Condition-based maintenance optimization based on matrix algebra

de Jonge, Bram

Published in:
Operations Research Letters

DOI:
10.1016/j.orl.2021.08.005

Operations Research Letters 49 (2021) 741–747

Contents lists available at ScienceDirect

Operations Research Letters

www.elsevier.com/locate/orl

Condition-based maintenance optimization based on matrix algebra

Bram de Jonge [1]

Department of Operations, Faculty of Economics and Business, University of Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 28 February 2021
Received in revised form 17 June 2021
Accepted 11 August 2021
Available online 17 August 2021

Keywords:
Maintenance
Condition-based maintenance
Planning time
Matrix algebra
Markov chain

ABSTRACT

We develop a matrix algebra approach for optimizing condition-based maintenance with a planning time for gradually deteriorating, continuously monitored equipment. This avoids the calculation times of simulation and the approximations of more direct approaches. We consider a model in which maintenance after failure should be planned as well, resulting in downtime, and a model with emergency repairs upon failure. Numerical examples are provided to compare the optimal policies of the two models and their performances under different parameter settings.

1. Introduction

Industrial equipment is subject to deterioration because of usage and exposure to the environment. Without intervening, this deterioration ultimately leads to failure of the equipment, after which a corrective maintenance action is required. However, in most cases, it is preferred to avoid failures by carrying out maintenance preventively. Preventive maintenance interventions can be prescribed based on the equipment age. However, ongoing developments in sensor technologies are enabling the possibility to monitor the condition of industrial equipment in real-time [6]. The resulting flow of information can be used to carry out maintenance based on the actual state of equipment; this type of maintenance is called condition-based maintenance (CBM).

Much research has been done in the area of maintenance optimization. We refer to [4] for a recent broad review and an explanation of the most important terminology. In this study we consider condition-based maintenance optimization for single-unit systems, i.e., maintenance interventions apply to the system as a whole. Various approaches exist to model the deterioration in such a setting, ranging from the most simple delay-time model that only adds one additional state in between the functioning state and the failed state, to detailed models with a deterioration level that is expressed on a continuous scale. In this study we adopt the latter, and we consider gradual and stochastic deterioration. We assume that a sensor is installed that allows for continuous monitoring of the deterioration level. Furthermore, we make the realistic assumption that maintenance requires a planning time. This is often realistic in practice, for instance because personnel must be scheduled, tools need to be arranged, and spare parts need to be ordered [8,12].

The number of existing studies that consider continuous monitoring of gradually deteriorating equipment is limited. Studies that do consider such a setting assume that preventive maintenance is scheduled when a deterioration threshold is exceeded, and that preventive maintenance is carried out after a fixed planning time. Bérenguer et al. [2] assume that corrective maintenance can also only be carried out at the end of the planning time, implying that failure during the planning time results in downtime. Maintenance durations are assumed to be stochastic and to depend on the deterioration level at the moment of maintenance. The aim is to minimize the long-run unavailability. Because exact analysis requires integration that is numerically burdensome, approximations are used. Saassouh et al. [17] extend this with a switch to a deterioration mode with a higher deterioration rate after a random amount of time. They distinguish a separate deterioration threshold for the higher deterioration rate. Simulation is used to approximate the two deterioration thresholds that simultaneously minimize the unavailability. Mercier and Castro [15] assume that an imperfect repair is carried out after the planning time. When failure occurs before the end of the planning time, or when the effect of the imperfect repair is deemed too small, the system is replaced. Markov renewal equations are derived to determine the long-run cost rate. Because solving these equations turned out to be too complicated, simulation is used. De Jonge et al. [5] consider negligible maintenance durations and aim to minimize the long-run cost rate. They study the effect of various practical factors on the performance of CBM; one of these factors is a required fixed

E-mail address: b.de.jonge@rug.nl.
[1] Postal address: P.O. Box 800, 9700 AV Groningen, The Netherlands.

https://doi.org/10.1016/j.orl.2021.08.005

planning time for carrying out preventive maintenance. An immediate emergency repair is assumed to be carried out when failure occurs before the end of the planning time. Simulation is used to approximate the optimal preventive maintenance threshold.

