nothing otherwise.*

PROOF: The proof of this proposition is identical in structure to the proof of Case (i) in Theorem 7.1 and for this reason it is omitted. □

One could also argue that the cost for preventive maintenance at a USO is actually less than the cost at an SO since there is already a cost attached to the opportunity at hand (e.g., service engineers are already at a wind park and they can at a small extra cost repair other systems in close proximity as well). In this case, the optimal control policy also reduces to a stationary control limit policy, which is described in Theorem 7.2.

Theorem 7.2. *Under the assumption that* $c_{pm}^{so} > c_{pm}^{uso}$ *and given the imperfect preventive maintenance probability* $1-p \in (0, 1]$*, the optimal policy under the average*

cost criterion is: For state 2 to do nothing. For state 1 to perform preventive maintenance at an unscheduled opportunity if $\mu_1 c_{cm} > (\mu_1 + \mu_2)\frac{c_{pm}^{uso}}{p}$, *and to do nothing otherwise, and to perform preventive maintenance at an SO if* $\mu_1 c_{cm} > (\mu_1 + \mu_2)\frac{c_{pm}^{so}}{p} + \lambda(c_{pm}^{so} - c_{pm}^{uso})$, *and to do nothing otherwise.*

PROOF: See Appendix 7.D. □

7.4.2 Long-run rate of cost per time unit

In the previous section, we characterized the structure of the optimal policy using the average cost criterion. This policy can be viewed as a control limit policy, with the control limit depending on the time until the next SO. In this section, we consider such a policy and we compute the long-run rate of cost per time unit. More concretely, we consider a policy under which in state 2 we do not perform preventive maintenance (i.e. we do nothing), and in state 1 we always perform preventive maintenance at SOs and we perform preventive maintenance at USOs if the remaining time till the next SO is greater than $\tilde{t}$, for some given value $\tilde{t} \in (0, \tau)$. The results obtained in this section are directly applicable to the results of Section 7.4.1, by setting $\tilde{t} = t^*$, cf. Theorem 7.1.

For the computation of the long-run rate of cost per time unit, we employ the theory of regenerative-like processes, also called stationary-cycle processes, described in Section 2.19 of Serfozo (2009). To this purpose, we consider the inter-regeneration times created by the SOs $\{\tau, 2\tau, 3\tau, \ldots\}$. For the cost computation, we assume that, at the SOs, the system is in state 1 or 2 according to a stationary probability $p_1(0)$ and $p_2(0)$, respectively. The long-run rate of cost per time unit is calculated as the expected total cost incurred between consecutive SOs divided by τ.

Let $p_i(t)$ be the probability that the system is in state $i \in \{1, 2\}$ given that the time until the next SO is $t \in [0, \tau)$, then the long-run rate of cost per time unit for this control limit policy (depending on the remaining time until the next planned maintenance) for any given time threshold is given in the next theorem.

Theorem 7.3. *Consider a given policy under which in state* 2 *we opt for the action do nothing, and in state* 1 *we repair at scheduled opportunities and at unscheduled opportunities for which the remaining time until the next scheduled opportunity is greater than* $\tilde{t} \in (0, \tau)$, *and we do nothing otherwise. Under this policy, the long-run*

rate of cost per time unit equals

$$\frac{c_{pm}^{so} p_1(0) + c_{pm}^{uso} \lambda \int_{\tilde{t}}^{\tau} p_1(t)\,\mathrm{d}t + c_{cm}\mu_1 \int_0^{\tau} p_1(t)\,\mathrm{d}t}{\tau}, \tag{7.2}$$

with

$$p_1(t) = \begin{cases} \dfrac{\mu_2}{\mu_1+\mu_2} + C_1\, e^{(\mu_1+\mu_2)t}, \ t \in [0,\tilde{t}), & (7.3) \\ \dfrac{\mu_2}{\mu_1+\mu_2+\lambda p} + C_2\, e^{(\mu_1+\mu_2+\lambda p)t}, \ t \in [\tilde{t},\tau), & (7.4) \end{cases}$$

where the constants C_1 and C_2 are obtained as follows

$$C_1 = C_2\, e^{\lambda p \tilde{t}} - \frac{\mu_2}{\mu_1+\mu_2}\frac{\lambda p}{\mu_1+\mu_2+\lambda p} e^{-(\mu_1+\mu_2)\tilde{t}},$$

$$C_2 = \frac{\frac{\mu_2}{\mu_1+\mu_2}\left(1 - e^{-(\mu_1+\mu_2)\tilde{t}}\right) + \frac{\mu_2}{\mu_1+\mu_2+\lambda p}\left(\frac{1}{1-p} - e^{-(\mu_1+\mu_2)\tilde{t}}\right)}{\frac{1}{1-p} e^{(\mu_1+\mu_2+\lambda p)\tau} - e^{\lambda p \tilde{t}}}.$$

Proof: The expected total cost incurred in one cycle consists of three parts (cf. Equation (7.2)), which are related to the expected cost associated with preventive maintenance at SOs, with preventive maintenance at USOs and with corrective maintenance, respectively. It is now sufficient to derive $p_i(t)$ for $t \in [0,\tau)$, $i \in \{1,2\}$.

For $t \in [\tilde{t},\tau)$, the time-dependent behavior of $p_1(t)$ is governed by

$$p_1(t) = p_1(t+\mathrm{d}t)(1 - (\mu_1+\lambda p)\,\mathrm{d}t) + p_2(t+\mathrm{d}t)\mu_2\,\mathrm{d}t. \tag{7.5}$$

Equation (7.5) is easily obtained by considering a small time interval of length $\mathrm{d}t$, and noticing that at time t we are in state 1 either due to a transition from state 2 with infinitesimal probability $\mu_2\,\mathrm{d}t$ or we have remained in state 1 with infinitesimal probability $1-(\mu_1+\lambda p)\,\mathrm{d}t$. Subtracting $p_1(t+\mathrm{d}t)$ from both sides of Equation (7.5), after some straightforward computations, yields

$$p_1(t+\mathrm{d}t) - p_1(t) = p_1(t+\mathrm{d}t)(\mu_1+\lambda p)\,\mathrm{d}t - p_2(t+\mathrm{d}t)\mu_2\,\mathrm{d}t.$$

Dividing this expression by $\mathrm{d}t$ and letting $\mathrm{d}t \to 0$ results in

$$p_1'(t) = p_1(t)(\mu_1+\lambda p) - p_2(t)\mu_2.$$

Following a similar analysis for $p_2(t)$, yields the following system of differential

equations, for $t \in [\tilde{t}, \tau)$,

$$\begin{bmatrix} p_1'(t) \\ p_2'(t) \end{bmatrix} = \begin{bmatrix} \mu_1 + \lambda p & -\mu_2 \\ -(\mu_1 + \lambda p) & \mu_2 \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix}, \qquad t \in [\tilde{t}, \tau). \tag{7.6}$$

Similarly, for $t \in [0, \tilde{t})$, we have

$$\begin{bmatrix} p_1'(t) \\ p_2'(t) \end{bmatrix} = \begin{bmatrix} \mu_1 & -\mu_2 \\ -\mu_1 & \mu_2 \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix}, \qquad t \in [0, \tilde{t}). \tag{7.7}$$

Solving the system of differential Equations (7.6) and (7.7) leads to the desired solutions (7.3) and (7.4), respectively. In this process, we would need to compute four unknown constants. This is achieved by using: (i) the normalizing condition, i.e. $p_1(t) + p_2(t) = 1$ for all $t \in [0, \tau)$, (ii) the continuity condition at $\tilde{t}$, i.e. $\lim_{t \to \tilde{t}^-} p_i(t) = p_i(\tilde{t})$ for $i \in \{1, 2\}$, and (iii) the boundary condition at the SOs imposed by the policy and the imperfect maintenance probability, i.e. $(1 - p)p_1(0) = \lim_{t \to \tau^-} p_1(t)$. □

7.4.2.1 Special cases

In case of only *scheduled opportunities*, which corresponds to the case $\tilde{t} \to \tau$ or, equivalently, to the case $\lambda \to 0$, the probabilities $p_i(t)$ for $i \in \{1, 2\}$ are derived from the system of linear equations in (7.7) plus the normalizing condition, i.e. $p_1(t) + p_2(t) = 1$ for all $t \in [0, \tau)$. This yields

$$p_1(t) = \frac{\mu_2}{\mu_1 + \mu_2} \left(1 - \frac{p e^{(\mu_1 + \mu_2)t}}{e^{(\mu_1 + \mu_2)\tau} - 1 + p} \right), \ t \in [0, \tau).$$

Plugging the above result into Equation (7.2), after appropriately considering in Equation (7.2) only the costs related to preventive maintenance at SOs and corrective maintenance

$$\frac{c_{\mathrm{pm}}^{\mathrm{so}} p_1(0) + c_{\mathrm{cm}} \mu_1 \int_0^\tau p_1(t)\, \mathrm{d}t}{\tau},$$

leads to the long-run rate of cost per time unit in the case of only SOs.

In case of *perfect maintenance*, i.e. in case $p = 1$, the boundary condition at the SOs imposed by the policy and the imperfect maintenance in the proof of Theorem 7.3 reduces to $\lim_{t \to \tau^-} p_1(t) = 0$, as immediately after an SO, the system is restored to state 2 with probability 1. This enables us to explicitly solve the system of linear Equations

(7.6) and (7.7), yielding

$$p_1(t) = \frac{\mu_2}{\mu_1+\mu_2} + e^{\frac{\lambda-\Lambda(t)}{\lambda}(\mu_1+\mu_2)(t-\tilde{t})} \cdot \left(\frac{\mu_2}{\lambda+\mu_1+\mu_2} - \frac{\mu_2}{\mu_1+\mu_2} - \frac{\mu_2}{\lambda+\mu_1+\mu_2} e^{(\lambda+\mu_1+\mu_2)(\tilde{t}-\tau+\frac{\Lambda(t)}{\lambda}(t-\tilde{t}))} \right),$$

where

$$\Lambda(t) = \begin{cases} 0, & \text{if } 0 \le t < \tilde{t}, \\ \lambda, & \text{if } \tilde{t} \le t < \tau. \end{cases}$$

Combining this expression with Equation (7.2), results in the long-run rate of cost per time unit in the case of perfect maintenance.

In case of only *unscheduled opportunities*, which is equivalent to considering $\tau \to \infty$, the condition of the system can be fully described using a double descriptor $\mathcal{S} = \big\{(i,j) : \ i \in \{1,2\}, \ j \in \{\text{SC}, \text{USO}\}\big\}$ which is independent of time, and thus the new model formulation falls into the framework of regular MDPs. It can be easily shown that: For state 2, the optimal policy is to do nothing, and, for state 1, the optimal policy is to repair if $\frac{(\mu_1+\mu_2)c_{\text{pm}}^{\text{uso}}}{p} < \mu_1 c_{\text{cm}}$ and to do nothing otherwise. Furthermore, under the optimal policy the average long-run rate of cost is equal to

$$\frac{c_{\text{pm}}^{\text{uso}}\lambda\mu_2 + c_{\text{cm}}\mu_1\mu_2}{\lambda p + \mu_1 + \mu_2}.$$

In case of only *corrective replacements*, the long-run rate of cost is equal to

$$c_{\text{cm}}\frac{\mu_1\mu_2}{\mu_2+\mu_1}.$$

7.5. Deferring planned maintenance

In this section, we consider that upon a successful maintenance activity (preventively, at an SO or at a USO, or correctively), the upcoming planned maintenance is deferred for a period of length τ, i.e. at the instances of successful maintenance the remaining time till the next SO is set equal to τ. We are interested in computing the long-run rate of cost under deferred maintenance and, in Section 7.6.3, using the results of this section and of the previous sections, in investigating the economical benefits of deferring planned maintenance.

Analogously to the analysis of Section 7.4.2, we derive the long-run rate of cost using renewal theory, see, e.g., (Ross, 2014, Proposition 7.3, page 433). In this case, we consider the renewal points to be the instances at which there was a successful maintenance activity, i.e. the SOs or USOs at which the preventive maintenance was perfect, or the epochs at which corrective maintenance is performed. Note that the underlying stochastic process that governs the condition of the system, regenerates after each successful maintenance activity. That is, after each successful maintenance activity the underlying stochastic process is in state 2 with probability 1. The long-run rate of cost per time unit for a policy in the class of optimal policies is given in the next theorem. As the expressions appearing in the theorem do not simplify upon further computations, we choose to present them in the form of probabilities and expectations associated with the Exponential distribution, as these expressions are straightforward (though cumbersome to compute) and shed insight on each of the individual events participating in the final expression, cf. Equation (7.8).

Theorem 7.4. *Consider a given policy under which in state* 2 *we do nothing, and in state* 1 *we repair at scheduled opportunities and at unscheduled opportunities for which the remaining time until the next scheduled opportunity is greater than* $\tilde{t} \in (0, \tau)$*, and we do nothing otherwise. Furthermore, consider that planned maintenance is deferred after a successful maintenance. Under this setting, the long-run rate of cost per time unit equals*

$$\begin{aligned}\frac{\mathbb{E}\left[\textit{Total cycle cost}\right]}{\mathbb{E}\left[\textit{Total cycle length}\right]} &= \frac{\mathbb{E}\left[CC\right]}{\frac{1}{\mu_2} + \mathbb{E}\left[CL\right]} \\ &= \frac{\mathbb{E}\left[CC\,\mathbb{1}_{\{CL \leq Y\}}\right] + \mathbb{E}\left[CC\,\mathbb{1}_{\{CL > Y\}}\right]}{\frac{1}{\mu_2} + \mathbb{E}\left[CL\,\mathbb{1}_{\{CL \leq Y\}}\right] + \mathbb{E}\left[CL\,\mathbb{1}_{\{CL > Y\}}\right]},\end{aligned} \tag{7.8}$$

with

$$\mathbb{E}\left[CL\,\mathbb{1}_{\{CL\leq Y\}}\right] = \mathbb{E}\left[CL\,\mathbb{1}_{\{USO[\tau-Y,\tau-\tilde{t}]\}}\right] + \mathbb{E}\left[CL\,\mathbb{1}_{\{SO[\tau-Y,\tau]\}}\right] + \mathbb{E}\left[CL\,\mathbb{1}_{\{CM[\tau-Y,\tau]\}}\right], \tag{7.9}$$

$$\mathbb{E}\left[CL\,\mathbb{1}_{\{CL>Y\}}\right] = (1-p)\mathbb{P}\left[SO[\tau-Y,\tau]\right]\left(\mathbb{E}[Y] + \frac{\tau(1-p)\mathbb{P}\left[SO[0,\tau]\right]}{1-(1-p)\mathbb{P}\left[SO[0,\tau]\right]} + \mathbb{E}\left[CL'\,\mathbb{1}_{\{CL'\leq Y\}}\,|\,Y=\tau\right]\right), \tag{7.10}$$

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{CL\leq Y\}}\right] = \mathbb{E}\left[CC\,\mathbb{1}_{\{USO[\tau-Y,\tau-\tilde{t}]\}}\right] + \mathbb{E}\left[CC\,\mathbb{1}_{\{SO[\tau-Y,\tau]\}}\right] + \mathbb{E}\left[CC\,\mathbb{1}_{\{CM[\tau-Y,\tau]\}}\right], \tag{7.11}$$

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{CL>Y\}}\right] = (1-p)\mathbb{P}\left[SO[\tau-Y,\tau]\right]\left(\mathbb{E}\left[CC\,\mathbb{1}_{\{SO[\tau-Y,\tau]\}}\right] + \frac{(\lambda(1-p)(\tau-\tilde{t})c_{pm}^{uso}+c_{pm}^{so})(1-p)\mathbb{P}\left[SO[0,\tau]\right]}{1-(1-p)\mathbb{P}\left[SO[0,\tau]\right]} + \mathbb{E}\left[CC\,\mathbb{1}_{\{CL'\leq Y\}}\,|\,Y=\tau\right]\right), \tag{7.12}$$

where the density of the Truncated Exponential random variable Y is given by

$$f_Y(y) = \mu_2\frac{e^{-\mu_2(\tau-y)}}{1-e^{-\mu_2\tau}},\ y\in[0,\tau), \tag{7.13}$$

and with, for $0 \leq y \leq \tau$,

$$\mathbb{1}_{\{SO[\tau-y,\tau]\}} \stackrel{d}{=} \mathbb{1}_{\{y<\min\{T_{\lambda p},T_{\mu_1}\}\}} + \mathbb{1}_{\{T_{\lambda p}<y<\min\{T_{\mu_1},\tilde{t}\}\}}\mathbb{1}_{\{y<\tilde{t}\}} + \mathbb{1}_{\{y-\tilde{t}\leq T_{\lambda p}<y, y\leq T_{\mu_1}\}}\mathbb{1}_{\{y\geq\tilde{t}\}}, \tag{7.14}$$

$$\mathbb{1}_{\{USO[\tau-y,\tau-\tilde{t}]\}} \stackrel{d}{=} \mathbb{1}_{\{T_{\lambda p}<\min\{T_{\mu_1},y-\tilde{t}\}\}}\mathbb{1}_{\{y\geq\tilde{t}\}}, \tag{7.15}$$

$$\mathbb{1}_{\{CM[\tau-y,\tau]\}} \stackrel{d}{=} \mathbb{1}_{\{T_{\mu_1}<\min\{y,T_{\lambda p}\}\}} + \mathbb{1}_{\{T_{\lambda p}<T_{\mu_1}<y\}}\mathbb{1}_{\{y<\tilde{t}\}} + \mathbb{1}_{\{T_{\lambda p}<T_{\mu_1}<y,\, T_{\lambda p}\geq y-\tilde{t}\}}\mathbb{1}_{\{y\geq\tilde{t}\}}, \tag{7.16}$$

$$\mathbb{E}\left[CL\,\mathbb{1}_{\{USO[\tau-y,\tau-\tilde{t}]\}}\right] = \mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{USO[\tau-y,\tau-\tilde{t}]\}}\right], \tag{7.17}$$

$$\mathbb{E}\left[CL\,\mathbb{1}_{\{SO[\tau-y,\tau]\}}\right] = yp\mathbb{P}\left[SO[\tau-y,\tau]\right], \tag{7.18}$$

$$\mathbb{E}\left[CL\,\mathbb{1}_{\{CM[\tau-y,\tau]\}}\right] = \mathbb{E}[T_{\mu_1}\mathbb{1}_{\{CM[\tau-y,\tau]\}}], \tag{7.19}$$

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{USO[\tau-y,\tau-\tilde{t}]\}}\right] = c_{pm}^{uso}\,\mathbb{P}\left[USO[\tau-y,\tau-\tilde{t}]\right] + \lambda(1-p)c_{pm}^{uso}\mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{USO[\tau-y,\tau-\tilde{t}]\}}\right], \tag{7.20}$$

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{SO[\tau-y,\tau]\}}\right] = \left(c_{pm}^{so} + \lambda(1-p)c_{pm}^{uso}\max\left\{y-\tilde{t},0\right\}\right)\mathbb{P}\left[SO[\tau-y,\tau]\right], \tag{7.21}$$

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{CM[\tau-y,\tau]\}}\right] = c_{cm}\mathbb{P}\left[CM[\tau-y,\tau]\right] + \lambda(1-p)c_{pm}^{uso}\mathbb{E}\left[\min\left\{T_{\mu_1},\max\left\{y-\tilde{t},0\right\}\right\}\,\mathbb{1}_{\{CM[\tau-y,\tau]\}}\right], \tag{7.22}$$

where $\mathbb{1}_{\{x\}}$ *is an indicator function taking value 1 if event* x *occurs, and it is zero otherwise,* $T_{\mu_1} \sim Exp(\mu_1)$, $T_{\lambda p} \sim Exp(\lambda p)$, $\mathbb{P}[\,\cdot\,] = \mathbb{E}[\mathbb{1}_{\{\cdot\}}]$ *for all events in Equations* (7.14)–(7.16)*, and* $CL \stackrel{d}{=} CL'$.

Proof: See Appendix 7.E. □

7.6. Numerical results

Using the results and the analyzes of the previous sections, in this section, we illustrate through a few well chosen examples the effect of the various parameters in the long-run rate of cost. In these examples, we investigate the financial advantage of the optimal policy, when compared to other (suboptimal) policies. Furthermore, we highlight

the financial benefit of perfect maintenance by comparing the long-run rate of cost for the perfect maintenance model ($p = 1$) to that of the imperfect maintenance model ($p \in (0, 1)$). Here, we also show the influence of imperfect maintenance on the maintenance planning. In addition, we illustrate the change introduced by the action of deferring planned maintenance after the occurrence of a successful maintenance. To illustrate the financial effects in a realistic context and to connect our analysis with the practice, we use values and data stemming from the wind industry.

