CBM for a series system with economic dependence

Parts of this chapter have been published as: M.C.A. Olde Keizer and R.H. Teunter, *Opportunistic condition-based maintenance and aperiodic inspections for a two-unit series system*, SOM Research Report 14033-OPERA, University of Groningen (2014).
Abstract

This chapter considers a two-component system subject to structural dependence (through a series structure) and (positive) economic dependence. The aperiodic inspection moments are optimized simultaneously with the critical levels at which maintenance is performed in order to minimize cost or maximize availability. For this purpose, a stochastic model is developed based on semi-regenerative properties of the maintained system state. We build on the work of B. Castanier, A. Grall, C. Bérenguer, A condition-based maintenance policy with non-periodic inspections for a two-unit series system, Reliability Engineering & System Safety 87 (1) (2005) 109-120, by fully including all opportunistic maintenance opportunities, determining the system unavailability time more accurately, and providing a more extensive performance evaluation. Results indicate that the accuracy with which the unavailability time is approximated has a great impact on the resulting optimal maintenance strategy. In addition, our model is shown to significantly outperform several classical maintenance policies, such as failure-based maintenance and block replacement.
3.1. Introduction

In this chapter, we focus on economic dependence, which applies when combining maintenance actions on several components yields a lower cost than maintaining each component separately. This is for example the case when shared set-up costs are involved. Since these fixed costs are independent of the number of components that require maintenance, it can be profitable to opportunistically replace components when another component requires immediate maintenance. Furthermore, we include structural dependence by considering two components that are functioning in series (i.e., both components are critical to the system).

Few articles consider CBM for systems consisting of multiple components subject to economic dependence \([1\text{–}3]\). In \([2]\), dynamic maintenance grouping is applied to optimize the maintenance costs on a rolling horizon for a system with periodic inspections. In \([3]\), a CBM policy is proposed based on the proportional hazards model, where the periodic inspection moments are fixed in advance. A CBM policy for a two-component series system with aperiodic inspections is developed by Castanier et al. \([1]\), where the maintenance costs are obtained by using semi-regenerative properties of the maintained system state. The latter model is advanced in that the inspection moments and the maintenance thresholds are optimized simultaneously.

In practice, the critical level at which preventive maintenance is initiated is usually based on recommendations of suppliers and manufacturers of condition-monitoring equipment rather than on incentives to save costs or improve reliability \([4]\). Justifications for the selected inspection moments are also frequently lacking, and usually based on a simple rule-of-thumb. Due to safety concerns, the critical level is likely to be set too low, while inspections may be scheduled more often than actually required. To overcome these problems, most existing literature on CBM considers either a model that optimizes the critical level at which a preventive replacement should be initiated given the (periodic) inspection intervals, or a model that optimizes the inspection intervals given the critical level. Besides \([1]\), joint optimization has only been studied in \([4\text{–}6]\). A CBM model based on the random coefficient growth model is described in \([4]\), while a multi-threshold CBM policy is considered in \([1, 5, 6]\). In fact, \([1]\) and \([6]\) both extend the model of \([5]\), where a CBM policy is developed for a system consisting of a single component and failures are noticed immediately. Partial repairs and durations of maintenance activities are considered in \([6]\), assuming that failures can only be noticed upon inspection. A system consisting of two components that are functioning in series is analyzed in \([1]\).
In this chapter, we extend the work of [1] that is unique in its simultaneous
optimization of inspection and maintenance decisions for a system consisting
of multiple components. We determine the unavailability costs more accurately
and include all possibilities for opportunistic replacements. In addition, we
consider availability as an additional performance criterion and we perform a
more extensive comparative cost assessment. We provide insights into the optimal
policy structure, and compare our results with those of classical maintenance
policies, such as failure-based maintenance and block replacement.

The remainder of this chapter is organized as follows. Section 3.2 describes
the deterioration model, followed by the multi-threshold maintenance policy in
Section 3.3. Next, the definition and evaluation of the performance criteria are
given in Section 3.4, while Section 3.5 contains the performance assessments
and a comparison with classical maintenance policies. Section 3.6 concludes the
chapter.

3.2. System description

3.2.1. Deterioration model

The discrete-time system under consideration consists of two components, which
independently suffer from increasing wear. The condition of component \( i \) at
time \( k \) can be described by a random variable \( X^i_k \), for \( i = 1, 2 \) and \( k \in \mathbb{N} \). Compo-
nent \( i \) will fail as soon as its deterioration level exceeds a preset failure level \( L^i \),
but such a failure can only be noticed upon inspection. This means that the
component remains in the failed state until the next planned inspection. This
assumption applies for example in a standby system or in a system where failure
does not imply a system stop, but where the quality of the produced items is
reduced.

An inspection can be performed at the start of each time unit. Both at the start
of the process (at time 0) and after each replacement (say at time \( t_r \)), component \( i \)
is assumed to be as-good-as-new, i.e., \( X^i_0 = 0 \) and \( X^i_{t_r} = 0 \), for \( i = 1, 2 \). The degrada-
tion of the global system is given by \( (X_k)_{k \in \mathbb{N}} = (X^1_k, X^2_k)_{k \in \mathbb{N}} \). Since we assume that
degradation will increase over time, we require the random deterioration incre-
ments \( \Delta_{(k,k+1)}X^i = X^i_{k+1} - X^i_k \) in each time interval to be nonnegative. In addition,
we assume that the increments are stationary and exchangeable, so the degrada-
tion increments satisfy the memoryless property. Let \( f^i \) denote the probability
density function of the deterioration increments \( \Delta_{(k,k+1)}X^i \) of component \( i \), for
\( i = 1, 2 \) and for all \( k \in \mathbb{N} \). From the assumption that the deterioration increments
are stationary and exchangeable, it follows that the distribution functions \( f^i \) are
Nomenclature

$\pi(x_1, x_2)$ Long-run probability that components 1 and 2 are in states $x_1$ and $x_2$, respectively, at the start of an inspection