The main contribution of this paper is methodological. Whereas the above mentioned studies on maintenance optimization for a continuously monitored gradually deteriorating unit either use simulation or an approximated model, we point out how matrix algebra can be used to determine the optimal deterioration threshold for planning preventive maintenance. The main advantage of using matrix algebra is that it results in a much simpler analysis, making it mathematically more appealing. From a practical point of view, it facilitates the implementation of condition-based maintenance in decision support systems that automatically notify when maintenance should be planned based on real-time sensor measurements [1]. This study thereby contributes to a further application of condition-based maintenance in practice. Furthermore, the use of matrix algebra avoids the long computation times and the stability issues that simulations involve. We note that our approach is exact if we assume a discrete-time deterioration model with a finite number of states, whereas an approximation is obtained if we discretize a continuous-time continuous-state deterioration model. However, the quality of this approximation can be improved by using smaller time steps and a larger number of deterioration states in the discretization.

Because we are the first to analyze condition-based maintenance with a planning time by using matrix algebra, we consider two relatively simple settings. In the first setting corrective maintenance also needs to be planned; in the second setting an immediate emergency repair is carried out upon failure. We believe that future studies can extend the use of matrix algebra to more comprehensive condition-based maintenance models.

The remainder of this paper is organized as follows. In Section 2, we formally describe the problem that we consider. Inspired by the aforementioned studies, we consider both the policy for which corrective maintenance also requires a planning time, and the policy with an emergency repair upon failure. In Section 3, we consider the setting in which maintenance can be carried out without a planning time. This is a prelude to the main analysis with a planning time in Section 4. In Section 5, we apply the analysis to deterioration according to the gamma process, which is commonly used to model gradual deterioration. This section includes a comparison of the two policies and an analysis on the effects of both the length of the required planning time and the amount of volatility in the deterioration process. We end with some concluding remarks in Section 6.

2. Problem description

We consider condition-based maintenance optimization for a single-unit system in a discrete-time setting. The unit has $m+1$ ordered deterioration states and deteriorates according to a discrete-time Markov chain. State 1 is the best (i.e. as-good-as-new) state, state m the most deteriorated but functioning state, and state $m+1$ the failed state. The state of the unit is fully described by its deterioration state. The transition probability matrix of this Markov chain is denoted by the $(m+1)\times(m+1)$ matrix P. The state of the unit cannot improve without carrying out maintenance, implying that P is upper diagonal. Furthermore, the Markov chain is an absorbing Markov chain and state $m+1$ is the absorbing state.

The unit fails when the final deterioration state $m+1$ is reached, after which corrective maintenance is required. Preventive maintenance can be carried out as long as failure has not occurred. Both maintenance types are assumed to require a negligible amount of time and to be perfect, i.e., they bring the unit back to the as-good-as-new state. The deterioration process of the unit is monitored continuously. In a discrete-time setting this means that the deterioration level is observed at the start of each time period. On the other hand, a condition-based maintenance model with inspections in a discrete-time setting would imply that it has to be decided whether to inspect at the start of each time period. Any upper diagonal transition probability matrix P of the Markov chain that models deterioration can be chosen. In Section 5 we specify this matrix by discretizing continuous-time continuous-state stationary gamma processes. We use such a gamma process because it is commonly used to model deterioration. Furthermore, because it is fully described by two parameters, it eases the specification of the transition probability matrix of the Markov chain.

The preventive maintenance policy that we consider is the control-limit policy with preventive maintenance threshold M [9]. In Section 3 we assume that maintenance can be carried out without a planning time. In this case the control-limit policy means that preventive maintenance is carried out when deterioration level M is reached or exceeded. Failure occurs when the deterioration level jumps from a deterioration state below M to the failed state $m+1$, followed by immediate corrective maintenance. The cost of preventive maintenance is denoted by c_{pm}, and the cost of corrective maintenance by c_{cm}.