7.6.1 Comparison of the optimal policy to suboptimal policies

In this section, we compute, in the context of the wind industry example, the long-run rate of cost under the optimal policy and we examine how it is affected by varying one by one the parameters τ, λ and $c_{\text{pm}}^{\text{uso}}$, while keeping all other parameters fixed. For the determination of the values used in the numerical computations of this section, we consider the gearbox of a wind turbine. Statistics from a recent field study by Ribrant and Bertling (2007) on Swedish wind parks in the period 1997-2005 showed that the gearbox is the most critical unit of a wind turbine. The notion of criticality is determined by the fact that a failure of the gearbox leads to the highest downtime when compared to all other wind turbine components, but also by the fact that this component has the highest failure rate among all wind turbine components (Ribrant and Bertling, 2007; Tavner et al., 2007; Spinato et al., 2009). Due to its extended downtime after a failure (which is captured in the corresponding maintenance cost), the corrective cost of a gearbox is relatively high compared to preventive maintenance costs, see, e.g., Nilsson and Bertling (2007). Based on the values reported in the aforementioned studies, we set $c_{\text{cm}} = 300000$, $c_{\text{pm}}^{\text{so}} = 1000$, $\mu_2 = 0.31$, $\mu_1 = 0.31$ and $p = 0.6$. In this case, the long-run rate of cost (in euros per year) in case of only corrective replacements is equal to 46500. Furthermore, motivated by the wind industry practice, we choose three different values for τ, that is $\tau \in \{0.25, 0.5, 1\}$ (years). Next, we consider three different values for $c_{\text{pm}}^{\text{uso}}$, i.e. $c_{\text{pm}}^{\text{uso}} \in \{2000, 3000, 4000\}$. Finally, with regard to λ, we consider four different values, i.e. $\lambda \in \{0.5, 1, 2, 4\}$.

In Table 7.1, we depict the long-run rate of cost for the above mentioned values under four different policies: The first policy corresponds to replacements only at USOs (π_{uso}). The second policy corresponds to replacements only at SOs (π_{so}). The third policy is the optimal policy (π_{opt}), which is derived in Theorem 7.1. Note, that it is numerically easier to obtain the optimal $\tilde{t}$ by minimizing the long-run rate of cost in Theorem 7.3, instead of the closed-form expression in Theorem 7.1, as the latter requires the derivation of a root solution. The fourth policy concerns the optimal policy, but for $p = 1$. This assumption is motivated from the practice, as it

is oftentimes difficult to exactly determine the value of p and it is typically assumed that after a maintenance the component is restored to a perfect state. This policy is denoted by π'_{opt}.

In Table 7.1, we observe, across all instances, that incorporating planned maintenance can significantly reduce costs compared to only corrective maintenance, which can be reduced even further by adding opportunistic maintenance. Intuitively, due to the cost structure, only planned maintenance at SOs can considerably improve the long-term rate of cost when compared to performing only opportunistic maintenance at USOs. Finally, if we compare π_{opt} with π'_{opt} we do not, despite the low value for p, observe significant differences. From an operational management perspective, this clearly implies that, if decision makers do not have any knowledge about the value of p and given a similar cost structure as in the gearbox case, assuming perfect maintenance will result in a long-run rate of cost that is close to optimal regardless of the true value of p. This will be valid as long as the preventive maintenance cost (at both opportunities) is very small in comparison to the corrective maintenance cost, as is the case of the gearbox costs. As a rule of thumb, one can easily compute the expected number of maintenance activities (planned or opportunistic) required for a successful preventive maintenance and based on this compute the long-run rate of preventive maintenance cost (approximately in the order of $\max\{c_{\text{pm}}^{\text{so}}, c_{\text{pm}}^{\text{uso}}\}/p$) and compare it with the corrective cost. If the corrective cost is significantly higher, then one may assume that there is no significant difference between π_{opt} and π'_{opt}, and as a consequence there is no significant difference in the values of the optimal policies under the imperfect and perfect maintenance. In the next section, we investigate the savings that can be obtained by improving the performance of a repair when a decision maker has some knowledge regarding the value of p.

7.6.2 Influence of imperfect maintenance

Let $\pi_{\text{opt}}^{(p)}$ represent the optimal policy as a function of the successful preventive maintenance probability p and let $C(\pi_{\text{opt}}^{(p)})$ denote the long-run rate of cost when the policy is $\pi_{\text{opt}}^{(p)}$. To demonstrate the effect of p in the rate of cost, we compute the relative difference in the cost of not having a perfect preventive maintenance as a function of p. This relative difference is denoted by $\delta(p)$ and it is equal to

$$\delta(p) = \frac{C(\pi_{\text{opt}}^{p}) - C(\pi_{\text{opt}}^{1})}{C(\pi_{\text{opt}}^{1})} \cdot 100\%.$$

Table 7.1 Long-run rate of cost varying λ, τ and $c_{\mathsf{pm}}^{\mathsf{uso}}$, while keeping all other parameters fixed for four policies.

		$\tau = 0.25$				$\tau = 0.5$				$\tau = 1$			
$c_{\mathrm{pm}}^{\mathrm{uso}}$	λ	π_{uso}	π_{so}	π_{opt}	π'_{opt}	π_{uso}	π_{so}	π_{opt}	π'_{opt}	π_{uso}	π_{so}	π_{opt}	π'_{opt}
2000	0.5	31674	7624	7193	7208	31674	12927	11627	11647	31674	20301	17134	17156
	1	24139	7624	6815	6840	24139	12927	10583	10614	24139	20301	14855	14886
	2	16522	7624	6183	6221	16522	12927	9007	9049	16522	20301	11794	11828
	4	10368	7624	5258	5307	10368	12927	7023	7067	10368	20301	8469	8498
3000	0.5	31842	7624	7230	7255	31842	12927	11687	11725	31842	20301	17224	17265
	1	24393	7624	6883	6927	24393	12927	10691	10751	24393	20301	15010	15068
	2	16863	7624	6304	6372	16863	12927	9188	9267	16863	20301	12034	12100
	4	10778	7624	5456	5543	10778	12927	7294	7375	10778	20301	8800	8855
4000	0.5	32011	7624	7266	7299	32011	12927	11748	11800	32011	20301	17314	17374
	1	24648	7624	6951	7009	24648	12927	10799	10883	24648	20301	15164	15248
	2	17203	7624	6424	6513	17203	12927	9368	9479	17203	20301	12274	12368
	4	11189	7624	5653	5766	11189	12927	7565	7677	11189	20301	9132	9208

$\delta(p)$ indicates how much extra cost is incurred due to imperfect maintenance, and thus shows the benefit of improving the probability of executing a perfect maintenance.

In this numerical example, similarly to before we choose $\mu_2 = 0.31$, and $\mu_1 = 0.31$. Furthermore, we set $\lambda = 4$ and $\tau = 1$. Figure 7.3 shows $\delta(p)$ for $p \in [0.5, 1]$ under two different cost structures (denoted by $\delta(p)^1$ and $\delta(p)^2$, respectively). Figure 7.4 depicts the corresponding optimal values for $\tilde{t}$ for both cost structures, denoted by t^1 and t^2, respectively. We use the same cost structure as in the previous section, i.e. for $\delta(p)^1$, we consider $c_{\text{pm}}^{\text{so}} = 1000, c_{\text{pm}}^{\text{uso}} = 2000$ and $c_{\text{cm}} = 300000$, whereas, for $\delta(p)^2$, we consider $c_{\text{pm}}^{\text{so}} = 26500, c_{\text{pm}}^{\text{uso}} = 28800$ and $c_{\text{cm}} = 75500$. The choice for the preventive maintenance cost at SOs and USOs in the second cost structure is common in the lithography industry (see Zhu et al. (2017)). Based on Figure 7.3, we can conclude that, under both cost structures, significant costs can be saved by improving the probability of executing a perfect preventive maintenance (e.g., by training).

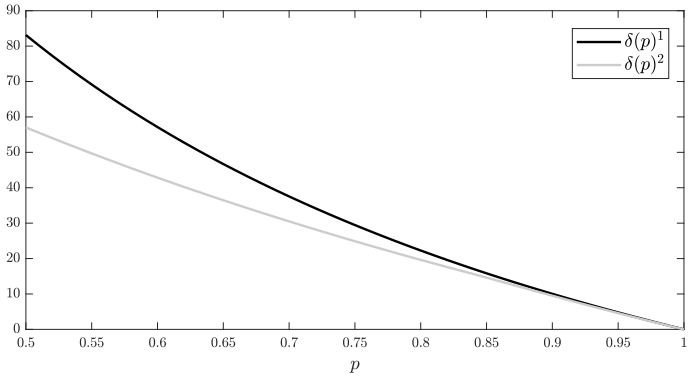


Figure 7.3 Plot of $\delta(p)^1$ and $\delta(p)^2$ for $p \in [0.5, 1]$ with $c_{\text{pm}}^{\text{so}} = 1000, c_{\text{pm}}^{\text{uso}} = 2000$ and $c_{\text{cm}} = 300000$ for $\delta(p)^1$, and $c_{\text{pm}}^{\text{so}} = 26500, c_{\text{pm}}^{\text{uso}} = 28800$ and $c_{\text{cm}} = 75500$ for $\delta(p)^2$.

The optimal policy $(\tilde{t})$, denoted by t^1 and t^2, under the first and second cost structure, respectively, is equal to $t^1 \approx 0.08$ and $t^2 \approx 0.39$ in case of perfect repairs. In Figure 7.4, where we plot t^1 and t^2 as a function of p, we observe the following regarding the influence of p on the maintenance planning: If the preventive maintenance cost (at both opportunities) is very small compared to the cost of corrective maintenance, the order of the total preventive maintenance cost incurred until a successful preventive maintenance compared to the corrective maintenance cost is still maintained. Therefore, the maintenance planning does not alter that much regardless of the value of p, where the optimal policy is to almost always perform preventive maintenance at USOs for all values of $p \in [0.5, 1]$. This also explains the small discrepancy between π_{opt} and π'_{opt} in Table 7.1. This is different in the case of

the second cost structure, where the maintenance planning changes substantially as a function of p. Whereas in the perfect case, the optimal policy is to perform preventive maintenance at a USO if the residual time until the next SO is larger than 0.39, for $p \lessapprox 0.83$, it is optimal to never perform preventive maintenance at a USO. Here, the order of the total preventive maintenance cost incurred until a successful preventive maintenance compared to the corrective maintenance cost is not maintained.

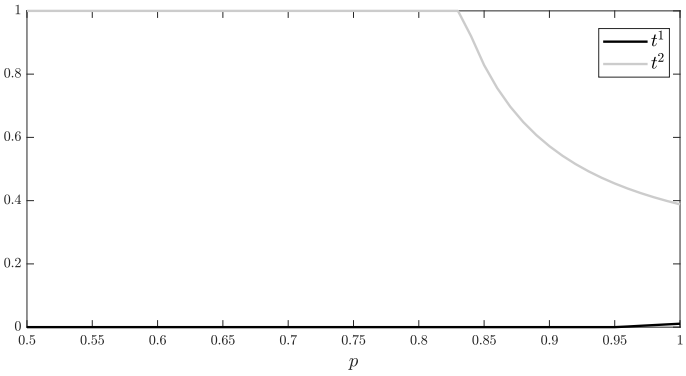


Figure 7.4 Plot of t^1 and t^2 for $p \in [0.5, 1]$ with $c_{\mathsf{pm}}^{\mathsf{so}} = 1000, c_{\mathsf{pm}}^{\mathsf{uso}} = 2000$ and $c_{\mathsf{cm}} = 300000$ for t^1, and $c_{\mathsf{pm}}^{\mathsf{so}} = 26500, c_{\mathsf{pm}}^{\mathsf{uso}} = 28800$ and $c_{\mathsf{cm}} = 75500$ for t^2.

Also in the opposite cost structure, i.e. $c_{\mathrm{pm}}^{\mathrm{uso}} < c_{\mathrm{pm}}^{\mathrm{so}}$ (similar examples can be found for $c_{\mathrm{pm}}^{\mathrm{uso}} = c_{\mathrm{pm}}^{\mathrm{so}}$), the maintenance planning can be influenced significantly by the imperfect repair probability. For instance, consider the setting with $\mu_1 = 1.1, \mu_2 = 0.9$, $c_{\mathrm{pm}}^{\mathrm{so}} = 4500$, $c_{\mathrm{pm}}^{\mathrm{uso}} = 4000$, $c_{\mathrm{cm}} = 10000$, and $\lambda = 0.5$. In case of perfect repairs (i.e. $p = 1$), the optimal policy is to perform preventive maintenance in state 1 at both SOs and USOs, and to do nothing otherwise (cf. Theorem 7.2). However, if $0.72 \lessapprox p \lessapprox 0.83$, the optimal policy is to only perform preventive maintenance at USOs and if $p \lessapprox 0.72$, then the optimal policy is to never perform PM. This example illustrates the influence of the imperfect repair probability on the maintenance planning.

7.6.3 Deferring of planned maintenance

In this section, we illustrate the change introduced by the action of deferring planned maintenance after the occurrence of a successful maintenance in three numerical examples that relate to the wind industry, the lithography industry, and to an artificially created example.

Figure 7.5 shows the long-run rate of cost for both the deferral and no deferral case for the example with data stemming from the wind industry. Again, with regard to

the cost parameters, we used $c_{\text{pm}}^{\text{so}} = 1000, c_{\text{pm}}^{\text{uso}} = 2000$ and $c_{\text{cm}} = 300000$. With regard to the other parameters, we set $\lambda = 4$, $\tau = 1$, $\mu_1 = 0.31$, $\mu_2 = 0.31$ and $p = 0.6$. We can observe that deferring the planned maintenance both significantly increases the long-run rate of cost under the optimal policy (an increase of 28.14% from 8468.87 to 10852.15) and changes the value connected to the optimal policy, $\tilde{t}$ from 0.112 to 0.

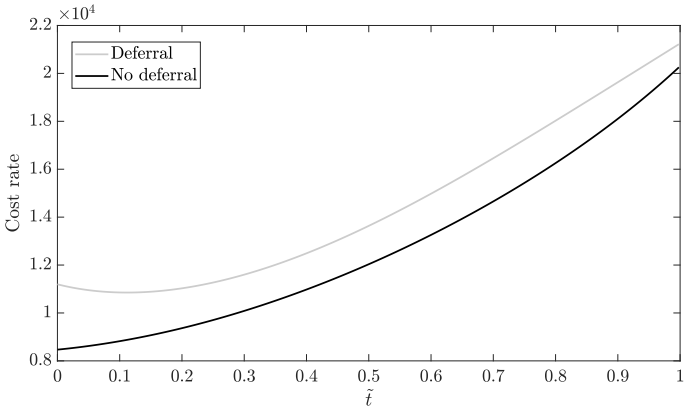
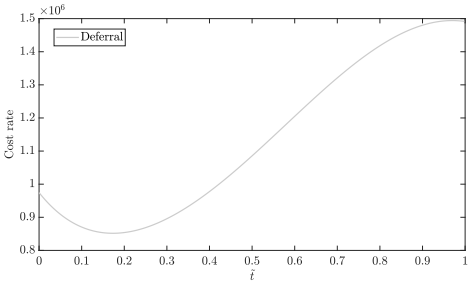


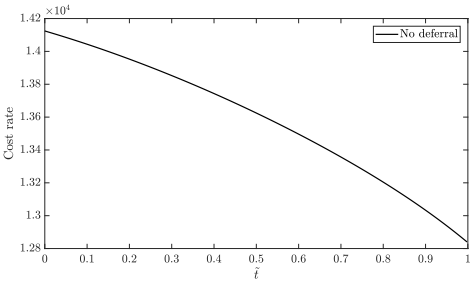
Figure 7.5 Cost rate in case of deferral and of no deferral for wind industry example. Optimal $\tilde{t}$ is equal to 0.112 and 0 for deferral and no deferral, respectively.

Figure 7.6a and Figure 7.6b depict the long-run rate of cost for both the deferral and the no deferral case, respectively, based on the values of the lithography industry example. We use the same cost parameters as in Section 7.6.2, that is $c_{\text{pm}}^{\text{so}} = 26500, c_{\text{pm}}^{\text{uso}} = 28800$ and $c_{\text{cm}} = 75500$. The other parameters remain unchanged, i.e. $\lambda = 4$, $\tau = 1$, $\mu_1 = 0.31$, $\mu_2 = 0.31$ and $p = 0.6$. Again, we observe the same influence of deferring the planned maintenance on both the long-run rate of cost under the optimal policy (an increase of 6533.3 % from 12840.12 to 851727.53) and on the value of $\tilde{t}$ associated with the optimal policy (from 1 to 0.175) , similarly to the numerical example for the wind industry. The drastic increase is due to the cost structure, and more explicitly, it is due to the preventive maintenance costs values (both at scheduled and unscheduled opportunities), which are relatively much closer to the corrective maintenance cost in comparison to the wind industry example.

To illustrate that the opposite effect (albeit to a much lesser degree than in the previous two examples) can also hold, we create an artificial example where we set $c_{\text{pm}}^{\text{so}} = 5000, c_{\text{pm}}^{\text{uso}} = 10000$ and $c_{\text{cm}} = 19000$, and $\lambda = 4$,$\tau = 4$,$\mu_1 = 1$, $\mu_2 = 0.4$ and $p = 0.5$. Figure 7.7 depicts the long-run rate of cost for both the deferral and the no deferral case for this example. Here we observe that for all values of $\tilde{t}$, cost savings can be obtained by deferring planned maintenance after the occurrence of a successful



(a) Cost rate in case of deferral. Optimal $\tilde{t}$ is equal to 0.175.



(b) Cost rate in case of no deferral. Optimal $\tilde{t}$ is equal to 1.

Figure 7.6 Cost rate for lithography industry example with deferral (a) and no deferral (b).

opportunistic maintenance. More specifically, whereas the optimal value of $\tilde{t}$ is equal to 1 for both cases, the long-run rate of cost under the optimal policy decreases with 0.88% from 6458.97 to 6402.44, when deferring planned maintenance.

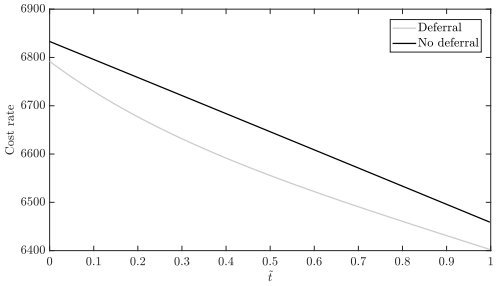


Figure 7.7 Cost rate in case of deferral and no deferral for artificial example. Optimal $\tilde{t}$ is equal to 1 for both deferral and no deferral.

7.7. Conclusion

In this chapter, we have considered the maintenance policy for a 3-state component degrading over time with corrective replacements at failures and preventive replacements at both scheduled and unscheduled opportunities under imperfect repair. By formulating this problem as a semi-MDP, we were able to characterize the structure of the optimal maintenance policy as a control limit policy, where the control limit depends on the time until the next planned maintenance opportunity. Using this approach, a closed-form expression for the optimal control limit has been derived.

Within this class of control limit policies, we have derived, using the theory of regenerative processes, an explicit expression for the long-run rate of cost. Using a similar approach based on renewal theory, we have derived an expression for the long-run rate of cost in the case when planned maintenance is deferred after the occurrence of a successful opportunistic maintenance.

A cost comparison with other suboptimal policies was made, which illustrated the benefits of optimizing the maintenance policy. Specifically, it was found that incorporating planned maintenance can significantly reduce costs compared to only corrective maintenance, which can be reduced even further by adding opportunistic maintenance. Moreover, numerical results indicate that the extent of the impact of the perfect repair probability on the optimal policy depends on the underlying cost structure. It was also shown that substantial cost savings can be obtained by improving the perfect repair probability. Finally, our numerical examples indicate that the deferral of planned maintenance after the occurrence of a successful opportunistic maintenance may impact the total cost in both a negative and positive way.

There is a number of extensions and topics for future research. The most important direction, and perhaps most obvious in light of the previous five chapters, is to include parameter uncertainty. Another interesting direction is to include network dependency on the level of degradation and failure dependencies, i.e. to consider a multi-dimensional process that captures the degradation of the various assets in the network. Such a future direction would be particularly interesting in the case of a small number of assets for which the Poisson approximation for the opportunistic maintenance may not be accurate. In addition, another very interesting research direction would be to consider a more general model in which the condition of the system degrades through $N > 2$ states. Next, in this analysis, we have assumed that the condition of the system is fully observable. However, in many real applications, condition monitoring data such as spectrometric oil data or vibration data gives only partial information about the underlying state of the system. From this perspective, it would be interesting to extend the model at hand to a partially observable model in which the condition monitoring data are stochastically related to the true system state. Finally, the results in this chapter are valid for systems with Hypo-Exponentially distributed lifetimes. Future research could relax this assumption by considering a phase-type lifetime distribution.