$\xi^{(i)}$ Inspection thresholds for component $i$ ($j = 0, 1, \ldots, n - 1$)

$S_n$ Preventive replacement threshold for component $i$

$\zeta_i$ Opportunistic replacement threshold for component $i$

$A(t)$ System availability up to time $t$

$A_\infty$ Long-run average system availability

$C(t)$ Cumulative operating cost up to time $t$

$C_\infty$ Long-run average operating cost per time unit

$c^{(i)}_C$ Cost of a corrective replacement for component $i$

$C^C_{(i)}(t)$ Cumulative corrective replacement cost of component $i$ up to time $t$

$c_k$ Cost of a preventive replacement for component $i$

$C^P_{(i)}(t)$ Cumulative preventive replacement cost of component $i$ up to time $t$

$c_s$ Fixed set-up cost for a replacement

$C_S(t)$ Cumulative set-up cost up to time $t$

$c_u$ Unavailability cost rate, per time unit that the system is unavailable

$C_U(t)$ Cumulative system unavailability cost up to time $t$

$E_\pi[\cdot]$ Expected value with respect to the stationary law $\pi$

$f_{_{i}}(\cdot)$ Pdf of the deterioration increments of component $i$

$f_{_{i}}^{(l)}(\cdot)$ Pdf of the cumulative deterioration increments over $l$ time units for component $i$

$L_i$ Failure-level of component $i$

$n$ Number of inspection threshold values for each component

$S$ Length of a semi-regeneration cycle in steady state

$X^i_k$ Condition of component $i$ at time $k$

infinitely divisible [7]. This is for example the case for all gamma distributions. The property of infinitely divisible increments makes a gamma process suitable for describing deterioration caused by continuous use [8]. In particular, it is very suitable for describing the steady evolution of wear between the start-up period
and the wear accumulation at the end of the system’s life [9]. As in [1], we therefore select \( f_i \) as the exponential distribution with rate parameter \( \alpha_i \) for component \( i \), \( i = 1, 2 \). Next, let \( f_i^{(l)} \) denote the probability density function of the cumulative deterioration increments during \( l \) consecutive time units for component \( i \). Since the sum of \( l \) exponential distributions with parameter \( \alpha_i \) (which is the increase in deterioration during \( l \) time units) follows an Erlang distribution with parameters \( \alpha_i \) and \( l \), it follows immediately that \( f_i^{(l)} \) is Erlang distributed with parameters \( \alpha_i \) and \( l \).

### 3.2.2. Maintenance actions and corresponding costs

At the start of each time unit, a decision is needed on whether or not to perform an inspection, while at each inspection, a decision is needed on what components to replace. Obviously, failed components should be replaced correctively. Moreover, a functioning component may be replaced preventively if it is close to failure, or opportunistically if the other component is replaced as well. The latter can be cost efficient due to the economic dependence.

The costs of the different maintenance operations are given in Table 3.1. It is realistic to assume that the fixed cost of an inspection (resulting from planning, transportation, and possibly a shutdown) are relatively large compared to the variable cost depending on the number of components that are inspected, as no material or repair costs are involved. For this reason, we assume that all components are inspected at every inspection, i.e., that there are common inspection moments, at which a cost \( c_n \) is incurred. Furthermore, a corrective replacement on component \( i \) incurs a cost \( c_c^{(i)} \), while a preventive (or opportunistic) replacement incurs a cost \( c_p^{(i)} \). Typically, replacing a failed component is more expensive than replacing a functioning component, i.e., \( c_c^{(i)} > c_p^{(i)} \). The economic dependence is incorporated through a shared set-up cost for maintenance. This set-up cost is incurred once if maintenance is performed, independent of the number of components that are replaced. In practice, shared set-up costs can arise from traveling to the right location, scheduling personnel, ordering spare parts, or doing

<table>
<thead>
<tr>
<th>Maintenance action</th>
<th>Corresponding costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspection (both components)</td>
<td>( c_n )</td>
</tr>
<tr>
<td>Corrective replacement (component ( i ))</td>
<td>( c_c^{(i)} )</td>
</tr>
<tr>
<td>Preventive replacement (component ( i ))</td>
<td>( c_p^{(i)} )</td>
</tr>
<tr>
<td>Shared set-up cost replacement</td>
<td>( c_s )</td>
</tr>
</tbody>
</table>

Table 3.1. An overview of the costs associated with the different maintenance operations.
paperwork. Since failures can only be noticed upon inspection, the system may spend some time in the failed state before it will be inspected and maintained. In this case, the so-called unavailability cost rate $c_u$ is incurred per unit of time that the system is unavailable. The amount of time that a component is unavailable cannot be measured exactly, because failures are only noticed upon inspection. As discussed in detail in Section 3.4.2, an upper bound for the unavailability time is used in [1], while we also present a more accurate approximation and show that this significantly affects the results.

### 3.2.3. Performance criteria

In practice, both the long-run average maintenance cost per period and the long-run average availability are important performance criteria. Whereas only the former is considered in [1], we consider both. Our multi-threshold maintenance policy aims at optimizing all threshold values simultaneously in order to minimize the maintenance costs, or to maximize the system availability.

### 3.3. Multi-threshold maintenance policy

Let us define $\xi_0^{(i)}, \xi_1^{(i)}, \ldots, \xi_n^{(i)}$ (with $\xi_0^{(i)} \leq \xi_1^{(i)} \leq \ldots \leq \xi_n^{(i)} \leq L_i$) as the threshold values of component $i$, where $n$ denotes the fixed number of inspection threshold values for each component. The lowest threshold value is that of a new component, and it is normalized to zero, i.e., $\xi_0^{(i)} := 0$. The notation $\xi_0^{(i)}$ is not strictly needed but will turn out to be helpful in representing the maintenance policy. At the moment, say $k$, of an inspection, the deterioration level $X_k^i$ of each component $i$ is revealed. Comparison of these levels with the threshold values then indicates whether or not a replacement should be performed for each component, and when to perform the next inspection. The maintenance policy thus consists of the following three steps.