In Section 4 we assume that a fixed planning time s is required for carrying out preventive maintenance. In this case, M denotes the deterioration level at which preventive maintenance is planned. We distinguish two versions of this model. In the first version we assume that corrective maintenance can also only be carried out after the planning time. If maintenance has not been planned at the moment of failure, the planning time starts upon failure. In this version a downtime cost c_{d} is incurred per time period of downtime between failure and corrective maintenance. In the second version we assume that an immediate emergency repair is carried out upon failure. The cost of such an emergency repair is denoted by c_{er}, and is typically higher than the corrective maintenance cost c_{cm}.

In all versions of the model we consider the aim is to minimize the long-run cost rate by choosing the optimal value for the preventive maintenance threshold M. Because each maintenance action brings the unit back to the as-good-as-new state, the long-run cost rate for given M can be calculated based on standard renewal theory [16]. We refer to the time between two consecutive maintenance actions as a cycle, and we calculate the long-run cost rate $\eta(M)$ as a function of M as the mean cost per cycle, denoted by $C(M)$, divided by the mean cycle length, denoted by $D(M)$. I.e., $\eta(M) = C(M)/D(M)$.

3. Instantaneous maintenance

The setting without a planning time is considered by De Jonge [3]. In this section we briefly summarize the relevant analysis and we introduce additional notation. For more details and numerical examples we refer to [3]. The setting with a planning time that we consider in the next section builds on the analysis and (new) notation in this section.

The $(m+1)\times(m+1)$ transition probability matrix P of the Markov chain that models the deterioration of the unit can be written as

$$P = \begin{pmatrix} Q & \mathbf{r} \\ \mathbf{0} & 1 \end{pmatrix}.$$

Herein, Q is an $m\times m$ matrix, $\mathbf{r}$ a column vector of length m, and $\mathbf{0}$ a row vector with zeros of length m. The Markov chain is an absorbing Markov chain with state $m+1$ the absorbing state. The other states are transient states and the corresponding fundamental matrix R equals

$$R = \sum_{k=0}^{\infty} Q^k = (I_m - Q)^{-1},$$

in which I_m is the $m \times m$ identity matrix. Entry (i, j) of matrix R equals the expected number of time periods that we are in transient state j starting from transient state i, before reaching the absorbing state. The first row of matrix R is important for our analysis. Because the deterioration process is non-decreasing and because maintenance brings us back to the as-good-as-new state, we have that, for any preventive maintenance threshold M, entry $(1, j)$, $j \in \{1, \ldots, M-1\}$, of matrix R equals the expected number of time periods that the deterioration state is j before reaching or exceeding threshold M.

We introduce the m-dimensional column vector $\mathbf{h}$ of which we let entry $M \in \{1, \ldots, m\}$ equal the expected number of time periods, starting from the as-good-as-new state, until preventive maintenance threshold M is reached or exceeded:

$$h_M = \sum_{j<M} R_{1,j}.$$

For $M = 1$ we get an empty sum, implying that $h_1 = 0$. In the current setting in which maintenance does not require a planning time, we have that the mean cycle length as a function of the preventive maintenance threshold M equals $D(M) = h_M$.

Furthermore, we introduce the m-dimensional column vector $\mathbf{q}$ of which entry $M \in \{1, \ldots, m\}$ equals the probability that a cycle ends with failure, given preventive maintenance threshold M. Thus, it is the probability of a one-step transition from a deterioration state below M to the failed state $m+1$, i.e.,

$$q_M = \sum_{j<M} R_{1,j} P_{j,m+1} = \sum_{j<M} R_{1,j} r_j.$$

Subsequently, a cycle ends with preventive maintenance with probability $1 - q_M$. This implies that the mean cost per cycle as a function of M equals

$$C(M) = c_{\text{cm}} q_M + c_{\text{pm}}(1 - q_M) = c_{\text{pm}} + (c_{\text{cm}} - c_{\text{pm}}) q_M.$$

We now have that the cost rate $\eta(M)$, as a function of the preventive maintenance threshold M, equals

$$\eta(M) = \frac{C(M)}{D(M)} = \frac{c_{\text{pm}} + (c_{\text{cm}} - c_{\text{pm}}) q_M}{h_M}.$$

Because $h_1 = 0$, this cost rate is not defined for preventive maintenance threshold $M = 1$. This would imply maintenance to be carried out immediately after the previous maintenance action, resulting in infinite costs. We let $M^* \in \arg\min_M \eta(M)$ denote the optimal preventive maintenance threshold, and $\eta^* = \eta(M^*)$ the corresponding optimal cost rate.