7.A. Optimality equations for semi-Markov decision process

We consider the so-called ratio-average cost for a controlled semi-MDP, which corresponds to the limes superior of the expected total cost over a finite number of jumps divided by the expected cumulative time of these jumps, see Ross (1970); Feinberg (1994); Schäl (1992), for instance.

We shall use here the definition of a controlled semi-MDP from Lippman (1975); Yushkevich (1982); Jaśkiewicz (2004). A controlled semi-MDP is specified by five objects: a Borel state space $\mathcal{S}$, a Borel action space $\mathcal{A}$, a law of motion – a measurable projection determining the state as a function of an action, a transition function (transition law) $\mathcal{P}$ – a probability measure depending measurably on the state and the action, and a reward (or cost) function c.

The process is observed at time $t = 0$ to be in some state $x_0 \in \mathcal{S}$. At that time an action $a_0 \in \mathcal{A}_{x_0}$ is chosen, where $\mathcal{A}_{x_0}$ is a compact set of actions available in state x_0. The set of all actions is $\mathcal{A}$ and is also assumed to be a Borel state space.

For the problem at hand, the state space is

$$\mathcal{S} = \big\{(i,j,t):\ i \in \{1,2\},\ j \in \{\mathrm{SC},\mathrm{USO}\},\ t \in (0,\tau)\big\} \cup \big\{(i,\mathrm{SO},0):\ i \in \{1,2\}\big\}\big\}$$

and the action space is $\mathcal{A} = \{\text{perform PM, do nothing, perform CM}\}$, cf. Section 7.4.1.

If the current state is x_0 and action a_0 is selected, then the immediate cost $c(x_0;a_0)$ is incurred, and the system remains in state x_0 for a random time t_1, with the cumulative distribution depending only on x_0 and a_0. Afterward, the system jumps to the state x_1 according to the probability measure (transition law) $\mathcal{P}(\cdot \,|\, x_0, a_0, t_1)$. This procedure yields a trajectory $(x_0, a_0, t_1, x_1, a_1, t_2, \ldots)$ of some stochastic process, where x_n is the state, a_n is the control variable and t_n is the time of the n-th transition, $n = 0, 1, \ldots$. In the sequel, we shall refer to the corresponding random variables by means of their capital letters: T_n – the random time of n-th transition, for $n = 1, 2, \ldots$ with $T_0 \triangleq 0$, X_n – the state at time T_n, and A_n – the action at time T_n.

Let H_n be the space of admissible histories up to the n-th transition, $H_n \triangleq (\mathcal{S} \times \mathcal{A} \times [0,\infty))^n \times \mathcal{S}$ and $H_0 \triangleq \mathcal{S}$. An element h_n of H_n is called a partial history of the process and is of the form $h_n = (x_0, a_0, t_1, \ldots, x_{n-1}, a_{n-1}, t_n, x_n)$. A control policy (or policy) is a sequence $\{\pi_n\}$, where each π_n is a conditional probability $\pi_n(\cdot \,|\, h_n)$ on the control set $\mathcal{A}_{x_n}$, given the entire history h_n such that $\pi_n(\mathcal{A}_{x_n} \,|\, h_n) = 1$, for

all h_n, $n = 1, 2, \ldots$. The class of all policies is denoted by Π and let Π_{DS} denote the class of all deterministic stationary policies.

For each initial state $x_0 \in \mathcal{S}$ and for each policy $\pi \in \Pi$, there exists a unique probability measure $\mathbb{P}^{\pi}_{x_0}$ such that

$$\begin{aligned}
\mathbb{P}^{\pi}_{x_0}[A_n \in A \,|\, h_n] &= \pi_n[A \,|\, h_n], \text{ for a Borel set } A \subset \mathcal{A},\\
\mathbb{P}^{\pi}_{x_0}[T_{n+1} - T_n \in S, X_{n+1} \in X \,|\, h_n, a_n] &= \mathcal{P}^{a_n}_{x_n}[S, X], \text{ for Borel sets } X \subset \mathcal{S}, S \subset \mathbb{R},\\
\mathbb{P}^{\pi}_{x_0}[X_{n+1} \in X \,|\, h_n, a_n, T_{n+1} - T_n = s] &= \mathcal{P}^{a_n}_{x_n}[X \,|\, s], \text{ for a Borel set } X \subset \mathcal{S},\\
\mathbb{P}^{\pi}_{x_0}[T_{n+1} - T_n \le s \,|\, h_n, a_n] &= F^{a_n}_{x_n}(s), s \in \mathbb{R}.
\end{aligned}$$

Further, let $\tau(x, a)$ denote the conditional mean sojourn (holding) time spent in state x under action a, i.e.

$$\tau(x; a) \triangleq \int_0^{\infty} s \mathrm{d}F^a_x(s),$$

and let $\tilde{F}^a_x(\alpha)$ denote the Laplace-Stieltjes transform of the sojourn time spent in state x under action a, i.e.

$$\tilde{F}^a_x(\alpha) \triangleq \int_0^{\infty} e^{-\alpha s} \mathrm{d}F^a_x(s).$$

For the problem at hand, the cost function is defined as follows

$$c(x; a) = \begin{cases} 0, & \text{if } x \in \mathcal{S}, a = \{\text{do nothing}\},\\ c_{\text{cm}}, & \text{if } x = (1, \text{SC}, t), t \in (0, \tau), a = \{\text{perform CM}\},\\ c^{\text{uso}}_{\text{pm}}, & \text{if } x = (i, \text{USO}, t), i = 1, 2, t \in (0, \tau), a = \{\text{perform PM}\},\\ c^{\text{so}}_{\text{pm}}, & \text{if } x = (i, \text{SO}, 0), i = 1, 2, a = \{\text{perform PM}\}. \end{cases}$$

Let $\mathcal{P}^{a_n}_{x_n}(s, x_{n+1})$ denote the joint density/mass distribution of the transition time $T_{n+1} - T_n$ and the allowed next state X_{n+1}, given the current state $X_n = x_n$ and the allowed action a_n. For $x_n = (1, \text{SC}, t)$, $t \in (0, \tau)$, and $a_n = \{\text{perform CM}\}$,

$$\begin{aligned}
\mathcal{P}^{a_n}_{x_n}\big(s, (2, \text{SC}, t - s)\big) &= \frac{\mu_2}{\lambda + \mu_2}(\lambda + \mu_2)e^{-(\lambda+\mu_2)s} = \mu_2\, e^{-(\lambda+\mu_2)s},\ s \in [0, t)\\
\mathcal{P}^{a_n}_{x_n}\big(s, (2, \text{USO}, t - s)\big) &= \lambda\, e^{-(\lambda+\mu_2)s},\ s \in [0, t)\\
\mathcal{P}^{a_n}_{x_n}\big(t, (2, \text{SO}, 0)\big) &= e^{-(\lambda+\mu_2)t}.
\end{aligned}$$

For the derivation of the above probabilities, it suffices to note that there are three possible evolutions in terms of the state of the system: either an SO or an SC or

a USO, where the time till the SO is equal to t, while the times till the next SC and the USO are Exponentially distributed with rates μ_2 and λ, respectively. The probabilities for $x_n = (2, \text{USO}, t)$ and $a_n = \{\text{do nothing}\}$ or $a_n = \{\text{perform PM}\}$ are identical. The remaining probabilities are obtained using very similar arguments. For $x_n = (2, \text{SC}, t)$ or $x_n = (1, \text{USO}, t)$ and $a_n = \{\text{do nothing}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{SC}, t-s) &= \mu_1 \, e^{-(\lambda+\mu_1)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{USO}, t-s) &= \lambda \, e^{-(\lambda+\mu_1)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(t, 1, \text{SO}, 0) &= e^{-(\lambda+\mu_1)t}.
\end{aligned}$$

For $x_n = (1, \text{SO}, 0)$ and $a_n = \{\text{do nothing}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{SC}, \tau-s) &= \mu_1 \, e^{-(\lambda+\mu_1)s}, \;\; s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{USO}, \tau-s) &= \lambda \, e^{-(\lambda+\mu_1)s}, \;\; s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(\tau, 1, \text{SO}, 0) &= e^{-(\lambda+\mu_1)\tau}.
\end{aligned}$$

For $x_n = (2, \text{SO}, 0)$ and $a_n = \{\text{do nothing}\}$ or $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s, 2, \text{SC}, \tau-s) &= \mu_2 \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(s, 2, \text{USO}, \tau-s) &= \lambda \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(\tau, 2, \text{SO}, 0) &= e^{-(\lambda+\mu_2)\tau}.
\end{aligned}$$

For $x_n = (1, \text{USO}, t)$ and $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s, 2, \text{SC}, t-s) &= p\,\mu_2 \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(s, 2, \text{USO}, t-s) &= p\,\lambda \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(t, 2, \text{SO}, 0) &= p \, e^{-(\lambda+\mu_1)t} \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{SC}, t-s) &= q\,\mu_2 \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \text{USO}, t-s) &= q\,\lambda \, e^{-(\lambda+\mu_2)s}, \;\; s \in [0, t) \\
\mathcal{P}_{x_n}^{a_n}(t, 1, \text{SO}, 0) &= q \, e^{-(\lambda+\mu_1)t}.
\end{aligned}$$

For $x_n = (1, \mathrm{SO}, 0)$ and $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s, 2, \mathrm{SC}, \tau - s) &= p\,\mu_2\, e^{-(\lambda+\mu_2)s}, \quad s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(s, 2, \mathrm{USO}, \tau - s) &= p\,\lambda\, e^{-(\lambda+\mu_2)s}, \quad s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(\tau, 2, \mathrm{SO}, 0) &= p\, e^{-(\lambda+\mu_1)\tau} \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \mathrm{SC}, \tau - s) &= q\,\mu_2\, e^{-(\lambda+\mu_2)s}, \quad s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(s, 1, \mathrm{USO}, \tau - s) &= q\,\lambda\, e^{-(\lambda+\mu_2)s}, \quad s \in [0, \tau) \\
\mathcal{P}_{x_n}^{a_n}(\tau, 1, \mathrm{SO}, 0) &= q\, e^{-(\lambda+\mu_1)\tau}.
\end{aligned}$$

From the joint distributions, the marginal cumulative distribution of the transition time $T_{n+1} - T_n$ can be immediately derived as follows, for $x_n = (1, \mathrm{SC}, t)$ and $a_n = \{\text{perform CM}\}$,

$$\begin{aligned}
F_{x_n}^{a_n}(s) &= 1 - e^{-(\lambda+\mu_2)s},\ s \in (0, t), \\
F_{x_n}^{a_n}(s) &= 1,\ s \geq t.
\end{aligned}$$

The distribution of the transition time from state $x_n = (2, \mathrm{USO}, t)$ under actions $a_n = \{\text{do nothing}\}$ or $a_n = \{\text{perform PM}\}$ are identical. The rest of the marginal cumulative distributions for the other states and actions follow analogously. For $x_n = (2, \mathrm{SC}, t)$ or $x_n = (1, \mathrm{USO}, t)$ and $a_n = \{\text{do nothing}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}[(1, \mathrm{SC}, t - s) \,|\, T_{n+1} - T_n = s] &= \frac{\mu_1}{\lambda + \mu_1}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(1, \mathrm{USO}, t - s) \,|\, T_{n+1} - T_n = s] &= \frac{\lambda}{\lambda + \mu_1}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(1, \mathrm{SO}, 0) \,|\, T_{n+1} - T_n = t] &= 1,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s) &= 1 - e^{-(\lambda+\mu_1)s}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}(s) &= 1, s \geq t.
\end{aligned}$$

For $x_n = (1, \text{USO}, t)$ and $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}[(2, \text{SC}, t-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{successful PM}\}}] &= \frac{\mu_2}{\lambda + \mu_2}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(2, \text{USO}, t-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{successful PM}\}}] &= \frac{\lambda}{\lambda + \mu_2}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(2, \text{SO}, 0) \,|\, T_{n+1} - T_n = t, \mathbb{1}_{\{\text{successful PM}\}}] &= 1, \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{SC}, t-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= \frac{\mu_1}{\lambda + \mu_1}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{USO}, t-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= \frac{\lambda}{\lambda + \mu_1}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{SO}, 0) \,|\, T_{n+1} - T_n = t, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= 1,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{successful PM}) &= 1 - e^{-(\lambda+\mu_2)s}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{successful PM}) &= 1, s \geq t, \\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{unsuccessful PM}) &= 1 - e^{-(\lambda+\mu_1)s}, \quad s \in (0, t), \\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{unsuccessful PM}) &= 1, s \geq t,
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}[\mathbb{1}_{\{\text{successful PM}\}} \,|\, X_n = (1, \text{USO}, t), a_n = \{\text{perform PM}\}] \\
&= \mathbb{P}[\mathbb{1}_{\{\text{successful PM}\}} \,|\, X_n = (1, \text{SO}, t), a_n = \{\text{perform PM}\}] = p.
\end{aligned}$$

For $x_n = (1, \text{SO}, 0)$ and $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}[(2, \text{SC}, \tau-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{successful PM}\}}] &= \frac{\mu_2}{\lambda + \mu_2}, \quad s \in (0, \tau), \\
\mathcal{P}_{x_n}^{a_n}[(2, \text{USO}, \tau-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{successful PM}\}}] &= \frac{\lambda}{\lambda + \mu_2}, \quad s \in (0, \tau), \\
\mathcal{P}_{x_n}^{a_n}[(2, \text{SO}, 0) \,|\, T_{n+1} - T_n = \tau, \mathbb{1}_{\{\text{successful PM}\}}] &= 1, \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{SC}, \tau-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= \frac{\mu_1}{\lambda + \mu_1}, \quad s \in (0, \tau), \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{USO}, \tau-s) \,|\, T_{n+1} - T_n = s, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= \frac{\lambda}{\lambda + \mu_1}, \quad s \in (0, \tau), \\
\mathcal{P}_{x_n}^{a_n}[(1, \text{SO}, 0) \,|\, T_{n+1} - T_n = \tau, \mathbb{1}_{\{\text{unsuccessful PM}\}}] &= 1,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{successful PM}) &= 1 - e^{-(\lambda+\mu_2)s}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{successful PM}) &= 1, s \geq \tau,\\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{unsuccessful PM}) &= 1 - e^{-(\lambda+\mu_1)s}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}(s \,|\, \text{unsuccessful PM}) &= 1, s \geq \tau.
\end{aligned}$$

For $x_n = (1, \text{SO}, 0)$ and $a_n = \{\text{do nothing}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}[(1, \text{SC}, \tau - s) \,|\, T_{n+1} - T_n = s] &= \frac{\mu_1}{\lambda + \mu_1}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}[(1, \text{USO}, \tau - s) \,|\, T_{n+1} - T_n = s] &= \frac{\lambda}{\lambda + \mu_1}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}[(1, \text{SO}, 0) \,|\, T_{n+1} - T_n = \tau] &= 1,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s) &= 1 - e^{-(\lambda+\mu_1)s}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}(s) &= 1, s \geq \tau.
\end{aligned}$$

For $x_n = (2, \text{SO}, 0)$ and $a_n = \{\text{do nothing}\}$ or $a_n = \{\text{perform PM}\}$,

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}[(2, \text{SC}, \tau - s) \,|\, T_{n+1} - T_n = s] &= \frac{\mu_2}{\lambda + \mu_2}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}[(2, \text{USO}, \tau - s) \,|\, T_{n+1} - T_n = s] &= \frac{\lambda}{\lambda + \mu_2}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}[(2, \text{SO}, 0) \,|\, T_{n+1} - T_n = \tau] &= 1,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{x_n}^{a_n}(s) &= 1 - e^{-(\lambda+\mu_2)s}, \;\; s \in (0,\tau),\\
\mathcal{P}_{x_n}^{a_n}(s) &= 1, s \geq \tau.
\end{aligned}$$

Having fully defined the probabilities for the problem at hand, we proceed in providing, following the proofs in Bhattacharya and Majumdar (1989), the proposition below that guarantees that (1) a dynamic programming equation holds for the optimal reward (this equation is typically referred to as the average optimality equality or as the Bellman equation), and (2) a deterministic stationary policy (optimal for long-run average reward) is provided by this equation.

Proposition 7.3. *For the model at hand, there exist a bounded function $V(\cdot)$ and a constant g such that*

$$V(x) = \min_{a \in \mathcal{A}_x} \Big\{ c(x;a) + \int_y V(y) \mathcal{P}_x^a(\mathrm{d}y) - g\,\tau(x;a) \Big\}, \ \forall x \in \mathcal{S}. \tag{7.23}$$

Moreover, the deterministic stationary policy $\pi^{*(\infty)} \in \Pi_{DS}$ is optimal for the ratio-average cost criterion with

$$g = \inf_{\pi \in \Pi_{DS}} J(x,\pi) \triangleq J^*(x)$$

where, for $\pi \in \Pi_{DS}$,

$$J(x,\pi) \triangleq \limsup_{n\to\infty} \frac{\mathbb{E}_x^\pi \left[\sum_{k=0}^{n-1} c(X_k, A_k)\right]}{\mathbb{E}_x^\pi \left[T_n\right]} \equiv \limsup_{n\to\infty} \frac{\mathbb{E}_x^\pi \left[\sum_{k=0}^{n-1} c(X_k, A_k)\right]}{\mathbb{E}_x^\pi \left[\sum_{k=0}^{n-1} \tau(X_k, A_k)\right]}. \tag{7.24}$$

Proof: The proof of the proposition relies on the fact that the costs $c(x;a)$ are non-negative and upper bounded by c_{cm}. We follow here the ideas presented in Bhattacharya and Majumdar (1989) and in Theorems 10.3.1 & 10.3.6 in (Hernández-Lerma and Lasserre, 2012, Sections 10.4 and 10.5). Following the ideas therein, we consider the corresponding α-discounted cost criterion

$$V_\alpha(x,\pi) = \mathbb{E}_x^\pi \left[\sum_{k=0}^{\infty} e^{-\alpha T_k} c(X_k, A_k) \right]$$

and $V_\alpha(x) = \inf_{\pi\in\Pi} V_\alpha(x,\pi)$. The main steps in the proof of the proposition are

Step 1: Show that the optimal reward $V_\alpha(x)$ under discounting is continuous and bounded. The latter follows easily by noting that $V_\alpha(x)$ is bounded by $V_\alpha(\pi_{\mathrm{DN}}, x)$, where π_{DN} denotes the policy of doing nothing at all opportunities, unless the component fails, in which case it is mandatory to do corrective maintenance. This yields

$$V_\alpha(x) \leq c_{\mathrm{cm}} \frac{\frac{\mu_1}{\mu_1+\alpha}}{1 - \frac{\mu_1}{\mu_1+\alpha}\frac{\mu_2}{\mu_2+\alpha}}, \ \forall x \in \mathcal{S}.$$

Analogously,

$$g \equiv J^*(x) \leq c_{\mathrm{cm}} \frac{\mu_1\mu_2}{\mu_2+\mu_1}.$$

See Appendix 7.A.1 for further details.

Step 2: Show that the discounted Bellman equation

$$V_\alpha(x) = \min_{a \in \mathcal{A}_x} \left\{ c(x;a) + \int_s \int_y e^{-\alpha s} V(y) \mathcal{P}_x^a(\mathrm{d}y \mid s)\, \mathrm{d}F_x^a(s) \right\}, x \in \mathcal{S}, \qquad (7.25)$$

holds. Also, there exists a Borel measurable function that minimizes the right side of the discounted Bellman equation for every $x \in \mathcal{S}$. The deterministic stationary policy is optimal under discounting. The proof follows verbatim the steps in (Bhattacharya and Majumdar, 1989, Theorem 3.1 on page 227).

Step 3: Choose $z \in \mathcal{S}$, then for all $x \in \mathcal{S}$, show that $|V_\alpha(x) - V_\alpha(z)|$ is bounded for all $\alpha > 0$. This follows oftentimes by the geometric ergodicity of the underlying Markov controlled model. In the case under consideration, this is proven by noting that from all states $x = (i, j, t) \in \mathcal{S}$, after time t the system is in an SO state with probability 1. This yields

$$|V_\alpha(x) - V_\alpha(z)| \leq c_{\mathrm{cm}} \left(2 + \left(\lambda + \mu_1 + \mu_2 + 1 + \frac{\mu_1 \mu_2}{\mu_1 + \mu_2} \right) (\tau_{x,1} + \tau_{z,1}) \right),$$

with $\tau_{x,1} < \infty$ denoting the expectation of the first passage time from state $x \in \mathcal{S}$ to state $(1, \mathrm{SO}, 0)$. See Appendix 7.A.2 for further details. A consequence of the above finding is that, for all deterministic stationary policies $\pi \in \Pi_{DS}$, the expected average cost in (7.24) is independent of x.