**Step 1: Corrective and preventive replacements.** In this step, we determine for each component separately whether it requires maintenance. Component $i$ ($i = 1, 2$) is inspected, and the observed deterioration level $X_k^i$ is compared with the threshold values as follows.

- **If** $0 \leq X_k^i < \xi_n^{(i)}$: Component $i$ does not require maintenance yet.
- **If** $\xi_n^{(i)} \leq X_k^i < L_i$: Immediately replace component $i$ preventively.
- **If** $X_k^i \geq L_i$: Component $i$ has failed. Replace component $i$ correctively.
Step 2: Opportunistic replacements. If the first step indicates that one component must be replaced, and so set-up costs have to be paid anyway, then it may be cost efficient to replace the other component as well. We therefore introduce additional threshold values $\zeta_i$ (with $0 \leq \zeta_i \leq \xi^{(i)}_n$ for $i = 1, 2$) and the following opportunistic replacement strategy.

If $0 \leq X^i_k < \zeta_i$: Component $i$ does not require maintenance yet.

If $\zeta_i \leq X^i_k < \xi^{(i)}_n$ and $X^j_k \geq \xi^{(j)}_n$ (for $j \neq i$) Replace component $i$ preventively (opportunistically).

Step 3: Next inspection moment. The third step consists of determining the next inspection moment, taking into account the decisions taken in steps 1 and 2. Note that after component $i$ has been replaced, it is as-good-as-new again, and $X^i_k$ becomes zero ($i = 1, 2$).

If $\xi^1_{l_1} \leq X^1_k < \xi^1_{l_1+1}$ and $\xi^2_{l_2} \leq X^2_k < \xi^2_{l_2+1}$ for $l_1, l_2 \in \{0, 1, \ldots, n - 1\}$: The next system inspection is scheduled at $\min\{n - l_1, n - l_2\}$ periods from now.

The system will thus be inspected more often when one of its components is approaching failure.

In Figure 3.1, an example of the wear patterns and corresponding inspection and maintenance actions of a two-component series system is given. Here $n = 2$, implying that the next inspection is always scheduled at most two time units later. At time 0, both components are as-good-as-new, and the next inspection is scheduled two time units later. As soon as at least one component exceeds its inspection threshold $\xi^{(i)}_1$, the next inspection is scheduled one time unit later. At times 6 and 13, one of the components requires a preventive replacement, while the other component is replaced opportunistically. This results in a complete system replacement. Furthermore, at time 10, only component 1 is replaced, since the deterioration level of component 2 does not exceeds its opportunistic replacement threshold $\zeta_2$. 
3.4. Evaluation of the performance criteria

The deterioration process contains (semi-)regenerative properties, implying that the costs (and system availability) can be evaluated during a single inspection cycle [1, 6]. To this end, a stationary law $\pi(x_1, x_2)$ can be constructed, denoting the probability density function of being in state $(x_1, x_2)$ at the start of an inspection.

3.4.1. Stationary law

The stationary law $\pi(x_1, x_2)$ can be obtained as the probability density of being in state $(y_1, y_2)$ multiplied by the probability of moving from state $(y_1, y_2)$ to state $(x_1, x_2)$ during an inspection cycle, integrated over all possible states $(y_1, y_2)$. Since these probabilities depend on whether or not components 1 and 2 are replaced after the previous inspection, we distinguish four cases:

- **Both components replaced.** If $y_1 \in [\xi_1^n, \infty)$ and $y_2 \in [\zeta_2, \infty)$ or if $y_1 \in [\xi_1^n, \zeta_1]$, and $y_2 \in [\xi_2^n, \infty)$, then both components are replaced. Moreover, the next inspection is scheduled $n$ periods later, which implies that the increases in deterioration levels from 0 to $x_1$ and from 0 to $x_2$ have densities of $f_1^{(n)}(x_1)$ and $f_2^{(n)}(x_2)$, respectively.
Only component 1 replaced. If \( y_1 \in [\xi_{l_1}^1, \infty) \) and \( y_2 < \zeta_2 \), then only component 1 will be replaced and hence start with a deterioration level of zero. Moreover, if \( y_2 \in [\xi_{l_2}^2, \xi_{l_2}^2] \) with \( l_2 \in \{0, 1, \ldots, n-1\} \), then the next inspection is scheduled \( n - l_2 \) periods later. The increases in deterioration levels from 0 to 0, and from \( y_2 \) to \( x_2 \) have densities of \( f_{1}^{(n-l_{2})}(x_1) \) and \( f_{2}^{(n-l_{2})}(x_2 - y_2) \), respectively.

Only component 2 replaced. If \( y_1 < \zeta_1 \) and \( y_2 \in [\xi_{l_2}^2, \infty) \), then only component 2 will be replaced and hence start with a deterioration level of zero. Moreover, if \( y_1 \in [\xi_{l_1}^1, \xi_{l_1}^1] \) with \( l_1 \in \{0, 1, \ldots, n-1\} \), then the next inspection is scheduled \( n - l_1 \) periods later. The increases in deterioration levels from 0 to 0, and from \( y_2 \) to \( x_2 \) have densities of \( f_{1}^{(n-l_{1})}(x_1 - y_1) \) and \( f_{2}^{(n-l_{1})}(x_2) \), respectively.

No replacement. If \( y_1 \in [\xi_{l_1}^1, \xi_{l_1}^1] \) and \( y_2 \in [\xi_{l_2}^2, \xi_{l_2}^2] \), with \( l_1, l_2 \in \{0, 1, \ldots, n-1\} \), then no replacement will be performed, and the next inspection is scheduled \( \min(n - l_1, n - l_2) = n - \max(l_1, l_2) \) periods later. The increases in deterioration levels from 0 to 0, and from \( y_2 \) to \( x_2 \) have densities of \( f_{1}^{(n-\max(l_{1}, l_{2}))}(x_1 - y_1) \) and \( f_{2}^{(n-\max(l_{1}, l_{2}))}(x_2 - y_2) \), respectively.