4. Planned maintenance

In the setting with a planning time, preventive maintenance is planned when deterioration level M is reached or exceeded. Because the deterioration process makes jumps and typically jumps over level M, the planning time generally starts at a deterioration level above M. We will use the $m \times m$ matrix V in which entry (M, j), $M, j \in \{1, \ldots, m\}$, equals the probability that, given preventive maintenance threshold M, preventive maintenance will be planned (i.e., there is no sudden failure without preventive maintenance having been planned already) and the planning time starts in deterioration state j. This means that our analysis considers all possible preventive maintenance thresholds M simultaneously. If $M = 1$, preventive maintenance will be planned immediately after the previous preventive maintenance action, implying that the deterioration level will always be 1 at the start of the planning time. If $M > 1$, the probability that the planning time starts at a deterioration level below M is obviously 0. Furthermore, the values $R_{1,i}$, $i = 1, \ldots, M-1$, indicate the expected number of times that we are in deterioration state i before the planning time starts. The matrix P specifies the probabilities of moving from one of these deterioration states to a deterioration state of at least M. It follows that V is specified as

$$V_{M,j} = \begin{cases} 1, & \text{if } M = 1 \text{ and } j = 1, \\ \sum_{i=1}^{M-1} R_{1,i} P_{i,j}, & \text{if } M > 1 \text{ and } j \geq M, \\ 0, & \text{otherwise.} \end{cases}$$

We note that the row sums of V are smaller than 1 because it is also possible to transit from a deterioration state below M to the failed state, i.e., without preventive maintenance having been planned. The corresponding probabilities are denoted by q_M (see Section 3), and we thus have

$$\sum_{j=1}^{m} V_{M,j} + q_M = \sum_{j=M}^{m} V_{M,j} + q_M = 1,$$

for all $M \in \{1, \ldots, m\}$, or, using matrix notation, $V\mathbf{1} + \mathbf{q} = \mathbf{1}$, in which $\mathbf{1}$ is the m-dimensional column vector of ones. We continue to introduce the $m \times m$ matrix S in which entry (i, j) indicates the expected number of time periods that we are in deterioration state j during the planning time, given that the deterioration state at the start of the planning time is i. This matrix equals

$$S = \sum_{i=0}^{s-1} Q^i = (I_m - Q)^{-1}(I - Q^s).$$

The definition of S is very similar to that of R (see Section 3); however, we sum to $s-1$ instead of ∞ because of the finite planning time s. It follows that entry (M, j) of the $m \times m$ matrix product VS equals the expected number of time periods that we are in deterioration state j given a preventive maintenance threshold M. Each time period that the deterioration state is j, the probability that we transit to the failed state equals r_j. It follows that the m-dimensional column vector $VS\mathbf{r}$ contains the probabilities of failure during the planning time for each preventive maintenance threshold M. Thus, for preventive maintenance threshold M, the probability that preventive maintenance will be planned and that failure occurs during the planning time equals

$$(VS\mathbf{r})_M.$$

Likewise, the column vector $VS\mathbf{1}$ contains the expected number of time periods that the unit is functioning during the planning time. Thus, for preventive maintenance threshold M, this expected duration equals

$$(VS\mathbf{1})_M.$$

We continue to distinguish two versions of the model with a planning time, as explained in Section 2. In Section 4.1 we assume that corrective maintenance can also only be carried out after the planning time. Thereafter, in Section 4.2, we assume that an emergency repair is carried immediately after failure.

4.1. Corrective maintenance after the planning time

In this version of the model there will be costs c_{pm} and c_{cm} for preventive and corrective maintenance actions, respectively.