Step 4: Show that there exists a solution say g to the average optimality equality (7.23). There exists a Borel measurable function π^* on $\mathcal{S}$ into $\mathcal{A}$ such that the maximum on the right side of (7.23) is attained at $\pi^*(x)$, $x \in \mathcal{S}$. The proof follows verbatim the steps in (Bhattacharya and Majumdar, 1989, Theorem 3.2 (a) & (b) on page 228).

Step 5: Show that the stationary policy $\pi^{*(\infty)}$ is optimal for the long-run average reward and g is the optimal reward, with $g = \limsup_{\alpha \to 0^+} \alpha V_\alpha(x)$. See Appendix 7.A.3 for further details.

□

Equivalent propositions (based on different methods, but more importantly based on different assumptions regarding the geometric ergodicity) can be found for example in Jaśkiewicz (2001); Vega-Amaya and Luque-Vásquez (2000); Jaśkiewicz (2004).

7.A.1 Proof of Step 1

Under the policy of doing nothing at all opportunities, unless the component fails in which case it is mandatory to do corrective maintenance, say π_{DN}, $V_\alpha(\pi_{\mathrm{DN}}, x)$ can be computed using first step analysis. Note that under this policy, it is not required to keep track of the remaining time to the next SO opportunity. Say $x = (i, j, \cdot)$. If $i = 1$, then after an Exponentially distributed time with rate μ_1, say T_{μ_1}, the component will fail and a cost c_{cm} will be incurred. If $i = 2$, then after a Hypo-Exponentially distributed time with rates (μ_2, μ_1), say $T_{\mu_1} + T_{\mu_2}$ (the two random times are independent), the component will fail and a cost c_{cm} will be incurred. All in all,

$$V_\alpha(x, \pi_{\mathrm{DN}}) = \mathbb{E}[e^{-\alpha(T_{\mu_1} + T_{\mu_2} \mathbb{1}_{\{i=2\}})}] \left(c_{\mathrm{cm}} + \mathbb{E}[V_\alpha((1, \mathrm{SC}, \cdot), \pi_{\mathrm{DN}})]\right). \tag{7.26}$$

Similarly,

$$V_\alpha((1, \mathrm{SC}, \cdot), \pi_{\mathrm{DN}}) = \mathbb{E}[e^{-\alpha(T_{\mu_1} + T_{\mu_2})}] \left(c_{\mathrm{cm}} + \mathbb{E}[V_\alpha((1, \mathrm{SC}, \cdot), \pi_{\mathrm{DN}})]\right),$$

which yields upon solving for $V_\alpha((1, \mathrm{SC}, \cdot), \pi_{\mathrm{DN}})$ and substituting that $\mathbb{E}[e^{-\alpha T_{\mu_i}}] = \frac{\mu_i}{\mu_i + \alpha}$, $i = 1, 2$,

$$V_\alpha((1, \mathrm{SC}, \cdot), \pi_{\mathrm{DN}}) = c_{\mathrm{cm}} \frac{\frac{\mu_1}{\mu_1+\alpha} \frac{\mu_2}{\mu_2+\alpha}}{1 - \frac{\mu_1}{\mu_1+\alpha} \frac{\mu_2}{\mu_2+\alpha}}.$$

Combining the last equation with (7.26) yields

$$V_\alpha(x, \pi_{\mathrm{DN}}) = c_{\mathrm{cm}} \frac{\frac{\mu_1}{\mu_1+\alpha} \left(\frac{\mu_2}{\mu_2+\alpha} \mathbb{1}_{\{i=2\}} + \mathbb{1}_{\{i \neq 2\}} \right)}{1 - \frac{\mu_1}{\mu_1+\alpha} \frac{\mu_2}{\mu_2+\alpha}} \leq c_{\mathrm{cm}} \frac{\frac{\mu_1}{\mu_1+\alpha}}{1 - \frac{\mu_1}{\mu_1+\alpha} \frac{\mu_2}{\mu_2+\alpha}}.$$

The proof for the long-run average cost follows by employing a simple renewal argument.

7.A.2 Proof of Step 3

Choose $x = (i, j, t) \in \mathcal{S}$. Let $T_{x,1}$ denote the first passage time from state x to state $(1, \mathrm{SO}, 0)$, and $\tilde{F}_{x,1}(\alpha) = \mathbb{E}[e^{-\alpha T_{x,1}}]$ and $\tau_{x,1} = \mathbb{E}[T_{x,1}]$.

Starting from state x, after time $t \in [0, \tau)$, the system is in an SO state with probability 1. More concretely, under the optimal policy (which is deterministic stationary), say $\pi_\alpha^{(\infty)}$, starting in state $x = (i, j, t)$, it will end up in state $(1, \mathrm{SO}, 0)$

after time t with probability p_x, and in state $(2, \mathrm{SO}, 0)$ with probability $1 - p_x$. In case state x coincides with an SO state then $p_x = 0$ or $p_x = 1$. Once in an SO state, the system state observed at only the SO epochs behaves like a discrete time (irreducible and aperiodic) Markov chain with only states $(1, \mathrm{SO}, 0)$ and $(2, \mathrm{SO}, 0)$. Thus, $\tau_{x,1} = \mathbb{E}[T_{x,1}] = \lim_{\alpha \to 0^+} \frac{1 - \tilde{F}_{x,1}(\alpha)}{\alpha} < \infty$.

From the above

$$V_\alpha(x) = \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[\alpha\text{-cost from } x \text{ to state } (1, \mathrm{SO}, 0) \text{ in } T_{x,1}] + \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}] V_\alpha(1, \mathrm{SO}, 0).$$

Note that $\mathbb{E}_x^{\pi_\alpha^{(\infty)}}[\alpha\text{-cost from } x \text{ to state } (1, \mathrm{SO}, 0) \text{ in } T_{x,1}]$ is equal to: (1) the expected discounted cost incurred directly in state x, which is upper bounded by c_{cm}, (2) the total expected discounted cost of all the SOs that occur in time $T_{x,1}$, which is upper bounded by $c_{\mathrm{cm}} \tau_{x,1}$, (3) the total expected discounted cost of all the USOs that occur in time $T_{x,1}$, which is upper bounded by $c_{\mathrm{cm}} \lambda \tau_{x,1}$, and (4) the total expected discounted cost of all the corrective maintenance opportunities that occur in time $T_{x,1}$, which is upper bounded by $c_{\mathrm{cm}}(\mu_1 + \mu_2)\tau_{x,1}$. All in all,

$$\mathbb{E}_x^{\pi_\alpha^{(\infty)}}[\alpha\text{-cost from } x \text{ to state } (1, \mathrm{SO}, 0) \text{ in } T_{x,1}] \leq c_{\mathrm{cm}}(1 + \tau_{x,1} + (\lambda + \mu_1 + \mu_2)\tau_{x,1}).$$

Then, straightforward computations yield

$$\begin{aligned}
\left|V_\alpha(x) - V_\alpha(1, \mathrm{SO}, 0)\right| &= \Big|\mathbb{E}_x^{\pi_\alpha^{(\infty)}}[\alpha\text{-cost from } x \text{ to state } (1, \mathrm{SO}, 0) \text{ in } T_{x,1}] \\
&\qquad + \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}] V_\alpha(1, \mathrm{SO}, 0) - V_\alpha(1, \mathrm{SO}, 0)\Big| \\
&\leq \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[\alpha\text{-cost from } x \text{ to state } (1, \mathrm{SO}, 0) \text{ in } T_{x,1}] \\
&+ \left|1 - \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}]\right| V_\alpha(1, \mathrm{SO}, 0) \\
&\leq c_{\mathrm{cm}}(1 + \tau_{x,1} + (\lambda + \mu_1 + \mu_2)\tau_{x,1}) \\
&+ \left(1 - \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}]\right) V_\alpha(1, \mathrm{SO}, 0).
\end{aligned}$$

Similarly, for $z = (i', j', t') \in \mathcal{S}$,

$$\begin{aligned}
\left|V_\alpha(z) - V_\alpha(1, \mathrm{SO}, 0)\right| &\leq c_{\mathrm{cm}}(1 + \tau_{x,1} + (\lambda + \mu_1 + \mu_2)\tau_{z,1}) \\
&+ \left(1 - \mathbb{E}_z^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{z,1}}]\right) V_\alpha(1, \mathrm{SO}, 0).
\end{aligned}$$

Then,

$$\begin{aligned}\big|V_\alpha(x) - V_\alpha(z)\big| &\le \big|V_\alpha(x) - V_\alpha(1,\mathrm{SO},0)\big| + \big|V_\alpha(z) - V_\alpha(1,\mathrm{SO},0)\big| \\ &\le c_{\mathrm{cm}}(2 + \tau_{x,1} + \tau_{z,1} + (\lambda+\mu_1+\mu_2)(\tau_{x,1}+\tau_{z,1})) \\ &\quad + \left(1 - \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}] + 1 - \mathbb{E}_z^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{z,1}}]\right) V_\alpha(1,\mathrm{SO},0).\end{aligned}$$

Combining the above with Step 1 yields

$$\begin{aligned}|V_\alpha(x) - V_\alpha(z)| &\le c_{\mathrm{cm}}(2 + \tau_{x,1} + \tau_{z,1} + (\lambda+\mu_1+\mu_2)(\tau_{x,1}+\tau_{z,1})) \\ &\quad + \left(1 - \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}] + 1 - \mathbb{E}_z^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{z,1}}]\right) \frac{\frac{\mu_1}{\mu_1+\alpha}}{1 - \frac{\mu_1}{\mu_1+\alpha}\frac{\mu_2}{\mu_2+\alpha}}.\end{aligned}$$

Lastly, note that

$$\left(1 - \mathbb{E}_x^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{x,1}}] + 1 - \mathbb{E}_z^{\pi_\alpha^{(\infty)}}[e^{-\alpha T_{z,1}}]\right) \frac{\frac{\mu_1}{\mu_1+\alpha}}{1 - \frac{\mu_1}{\mu_1+\alpha}\frac{\mu_2}{\mu_2+\alpha}} \le (\tau_{x,1}+\tau_{z,1})\frac{\mu_1\mu_2}{\mu_1+\mu_2},$$

which yields

$$|V_\alpha(x) - V_\alpha(z)| \le c_{\mathrm{cm}}\left(2 + \left(\lambda + \mu_1 + \mu_2 + 1 + \frac{\mu_1\mu_2}{\mu_1+\mu_2}\right)(\tau_{x,1}+\tau_{z,1})\right).$$

7.A.3 Proof of Step 5

To prove this step, we follow to a large extent the approach in (Bhattacharya and Majumdar, 1989, Theorem 3.2 (c)–(d)). Consider the average optimality equality (7.23), this yields for an arbitrary policy π,

$$V(X_k) \le c(X_k;a_k) + \mathbb{E}_x^\pi[V(X_{k+1}) \,|\, X_k, a_k] - g\,\tau(X_k;a_k),\ k = 0,1,\ldots,$$

which can be equivalently written as

$$c(X_k;a_k) \ge g\,\tau(X_k;a_k) + V(X_k) - \mathbb{E}_x^\pi[V(X_{k+1}) \,|\, X_k, a_k],\ k = 0,1,\ldots.$$

Taking expectations on both sides one gets

$$\mathbb{E}_x^\pi[c(X_k;a_k)] \ge g\,\mathbb{E}_x^\pi[\tau(X_k;a_k)] + \mathbb{E}_x^\pi[V(X_k)] - \mathbb{E}_x^\pi[V(X_{k+1})],\ k = 0,1,\ldots.$$

Summing both sides of the above equation over $k = 0, 1, ..., N-1$, and dividing by $\mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} \tau(X_k; a_k)\right]$ one has

$$\frac{\mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} c(X_k; a_k)\right]}{\mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} \tau(X_k; a_k)\right]} \geq g + \frac{V(x) - \mathbb{E}_x^\pi[V(X_N)]}{\mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} \tau(X_k; a_k)\right]}. \tag{7.27}$$

Note that as $N \to \infty$, for $x = (i, j, t)$,

$$\frac{t + (N-1)\tau}{1 + \lambda\tau + \frac{\mu_1\mu_2}{\mu_1+\mu_2}\tau} \leq \mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} \tau(X_k; a_k)\right] . \leq \frac{t + (N-1)\tau}{1 + \lambda\tau + (\mu_1 + \mu_2)\tau}.$$

As such, $\mathbb{E}_x^\pi \left[\sum_{k=0}^{N-1} \tau(X_k; a_k)\right]$ is bounded from below for large values of N. Taking $\limsup_{N\to\infty}$ on both sides of Equation (7.27) yields $J(x, \pi) \geq g$.

Since, for $\pi^{*(\infty)}$, the above analysis holds with an equality, it is evident that $J(x, \pi^{*(\infty)}) = g$. Note that g is an arbitrary limit point of $\alpha V_\alpha(x)$ as $\alpha \to 0^+$. Furthermore, since $\alpha|V_\alpha(x) - V_\alpha(z)| \to 0$ as $\alpha \to 0^+$ for all x and for all z, it is now evident that $g = \limsup_{\alpha\to 0^+} \alpha V_\alpha(x)$ for all $x \in \mathcal{S}$.

To complete the proof, we need to show that the deterministic stationary policy is optimal. To this purpose, note that for $v_\alpha(x) = V_\alpha(x) - V_\alpha(z)$, for some arbitrary choice of $z \in \mathcal{S}$, the average optimality equation assumes the form

$$v_\alpha(x) + (1 - \tilde{F}_x^{\pi^*}(\alpha))V_\alpha(z) = \min_{a\in\mathcal{A}_x} \Big\{c(x; a) + \int_s \int_y e^{-\alpha s} v(y) \mathcal{P}_x^a(\mathrm{d}y \,|\, s)\, \mathrm{d}F_x^a(s)\Big\}.$$

Moreover,

$$\limsup_{\alpha\to 0^+} (1 - \tilde{F}_x^{\pi^*}(\alpha))V_\alpha(z) = \limsup_{a\to 0^+} \alpha V_\alpha(z) \lim_{a\to 0^+} \frac{1 - \tilde{F}_x^{\pi^*}(\alpha)}{\alpha} = g\tau(x; \pi^*).$$

So there exists a sequence of discounting factors $\alpha(n) \to 0^+$ such that

$$\limsup_{\alpha(n)\to 0^+} (1 - \tilde{F}_x^{\pi^*}(\alpha(n)))V_{\alpha(n)}(z) = g\tau(x; \pi^*)$$

and

$$h(x) = \liminf_{n\to\infty} v_{\alpha(n)}(x)$$

yielding

$$v_{\alpha(n)}(x)+(1-\tilde{F}_x^{\pi^*}(\alpha(n)))V_{\alpha(n)}(z) = \min_{a\in\mathcal{A}_x}\Big\{c(x;a)+\int_s\int_y e^{-\alpha(n)s}v(y)\mathcal{P}_x^a(\mathrm{d}y\,|\,s)\,\mathrm{d}F_x^a(s)\Big\}.$$

Let $\pi_{\alpha(n)}(x) \in \mathcal{A}_x$ be the policy that the above equation attains the minimum, so that

$$v_{\alpha(n)}(x) + (1 - \tilde{F}_x^{\pi^*}(\alpha(n)))V_{\alpha(n)}(z) = c(x;\pi_n)+ \int_s\int_y e^{-\alpha(n)s}v(y)\mathcal{P}_x^{\pi_{\alpha(n)}}(\mathrm{d}y\,|\,s)\,\mathrm{d}F_x^{\pi_{\alpha(n)}}(s)\Big\}.$$

Taking the limit as $n \to \infty$ yields

$$h(x) + g\tau(x;\pi^*) = c(x;\pi^*) + \int_y h(y)\mathcal{P}_x^{\pi^*}(\mathrm{d}y).$$

This proves that π^* is optimal.

7.B. Average cost equalities – Bellman equations

We proceed writing down the average cost equalities for the model at hand, cf. Proposition 7.3. More concretely, for $t \in [0,\tau)$, let $V(i,j,t)$ be the value function when the state of the system is $(i,j,t) \in \mathcal{S}$. The average optimality equations read as follows:

$$\begin{aligned} V(2,\mathrm{SC},t) &= 0 - g\int_0^t e^{-(\mu_1+\lambda)x}\,\mathrm{d}x + V(1,\mathrm{SO},0)\int_t^\infty(\mu_1+\lambda)e^{-(\mu_1+\lambda)x}\,\mathrm{d}x \\ &\quad + \int_0^t\left(\frac{\mu_1}{\mu_1+\lambda}V(1,\mathrm{SC},t-x) + \frac{\lambda}{\mu_1+\lambda}V(1,\mathrm{USO},t-x)\right)(\mu_1+\lambda)e^{-(\mu_1+\lambda)x}\,\mathrm{d}x \\ &= e^{-(\mu_1+\lambda)t}\left(\int_0^t\big(\mu_1V(1,\mathrm{SC},y)+\lambda V(1,\mathrm{USO},y)-g\big)\,e^{(\mu_1+\lambda)y}\,\mathrm{d}y + V(1,\mathrm{SO},0)\right), \end{aligned} \quad (7.28)$$

$$\begin{aligned} V(1,\mathrm{SC},t) &= c_c \\ &\quad + e^{-(\mu_2+\lambda)t}\left(\int_0^t\big(\mu_2V(2,\mathrm{SC},y)+\lambda V(2,\mathrm{USO},y)-g\big)\,e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\right), \end{aligned}$$

$$V(2,\mathrm{USO},t) = \min\Bigg\{ c_p^{\mathrm{uso}} + e^{-(\mu_2+\lambda)t}\Bigg(\int_0^t \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\Bigg);$$
$$e^{-(\mu_2+\lambda)t}\Bigg(\int_0^t \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\Bigg)\Bigg\}, \quad (7.29)$$

$$V(2,\mathrm{SO},0) = \min\Bigg\{ c_p^{\mathrm{so}} + e^{-(\mu_2+\lambda)\tau}\left(\int_0^\tau \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\right);$$
$$e^{-(\mu_2+\lambda)\tau}\left(\int_0^\tau \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\right)\Bigg\}, \quad (7.30)$$

$$V(1,\mathrm{USO},t) = \min\Bigg\{ c_p^{\mathrm{uso}} + pe^{-(\mu_2+\lambda)t}\left(\int_0^t \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\right)$$
$$+ qe^{-(\mu_1+\lambda)t}\left(\int_0^t \big(\mu_1 V(1,\mathrm{SC},y) + \lambda V(1,\mathrm{USO},y) - g\big)\, e^{(\mu_1+\lambda)y}\,\mathrm{d}y + V(1,\mathrm{SO},0)\right);$$
$$e^{-(\mu_1+\lambda)t}\left(\int_0^t \big(\mu_1 V(1,\mathrm{SC},y) + \lambda V(1,\mathrm{USO},y) - g\big)\, e^{(\mu_1+\lambda)y}\,\mathrm{d}y + V(1,\mathrm{SO},0)\right)\Bigg\}, \quad (7.31)$$

$$V(1,\mathrm{SO},0) = \min\Bigg\{ c_p^{\mathrm{so}} + pe^{-(\mu_2+\lambda)\tau}\left(\int_0^\tau \big(\mu_2 V(2,\mathrm{SC},y) + \lambda V(2,\mathrm{USO},y) - g\big)\, e^{(\mu_2+\lambda)y}\,\mathrm{d}y + V(2,\mathrm{SO},0)\right)$$
$$+ qe^{-(\mu_1+\lambda)\tau}\left(\int_0^\tau \big(\mu_1 V(1,\mathrm{SC},y) + \lambda V(1,\mathrm{USO},y) - g\big)\, e^{(\mu_1+\lambda)y}\,\mathrm{d}y + V(1,\mathrm{SO},0)\right);$$
$$e^{-(\mu_1+\lambda)\tau}\left(\int_0^\tau \big(\mu_1 V(1,\mathrm{SC},y) + \lambda V(1,\mathrm{USO},y) - g\big)\, e^{(\mu_1+\lambda)y}\,\mathrm{d}y + V(1,\mathrm{SO},0)\right)\Bigg\}. \quad (7.32)$$

In this paragraph, we explain in detail how Equation (7.28) is obtained. State $(2,\mathrm{SC},t)$ is associated with only the decision "do nothing". Therefore, there is no minimum operator appearing on the right hand side of Equation (7.28) and the

corresponding cost is equal to zero. For the other terms appearing on the right hand side of Equation (7.28), it suffices to note that there are three possible evolutions in terms of the state of the system: either an SO or an SC or a USO, where the time till the next SO is equal to t, while the times till the SC and USO are Exponentially distributed with rates μ_1 and λ, respectively. In particular, the expected sojourn time of the semi-MDP in state $(2, \mathrm{SC}, t)$ can be calculated as the expectation of the minimum of a deterministic time t and two Exponentially distributed times, which can be easily verified to be equal to $\int_0^t e^{-(\mu_1+\lambda)x}\,\mathrm{d}x$.

$$\begin{aligned} s(2,\mathrm{SC},t) &= t\int_t^\infty (\mu_1+\lambda)e^{-(\mu_1+\lambda)x}\,\mathrm{d}x + \int_0^t x(\mu_1+\lambda)e^{-(\mu_1+\lambda)x}\,\mathrm{d}x \\ &= \int_0^t e^{-(\mu_1+\lambda)x}\,\mathrm{d}x. \end{aligned} \tag{7.33}$$

The set of optimality equations for the remaining states can be obtained using very similar arguments. Note that in Equations (7.29)–(7.32), inside the minimum, the left term corresponds to the action 'perform preventive maintenance', while the right terms correspond to the action 'do nothing'.