Using this information, the probability law \( \pi(x_1, x_2) \) can be constructed as follows.

\[
\pi(x_1, x_2) = \left( \int_{\xi_{n}^1}^{\infty} \int_{\xi_{2}^2}^{\infty} \pi(y_1, y_2)dy_2dy_1 + \int_{\xi_{n}^1}^{\infty} \int_{\xi_{2}^2}^{\infty} \pi(y_1, y_2)dy_2dy_1 \right) \int_{1}^{(n)}(x_1) f_{2}^{(n)}(x_2) \\
+ \sum_{l_2=0}^{n-1} \int_{\xi_{n}^1}^{\infty} \int_{\xi_{n}^1}^{\min(\xi_{l_2}^2, \xi_2^2)} \pi(y_1, y_2) f_{1}^{(n-l_{2})}(x_1) f_{2}^{(n-l_{2})}(x_2 - y_2)dy_2dy_1 \\
+ \sum_{l_1=0}^{n-1} \int_{\min(\xi_{l_1}^1, \xi_1)}^{\min(l_{1}^1, \xi_1)} \int_{\xi_{n}^2}^{\infty} \pi(y_1, y_2) f_{1}^{(n-l_{1})}(x_1 - y_1) f_{2}^{(n-l_{1})}(x_2)dy_2dy_1 \\
+ \sum_{l_1=0}^{n-1} \sum_{l_2=0}^{n-1} \int_{\xi_{l_1}^1}^{\xi_{l_1}^1} \int_{\xi_{l_2}^{2+1}}^{\xi_{l_2}^{2+1}} \pi(y_1, y_2) f_{1}^{(n-\max(l_{1}, l_{2}))}(x_1 - y_1) f_{2}^{(n-\max(l_{1}, l_{2}))}(x_2 - y_2)dy_2dy_1 
\]

This expression can be rewritten as a nonhomogeneous linear Fredholm integral equation of the second kind:

\[
\pi(x_1, x_2) = F(x_1, x_2) + \int_{S} \pi(y_1, y_2)K(x_1, x_2, y_1, y_2)dy_2dy_1, 
\]
where

\[ F(x_1, x_2) = f_1^{(n)}(x_1) f_2^{(n)}(x_2), \]

and, for \( k = 0, 1, \ldots, n - 1, \)

\[
\begin{align*}
S_k^1 &= \{(y_1, y_2) : y_1 \in [\min(\xi^1_k, \xi_1), \min(\xi^1_{k+1}, \xi_1)] \cap y_2 \in [\xi_n^2, \infty]\}, \\
S_k^2 &= \{(y_1, y_2) : y_1 \in [\xi_1^k, \infty) \cap y_2 \in [\min(\xi^2_k, \xi_2), \min(\xi^2_{k+1}, \xi_2)]\}, \\
S_k^3 &= \{(y_1, y_2) : (y_1 \in [0, \xi^1_k) \cap y_2 \in [\xi^2_k, \xi^2_{k+1}]) \cup (y_1 \in [\xi^1_k, \xi^1_{k+1}] \cap y_2 \in [0, \xi^2_k])\},
\end{align*}
\]

and

\[ S = \bigcup_{k=0}^{n-1} (S_k^1 \cup S_k^2 \cup S_k^3), \]

\[ K(x_1, x_2, y_1, y_2) =
\begin{align*}
&f_1^{(n-k)}(x_1 - y_1) f_2^{(n-k)}(x_2) - F(x_1, x_2), &\text{if } (y_1, y_2) \in S_k^1, k = 0, 1, \ldots, n - 1, \\
&f_1^{(n-k)}(x_1) f_2^{(n-k)}(x_2 - y_2) - F(x_1, x_2), &\text{if } (y_1, y_2) \in S_k^2, k = 0, 1, \ldots, n - 1, \\
&f_1^{(n-k)}(x_1 - y_1) f_2^{(n-k)}(x_2 - y_2) - F(x_1, x_2), &\text{if } (y_1, y_2) \in S_k^3, k = 0, 1, \ldots, n - 1.
\end{align*}
\]

The densities \( f_i^{(l)} \) are regular [1], which implies that we can solve the equation above by applying the method of successive approximations [10]. Due to the high level of complexity, we will approximate the integrals numerically in our experiments in Section 3.5 by applying the extended midpoint rule [11] to the two-dimensional case, by dividing the area into \( 30 \times 30 \) parts. Note that we cannot use infinite upper bounds while using the extended midpoint rule, but instead assume that the deterioration level of component \( i \) will never exceed a value of \( 1.5 \) times \( L_i \), for \( i = 1, 2 \). Initial testing showed that this value is sufficiently large to ensure that the results are not affected.

### 3.4.2. Long-run average cost per time unit and system availability

Let \( C(t) \) denote the cumulative operating cost up to time \( t \), consisting of costs arising from maintenance activities and down-time of the system, and let \( A(t) \) denote the total time that the system is available up to time \( t \). Because of the semi-regenerative properties of the multi-threshold maintenance policy, we can consider the long-run average maintenance cost \( C(S) \) during an inspection cycle, with length \( S \). This cost can then be divided by the long-run average length of an inspection cycle to obtain the long-run average maintenance cost per period. Similar logic applies to the availability criterion. In the following, let \( E_\pi[\cdot] \) denote

67
the expected value with respect to the stationary law $\pi$. The different components of the cost and availability function are specified below.