Corrective maintenance can also only be carried out after the planning time, and a downtime cost c_d is incurred per time period in between failure and corrective maintenance. In this setting with planned corrective maintenance (pcm) we let $C_{\text{pcm}}(M)$ denote the mean cost per cycle, $D_{\text{pcm}}(M)$ the mean cycle length, and $\eta_{\text{pcm}}(M)$ the cost rate, all as functions of the preventive maintenance threshold M.

For preventive maintenance threshold M, we have that the probability of a sudden failure before preventive maintenance has been planned equals q_M, and that the probability of failure during the planning time equals $(VS\mathbf{r})_M$. This implies that the probability that a cycle ends with failure equals $q_M + (VS\mathbf{r})_M$. Furthermore, the expected amount of downtime during a cycle equals $s - (VS\mathbf{1})_M$. It follows that the mean cost per cycle equals

$$C_{\text{pcm}}(M) = c_{\text{pm}} + (c_{\text{cm}} - c_{\text{pm}})(q_M + (VS\mathbf{r})_M) + c_d(s - (VS\mathbf{1})_M).$$

The mean cycle length equals the expected amount of time until reaching or exceeding deterioration level M, plus the planning time s:

$$D_{\text{pcm}}(M) = h_M + s.$$

It follows that the cost rate as a function of the preventive maintenance threshold M equals

$$\eta_{\text{pcm}}(M) = \frac{c_{\text{pm}} + (c_{\text{cm}} - c_{\text{pm}})(q_M + (VS\mathbf{r})_M) + c_d(s - (VS\mathbf{1})_M)}{h_M + s}. \quad (1)$$

We let $M^*_{\text{pcm}} \in \arg\min_M \eta_{\text{pcm}}(M)$ denote the optimal preventive maintenance threshold for this policy, and $\eta^*_{\text{pcm}} = \eta_{\text{pcm}}(M^*_{\text{pcm}})$ the corresponding optimal cost rate.

4.2. Emergency repair upon failure

In this version of the model an immediate emergency repair will be carried out after failure, implying that there is no downtime. Such an emergency repair also brings the unit back to the as-good-as-new state. The cost of an emergency repair is denoted by c_{er}, and that of preventive maintenance again by c_{pm}. In this setting we let $C_{\text{er}}(M)$ denote the mean cost per cycle, $D_{\text{er}}(M)$ the mean cycle length, and $\eta_{\text{er}}(M)$ the cost rate, again all as a function of the preventive maintenance threshold M.

Similar to the model with planned corrective maintenance, we have that the probability that a cycle ends with failure given preventive maintenance threshold M equals $q_M + (VS\mathbf{r})_M$. This implies that the mean cost per cycle equals

$$C_{\text{er}}(M) = c_{\text{pm}} + (c_{\text{er}} - c_{\text{pm}})(q_M + (VS\mathbf{r})_M).$$

The mean cycle length equals the expected amount of time until reaching or exceeding deterioration level M, plus the expected amoumt of time between planning preventive maintenance and either failure or the end of the planning time:

$$D_{\text{er}}(M) = h_M + (VS\mathbf{1})_M.$$

It follows that the cost rate as a function of the preventive maintenance threshold M equals

$$\eta_{\text{er}}(M) = \frac{c_{\text{pm}} + (c_{\text{er}} - c_{\text{pm}})(q_M + (VS\mathbf{r})_M)}{h_M + (VS\mathbf{1})_M}. \quad (2)$$

We let $M^*_{\text{er}} \in \arg\min_M \eta_{\text{er}}(M)$ denote the optimal preventive maintenance threshold for this policy, and $\eta^*_{\text{er}} = \eta_{\text{er}}(M^*_{\text{er}})$ the corresponding optimal cost rate.