We observe that, since $c_{\mathrm{pm}}^{\mathrm{so}}, c_{\mathrm{pm}}^{\mathrm{uso}} > 0$ and $p + q = 1$, Equations (7.29) and (7.30) yield that it is never optimal to perform preventive maintenance in state 2 in both USOs and SOs, respectively.

We define the following auxiliary functions, for $t \in [0, \tau)$ and $i \in \{1, 2\}$,

$$\begin{aligned} F_i(t) \triangleq e^{-(\mu_i+\lambda)t}\Bigg(\int_0^t \left(\mu_i V(i,\mathrm{SC},y) + \lambda V(i,\mathrm{USO},y) - g\right) e^{(\mu_i+\lambda)y}\,\mathrm{d}y \\ + V(i,\mathrm{SO},0)\Bigg), \end{aligned} \tag{7.34}$$

so that Equations (7.28)-(7.32) reduce to

$$V(1,\mathrm{SC},t) = c_{\mathrm{cm}} + F_2(t), \quad V(2,\mathrm{SC},t) = F_1(t), \quad t \in [0,\tau), \tag{7.35}$$

$$V(i,\mathrm{USO},t) = \min\left\{c_{\mathrm{pm}}^{\mathrm{uso}} + pF_2(t) + qF_i(t), F_i(t)\right\}, \quad i \in \{1,2\}, t \in [0,\tau), \tag{7.36}$$

$$V(i,\mathrm{SO},0) = \min\left\{c_{\mathrm{pm}}^{\mathrm{so}} + pF_2(\tau) + qF_i(\tau), F_i(\tau)\right\}, \quad i \in \{1,2\}, t \in [0,\tau). \tag{7.37}$$

7.C. Proof of Theorem 7.1

PROOF: We distinguish four cases, each corresponding to a different set of actions. Case (i): $F_1(\tau) - F_2(\tau) \leq \frac{c_{\mathrm{pm}}^{\mathrm{so}}}{p}$; Case (ii): $\frac{c_{\mathrm{pm}}^{\mathrm{so}}}{p} < F_1(\tau) - F_2(\tau) < \frac{c_{\mathrm{pm}}^{\mathrm{uso}}}{p}$; Case (iii): $\frac{c_{\mathrm{pm}}^{\mathrm{uso}}}{p} < F_1(\tau) - F_2(\tau)$; Case (iv): $F_1(\tau) - F_2(\tau) = \frac{c_{\mathrm{pm}}^{\mathrm{uso}}}{p}$.

Case (i): In state $(2, \mathrm{SO}, 0)$, it is optimal to not perform preventive maintenance. Furthermore, from the assumption

$$F_1(\tau) - F_2(\tau) \leq \frac{c_{\mathrm{pm}}^{\mathrm{so}}}{p} \tag{7.38}$$

and Equation (7.37) for $i = 1$, it becomes evident that it is also optimal to not perform preventive maintenance in state $(1, \mathrm{SO}, 0)$. Since the function $F_1(t) - F_2(t)$ is, by definition, a continuous function in $t \in [0, \tau]$, $c_{\mathrm{pm}}^{\mathrm{so}} < c_{\mathrm{pm}}^{\mathrm{uso}}$, and taking into account Equation (7.38), it is evident that there exists an $\varepsilon > 0$, such that

$$F_1(t) - F_2(t) \leq \frac{c_{\mathrm{pm}}^{\mathrm{uso}}}{p}, \text{ for all } t \in (\tau - \epsilon, \tau]. \tag{7.39}$$

Equation (7.39), in light of Equation (7.36), implies that if the elapsed time from the SO is less than ϵ, then, under the assumption it is optimal to not perform preventive maintenance on the system in state $(i, \mathrm{SO}, 0)$, it is also not optimal to perform preventive maintenance at a USO. In this case, for $t \in (\tau - \epsilon, \tau]$, we have that $V(1, \mathrm{USO}, t) = F_1(t)$ and $V(2, \mathrm{USO}, t) = F_2(t)$, cf. Equation (7.36). Taking the derivative with respect to t in Equation (7.34) and substituting the above obtained values for $V(1, \mathrm{USO}, t)$ and $V(2, \mathrm{USO}, t)$ yields, for $t \in (\tau - \epsilon, \tau]$,

$$\begin{aligned} F_1'(t) - F_2'(t) &= -(\mu_1 + \lambda)F_1(t) + \mu_1 V(1, \mathrm{SC}, t) + \lambda V(1, \mathrm{USO}, t) \\ &\quad + (\mu_2 + \lambda)F_2(t) - \mu_2 V(2, \mathrm{SC}, t) - \lambda V(2, \mathrm{USO}, t) \\ &= -(\mu_1 + \lambda)F_1(t) + \mu_1(c_{\mathrm{cm}} + F_2(t)) + \lambda F_1(t) \\ &\quad + (\mu_2 + \lambda)F_2(t) - \mu_2 F_2(t) - \lambda F_2(t), \\ &= -(\mu_1 + \mu_2)F_1(t) + (\mu_1 + \mu_2)F_2(t) + \mu_1 c_{\mathrm{cm}}. \end{aligned}$$

The solution to the above differential equation reads, for $t \in (\tau - \epsilon, \tau]$,

$$F_1(t) - F_2(t) = \frac{\mu_1 c_{\mathrm{cm}}}{\mu_1 + \mu_2} + \left(F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\mathrm{cm}}}{\mu_1 + \mu_2} \right) e^{(\mu_1 + \mu_2)(\tau - t)}. \tag{7.40}$$

If $F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\text{cm}}}{\mu_1+\mu_2} \neq 0$, it follows that, for $t \in (\tau - \epsilon, \tau]$, the function $F_1(t) - F_2(t)$ is strictly monotone. In this case, by extending the previous analysis to the entire domain, which would maintain the strict monotonicity of the function $F_1(t) - F_2(t)$, we would reach a contradiction: For $t = 0$, Equation (7.34) yields $F_1(0) = V(1, \text{SO}, 0) \overset{(7.37)}{=} F_1(\tau)$ and $F_2(0) = V(2, \text{SO}, 0) \overset{(7.37)}{=} F_2(\tau)$, where $\overset{(\cdot)}{=}$ denotes that the equality follows from Equation $(\cdot)$. We thus have

$$F_1(0) - F_2(0) = F_1(\tau) - F_2(\tau). \tag{7.41}$$

Due to (7.41), it is evident that $F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\text{cm}}}{\mu_1+\mu_2} = 0$, thus the function $F_1(t) - F_2(t)$ satisfying Equation (7.40) is a constant function, i.e.

$$F_1(t) - F_2(t) = \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2}, \ t \in (0, \tau]. \tag{7.42}$$

Combining Equation (7.38) with Equation (7.42) leads to the optimality condition for Case (i). That is, if

$$\mu_1 c_{\text{cm}} \leq (\mu_1 + \mu_2) \frac{c_{\text{pm}}^{\text{so}}}{p},$$

we do not perform preventive maintenance at any opportunity.

Case (ii): In state $(2, \text{SO}, 0)$, similarly to the previous case, it is optimal to not perform preventive maintenance. However, from the assumption

$$\frac{c_{\text{pm}}^{\text{so}}}{p} < F_1(\tau) - F_2(\tau) < \frac{c_{\text{pm}}^{\text{uso}}}{p} \tag{7.43}$$

and Equation (7.37) for $i = 1$, it becomes evident that it is optimal to perform preventive maintenance on the system in state $(1, \text{SO}, 0)$. Similarly to Case (i), as $F_1(\tau) - F_2(\tau) < \frac{c_{\text{pm}}^{\text{uso}}}{p}$, there exists an $\varepsilon > 0$ for which (7.39) holds.

Repeating the same analysis as in Case (i), we can show that, for $t \in [0, \tau]$, the function $F_1(t) - F_2(t)$ satisfies Equation (7.40) and that it is a non-decreasing function if

$$F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} < 0. \tag{7.44}$$

However, for $t = 0$, we now have that

$$\begin{aligned} F_1(0) - F_2(0) &\overset{(7.34)}{=} V(1, SO, 0) - V(2, SO, 0) \\ &= c_{\text{pm}}^{\text{so}} + pF_2(\tau) + (1-p)F_1(\tau) - F_2(\tau) \\ &= c_{\text{pm}}^{\text{so}} + (1-p)(F_1(\tau) - F_2(\tau)). \end{aligned} \tag{7.45}$$

Combining (7.45) with (7.40) (on the domain $t \in [0, \tau]$) yields

$$\begin{aligned} c_{\text{pm}}^{\text{so}} + (1-p)(F_1(\tau) - F_2(\tau)) &= \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} + \left(F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} \right) e^{(\mu_1+\mu_2)(\tau)} \\ \Leftrightarrow \left(1 - p - e^{(\mu_1+\mu_2)(\tau)}\right) \left(F_1(\tau) - F_2(\tau)\right) &= \left(1 - e^{(\mu_1+\mu_2)(\tau)}\right) \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} - c_{\text{pm}}^{\text{so}} \\ \Leftrightarrow F_1(\tau) - F_2(\tau) &= \frac{\left(1 - e^{(\mu_1+\mu_2)\tau}\right) \frac{\mu_1 c_{\text{cm}}}{\mu_1+\mu_2} - c_{\text{pm}}^{\text{so}}}{1 - p - e^{(\mu_1+\mu_2)\tau}}. \end{aligned} \tag{7.46}$$

Combining Equations (7.43), (7.44), and (7.46) leads to the optimality condition for Case (ii). That is, if

$$(\mu_1 + \mu_2)\frac{c_{\text{pm}}^{\text{so}}}{p} < \mu_1 c_{\text{cm}} < \left(\frac{c_{\text{pm}}^{\text{uso}}}{p} - \frac{c_{\text{pm}}^{\text{uso}} - c_{\text{pm}}^{\text{so}}}{e^{(\mu_1+\mu_2)\tau} - 1} \right) (\mu_1 + \mu_2),$$

we perform preventive maintenance on the system if it is in state 1 at an SO but not at a USO.

Case (iii): In state $(2, \text{SO}, 0)$, similarly to the previous case, it is optimal to not perform preventive maintenance. However, from the assumption

$$F_1(\tau) - F_2(\tau) > \frac{c_{\text{pm}}^{\text{uso}}}{p} > \frac{c_{\text{pm}}^{\text{so}}}{p} \tag{7.47}$$

and Equation (7.37) for $i = 1$, it becomes evident that it is optimal to perform preventive maintenance on the system in state $(1, \text{SO}, 0)$. Along the lines of the previous cases, as $F_1(\tau) - F_2(\tau) > \frac{c_{\text{pm}}^{\text{uso}}}{p}$, there exists an $\varepsilon > 0$ for which

$$F_1(t) - F_2(t) \geq \frac{c_{\text{pm}}^{\text{uso}}}{p}, \text{ for all } t \in (\tau - \epsilon, \tau]. \tag{7.48}$$

In this case, for $t \in (\tau - \epsilon, \tau]$, we have that $V(1, \text{USO}, t) = c_{\text{pm}}^{\text{uso}} + pF_2(t) + (1 - p)F_1(t)$ and $V(2, \text{USO}, t) = F_2(t)$ (cf. Equation (7.36)). Taking a derivative with respect to t in (7.34) and substituting the above obtained values for $V(1, \text{USO}, t)$

and $V(2, \mathrm{USO}, t)$ yields, for $t \in (\tau - \varepsilon, \tau]$,

$$\begin{aligned} F_1'(t) - F_2'(t) =& -(\mu_1 + \lambda) F_1(t) + \mu_1 V(1, \mathrm{SC}, t) + \lambda V(1, \mathrm{USO}, t) \\ &+ (\mu_2 + \lambda) F_2(t) - \mu_2 V(2, \mathrm{SC}, t) - \lambda V(2, \mathrm{USO}, t) \\ =& -(\mu_1 + \lambda) F_1(t) + \mu_1 (c_{\mathrm{cm}} + F_2(t)) \\ &+ \lambda (c_{\mathrm{pm}}^{\mathrm{uso}} + p F_2(t) + (1-p) F_1(t)) \\ &+ (\mu_2 + \lambda) F_2(t) - \mu_2 F_1(t) - \lambda F_2(t) \\ =& -(\mu_1 + \mu_2 + \lambda p)(F_1(t) - F_2(t)) + \mu_1 c_{\mathrm{cm}} \\ &+ \lambda c_{\mathrm{pm}}^{\mathrm{uso}}. \end{aligned} \tag{7.49}$$

The solution to the above differential equation reads for $t \in (\tau - \varepsilon, \tau]$,

$$\begin{aligned} F_1(t) - F_2(t) =& \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} \\ &+ \left(F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} \right) e^{(\mu_1 + \mu_2 + \lambda p)(\tau - t)}. \end{aligned} \tag{7.50}$$

Note that, if we assume that $F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} \geq 0$, then we can extend (7.50) on the entire domain $t \in [0, \tau]$, and the function $F_1(t) - F_2(t)$ is non-increasing. However, this is unfeasible. Note that, for $t = 0$, Equation (7.34) yields $F_1(0) = V(1, \mathrm{SO}, 0) \overset{(7.37)}{=} c_{\mathrm{pm}}^{\mathrm{so}} + p F_2(\tau) + q F_1(\tau)$ and $F_2(0) = V(2, \mathrm{SO}, 0) \overset{(7.37)}{=} F_2(\tau)$, thus

$$\begin{aligned} F_1(0) - F_2(0) = c_{\mathrm{pm}}^{\mathrm{so}} + q(F_1(\tau) - F_2(\tau)) &\geq F_1(\tau) - F_2(\tau) \\ \Leftrightarrow \; F_1(\tau) - F_2(\tau) &\leq \frac{c_{\mathrm{pm}}^{\mathrm{so}}}{p}, \end{aligned} \tag{7.51}$$

which contradicts Assumption (7.47). Due to this contradiction, it is necessary to assume that $F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} < 0$. This implies that the function $F_1(t) - F_2(t)$ is non-decreasing and we can extend (7.50) on the domain $t \in [t^*, \tau]$, where t^* is such that $F_1(t^*) - F_2(t^*) = \frac{c_{\mathrm{pm}}^{\mathrm{uso}}}{p}$, i.e. for $t \in [t^*, \tau]$,

$$\begin{aligned} F_1(t) - F_2(t) =& \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} \\ &+ \left(F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\mathrm{cm}} + \lambda c_{\mathrm{pm}}^{\mathrm{uso}}}{\mu_1 + \mu_2 + \lambda p} \right) e^{(\mu_1 + \mu_2 + \lambda p)(\tau - t)}. \end{aligned} \tag{7.52}$$

See Figure 7.8 for a visualization of $F_1(t) - F_2(t)$.

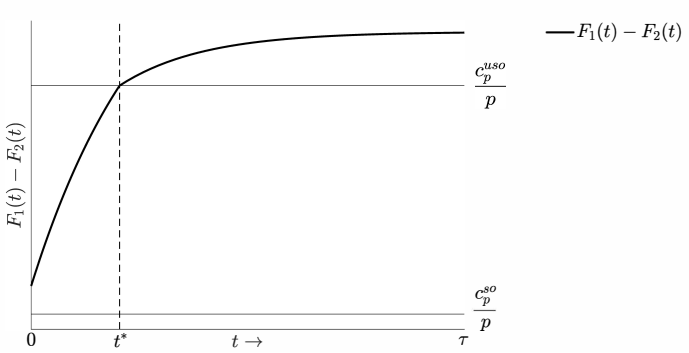


Figure 7.8 The case of maintaining a system at scheduled and unscheduled opportunities in $t \in [t^*, \tau)$.

From the definition of t^*, and the continuity of $F_1(t) - F_2(t)$, it follows that there exists an $\varepsilon > 0$, such that

$$F_1(t) - F_2(t) \leq \frac{c_{\text{pm}}^{\text{uso}}}{p}, \text{ for all } t \in (t^* - \epsilon, t^*]. \tag{7.53}$$

Note that if one were to assume that $F_1(t) - F_2(t) \geq \frac{c_{\text{pm}}^{\text{uso}}}{p}$, for all $t \in (t^* - \epsilon, t^*]$, then due to Equation (7.51), this would again contradict Assumption (7.47).

Now repeating the analysis performed in Case (i), albeit in a different domain, we can show that, for $t \in [0, t^*]$,

$$F_1(t) - F_2(t) = \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} + \left(F_1(t^*) - F_2(t^*) - \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} \right) e^{(\mu_1 + \mu_2)(t^* - t)}. \tag{7.54}$$

From the continuity of $F_1(t) - F_2(t)$ at $t = t^*$, we obtain

$$\begin{aligned} \frac{c_{\text{pm}}^{\text{uso}}}{p} =& \frac{\mu_1 c_{\text{cm}} + \lambda c_{\text{pm}}^{\text{uso}}}{\mu_1 + \mu_2 + \lambda p} \\ &+ \left(F_1(\tau) - F_2(\tau) - \frac{\mu_1 c_{\text{cm}} + \lambda c_{\text{pm}}^{\text{uso}}}{\mu_1 + \mu_2 + \lambda p} \right) e^{(\mu_1 + \mu_2 + \lambda p)(\tau - t^*)}. \end{aligned} \tag{7.55}$$

Furthermore, setting $t = 0$ in Equation (7.54) and using (7.51) yields

$$c_{\text{pm}}^{\text{so}} + (1-p)(F_1(\tau) - F_2(\tau)) = \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} + \left(\frac{c_{\text{pm}}^{\text{uso}}}{p} - \frac{\mu_1 c_{\text{cm}}}{\mu_1 + \mu_2} \right) e^{(\mu_1 + \mu_2) t^*}. \tag{7.56}$$

Note that Equations (7.55) and (7.56) form a system of two equations with two unknowns, which produce a unique solution for t^*, cf. Equation (7.1). Since $F_1(t) - F_2(t)$ is a continuous function throughout $[0, \tau)$, we can directly use the optimality condition for Case (ii) to state the optimality condition for this case. That is, if

$$\mu_1 c_{\text{cm}} > \left(\frac{c_{\text{pm}}^{\text{uso}}}{p} - \frac{c_{\text{pm}}^{\text{uso}} - c_{\text{pm}}^{\text{so}}}{e^{(\mu_1+\mu_2)\tau} - 1} \right) (\mu_1 + \mu_2),$$

we perform preventive maintenance on the system if it is in state 1 at an SO and at a USO for which the residual time until the next SO is in the interval $[\hat{t}, \tau)$, with $\hat{t} = \min\{\tau, \max\{0, t^*\}\}$.

Case (iv): This case follows evidently by performing again the steps of Case (iii) for $t^* = \tau$.