First, the long-run average length of an inspection cycle, denoted by $E_{\pi}[S]$, can be obtained by again distinguishing four different cases; both components replaced, only component 1 replaced, only component 2 replaced, and no replacement, respectively:

$$E_{\pi}[S] = n \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_2 dx_1 + \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_1 dx_2 \right)$$

$$+ \sum_{k=1}^{n} k \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_2 dx_1 \right)$$

$$+ \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_1 dx_2 \right)$$

$$+ \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_2 dx_1 + \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_1 dx_2 \right).$$

Next, the long-run average inspection cost per semi-regeneration cycle, which is incurred exactly once per inspection cycle, is equal to $c_p$. The long-run average preventive maintenance cost of component $i$ per inspection cycle (excluding the set-up cost), denoted by $E_{\pi}[C_{p}^{(i)}(S)]$, is given by

$$E_{\pi}[C_{p}^{(i)}(S)] = c_p \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_j dx_i + \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_i dx_j \right).$$

while the long-run average corrective maintenance cost of component $i$ per inspection cycle (excluding the set-up cost), denoted by $E_{\pi}[C_{C}^{(i)}(S)]$, can be obtained as

$$E_{\pi}[C_{C}^{(i)}(S)] = c_c \int_{L_i}^{\infty} \int_{L_i}^{\infty} \pi(x_1, x_2) dx_j dx_i.$$

In case at least one component is replaced, either preventively or correctively, the set-up cost for a replacement $c_s$ needs to be paid once. The long-run average cumulative set-up cost per inspection cycle, denoted by $E_{\pi}[C_{S}(S)]$, is thus given by

$$E_{\pi}[C_{S}(S)] = c_s \left( \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_2 dx_1 + \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} \pi(x_1, x_2) dx_1 dx_2 \right).$$
The long-run average cost incurred for the time that the system spends in the failed state during one inspection cycle, denoted by $E_\pi[C_U(S)]$, is given by

$$E_\pi[C_U(S)] = c_u E_\pi[D_U(S)],$$

where $E_\pi[D_U(S)]$ denotes the long-run average time that the system is unavailable during an inspection cycle. Since by assumption failures can only be noticed upon inspection, which are scheduled at discrete points in time, the exact moment at which a failure occurs is unknown. As an alternative to using an upper bound for the unavailability time $[1]$, we suggest to assume a linear increase in deterioration between two consecutive inspection moments. This provides a better approximation of the down-time of the system, as will be further explained on the next pages.

The long-run average maintenance cost per time unit can now be obtained as

$$C_\infty = \lim_{t \to \infty} \frac{E[C(t)]}{t} = \frac{E_\pi[C(S)]}{E_\pi[S]} = c_n + \sum_{i=1}^{2} E_\pi[C_p(i)\{S\}] + \sum_{i=1}^{2} E_\pi[C_C(i)\{S\}] + E_\pi[C_S(S)] + E_\pi[C_U(S)] \over E_\pi[S],$$

while the long-run average system availability can be obtained as

$$A_\infty = \frac{E_\pi[S] - E_\pi[D_U(S)]}{E_\pi[S]}.$$

**Long-run average system unavailability time**

Suppose that component $i$ fails at time $t_f(i)$, with $t_f(i) \in [k-1, k)$ and $1 \leq k \leq S$. Since the deterioration level of a component, and hence whether or not it has failed, can only be observed at the discrete inspection moments, the exact moment of failure $t_f(i)$ is unknown. Castanier et al. [1] suggest to approximate the component unavailability time $D_U(i)\{S\}$ by assuming that the failure occurred at time $k-1$. However, this approximation, which we denote by $\hat{D}_U(i)\{S\}$, is obvious positively biased and indeed an upper bound, which the authors acknowledge. Moreover, since maintenance policies typically try to achieve a high up-time, failures that do occur are likely to occur towards the end of a period. We therefore use an alternative, linear approximation of the unavailability time, denoted by $\tilde{D}_U(i)\{S\}$, as is illustrated in Figure 3.2. Note that in order for $\tilde{D}_U(i)\{S\}$ to equal $S - m$, we require an increase in deterioration of $s = \frac{L_i - y - u}{m - (k-1)} = \frac{L_i - y - u}{m - \lfloor m \rfloor}$ between times $k-1(= \lfloor m \rfloor)$ and $k(= \lceil m \rceil)$. 
For presentational ease, we introduce the binary variable \( q \), indicating whether the system unavailability time is approximated by using [1]'s upper bound (\( q = 0 \)), or by assuming a linear increase in deterioration between two consecutive inspection moments (\( q = 1 \)). In other words,

\[
E_{\pi, q}[D_U(S)] = \begin{cases} 
E_{\pi} [\hat{D}_U(S)], & \text{if } q = 0, \\
E_{\pi} [\tilde{D}_U(S)], & \text{if } q = 1. 
\end{cases}
\]

Next, we define the functions \( u_{1, q}(m, l) \) as the system unavailability time if only one component fails, (approximately) at time \( m \), in an inspection cycle of length \( l \) using method \( q \), and \( u_{2, q}(m_1, m_2, l) \) as the system unavailability time if components 1 and 2 fail at (approximate) times \( m_1 \) and \( m_2 \), respectively, in an inspection cycle of length \( l \) using method \( q \). Then

\[
u_{1, q}(m, l) = \begin{cases} 
\lfloor l - m \rfloor, & \text{if } q = 0, \\
l - m, & \text{if } q = 1, 
\end{cases}
\]

\[
u_{2, q}(m_1, m_2, l) = \begin{cases} 
\max\{\lfloor l - m_1 \rfloor, \lfloor l - m_2 \rfloor\}, & \text{if } q = 0, \\
\max\{l - m_1, l - m_2\}, & \text{if } q = 1. 
\end{cases}
\]