5. Numerical examples

In this section we use stationary gamma processes to model deterioration. The gamma process is rather flexible and applicable to model a wide variety of deterioration processes [23]. We use the following definition for the density function f of the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$:

$$f_{\alpha,\beta}(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} t^{\alpha-1} e^{-\frac{t}{\beta}}, \ \ t > 0,$$

in which $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}\,dz$ denotes the gamma function. The stationary gamma process has a shape function at with shape parameter $a > 0$ and a scale parameter $b > 0$. It is a stochastic process $\{X(t) : t \in \mathbb{R}_+\}$ with the following properties:

1. $X(0) = 0$ with probability 1;
2. $X(\tau) - X(t) \sim f_{a(\tau-t),b}$ for $\tau > t \geq 0$; and
3. $X(t)$ has independent increments.

We assume that failure occurs if a certain fixed deterioration level L is exceeded. Examples of other recent studies that also adopt a gamma deterioration process with a fixed failure level are [7,20,21].

We use the approach described in De Jonge [3] to discretize the stationary gamma process. This approach prescribes to choose a number of deterioration states before failure m and a time step Δt, and results in an $(m+1) \times (m+1)$ transition probability matrix P. Arbitrarily good approximations can be obtained by choosing m sufficiently large and Δt sufficiently small. In the continuous-time model we use s to refer to the planning time and c_d to refer to the downtime cost per unit of time. In the discretized model each time step represents a time period Δt in the continuous-time model. This means that, in the discretized model, the planning time equals $s/\Delta t$ time periods, and the downtime cost equals $c_d \Delta t$ per time period. The results in this section are obtained based on the discretized model and on the analysis in the previous section, and are presented after transforming them back to the continuous-time model. Furthermore, we present specific problem instances in this section to illustrate the phenomena and insights that we describe, but we want to stress that other problem instances result in similar insights.

For any preventive maintenance threshold M, the mean cycle length $D_{\text{pcm}}(M)$ of the policy with planned corrective maintenance is longer than the mean cycle length $D_{\text{er}}(M)$ of the policy with emergency repairs. Furthermore, if the cost of an emergency repair c_{er} equals the cost of planned corrective maintenance c_{cm}, and if the downtime cost equals $c_d = 0$, then, again for any preventive maintenance threshold M, the mean cost per cycle $C_{\text{pcm}}(M)$ of the policy with planned corrective maintenance equals that of the policy with emergency repairs $C_{\text{er}}(M)$. If $c_{\text{er}} > c_{\text{cm}}$ and $c_d = 0$, then $C_{\text{pcm}}(M) < C_{\text{er}}(M)$ for any M. It follows that if $c_d = 0$ and $c_{\text{er}} \geq c_{\text{cm}}$, for any M, the cost rate $\eta_{\text{pcm}}(M)$ in case of planned corrective maintenance is lower than the cost rate $\eta_{\text{er}}(M)$ in case of emergency repairs. This is then obviously also the case for the optimal cost rates: $\eta^*_{\text{pcm}} < \eta^*_{\text{er}}$ if $c_d = 0$ and $c_{\text{er}} \geq c_{\text{cm}}$.

The above is illustrated by Fig. 1, which shows cost rates of the two policies as functions of the preventive maintenance threshold M for a discretized gamma deterioration process with $a = 2$ and $b = 0.5$, failure level $L = 1$, planning time $s = 0.2$, preventive maintenance cost $c_{\text{pm}} = 1$, and corrective maintenance cost $c_{\text{cm}} = 3$. The gamma process is discretized by using $m = 100$ deterioration states and time step $\Delta t = 0.01$. The bottom line in the figure represents the policy with planned corrective maintenance with downtime cost $c_d = 0$, whereas the line directly above it represents the policy

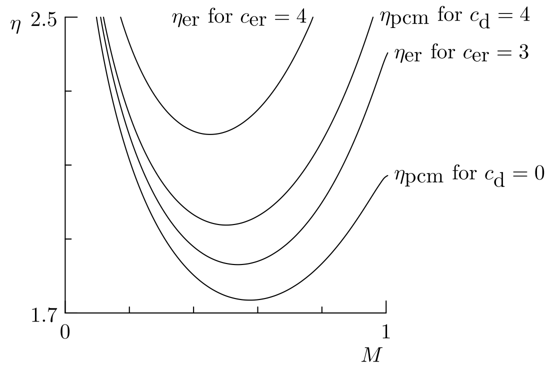

Fig. 1. Cost rates for the policies with planned corrective maintenance η_{pcm} and with emergency repairs η_{er} as a function of the maintenance threshold M; for a gamma deterioration process with $a = 2$, $b = 0.5$, $L = 1$, a planning time $s = 0.2$, and maintenance costs $c_{\text{pm}} = 1$, $c_{\text{cm}} = 3$.