□

7.D. Proof of Theorem 7.2

PROOF: Similarly to the proof of Theorem 7.1, we need to make certain assumptions here regarding the actions at the given opportunities. In particular, we distinguish four cases, each corresponding to a different set of actions: Case (i): $F_1(\tau) - F_2(\tau) \leq \frac{c_{\text{pm}}^{\text{uso}}}{p}$; Case (ii): $\frac{c_{\text{pm}}^{\text{uso}}}{p} < F_1(\tau) - F_2(\tau) < \frac{c_{\text{pm}}^{\text{so}}}{p}$; Case (iii): $\frac{c_{\text{pm}}^{\text{so}}}{p} < F_1(\tau) - F_2(\tau)$; Case (iv): $F_1(\tau) - F_2(\tau) = \frac{c_{\text{pm}}^{\text{so}}}{p}$. The proof of this theorem is similar in structure to the proof of Theorem 7.1 and for this reason it is omitted. □

7.E. Proof of Theorem 7.4

PROOF: We first focus on the derivation of the cycle length appearing in the denominator of Equation (7.8). Observe that the length of a renewal cycle consists of the time the system spends in state 2 plus the time from the state-change $2 \to 1$ until the first successful maintenance. To this purpose, let CL denote the length of the part of the renewal cycle that the underlying stochastic process spends in state 1. Furthermore, let Y denote the random amount of time from a state-change $2 \to 1$ to the first SO, we then have for the probability density function of Y that

$$f_Y(y) = f_{T_{\mu_2}}(\tau - y | T_{\mu_2} < \tau),$$

which leads to Equation (7.13). Conditioning on Y, a renewal cycle can either end before the first SO, or at the first SO, or after the first SO. Hence, we have that the expected cycle length is equal to

$$\frac{1}{\mu_2} + \mathbb{E}\left[CL\,\mathbb{1}_{\{CL\leq Y\}}\right] + \mathbb{E}\left[CL\,\mathbb{1}_{\{CL>Y\}}\right]. \tag{7.57}$$

We first focus on deriving expressions for the individual expectations in Equation (7.57). Note that the first successful maintenance can be of type $j \in \{\mathrm{SC}, \mathrm{SO}, \mathrm{USO}\}$ and may occur in the interval $[t, t']$, this is in short denoted by $j\,[t, t']$. Thus, rewriting the first part in Equation (7.57) results in (cf. Equation (7.9))

$$\begin{aligned}\mathbb{E}\left[CL\;\mathbb{1}_{\{CL\leq Y\}}\right] =& \mathbb{E}\left[CL\mathbb{1}_{\{\mathrm{USO}[\tau-Y,\tau-\tilde{t}]\}}\right] + \mathbb{E}\left[CL\mathbb{1}_{\{\mathrm{SO}[\tau-Y,\tau]\}}\right] \\ &+ \mathbb{E}\left[CL\mathbb{1}_{\{\mathrm{CM}[\tau-Y,\tau]\}}\right].\end{aligned} \tag{7.58}$$

For the second expectation in Equation (7.57), observe that the length of this part can be further decomposed: first the system goes through a geometric number of intervals of length τ in which no successful maintenance activity takes place, after which the system enters the last interval in which the successful maintenance activity takes place. To this end, let p_u be the probability that there is no successful maintenance activity in an arbitrary interval between two SOs (including the SO with which this interval ends) after the state change $2 \to 1$, i.e.

$$p_u \triangleq (1-p)\mathbb{P}[T_{\mu_1} > \tau, T_{\lambda p} > \tau - \tilde{t}] = (1-p)e^{-\mu_1\tau-\lambda p(\tau-\tilde{t})} = (1-p)\mathbb{P}\left[\mathrm{SO}[0,\tau]\right].$$

We then have, from the memoryless property of T_{μ_1} and $T_{\lambda p}$,

$$\begin{aligned}&\mathbb{E}\left[CL\,\mathbb{1}_{\{CL>Y\}}\right] \\ &= (1-p)\mathbb{P}[\mathrm{SO}_{[\tau-Y,\tau]}]\left(\mathbb{E}\left[Y\right] + \sum_{k=0}^{\infty} p_u^k(1-p_u)\left(\mathbb{E}\left[CL\,\mathbb{1}_{\{Y+k\tau\leq CL\leq Y+(k+1)\tau\}}\right]\right)\right) \\ &= (1-p)\mathbb{P}[\mathrm{SO}_{[\tau-Y,\tau]}]\left(\mathbb{E}\left[Y\right] + \sum_{k=0}^{\infty} p_u^k(1-p_u)\left(k\tau + \mathbb{E}\left[CL'\,\mathbb{1}_{\{CL'\leq Y\}}\,|\,Y=\tau\right]\right)\right) \\ &= (1-p)\mathbb{P}[\mathrm{SO}_{[\tau-Y,\tau]}]\left(\mathbb{E}\left[Y\right] + \frac{\tau p_u}{1-p_u} + \mathbb{E}\left[CL'\,\mathbb{1}_{\{CL'\leq Y\}}\,|\,Y=\tau\right]\right),\end{aligned}$$

where $\mathbb{E}\left[CL'\,\mathbb{1}_{\{CL'\leq Y\}}\,|\,Y=\tau\right]$ is the expected length of the last part of the renewal cycle, i.e. the interval in which the successful maintenance activity takes place.

Analogously to Equation (7.58), we can further decompose $\mathbb{E}\left[CL'\,\mathbb{1}_{\{CL'\leq Y\}} \,|\, Y=\tau\right]$ by conditioning on the type of the successful maintenance activity with which it ends.

We are now left with defining the events that lead to $j\,[t,t']$, such that we can calculate the expectations in Equations (7.17)-(7.19). With respect to $\mathrm{SO}[\tau-y,\tau]$, observe that if $y\in[0,\tilde{t})$, $\mathbb{1}_{\{\mathrm{SO}[\tau-y,\tau]\}}$ is equal to 1 if $T_{\mu_1}>y$, since we do not take any USOs. If $y\in[\tilde{t},\tau]$, no successful USOs in $[\tau-y,\tau-\tilde{t}]$ can occur and $T_{\mu_1}>y$ for $\mathbb{1}_{\{\mathrm{SO}[\tau-y,\tau]\}}$ to be equal to 1. Combining this leads to Equation (7.14). Equations (7.15) and (7.16) are obtained along similar lines. Note that all expectations and probabilities only involve Exponentially distributed random variables. Consequently, closed-form expressions can be obtained using straightforward calculus. However, for the sake of brevity, we have chosen to provide one closed-form expression and omit the rest (which can be obtained analogously). For Equation (7.17), we have for $y>\tilde{t}$:

$$
\begin{aligned}
&\mathbb{E}\left[CL\mathbb{1}_{\{\mathrm{USO}[\tau-y,\tau-\tilde{t}]\}}\right]\\
&=\mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{\mathrm{USO}[\tau-y,\tau-\tilde{t}]\}}\right]\\
&=\mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{T_{\lambda p}\leq\min\left\{y-\tilde{t},T_{\mu_1}\right\}\}}\right]\\
&=\int_0^{y-\tilde{t}}\mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{T_{\lambda p}\leq x\}}\right]\mu_1e^{-\mu_1x}\,\mathrm{d}x+\int_{y-\tilde{t}}^{\infty}\mathbb{E}\left[T_{\lambda p}\mathbb{1}_{\{T_{\lambda p}\leq y-\tilde{t}\}}\right]\mu_1e^{-\mu_1x}\,\mathrm{d}x\\
&=\int_0^{y-\tilde{t}}\int_0^{x}z\lambda pe^{-\lambda pz}\,\mathrm{d}z\mu_1e^{-\mu_1x}\,\mathrm{d}x+\int_{y-\tilde{t}}^{\infty}\int_0^{y-\tilde{t}}z\lambda pe^{-\lambda pz}\,\mathrm{d}z\mu_1e^{-\mu_1x}\,\mathrm{d}x\\
&=\frac{\lambda p}{\lambda p+\mu_1}\left(\frac{1-e^{-(\lambda p+\mu_1)(y-\tilde{t})}(1+(\lambda p+\mu_1)(y-\tilde{t}))}{\lambda p+\mu_1}\right).
\end{aligned}
$$

We now focus on the numerator of Equation (7.8), i.e. the expected cycle cost. To that end, let CC be the cost incurred in a renewal cycle. The analysis for the expected cycle cost, $\mathbb{E}\,[CC]$, is similar to the analysis of the expected cycle length. Again, we decompose the length of a renewal cycle into three parts (i.e. the interval after the state change until the first SO, the geometric number of intervals of length τ in which no successful maintenance activity takes place, and the last interval in which the successful maintenance activity takes place), and compute the conditional expected cycle costs in these parts (mainly consisting of costs incurred at unsuccessful maintenance activities). Thus,

$$
\mathbb{E}\,[CC]=\mathbb{E}\left[CC\,\mathbb{1}_{\{CL\leq Y\}}\right]+\mathbb{E}\left[CC\,\mathbb{1}_{\{CL>Y\}}\right]. \tag{7.59}
$$

We first focus on the first part in Equation (7.59) and condition further on the type of activity, which yields

$$\begin{aligned}\mathbb{E}\left[CC\,\mathbb{1}_{\{CL\leq Y\}}\right] =& \mathbb{E}\left[CC\,\mathbb{1}_{\{\mathrm{USO}[\tau-Y,\tau-\tilde{t}]\}}\right] + \mathbb{E}\left[CC\,\mathbb{1}_{\{\mathrm{SO}[\tau-Y,\tau]\}}\right] \\ &+ \mathbb{E}\left[CC\,\mathbb{1}_{\{\mathrm{CM}[\tau-Y,\tau]\}}\right].\end{aligned}$$

Analogous to the expected cycle length, the expected cost incurred during the geometric number of intervals of length τ, in which no successful maintenance activity takes place, is equal to

$$\sum_{k=0}^{\infty} p_u^k(1-p_u)k\left(\lambda(1-p)(\tau-\tilde{t})c_{\mathrm{pm}}^{\mathrm{uso}}+c_{\mathrm{pm}}^{\mathrm{so}}\right) = \frac{(\lambda(1-p)(\tau-\tilde{t})c_{\mathrm{pm}}^{\mathrm{uso}}+c_{\mathrm{pm}}^{\mathrm{so}})p_u}{1-p_u}.$$

Observe that the expected cost in the interval in which the successful maintenance activity takes place is composed of two parts regardless of the type of activity, i.e. the cost of the successful maintenance activity itself and the cost related to the unsuccessful USOs up to the successful maintenance activity (see Equations (7.20) - (7.22)). Again, all expectations and probabilities related to the costs only involve Exponentially distributed random variables, and again, for the sake of brevity, we have chosen to provide one closed-form expression and omit the rest (which can be obtained analogously). For Equation (7.21), we have

$$\mathbb{E}\left[CC\,\mathbb{1}_{\{\mathrm{SO}[\tau-y,\tau]\}}\right] = \left(c_{\mathrm{pm}}^{\mathrm{so}}+\lambda(1-p)c_{\mathrm{pm}}^{\mathrm{uso}}\max\left\{y-\tilde{t},0\right\}\right)\mathbb{P}\left[\mathrm{SO}[\tau-y,\tau]\right],$$

with

$$\begin{aligned}\mathbb{P}\left[\mathrm{SO}[\tau-y,\tau]\right] &= \mathbb{P}\left[T_{\mu_1}>y\right]\mathbb{1}_{\{y<\tilde{t}\}} + \mathbb{P}\left[T_{\lambda p}>y-\tilde{t}, T_{\mu_1}>y\right]\mathbb{1}_{\{y\geq\tilde{t}\}} \\ &= e^{-\mu_1 y}\,\mathbb{1}_{\{y<\tilde{t}\}} + e^{-(\mu_1 y+\lambda p(y-\tilde{t}))}\,\mathbb{1}_{\{y\geq\tilde{t}\}},\end{aligned}$$

which completes the proof. □

Chapter 8

Conclusion and future research

In this thesis, we have studied novel mathematical models for maintenance optimization, mainly motivated by the increasing and improved availability of real-time data that can be leveraged in such models. In Chapters 2-6, we have focused on the class of maintenance problems characterized by parameter uncertainty, while in Chapter 7, we have studied a maintenance model with parameter certainty. In the introductory chapter of this thesis, we have introduced a framework – in the form of a structured road map – to guide the development of mathematical models that integrate (i) learning from real-time data, and (ii) decision making, for maintenance models characterized by parameter uncertainty. In this last chapter of the thesis, we first summarize our work by revisiting this proposed framework and its application, and then identify interesting directions for future research.

8.1. Framework revisited

The proposed structured learning and decision making framework comprises five steps, see Figure 1.4 in the introductory chapter. In this section, we shortly summarize how each step was carried out in each individual chapter focused on parameter uncertainty. Since the maintenance models in these chapters differ significantly in virtually all aspects but parameter uncertainty, we demonstrate the unified applicability of our proposed framework.

Step 1: Identify conjugate pair for Bayesian inference

In Chapter 2, we have studied a compound Poisson model in which two parameters – of the arrival and jump process, respectively – need to be learned simultaneously from the same data. We showed that the joint posterior distribution of both parameters using the same observed data can be decomposed into two independent posterior distributions that take the same form as the prior distribution. This result allowed us to use conventional conjugate pairs for both parameters. Next, in Chapter 3, we used Bayesian regression theory to tractably infer the unknown drift of a Brownian motion using the Normal random variable as a prior distribution. We established in Chapter 4 that the Gamma distribution remains a conjugate prior for the Poisson distribution, even under condition-based production policies in which one controls the Poisson process. In Chapter 5, we used the same Gamma-Poisson model as in Chapter 4 but without controlling the deterioration. Finally, in Chapter 6, we considered a class of distributions (i.e., Newsboy distributions) famous for being the only class that preserves conjugacy (using a Gamma prior) under censored observations.

Step 2: Determine sufficient statistics

The set of sufficient statistics is usually a direct consequence – through the updating rules of each conjugate pair – of the previous step, but in some of our models it required a more careful treatment.

In Chapter 2, we assumed that the compounding distribution is a member of the non-negative exponential family. For this class, there is a general conjugate prior, but depending on the exact distribution, the sufficient statistic might differ. We restricted our analysis to members for which the sufficient statistic is linear in the data, so that we only needed to keep track of the current deterioration level, the number of shocks sustained, and the age of a component. In Chapter 3, the Bayesian regression framework implied that for the Brownian motion, we only needed to keep track of the current deterioration level and the age as sufficient statistics to infer the unknown drift. The required sufficient statistics also implied that in Chapter 2, we needed real-time information (i.e., continuous monitoring) because otherwise the number of shocks sustained cannot be measured, while in Chapter 3, one only needs to sample at discrete epochs. We showed in Chapter 4 that, in addition to the deterioration level and age of the component, one also needs to keep track of a statistic that captures the history of production decisions to infer the rate of the Poisson process. This is intuitive as these decisions impact the observed deterioration data. When pooling data stemming from multiple systems, we showed in Chapter 5 that one can simply

add all sufficient statistics (the ages and cumulative deterioration) for joint learning of a common, yet unknown Poisson rate. Finally, for the Newsboy distributions studied in Chapter 6, we needed to keep track of all observed lifetimes, and the number of uncensored observations to infer the unknown parameter of the lifetime distribution.

Step 3: Derive posterior predictive distribution and construct Bayesian MDP

In each chapter, we derived an analytical expression for the posterior predictive distribution as a function of the corresponding sufficient statistics, that we then used to construct the optimization problem.

In Chapter 2 and 3, we constructed an infinite-horizon discounted Bayesian Markov decision process (MDP). For the Bayesian MDP in Chapter 2, the restriction to linear sufficient statistics enabled us to keep the dimension of this Bayesian MDP to only 3 states. In Chapter 4, we studied production and maintenance decisions in continuous time. Integrating Bayesian learning directly into a continuous-time MDP is inherently complex and, even for numerical purposes, not tractable. Instead, we proposed a heuristic based on (i) the optimal policy when there is no parameter uncertainty, and (ii) the established Bayesian learning framework from the previous steps. This heuristic can effectively guide production decisions in continuous time when there is parameter uncertainty. In Chapter 5, we built a finite-horizon Bayesian MDP to explicitly take into account a finite lifespan over which multiple systems are maintained. Finally, in Chapter 6, we have constructed a finite-horizon Bayesian dynamic program (DP) instead of a discrete-time Bayesian MDP. The reason for this is that maintenance decisions in the model studied in this chapter are not made at equidistant epochs, but at random moments in time (a component is replaced after reaching its age-replacement threshold or upon failure, whichever comes first). Together with this Bayesian DP, we also proposed a heuristic that myopically integrates Bayesian learning.

Step 4: Analyze structural properties of the optimal policy

When establishing structural properties of the optimization problems, we proceeded as follows. We first studied stochastic ordering properties of the posterior predictive distribution with respect to its sufficient statistics. Here, the derived analytical expressions of the posterior predictive distributions from Step 3 served as input. We then showed that the value function of the optimization problem has certain monotonic properties with respect to the sufficient statistics using the stochastic ordering properties of the posterior predictive distribution. For this, we employed an

inductive argument on the well-known value iteration algorithm. Using the established properties of the value function, we then established structural properties of the optimal policy.

Specifically, in Chapter 2, we showed that the optimal policy is a control limit policy that is (i) non-decreasing in the age of a component, and (ii) depending on the number of shocks sustained. In Chapter 3, establishing structural properties of the proposed Bayesian MDP proved to be complex due to an absolute value operator which rendered the approach taken in Chapter 2 inapplicable. Instead, we established the equivalence between the original Bayesian MDP and a so-called folded representation, which allowed for structural analysis. Using this folded Bayesian MDP, we established the optimality of a bandwidth policy that is monotone in the age of the component. In Chapter 4, we characterized the optimal production control policy for the case without parameter uncertainty. By establishing monotonic properties with respect to the rate, we also gained insights into the behavior of our proposed heuristic for the case that this rate is in fact unknown. As we considered multiple systems in Chapter 5, the resulting Bayesian MDP suffered from the curse of dimensionality. As a remedy, we proved a decomposition result that established the equivalence between the original Bayesian MDP and multiple two-state MDPs with a binary action space, each focused on an individual system. Using this result, we were able to prove the optimality of a control limit policy for each system, where the control limit depends on the accumulated data of all systems. Finally in Chapter 6, we proved, again by first establishing stochastic ordering properties, some structural properties regarding the value function of the Bayesian DP. We also established that the proposed myopic policy is asymptotically optimal; that is, it almost surely learns the unknown parameter and converges to the optimal policy with full knowledge of the parameter.

Step 5: Compute optimal policy

The novel structural properties that we established in the previous step can be exploited to decrease the computational time complexity of numerically computing the optimal policy by employing existing algorithms that rely on such properties.

By computing optimal policies, we were able to compare these optimal policies with other policies in various settings, and evaluate the robustness of these optimal policies to conditions that violate their underlying model assumptions. This allowed us to numerically assess:

- the benefits of integrating learning and decision making (Chapter 2),

- the robustness of the optimal policy to noisy degradation data (Chapter 2) and its performance when degradation data is not relayed in real time (Chapter 2),
- the performance of the optimal policy when hyperparameters of the initial prior distributions need to be estimated from limited data (Chapter 2),
- the robustness of the optimal policy to misspecification of the initial prior distribution (Chapter 3),
- the optimality gaps with so-called clairvoyant Oracle policies that know the true values of unknown parameters (Chapters 2 - 4),
- the value of condition-based production policies and the benefits of integrating maintenance and production decisions (Chapter 4),
- the benefits of pooling data for the purpose of joint learning (Chapter 5), and
- the monotonic behavior of the optimal age-replacement threshold under censored learning when the number of replacements is finite (see Conjecture 6.1 in Chapter 6).

8.2. Future research

In the concluding sections of each individual chapter, we already addressed several possible suggestions for future research pertaining to the work in the corresponding chapter. In this last section, we outline some avenues for future research which are of interest for all of them, stretching well beyond each individual chapter.

Leaving the realm of conjugate pairs

The use of conjugate pairs in our framework facilitates a tractable way of integrating learning and decision making, and allows us to subsequently analyze structural properties of optimal policies. At the same time, however, it also limits this analysis to deterioration processes and lifetime distributions that are characterized by uncertain parameters that have such conjugacy properties. Future research could explore how our methodological framework can be extended to settings that lie outside conjugate families. In such extensions, Bayes' rule can still be applied to infer unknown parameters, but it will rely on tedious numerical integrations (see Equations (1.2) and (1.3) in Chapter 1) as opposed to easy updating rules that we constantly used throughout this thesis. As a direct consequence, the resulting optimization

problems that integrate this Bayesian learning will most likely be not tractable at all. We believe that these two obstacles – computational complexities associated with Bayesian learning and intractable optimization problems – can be overcome by applying novel techniques stemming from the area of machine learning, in particular from Bayesian reinforcement learning (RL) (Ghavamzadeh et al., 2015).