For the derivation of \( E_{\pi, q}[D_U(S)] \), we define the function \( h_i(m | y) \) as the probability density function of a failure of component \( i \) (\( i = 1, 2 \)) at time \( m \) (under the assumption of a linear increase in deterioration), given a previous deterioration level of \( y \). This function consists of two parts; we multiply the probability that a failure occurs during the \( \lfloor m \rfloor \)-th time unit with the probability of a failure at time
Given a failure during the \([m]\)-th time unit. We find that

\[
h_i(m|y) = \begin{cases} 
\int_{L_i-y}^{\infty} L_i - y f_i(s) ds \cdot \frac{f_i(\frac{L_i-y}{m-[m]})}{f_i([m]) f_i(\frac{L_i-y}{m-[m]})} dt', & \text{if } [m] = 1, \\
\int_{0}^{L_i-y} f_i([m])(u) \left( \int_{L_i-y-u}^{\infty} f_i(s) ds \right) du \cdot \\
\frac{\int_{0}^{L_i-y} f_i([m])(u) f_i(\frac{L_i-y-u}{m-[m]}) du}{f_i([m]) \int_{0}^{L_i-y} f_i([m])(u) f_i(\frac{L_i-y-u}{m-[m]}) du t'}, & \text{if } [m] > 1.
\end{cases}
\]

Furthermore, we define the function \(F_i(y, l)\) as the probability that component \(i\) will not fail during the first \(l\) time units starting with a deterioration level of \(y\):

\[
F_i(y, l) = \int_{0}^{L_i-y} f_i^{(l)}(u) du.
\]

We can now obtain an expression for the approximated system unavailability times for both methods by distinguishing two cases: either both components fail, or just one component fails. If both components fail, the average system unavailability time is obtained by multiplying \(u_2, q(m_1, m_2, \cdot)\) with the probability density functions \(h_1(m_1|\cdot)\) and \(h_2(m_2|\cdot)\), and integrating the resulting expression with respect to \(m_1\) and \(m_2\). Similarly, in case only component 1 (2) will fail, the average system unavailability time is obtained by multiplying \(u_1, q(m, \cdot)\) with the probability density function \(h_1(m|\cdot)\) (\(h_2(m|\cdot)\)) and the probability that component 2 (1) will not fail \(F_2(\cdot, \cdot)\) (\(F_1(\cdot, \cdot)\)), and integrating this expression with respect to \(m\). Similar to obtaining an expression for the probability law, this is done separately for each of the following four cases: both components are replaced, only component 1 is replaced, only component 2 is replaced, and no replacement is performed during the previous inspection. This gives

\[
E_{\pi,q}[D_U(S)] = \\
\left( \int_{\xi_1}^{\xi_2} \int_{\xi_1}^{\xi_2} \pi(y_1, y_2) dy_2 dy_1 + \int_{\xi_1}^{\xi_2} \int_{\xi_1}^{\xi_2} \pi(y_1, y_2) dy_2 dy_1 \right) \cdot \\
\left( \int_{0}^{n} \int_{0}^{n} u_2, q(m_1, m_2, n) \cdot h_1(m_1|0) \cdot h_2(m_2|0) dm_2 dm_1 + \\
\int_{0}^{n} u_1, q(m, n) \cdot \left( h_1(m|0) \cdot F_2(0, n) + F_1(0, n) \cdot h_2(m|0) \right) dm \right)
\]
will assume an unavailability cost of at least 100 times the inspection cost. This implies roughly that the probability of failure in the next period should not exceed 

\[ \int_0^l \int_0^l \pi(y_1, y_2) \cdot \left( \min{l_1, l_2} \cdot u_2, q(m, l_1) \cdot h_1(m_1 | y_1) \cdot h_2(m_2 | y_2) \right) \, dy_2 \, dy_1 \].

\( h_1(m_1 | 0) \cdot h_2(m_2 | y_2) \, dm_2 \, dm_1 + \int_0^l \int_0^l u_1, q(m, l_2) \cdot \left( h_1(m | 0) \cdot F_2(y_2, l_2) + F_1(0, l_2) \cdot h_2(m | y_2) \right) \, dm \) \, dy_2 \, dy_1

\( h_1(m_1 | y_1) \cdot h_2(m_2 | 0) \, dm_2 \, dm_1 + \int_0^l u_1, q(m, l_1) \cdot \left( h_1(m | y_1) \cdot F_2(0, l_1) + F_1(y_1, l_1) \cdot h_2(m | 0) \right) \, dm \) \, dy_2 \, dy_1

\[ + \sum_{l_1=1}^{\min{l_1, l_2}} \int_0^l \int_0^l \pi(y_1, y_2) \cdot \left( \min{l_1, l_2} \cdot u_2, q(m_1, m_2, l_1, l_2) \cdot h_1(m_1 | y_1) \cdot h_2(m_2 | y_2) \right) \, dy_2 \, dy_1 \]

\[ + \sum_{l_1=1}^{\min{l_1, l_2}} \int_0^l \int_0^l \pi(y_1, y_2) \cdot \left( \min{l_1, l_2} \cdot u_1, q(m, l_1, l_2) \right) \, dy_2 \, dy_1 \]

\[ + \sum_{l_1=1}^{\min{l_1, l_2}} \int_0^l \int_0^l \pi(y_1, y_2) \cdot \left( \min{l_1, l_2} \cdot F_1(y_1, \min{l_1, l_2}) \cdot h_2(m | y_2) \right) \, dm \] \, dy_2 \, dy_1.

3.5. Numerical investigation

For presentational purposes, we consider a system consisting of two identical components. In this way, the threshold values (which are the same for both components) are easier to optimize and the results are easier to interpret than for non-identical components. It also allows us to omit the superscripts denoting to which component a certain cost or threshold value corresponds. Figure 3.3 shows the failure probability over time of a component with \( \alpha = 3.5 \) and \( L = 2 \), provided that it does not undergo any maintenance actions and that it is as-good-as-new at time 0. It follows from Figure 3.3 that \( n \) (the number of inspection thresholds) should not be chosen too large, especially when considering a relatively high unavailability cost as is typically the case in practice. In our experiments, we will assume an unavailability cost of at least 100 times the inspection cost. This implies roughly that the probability of failure in the next period should not exceed
one percent. From Figure 3.3, we observe that the failure probability is much more than one percent for \( n \) larger than two. Hence, we set the number of inspection thresholds to \( n = 2 \), which means that the next inspection is always scheduled either one or two periods later.