Leveraging analytical results

In this thesis, we have obtained a good understanding of structural properties of value functions and optimal policies in stylized settings with assumptions – such as conjugate families – that could be violated in practice. In RL, such assumptions could indeed be relaxed, and we could additionally integrate our knowledge of structural properties into the algorithmic design to speed-up the development of such RL algorithms. Kunnumkal and Topaloglu (2008), for instance, show how one can exploit structural properties of the underlying MDP to improve the convergence behavior of the well-known Q-learning algorithm. Jiang and Powell (2015) propose a provably convergent approximate dynamic programming algorithm (ADP) that explicitly exploits the monotonicity of value functions. Other examples that highlight such an approach can be found in the inventory management literature. For instance, the value function in both the canonical lost sales and dual sourcing inventory problem is known to be $L^\natural$-convex (Zipkin, 2008; Hua et al., 2015; Chen et al., 2018), and this property has been used to successfully develop ADP policies that learn the parameters of an $L^\natural$-convex function to approximate the value function in both settings (Sun et al., 2014; Chen and Yang, 2019). Along these lines, we believe that a promising avenue for future research can be found in leveraging our established analytical results to guide the development of RL algorithms that can be used in real-life applications.

Combining types of uncertainty

In the introductory chapter, we already mentioned that maintenance problems can be characterized by types of uncertainty other than parameter uncertainty, namely state and population uncertainty. Research has predominantly focused on these sources of uncertainty from an isolated point of view, though it seems very interesting to study models that integrate multiple types of uncertainty. In Chapter 2, we numerically studied such an integration, where the efficacy of our model was assessed in the case that there is also state uncertainty (i.e., when the deterioration state is only partially observable). Studying such a combination analytically will provide insights into how our results for the parameter uncertainty setting generalize when adding

state uncertainty. The combination of population and parameter uncertainty could lead to models that match practice more closely. An example is where components originate from distinct and heterogeneous sub-populations, each of them having a common, yet unknown parameter.

Like parameter uncertainty, population and state uncertainty in isolation can be tractably dealt with using Bayesian learning. The main challenge when extending our framework to combinations of uncertainty lies in keeping the joint Bayesian learning tractable. Indeed, when multiple layers of uncertainty in deterioration data exist, it may lead to identifiability issues pertaining to learning each individual parameter of each source of uncertainty from that data.

Multi-component systems

In this thesis, we have studied optimal maintenance policies for single components, yet real-life systems generally consist of multiple components. Implementing decisions induced by such single-component models will likely lead to suboptimal decisions on a system-level; that is, it is costly and inconvenient to shut down an entire system each time a single component within a multi-component system needs maintenance. Future research could focus on optimal maintenance policies from a system perspective, where the deterioration processes or lifetime distributions of individual components are characterized by parameter uncertainty. To that end, the work in this thesis can be viewed as a stepping stone towards building such multi-component models.

When moving towards multi-component models, there are a few concrete avenues to consider. First, in Chapter 7, we showed that using opportunities – arising due to failures of other components in a system – for performing maintenance can lead to significant cost reductions compared to treating components in isolation. It would be interesting to see to what extent our established results generalize when parameter uncertainty (e.g., through the rate at which opportunities arrive) is introduced.

Next, in multi-component systems there can be stochastic dependence between multiple components (i.e., correlation between deterioration processes or lifetime distributions), so that data stemming from each component can be used to jointly learn the unknown parameters of all components. The numerical study in Chapter 5 showed that such joint learning can be very beneficial, and extending it to multivariate processes that explicitly take into account stochastic dependence seems an interesting direction. One such direction that seems particularly promising is to consider a multivariate Brownian motion with unknown drifts – thereby building further on the work of Chapter 3 – where the covariance matrix then models the stochastic

dependence. For this case, it is known that Bayesian learning of the unknown drifts remains tractable (see, e.g., Gelman et al., 1995, Chapter 3), so that our proposed framework might be applicable as well. Finally, the performance of a multi-component system will be determined by the performance of the components and their configuration within the system, where series, parallel, k-out-of-N, and redundancy systems are well-known examples (Olde Keizer et al., 2017). From this perspective, it seems interesting to study our proposed models in such configurations, and to see to which extent parameter uncertainty on a component-level impacts optimal decisions on a system-level.

Integration with logistical support

Maintenance decision making is usually studied in isolation, whereas in practice it often interacts with other processes. In Chapter 3 and Chapter 4, we studied the interaction with process control and production control, respectively, but there are other areas that should be considered as well. We believe that particularly the integration with the management of logistical support needed for performing maintenance activities is an interesting avenue. When a maintenance activity is performed, it generally requires a spare part to execute a replacement and a service engineer to carry out the replacement. As a maintenance activity is (to a large degree) leading in deciding when spare parts and service engineers are required, we believe that such optimal spare parts provisioning policies and service engineer management policies can build on our proposed models.

When considering spare parts management and maintenance jointly, one interesting topic for future research is to see which impact parameter uncertainty on a component-level has on the optimal control of spare parts. On a conceptual level, we believe that as a decision maker gradually learns that components are deteriorating at a faster (lower) rate, she will stock more (less) spare parts to protect against prolonged system downtime due to part unavailability. Such condition-based spare parts provisioning policies that integrate (i) learning from deterioration data, and (ii) stocking decisions for spare parts are worthwhile to investigate.

One application in spare parts management that seems particularly promising is when additive manufacturing (AM, also known as 3D printing) is used to print spare parts. Asset owners are increasingly using AM as a strategy to reduce system downtime by quickly printing a part when a component fails. However, this technique is still very much evolving and as a result, the quality of printed spare parts can be highly variable leading to parameter uncertainty in the deterioration processes when such

parts are installed (Song and Zhang, 2020, 2021). Our proposed models in this thesis can be used to cope with this parameter uncertainty and effectively guide spare parts printing decisions in advance of a replacement.

The systems for which we have developed maintenance models in this thesis are usually geographically distributed in a so-called service network (e.g., medical equipment in hospitals in a certain country). The management of service engineers then boils down to dispatching and repositioning engineers in these service networks such that fast responses are realized whenever an engineer is needed to perform a maintenance activity (Drent et al., 2020c). As a decision maker learns the deterioration behavior of components, it might be beneficial to dynamically adjust the positioning and dispatching of service engineers based on this accumulated data. To that end, our proposed models can guide these decisions to ensure the presence of a service engineer whenever a component in a system needs to be replaced.

Bibliography

D. T. Abdul-Malak, J. P. Kharoufeh, and L. M. Maillart. Maintaining systems with heterogeneous spare parts. *Naval Research Logistics*, 66(6):485–501, 2019.

P. Afèche and B. Ata. Bayesian dynamic pricing in queueing systems with unknown delay cost characteristics. *Manufacturing & Service Operations Management*, 15 (2):292–304, 2013.

R. Ahmad and S. Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*, 63 (1):135–149, 2012.

S. Alaswad and Y. Xiang. A review on condition-based maintenance optimization models for stochastically deteriorating systems. *Reliability Engineering & System Safety*, 157:54–63, 2017.

M. Albano, E. Jantunen, G. Papa, and U. Zurutuza. *The MANTIS Book: Cyber Physical System Based Proactive Collaborative Maintenance*. River Publishers Series in Automation, Control and Robotics, 2019.

V. F. Araman and R. Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations Research*, 57(5):1169–1188, 2009.

A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M. Ghosh, and S. Marcus. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.

J. J. Arts and R. J. I. Basten. Design of multi-component periodic maintenance programs with single-component models. *IISE Transactions*, 50(7):606–615, 2018.

R. Austin and S. Muktita. Indonesian report on Lion Air crash finds numerous problems. *New York Times*, 2019. URL `https://www.nytimes.com/2019/10/25/world/asia/lion-air-crash-report.html`.

G. Aydin and S. Ziya. Pricing promotional products under upselling. *Manufacturing & Service Operations Management*, 10(3):360–376, 2008.

G. Aydin and S. Ziya. Personalized dynamic pricing of limited inventories. *Operations Research*, 57(6):1523–1531, 2009.

K. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31(9):1150–1160, 1985.

K. Azoury and J. Miyaoka. Optimal policies and approximations for a Bayesian linear regression inventory model. *Management Science*, 55(5):813–826, 2009.

R. D. Baker. Data-based modeling of the failure rate of repairable equipment. *Lifetime Data Analysis*, 7(1):65–83, 2001.

R. D. Baker and A. H. Christer. Review of delay-time OR modelling of engineering aspects of maintenance. *European Journal of Operational Research*, 73(3):407–422, 1994.

R. Bapat and S. Kochar. On likelihood-ratio ordering of order statistics. *Linear Algebra and Its Applications*, 199:281–291, 1994.

R. Barlow and L. Hunter. Optimum preventive maintenance policies. *Operations Research*, 8(1):90–100, 1960.

H. Bastani, D. Simchi-Levi, and R. Zhu. Meta dynamic pricing: Transfer learning across experiments. *Management Science*, 68(3):1865–1881, 2022.

M. Bell. The true cost of poor quality. Technical report, PDL Group, 2020.

R. Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8):716, 1952.

R. Bellman. A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4):221–229, 1956.

M. Ben-Daya and M. Rahim. Effect of maintenance on the economic design of $\bar{x}$-control chart. *European Journal of Operational Research*, 120(1):131–143, 2000.

S. Benjaafar, W. L. Cooper, and J. S. Kim. On the benefits of pooling in production-inventory systems. *Management Science*, 51(4):548–565, 2005.

A. Bensoussan, M. Cakanyildirim, A. Royal, and S. P. Sethi. Bayesian and adaptive controls for a newsvendor facing exponential demand. *Risk and Decision Analysis*, 1(4):197–210, 2009.

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.

O. Berman, D. Krass, and M. Mahdi Tajbakhsh. On the benefits of risk pooling in inventory management. *Production and Operations Management*, 20(1):57–71, 2011.

D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific, 3rd edition, 2007.

D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific, 4th edition, 2012.

R. N. Bhattacharya and M. Majumdar. Controlled semi-Markov models under long-run average rewards. *Journal of Statistical Planning and Inference*, 22(2):223–242, 1989.

P. Billingsley. *Probability and Measure*. John Wiley & Sons, 3rd edition, 1995.

C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

A. Bisi, M. Dada, and S. Tokdar. A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing & Service Operations Management*, 13(4):525–533, 2011.

G. R. Bitran and S. V. Mondschein. Periodic pricing of seasonal products in retailing. *Management Science*, 43(1):64–79, 1997.

E. Blackwell, T. Gambell, V. Marya, and C. Schmitz. The great re-make: Manufacturing for modern times. Technical report, McKinsey, 2017.

P. J. Boland, E. El-Neweihi, and F. Proschan. Applications of the hazard rate ordering in reliability and order statistics. *Journal of Applied Probability*, 31(1):180–192, 1994.

D. J. Braden and M. Freimer. Informational dynamics of censored observations. *Management Science*, 37(11):1390–1404, 1991.

P. Bremaud. *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, NY, 1st edition, 1980.

J. M. Calabrese. Optimal workload allocation in open networks of multiserver queues. *Management Science*, 38(12):1792–1802, 1992.

J. M. Calabrese. Bayesian process control for attributes. *Management Science*, 41(4): 637–645, 1995.

P. Cao, J. Li, and H. Yan. Optimal dynamic pricing of inventories with stochastic demand and discounted criterion. *European Journal of Operational Research*, 217 (3):580–588, 2012.

B. P. Carlin and T. A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC, 2nd edition, 2000.

J. Chakravorty and A. Mahajan. Sufficient conditions for the value function and optimal strategy to be even and quasi-convex. *IEEE Transactions on Automatic Control*, 63(11):3858–3864, 2018.

X. Chao, H. Chen, and S. Zheng. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. *Operations Research*, 57(1):82–93, 2009.

L. Chen. Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Operations Research*, 58(2):396–413, 2010.

L. Chen and E. L. Plambeck. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Operations Management*, 10(2):236–256, 2008.

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

W. Chen and H. Yang. A heuristic based on quadratic approximation for dual sourcing problem with general lead times and supply capacity uncertainty. *IISE Transactions*, 51(9):943–956, 2019.

X. Chen, X. Gao, and Z. Pang. Preservation of structural properties in optimization with decisions truncated by random variables and its applications. *Operations Research*, 66(2):340–357, 2018.

W. Chiu. Economic design of np charts for processes subject to a multiplicity of assignable causes. *Management Science*, 23(4):404–411, 1976.

A. Christer and W. Waller. Delay time models of industrial inspection maintenance problems. *Journal of the Operational Research Society*, 35(5):401–406, 1984.

A. H. Christer. Modelling inspection policies for building maintenance. *Journal of the Operational Research Society*, 33(8):723–732, 1982.

A. H. Christer. Developments in delay time analysis for modelling plant maintenance. *Journal of the Operational Research Society*, 50(11):1120–1137, 1999.

C. Coleman, S. Damodaran, M. Chandramouli, and E. Deuel. Making maintenance smarter: Predictive maintenance and the digital supply network. Technical report, Deloitte, 2017.

P. W. Coopers. Sensing the future of the internet of things. Digital IQ snapshot, 2014.

E. Covington. Hot spot burnout of tungsten filaments. *Journal of the Illuminating Engineering Society*, 2(4):372–380, 1973.

A. DasGupta. *Fundamentals of Probability: A First Course*. Springer Science & Business Media, 2010.

S. Dayanik and Ü. Gürler. An adaptive Bayesian replacement policy with minimal repair. *Operations Research*, 50(3):552–558, 2002.

B. De Jonge and P. A. Scarf. A review on maintenance optimization. *European Journal of Operational Research*, 285(3):805–824, 2020.

B. De Jonge, A. S. Dijkstra, and W. Romeijnders. Cost benefits of postponing time-based maintenance under lifetime distribution uncertainty. *Reliability Engineering & System Safety*, 140:15–21, 2015.

M. H. DeGroot. *Optimal Statistical Decisions*. John Wiley & Sons, 2005.

L. Deprez, K. Antonio, J. J. Arts, and R. N. Boute. Data-driven preventive maintenance for a heterogeneous machine portfolio. Working paper, 2022.

C. Derman. Optimal replacement and maintenance under Markovian deterioration with probability bounds on failure. *Management Science*, 9(3):478–481, 1963.

X. Ding, M. L. Puterman, and A. Bisi. The censored newsvendor and the optimal acquisition of information. *Operations Research*, 50(3):517–527, 2002.

S. Dolinšek, B. Šuštaršič, and J. Kopač. Wear mechanisms of cutting tools in high-speed cutting processes. *Wear*, 250(1-12):349–356, 2001.

C. Drent and J. J. Arts. Optimal control of stochastically deteriorating manufacturing systems. Working paper, 2022.

C. Drent and S. Kapodistria. Bayesian process control for critical systems. Working paper. Available at: https://www.researchgate.net/publication/353462766, 2021.

C. Drent and G. J. Van Houtum. Data pooling: Shared learning benefits in condition-based maintenance. Working paper, 2022.

C. Drent, S. Kapodistria, and J. A. C. Resing. Condition-based maintenance

policies under imperfect maintenance at scheduled and unscheduled opportunities. *Queueing Systems*, 93(3):269–308, 2019.

C. Drent, M. Drent, J. J. Arts, and S. Kapodistria. Real-time integrated learning and decision making for deteriorating systems. Working paper. Available at: https://www.researchgate.net/publication/344563348, 2020a.

C. Drent, S. Kapodistria, and O. Boxma. Censored lifetime learning: Optimal Bayesian age-replacement policies. *Operations Research Letters*, 48(6):827–834, 2020b.

C. Drent, M. Olde Keizer, and G. J. Van Houtum. Dynamic dispatching and repositioning policies for fast-response service networks. *European Journal of Operational Research*, 285(2):583–598, 2020c.

L. Du, M. Hu, and J. Wu. Sales effort management under all-or-nothing constraint. *Management Science*, forthcoming, 2021.

A. J. Duncan. The economic design of X charts used to maintain current control of a process. *Journal of the American Statistical Association*, 51(274):228–242, 1956.

I. Dursun, A. Akcay, and G. J. Van Houtum. Age-based maintenance under population heterogeneity: Optimal exploration and exploitation. *European Journal of Operational Research*, 301(3):1007–1020, 2022.

ECRI. Healthcare product comparison system. Technical report, ECRI Institute, Plymouth Meeting, PA, 2013.

A. 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.

G. D. Eppen. Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5):498–501, 1979.

V. F. Farias and B. Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

E. A. Feinberg. Constrained semi-Markov decision processes with average rewards. *Zeitschrift für Operations Research*, 39(3):257–288, 1994.

E. A. Feinberg and A. Shwartz. *Handbook of Markov Decision Processes: Methods and Applications*, volume 40. Springer Science & Business Media, 2012.

Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner. Monitoring wind turbine gearboxes. *Wind Energy*, 16(5):728–740, 2013.

B. Fox. Age replacement with discounting. *Operations Research*, 14(3):533–537, 1966.

B. Fox. Adaptive age replacement. *Journal of Mathematical Analysis and Applications*, 18(2):365–376, 1967.

G. Gallego and G. Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

G. Gallego and G. Van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41, 1997.

N. Gebraeel, M. Lawley, R. Li, and J. Ryan. Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37(6): 543–557, 2005.

A. Gelman, J. B. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall, 1995.

M. Ghavamzadeh, S. Mannor, J. Pineau, and A. Tamar. Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, 8(5-6):359–483, 2015.

J. Ghosh, M. Delampady, and T. Samanta. *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Science & Business Media, 2007.

J. C. Gittins and K. Glazebrook. On Bayesian models in stochastic scheduling. *Journal of Applied Probability*, 14(3):556–565, 1977.

J. C. Gittins and D. M. Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

X. Guo and O. Hernández-Lerma. *Continuous-Time Markov Decision Processes: Theory and Applications*. Springer-Verlag, 2009.

V. Gupta and N. Kallus. Data pooling in stochastic optimization. *Management Science*, 68(3):1595–1615, 2022.

J. M. Harrison, N. B. Keskin, and A. Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58 (3):570–586, 2012.

O. Hernández-Lerma and J. B. Lasserre. *Further Topics on Discrete-Time Markov Control Processes*, volume 42. Springer Science & Business Media, 2012.

S. Hu, X. Hu, and Q. Ye. Optimal rebate strategies under dynamic pricing. *Operations Research*, 65(6):1546–1561, 2017.

Z. Hua, Y. Yu, W. Zhang, and X. Xu. Structural properties of the optimal policy for dual-sourcing systems with general lead times. *IIE Transactions*, 47(8):841–850, 2015.

K. Huang and J. Mi. Applications of likelihood ratio order in Bayesian inferences. *Probability in the Engineering and Informational Sciences*, 34(1):1–13, 2020.

V. Insinna and G. Ziezuliwicz. Investigation blames Air Force and Navy for systemic failures in fatal Marine Corps C-130 crash that killed 16, December 2018.

A. K. S. Jardine and A. H. C. Tsang. *Maintenance, Replacement, and Reliability: Theory and Applications*. CRC Press, 2005.

A. K. S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

A. Jaśkiewicz. An approximation approach to ergodic semi-Markov control processes. *Mathematical Methods of Operations Research*, 54(1):1–19, 2001.

A. Jaśkiewicz. On the equivalence of two expected average cost criteria for semi-Markov control processes. *Mathematics of Operations Research*, 29(2):326–338, 2004.

J. Jasper. Lion Air crash report 'criticises design, maintenance and pilot error'. *The Guardian*, 2019. URL `https://www.theguardian.com/world/2019/oct/25/lion-air-crash-report-criticises-design-maintenance-and-pilot-error`.

D. R. Jiang and W. B. Powell. An approximate dynamic programming algorithm for monotone value functions. *Operations Research*, 63(6):1489–1511, 2015.

L. Kallenberg. *Markov decision processes*. University of Leiden, 2020.

S. Kalosi, S. Kapodistria, and J. A. C. Resing. Condition-based maintenance at both scheduled and unscheduled opportunities. In P. Scarf, S. Wu, and P. Do, editors, *Proceedings of the 9th IMA International Conference on Modelling in Industrial Maintenance and Reliability*, ISBN: 978-0-905091-31-0, 2016. https://arxiv.org/abs/1607.02299.

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

S. Karlin and H. Rubin. Distributions possessing a monotone likelihood ratio. *Journal of the American Statistical Association*, 51(276):637–643, 1956.

H. Kawai, J. Koyanagi, and M. Ohnishi. Optimal maintenance problems for

Markovian deteriorating systems. In S. Osaki, editor, *Stochastic Models in Reliability and Maintenance*, pages 193–218. Springer-Verlag, 2002.

A. Khaleghei and M. J. Kim. Optimal control of partially observable semi-Markovian failing systems: An analysis using a phase methodology. *Operations Research*, 69 (4):1282–1304, 2021.

A. Y. Khinchin. Sequences of chance events without after-effects. *Theory of Probability & Its Applications*, 1(1):1–15, 1956.

J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3):502–506, 1953.

M. J. Kim. Robust control of partially observable failing systems. *Operations Research*, 64(4):999–1014, 2016.

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. J. Kim, R. Jiang, V. Makis, and C. G. Lee. Optimal Bayesian fault prediction scheme for a partially observable system subject to random failure. *European Journal of Operational Research*, 214(2):331–339, 2011.

A. J. Kleywegt and J. D. Papastavrou. The dynamic and stochastic knapsack problem with random sized items. *Operations Research*, 49(1):26–41, 2001.

P. Kolesar. Minimum cost replacement under Markovian deterioration. *Management Science*, 12(9):694–706, 1966.

S. Kunnumkal and H. Topaloglu. Exploiting the structural properties of the underlying Markov decision problem in the Q-learning algorithm. *INFORMS Journal on Computing*, 20(2):288–301, 2008.

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.

H. J. Kushner and P. G. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer, New York, NY, 1992.

H. D. Kwon and S. A. Lippman. Acquisition of project-specific assets with Bayesian updating. *Operations Research*, 59(5):1119–1130, 2011.

H. D. Kwon, S. A. Lippman, and C. S. Tang. Optimal markdown pricing strategy with demand learning. *Probability in the Engineering and Informational Sciences*, 26(1):77–104, 2012.

J. Y. J. Lam and D. Banjevic. A myopic policy for optimal inspection scheduling for condition based maintenance. *Reliability Engineering & System Safety*, 144:1–11, 2015.

D. Lamghari-Idrissi. *A New After-Sales Service Measure for Stable Customer Operations*. PhD thesis, Eindhoven University of Technology, 2021.

M. A. Lariviere and E. L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45(3):346–363, 1999.

B. Lee and M. Rahim. An integrated economic design model for quality control, replacement, and maintenance. *Quality Engineering*, 13(4):581–593, 2001.

P. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

R. Li and J. Ryan. A Bayesian inventory model using real-time condition monitoring information. *Production and Operations Management*, 20(5):754–771, 2011.

K. Linderman, K. McKone-Sweet, and J. Anderson. An integrated systems approach to process control and maintenance. *European Journal of Operational Research*, 164(2):324–340, 2005.

S. A. Lippman. On dynamic programming with unbounded rewards. *Management Science*, 21(11):1225–1233, 1975.

B. Liu, S. Wu, M. Xie, and W. Kuo. A condition-based maintenance policy for degrading systems with age- and state-dependent operating cost. *European Journal of Operational Research*, 263(3):879–887, 2017.

X. Liu, Q. Sun, Z. Ye, and M. Yildirim. Optimal multi-type inspection policy for systems with imperfect online monitoring. *Reliability Engineering & System Safety*, 207:107335, 2021.

X. Lu, J. S. Song, and K. Zhu. Analysis of perishable-inventory systems with censored demand data. *Operations Research*, 56(4):1034–1038, 2008.

T. Macquart and A. Maheri. A stall-regulated wind turbine design to reduce fatigue. *Renewable Energy*, 133:964–970, 2019.

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

L. M. Maillart and S. M. Pollock. Cost-optimal condition-monitoring for predictive maintenance of 2-phase systems. *IEEE Transactions on Reliability*, 51(3):322–330, 2002.

V. Makis. Multivariate Bayesian control chart. *Operations Research*, 56(2):487–496, 2008.

V. Makis and A. Jardine. Optimal replacement policy for a general model with imperfect repair. *Journal of the Operational Research Society*, 43(2):111–120, 1992.

V. Makis and X. Jiang. Optimal replacement under partial observations. *Mathematics of Operations Research*, 28(2):382–394, 2003.

V. Makis, J. Wu, and Y. Gao. An application of DPCA to oil data for CBM modeling. *European Journal of Operational Research*, 174(1):112–123, 2006.

A. Mandelbaum and M. I. Reiman. On pooling in queueing networks. *Management Science*, 44(7):971–981, 1998.

J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon. The internet of things: Mapping the value beyond the hype. Technical report, McKinsey, 2015.

J. S. Maritz and T. Lwin. *Empirical Bayes Methods*. Chapman and Hall/CRC, 2nd edition, 2018.

A. J. Mersereau. Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management*, 17(3):335–349, 2015.

C. Morris. Natural exponential families with quadratic variance functions. *The Annals of Statistics*, 10(1):65–80, 1982.

A. Müller. Stochastic ordering of multivariate normal distributions. *Annals of the Institute of Statistical Mathematics*, 53(3):567–575, 2001.

T. Nakagawa. Optimum policies when preventive maintenance is imperfect. *IEEE Transactions on Reliability*, 28(4):331–332, 1979a.

T. Nakagawa. Imperfect preventive-maintenance. *IEEE Transactions on Reliability*, 28(5):402–402, 1979b.

J. Nilsson and L. Bertling. Maintenance management of wind power systems using condition monitoring systems - life cycle cost analysis for two case studies. *IEEE Transactions on Energy Conversion*, 22(1):223–229, 2007.

NOS. NS-dubbeldekker ontspoorde door problemen met remsysteem, 2020. URL `https://nos.nl/artikel/2339898-ns-dubbeldekker-ontspoorde-door-problemen-met-remsysteem`.

M. Nourelfath and F. Yalaoui. Integrated load distribution and production planning in series-parallel multi-state systems with failure rate depending on load. *Reliability Engineering & System Safety*, 106:138–145, 2012.

M. Olde Keizer, S. D. P. Flapper, and R. H. Teunter. Condition-based maintenance policies for systems with multiple dependent components: A review. *European Journal of Operational Research*, 261(2):405–420, 2017.

T. L. Olsen and B. Tomlin. Industry 4.0: Opportunities and challenges for operations management. *Manufacturing & Service Operations Management*, 22(1):113–122, 2020.

R. Pacquin. Asset management: The changing landscape of predictive maintenance. AberdeenGroup, March 2014.

S. Panagiotidou and G. Tagaras. Statistical process control and condition-based maintenance: A meaningful relationship through data sharing. *Production and Operations Management*, 19(2):156–171, 2010.

Y. Peng, M. Dong, and M. J. Zuo. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425–438, 1996.

Philips. Allura Xper FD 10 - DS Interventional X-ray system, 2020. URL `https://www.usa.philips.com/healthcare/product/HC889021/diamondselectalluraxperfd10refurbishedxraysystem`.

E. Porteus and A. Angelus. Opportunities for improved statistical process control. *Management Science*, 43(9):1214–1228, 1997.

A. Prajapati, J. Bechtel, and S. Ganesan. Condition based maintenance: a survey. *Journal of Quality in Maintenance Engineering*, 18(4):384–400, 2012.

M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2005.

Y. Raaijmakers. *Job-Replication Trade-Offs: Performance Analysis of Redundancy Systems*. PhD thesis, Eindhoven University of Technology, 2021.

M. Rahim and P. Banerjee. A generalized model for the economic design of $\bar{x}$-control charts for production systems with increasing failure rate and early replacement. *Naval Research Logistics*, 40(6):787–809, 1993.

J. Ribrant and L. Bertling. Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005. *IEEE Transactions on Energy Conversion*, 22(1):167–173, 2007.

R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Springer Science & Business Media, 2nd edition, 2009.

D. Rosenfield. Markovian deterioration with uncertain information. *Operations Research*, 24(1):141–155, 1976.

S. M. Ross. A Markovian replacement model with a generalization to include stocking. *Management Science*, 15(11):702–715, 1969.

S. M. Ross. Average cost semi-Markov decision processes. *Journal of Applied Probability*, 7(3):649–656, 1970.

S. M. Ross. *Stochastic Processes*. John Wiley & Sons, 2nd edition, 1996.

S. M. Ross. *Introduction to Probability Models*. Academic Press, 11th edition, 2014.

S. M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press, 6th edition, 2020.

D. Russo. A note on the equivalence of upper confidence bounds and Gittins indices for patient agents. *Operations Research*, 69(1):273–278, 2021.

P. A. Scarf. On the application of mathematical models in maintenance. *European Journal of Operational Research*, 99(3):493–506, 1997.

M. Schäl. On the second optimality equation for semi-Markov decision models. *Mathematics of Operations Research*, 17(2):470–486, 1992.

M. Schäl. Average optimality in dynamic programming with general state space. *Mathematics of Operations Research*, 18(1):163–172, 1993.

J. Senoner, T. Netland, and S. Feuerriegel. Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science*, forthcoming, 2021.

R. Serfozo. *Basics of Applied Stochastic Processes*. Springer, corrected 2nd printing, 2012 edition, 2009.

M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer Science & Business Media, 2007.

X. Si, T. Li, Q. Zhang, and X. Hu. An optimal condition-based replacement method for systems with observed degradation signals. *IEEE Transactions on Reliability*,

67(3):1281–1293, 2018.

N. Singpurwalla. Survival in dynamic environments. *Statistical Science*, 10(1):86–103, 1995.

D. R. Smith and W. Whitt. Resource sharing for efficiency in traffic systems. *Bell System Technical Journal*, 60(1):39–55, 1981.

K. Sobczyk. Stochastic models for fatigue damage of materials. *Advances in Applied Probability*, 19(3):652–673, 1987.

J. S. Song and Y. Zhang. Stock or print? Impact of 3-D printing on spare parts logistics. *Management Science*, 66(9):3860–3878, 2020.

J. S. Song and Y. Zhang. Predictive 3D printing with IoT. *Working paper. Available at: SSRN 3895854*, 2021.

F. Spinato, P. J. Tavner, G. J. W. van Bussel, and E. Koutoulakos. Reliability of wind turbine subassemblies. *IET Renewable Power Generation*, 3(4):387–401, 2009.

P. Sun, K. Wang, and P. Zipkin. Quadratic approximation of cost functions in lost sales and perishable inventory control problems. *Fuqua School of Business Working Paper*, 2014.

G. Tagaras. An integrated cost model for the joint optimization of process control and maintenance. *Journal of the Operational Research Society*, 39(8):757–766, 1988.

G. Tagaras. A dynamic programming approach to the economic design of $\bar{x}$-charts. *IIE Transactions*, 26(3):48–56, 1994.

G. Tagaras. Dynamic control charts for finite production runs. *European Journal of Operational Research*, 91(1):38–55, 1996.

G. Tagaras. A survey of recent developments in the design of adaptive control charts. *Journal of Quality Technology*, 30(3):212–231, 1998.

G. Tagaras and Y. Nikolaidis. Comparing the effectiveness of various Bayesian $\bar{X}$-control charts. *Operations Research*, 50(5):878–888, 2002.

P. J. Tavner, J. Xiang, and F. Spinato. Reliability analysis for wind turbines. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 10(1):1–18, 2007.

H. Taylor. Markovian sequential replacement processes. *The Annals of Mathematical Statistics*, 36(6):1677–1694, 1965.

H. Taylor. Statistical control of a Gaussian process. *Technometrics*, 9(1):29–41, 1967.

D. M. Topkis. *Supermodularity and Complementarity*. Princeton University Press, 1998.

M. A. Uit het Broek, R. H. Teunter, B. De Jonge, J. Veldman, and N. Van Foreest. Condition-based production planning: Adjusting production rates to balance output and failure risk. *Manufacturing & Service Operations Management*, 22(4): 792–811, 2020.

M. A. Uit het Broek, R. H. Teunter, B. De Jonge, and J. Veldman. Joint condition-based maintenance and condition-based production optimization. *Reliability Engineering & System Safety*, 214:107743, 2021.

J. Van Noortwijk. A survey of the application of Gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1):2–21, 2009.

C. Van Oosterom, A. Elwany, D. Çelebi, and G. J. Van Houtum. Optimal policies for a delay time model with postponed replacement. *European Journal of Operational Research*, 232(1):186–197, 2014.

C. Van Oosterom, H. Peng, and G. J. Van Houtum. Maintenance optimization for a Markovian deteriorating system with population heterogeneity. *IISE Transactions*, 49(1):96–109, 2017.

H. E. Van Staden and R. N. Boute. The effect of multi-sensor data on condition-based maintenance policies. *European Journal of Operational Research*, 290(2):585–600, 2021.

E. Van Wingerden, T. Tan, and G. J. Van Houtum. Spare parts management under double demand uncertainty. *BETA Working Papers 534*, 2017.

O. Vega-Amaya and F. Luque-Vásquez. Sample-path average cost optimality for semi-Markov control processes on Borel spaces: unbounded costs and mean holding times. *Applicationes Mathematicae*, 27:343–367, 2000.

Wall Street Journal Custom Studios. How manufacturers achieve top quartile performance, 2017.

B. Wang and M. Wang. Stochastic ordering of folded normal random variables. *Statistics & Probability Letters*, 81(4):524–528, 2011.

H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489, 2002.

W. Wang. Delay time modelling. In *Complex System Maintenance Handbook*, pages 345–370. Springer, 2008.

W. Wang. An overview of the recent advances in delay-time-based maintenance modelling. *Reliability Engineering & System Safety*, 106:165–178, 2012.

M. Yang and V. Makis. ARX model-based gearbox fault detection and localization under varying load conditions. *Journal of Sound and Vibration*, 329(24):5209–5221, 2010.

A. A. Yushkevich. On semi-Markov controlled models with an average reward criterion. *Theory of Probability & Its Applications*, 26(4):796–803, 1982.

L. Zhang, Y. Lei, and H. Shen. How heterogeneity influences condition-based maintenance for Gamma degradation process. *International Journal of Production Research*, 54(19):5829–5841, 2016.

T. Zhang, R. Dwight, and K. El-Akruti. Condition based maintenance and operation of wind turbines. In *Engineering Asset Management-Systems, Professional Practices and Certification*, pages 1013–1025. Springer, 2015.

Q. Zhu, H. Peng, and G. J. Van Houtum. An age-based maintenance policy using the opportunities of scheduled and unscheduled system downs. *BETA Working Papers 498*, 2016.

Q. Zhu, H. Peng, B. Timmermans, and G. J. Van Houtum. A condition-based maintenance model for a single component in a system with scheduled and unscheduled downs. *International Journal of Production Economics*, 193:365–380, 2017.

P. Zipkin. On the structure of lost-sales inventory models. *Operations Research*, 56 (4):937–944, 2008.

Summary

The availability of advanced technical systems such as wind turbines, lithography systems, and X-ray systems, is crucial for the smooth operation of public services as well as for the primary processes of companies. Unavailability – especially when unplanned – and failures of these systems have severe consequences, both from a societal and financial point of view. Maintenance optimization deals with minimizing the risk of such failures, and otherwise mitigate their consequences, by performing preventive replacements on critical components when needed. However, early preventive replacements lead to high capital expenditures as the useful lifetimes of the systems is cut short. The most important challenge is therefore to trade-off two conflicting objectives: (i) minimize the risk of failure and unplanned downtime with all its adverse consequences and (ii) maximize the utilization of the useful lifetime of a component. This challenge is particularly difficult when components' deterioration processes or lifetime distributions are characterized by parameters that are a-priori unknown and need to be learned. Fortunately, there is an increasing and improved availability of real-time data stemming from ubiquitous sensors installed in modern systems, which provides enormous opportunities when leveraged in maintenance optimization characterized by such parameter uncertainty.

In this thesis, the overarching topic is the development of a general theory for a class of maintenance problems characterized by parameter uncertainty. We propose a unified framework in the form of a structured road map to develop mathematical models that integrate learning from real-time data with decision making, and demonstrate its wide applicability by adopting it for the analysis of various scenarios.

We first analyze a scenario in which a component's condition deteriorates according to a compound Poisson degradation with a fairly general compounding distribution, where the parameters of both the Poisson process and the compounding distribution are unknown. We propose a Bayesian framework to learn these unknown parameters independently from each other and integrate this into a Bayesian Markov decision

process (MDP). We establish the optimality of a control limit policy that depends on the entire deterioration path of the individual system. In a case study performed on real-life data from interventional X-ray machines from Philips Healthcare, we show that integration of learning and decision making leads to cost reductions of 11% relative to approaches that do not learn from real-time data and 4% relative to approaches that separate learning and decision making.

We then study the joint optimization of maintenance and process control, where a component's condition evolves according to a Brownian motion with unknown drift. We establish the optimality of a bandwidth policy that is monotone in the age of the process under both the discounted and average cost criterion. With this policy, the production process is continued if the condition is within a bandwidth that is described by both an upper and a lower control limit. We again build a Bayesian MDP, yet this formulation suffers from non-monotonic properties that render conventional proof techniques to establish structural properties not applicable. We overcome this challenge by translating the original Bayesian MDP into an alternative formulation which does allow for establishing structural properties of the optimal policy.

Subsequently, we consider a scenario where we do not only monitor deterioration, but where we can also control the deterioration by adjusting settings of the system (e.g., controlling a wind turbine's pitch to influence its deterioration). Dynamically adjusting settings can then be used to ensure that the useful lifetime of the system is maximized at precisely the time of a planned maintenance moment. We formulate this problem as a continuous-time MDP and characterize the monotonic behavior of the optimal setting policy. We further show that under optimal setting decisions, the length of the interval between maintenance moments can be easily optimized, thereby addressing the trade-off that arises due to maintenance costs at such moments. We also develop a Bayesian heuristic for the case that there is parameter uncertainty in the deterioration process, and show that it performs close to a clairvoyant policy.

We then analyze a scenario in which multiple systems must be maintained and where data stemming from each individual system can be pooled to jointly learn a common, yet unknown parameter. Each system has a critical component whose condition deteriorates according to a Poisson process with a common unknown rate. We formulate this problem as a finite-horizon Bayesian MDP, whose formulation suffers from the well-known curse of dimensionality. As a remedy, we prove a new decomposition result that establishes the equivalence between the original, high-dimensional MDP and multiple two-state MDPs with a binary action space, each focused on an individual system. We establish the optimality of a control limit policy for each system that depends on the entire deterioration paths of all systems. In a

numerical study, we show that savings because of pooling data can be large, but that the exact magnitude depends on the magnitude of the uncertainty in the parameter.

We then revisit the canonical age-based maintenance model, but assume that the systems' lifetime distributions are parametrized by an a-priori unknown parameter that needs to be learned. In this scenario, there is an additional exploration versus exploitation trade-off: Corrective replacements are uncensored but costly observations, whereas preventive replacements are censored but cheap observations of the lifetime distribution. We first analyze the optimal policy for a finite sequence of replacements and establish some properties. Subsequently, we propose a myopic Bayesian policy, for which we show that it almost surely learns the unknown parameter and converges to the optimal policy with full knowledge of the parameter.

The models considered in the previous scenarios are all characterized by parameter uncertainty. In the last scenario, we depart from this and model the system's deterioration behavior as a delay time model of which the parameters are fully known. We analyze a scenario with two types of preventive maintenance: Planned maintenance at periodic, scheduled opportunities, and opportunistic maintenance at unscheduled opportunities. The latter type of maintenance arises particularly in application areas with a network dimension (e.g., wind turbines in a wind farm), where if a system in the network fails, this constitutes an opportunity for maintenance for the other systems. We formulate this problem as a semi-MDP, which allows us to characterize the structure of the optimal policy under both the discounted and average cost criterion. This policy, depending on the deterioration of the component and the remaining time until the next planned maintenance, indicates when it is optimal to perform preventive maintenance at both scheduled and unscheduled opportunities.

About the author

Collin Drent was born in Geleen, the Netherlands, on October 3, 1993. He finished his pre-university education at Groenewald in Stein in 2012. In 2015, Collin obtained a Bachelor's degree in Industrial Engineering and in 2017, he received a Master's degree in Operations Management and Logistics. Both degrees were obtained cum laude from the Eindhoven University of Technology (TU/e), the Netherlands. During his Master's studies, Collin was awarded the ASML Technology Scholarship, and he spent a semester abroad at the National Taiwan University in Taipei, Taiwan. His Master thesis research addressed data-driven dispatching and repositioning policies in service logistics networks and was supervised by dr. Minou Olde Keizer and prof.dr.ir. Geert-Jan van Houtum. This thesis was a finalist in the 2017 Service Logistics Forum Thesis Awards and was later converted in a journal article (Drent et al., 2020c).

In January 2018, Collin started as a PhD candidate in the Stochastic Operations Research group in the Department of Mathematics and Computer Science at the TU/e. Under the supervision of prof.dr.ir. Onno Boxma, prof.dr.ir. Geert-Jan van Houtum and dr. Stella Kapodistria, he studied the integration of learning and decision making for deteriorating and aging systems in the context of maintenance. During his PhD project, Collin spent several research visits at the MIT Luxembourg Centre for Logistics and Supply Chain Management at the University of Luxembourg to work with dr.ir. Joachim Arts and dr.ir. Melvin Drent. The most important results of this PhD research are described in this thesis.

During his PhD candidacy, Collin co-organized the 13th Young European Queuing Theorists (YEQT) workshop and attended several national and international conferences to present his work.

Collin will defend his PhD thesis at the TU/e on July 1, 2022.