Note that our computations are made using R [12] on a computer with a 3.30 GHz quad core processor and 8.00 GB of RAM. Due to the high complexity and considerable computing time, it is convenient to precalculate the integrals with respect to \( m_1, m_2, \) and \( m \) in Equation (3.1). Because these are independent from the threshold values and the cost parameters, we only need to calculate them once for all \( y_i \in [0, L_i], i = 1, 2 \) (which takes about 55 hours) for this specific setting of \( n = 2, L = 2, \) and \( \alpha = 3.5 \). If we also calculate \( \pi \) (which is independent of the cost parameters too) for all combinations of the threshold values (approximately 87 hours), the cost and availability criteria can be calculated for all possible combinations of the threshold values within 40 seconds for any setting of the cost parameters.

The cost scenario that we consider is partially based on [1]; inspection cost normalized at \( c_n = 1 \), preventive replacement cost \( c_p = 40 \) per component, corrective replacement cost \( c_c = 100 \) per component, set-up cost \( c_s = 35 \) per (system) replacement, and unavailability cost rate \( c_u = 150 \) per time unit. We do a full grid search, and calculate the value of \( C_\infty \) for all \( \xi_1, \xi_2, \zeta \in \{0, 0.1, \ldots, L\} \), with \( 0 \leq \xi_1 \leq \xi_2 \) and \( 0 \leq \zeta \leq \xi_2 \) in order to obtain the cost-minimizing threshold values.
We assess the performances using both the upper bound on the unavailability time, and our linear approximation. We refer to these as ‘upper bound’ and ‘linear approximation’, respectively, in the remainder of this section. Results indicate that the minimal long-run average cost per period is located somewhere around 29.96 when using the upper bound, and around 25.99 when using the linear approximation. We remark that these cost figures are not readily comparable (for determining cost savings) as they are based on different cost approximations, but the large difference does show that using a more accurate approximation significantly alters the results. In addition, the corresponding optimal threshold values are given by $\xi_1 = 0$, $\xi_2 = 1.0$, and $\zeta = 1.0$ for the upper bound, and by $\xi_1 = 1.3$, $\xi_2 = 1.3$, and $\zeta = 0.8$ for the linear approximation. These two maintenance policies differ in many aspects. Whereas by using the upper bound it is optimal to inspect each time unit, to replace a component at a deterioration level of 1, and to never perform opportunistic replacements, by using the linear approximation we find that the system is inspected every other period, a component is replaced preventively at a level of 1.3, and opportunistic replacements are performed at a deterioration level of 0.8. This implies that the accuracy of approximating the unavailability time has a great impact on the resulting optimal maintenance strategy as well. To gain more insight into the behavior of $C_\infty$ and the differences between the upper bound and the linear approximation, Figure 3.4 shows the minimal value of $C_\infty$ for different (fixed) values of each one of the threshold values. From this figure, it appears that the minimal costs estimated by using both the upper bound and the linear approximation behave quite similarly, although the minimal costs obtained by the upper bound are clearly higher than those obtained by the linear approximation. Furthermore, we observe that increasing any one of the thresholds $\xi_1$, $\xi_2$, and $\zeta$ has a greater impact on the minimal costs based on the upper bound than those based on the linear approximation. This emphasizes the importance of approximating the unavailability time accurately. Besides, Figure 3.4 illustrates that both the inspection threshold $\xi_1$ and the opportunistic replacement threshold $\zeta$ should not be set too high. This can be explained by the fact that the preventive replacement threshold $\xi_2$ should exceed these two thresholds, forcing the number of preventive replacements to decrease as well. On the other hand, setting the preventive replacement threshold $\xi_2$ too low causes the maintenance cost to increase, as maintenance is then performed too often.
Figure 3.4. The minimal value of $C_\infty$ for different values of $\xi_1$, $\xi_2$, and $\zeta$. 
3.5.1. Comparison to classical maintenance policies

As noted in [1], many classical maintenance policies can be viewed as special cases of our multi-threshold maintenance policy. We compare the results of our policy with several of these for the particular example that we consider.

**No opportunistic replacements:** Set $\zeta$ equal to $\xi_2$ to omit opportunistic replacements.

**Periodic inspections:** Inspections are performed periodically (with periodicity $n = 2$) by setting $\xi_1$ equal to $\xi_2$.

**Failure-based maintenance:** Under the assumption that failures can only be noticed upon inspection, failure-based maintenance can be achieved by setting both $\xi_2$ and $\zeta$ equal to the failure level $L$ such that no preventive maintenance is performed.

**Block replacement:** Block replacement is a strategy in which all components are replaced periodically. This policy can be obtained by setting all thresholds to zero.

The minimal long-run average costs obtained with the linear approximation (upper bound), along with the optimal threshold values, for each of the above maintenance strategies are shown in Table 3.2 and summarized in Figure 3.5. If it is optimal to perform periodic inspections, the corresponding periodicity is also presented in the table. Note that for the upper bound, the optimal solution obtained with our multi-threshold maintenance policy does not deviate from the one obtained by omitting opportunistic replacements. This is due to the fact that in this particular example it is optimal to not include opportunistic replacements. Similarly, for the linear approximation it turns out that periodic inspections are optimal in this particular case. Including opportunistic maintenance, and therefore

<table>
<thead>
<tr>
<th>Maintenance policy</th>
<th>$C_{\infty}$</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$\zeta$</th>
<th>Periodicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-threshold policy</td>
<td>25.99 (29.96)</td>
<td>1.3 (0.0)</td>
<td>1.3 (1.0)</td>
<td>0.8 (1.0)</td>
<td>2 (-)</td>
</tr>
<tr>
<td>No opportunistic replacements</td>
<td>26.76 (29.96)</td>
<td>1.2 (0.0)</td>
<td>1.2 (1.0)</td>
<td>1.2 (1.0)</td>
<td>2 (-)</td>
</tr>
<tr>
<td>Periodic inspection</td>
<td>25.99 (30.25)</td>
<td>1.3 (1.1)</td>
<td>1.3 (1.1)</td>
<td>0.8 (0.7)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Failure-based maintenance</td>
<td>51.17 (72.53)</td>
<td>1.9 (2.0)</td>
<td>2.0 (2.0)</td>
<td>2.0 (2.0)</td>
<td>- (2)</td>
</tr>
<tr>
<td>Block replacement</td>
<td>58.78 (59.64)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>
Figure 3.5. The minimal value of $C_\infty$, obtained with different maintenance policies.

performing a system-wide optimization, is essential for this example.

### 3.5.2. Sensitivity analysis

**Influence of the set-up cost**

So far, we assumed a set-up cost of $c_s = 35$. To investigate the influence of the degree of economic dependence, we now vary this cost between 0 and 50. Figure 3.6 shows the minimal long-run average cost per period for these values of the set-up cost. As expected, the costs obtained by the linear approximation are lower than those obtained by using the upper bound. Furthermore, increasing the set-up cost has a larger effect on the costs obtained with the upper bound than with the linear approximation, because a higher set-up cost implies less preventive maintenance, and hence a higher unavailability time. In addition, Figure 3.7 shows the long-run average availability $A_\infty$ corresponding to the cost-minimizing threshold values for different values of the set-up cost. It decreases as the set-up cost increases, because maintenance will be performed less often.

Furthermore, Figure 3.8 shows the threshold values that minimize the long-run average cost for different values of the set-up cost, both using the upper bound and the linear approximation. This figure confirms that the preventive replacement threshold $\xi_2$ is consistently higher for the linear approximation than
for the upper bound. In the latter case, failures are penalized more severely due to the overestimation of the system downtime. Preventive maintenance is thus performed at an earlier stage. It also appears that the inspection threshold \( \xi_1 \) is equal to \( \xi_2 \) for a wider range of set-up costs under the linear approximation, meaning that fewer inspections are performed. Related, there are more opportunistic replacements under the linear approximation. Under this more accurate estimation of the downtime, the model thus shows an increased ability to respond to the degree of economic dependence, by clustering more maintenance activities.

**Influence of the unavailability cost rate**

Next, we vary the unavailability cost rate between 100 and 200. This results in the minimal costs \( C_\infty \) shown in Figure 3.9 with the corresponding system availability shown in Figure 3.10. Naturally, as all other cost parameters remain unchanged, the minimal long-run average cost per time unit increases as the unavailability cost rate increases. Furthermore, a higher unavailability cost rate implies more frequent preventive maintenance to avoid expensive downtime. This holds in particular for the upper bound, where the system downtime is overestimated. The long-run average system availability corresponding to the cost-minimizing threshold values thus increases as \( c_u \) increases, especially for the upper bound.

Furthermore, Figure 3.11 shows the threshold values that minimize the long-run average cost per period for the different values of the unavailability cost rate,
Figure 3.8. Cost-minimizing threshold values for different values of the set-up cost, obtained with the upper bound and the linear approximation.
obtained with the upper bound and the linear approximation. Similar to the case where we varied the set-up cost, the preventive replacement threshold is set higher in case the linear approximation is used than with the upper bound, resulting in less frequent preventive maintenance. The same holds for the opportunistic replacements, as long as the unavailability cost rate does not exceed 140. This causes the inspection threshold $\xi_1$ to drop to zero for the upper bound, implying no opportunistic replacements. For the linear approximation, however, inspections are performed every other time unit, and both preventive and opportunistic replacements are performed more often when the unavailability cost rate increases.

We remark that these sensitivity results were also observed for other values of the deterioration parameters $\alpha_i$, the failure levels $L_i$, and the cost parameters, for $i = 1, 2$.

### 3.6. Conclusion

In this chapter, we built on the work of Castanier et al. [1], who developed an advanced CBM policy for a two-component series system subject to economic dependence, where the aperiodic inspection moments are optimized simultaneously with the critical condition levels at which maintenance is performed.
Figure 3.11. Cost-minimizing threshold values for different values of the unavailability cost rate, obtained with the upper bound and the linear approximation.
Whereas only the long-run average maintenance cost per period was considered as a performance criterion in [1], we considered the long-run average system availability as well. Since the deterioration level of a component, and hence whether or not a failure has occurred, can only be observed at the inspection moments, the amount of time that a component is unavailable cannot be measured exactly. An upper bound is used in [1], but we approximated it more accurately by assuming a gradual, linear increase in deterioration between two consecutive time units. Results indicate that this greatly influences the resulting optimal maintenance strategy. Using an upper bound, the overestimated unavailability time causes both inspections and preventive replacements to be performed too often, reducing the profitability of opportunistic replacements.

A numerical sensitivity study revealed insights on the trade-off between different types of maintenance actions. Both the inspection thresholds and the opportunistic replacement threshold should not be set too high, as this forces the number of preventive replacements to reduce as well. At the same time, the preventive replacement threshold should not be set too low, since maintenance is then performed too often, which increases the maintenance costs. By selecting the right thresholds, our policy was shown to outperform simpler, classical maintenance policies. In fact, a number of these classical policies can be viewed as special cases of our policy, making it widely applicable and of value to the maintenance literature.

Since in practice systems often contain more than two components, for which different structural relations exist, a direction for future research is to extend this model to a $k$-out-of-$N$-system, i.e., the case where a system consisting of $N$ components functions as long as at least $k$ components function [13]. Other relevant extensions of the system considered here include uncertain deterioration failure levels, dependent deterioration processes for the different components, and the inclusion of predetermined periods during which maintenance activities are preferably scheduled such as turn arounds. However, the current analysis is already complex and has a considerable computing time. This is partly due to the fact that no efficient optimization approach exists to find the optimal threshold values, forcing us to do a full grid search. Although we can deal with the long computing time by dividing the calculations into different parts and running them separately, future research could also address alternative ways to analyze the stationary law.
References