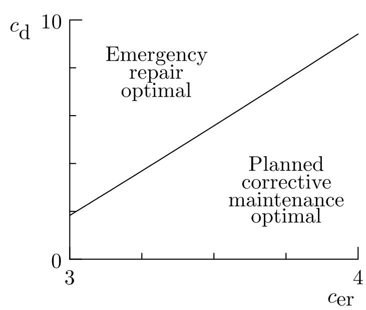

Fig. 2. Optimal policy for combinations of an emergency repair cost c_{er} and a downtime cost c_{d} for a gamma deterioration process with $a = 2$, $b = 0.5$, $L = 1$, a planning time $s = 0.2$, and maintenance costs $c_{\text{pm}} = 1$, $c_{\text{cm}} = 3$.

with emergency repairs with an emergency repair cost that equals the cost of planned corrective maintenance, i.e., $c_{\text{er}} = c_{\text{cm}} = 3$. The latter indeed results in a higher cost rate for all M.

The cost rate $\eta_{\text{pcm}}(M)$ of the policy with planned corrective maintenance is increasing in the downtime cost c_{d} for any M; see (1) and the line η_{pcm} for $c_{\text{d}} = 4$ in Fig. 1. This implies that the optimal cost rate η^*_{pcm} of this policy is also increasing in c_{d}. It follows that the policy with emergency repairs is preferred over the policy with planned corrective maintenance if the downtime cost c_{d} exceeds some threshold. In turn, the cost rate $\eta_{\text{er}}(M)$ of the policy with emergency repairs is increasing in the emergency repair cost c_{er} for any M; see (2) and the line η_{er} for $c_{\text{er}} = 4$ in Fig. 1. This implies that the optimal cost rate η^*_{er} of this policy is also increasing in c_{er}. Thus, the threshold value for c_{d} for which the policy with emergency repairs becomes optimal is increasing in the emergency repair cost c_{er}. Fig. 2 visualizes this for the problem instance at hand.

By using our matrix algebra approach it takes approximately 0.3 s to carry out the calculations for drawing one of the lines in Fig. 1, and this will increase to approximately 4.9 s if $m = 1{,}000$ deterioration states are considered instead of $m = 100$ states. The calculation time that is needed to do this by using simulation depends on the number of values for M that are considered, the simulation horizon, and the computer that is used, but this could easily take 15 min [3]. We note that these performances are similar for all four lines, and also for other problem instances.

Fig. 3 shows the effect of changing the planning time s on the optimal cost rates η^*_{pcm} and η^*_{er} of the two policies (a), and on the corresponding optimal preventive maintenance thresholds M^*_{pcm} and M^*_{er} (b). This figure is based on a gamma deterioration process with parameters $a = 1$ and $b = 1$, and with failure threshold $L = 1$. The maintenance costs are $c_{\text{pm}} = 1$, $c_{\text{cm}} = 3$, $c_{\text{d}} = 8$, and $c_{\text{er}} = 4$, i.e., an emergency repair is more expensive than corrective maintenance and there is a positive downtime cost.

If the time that is needed for planning preventive maintenance gets longer, preventive maintenance will obviously already be planned at a lower deterioration level. Furthermore, because the actual course of the deterioration cannot be used anymore during the planning time, the optimal cost rate is increasing in the length of the planning period. A longer planning time involves potentially long downtimes and thereby high failure costs for the policy with planned corrective maintenance. For the policy with emergency repairs, on the other hand, the cost of a failure is always the same, regardless of the amount of planning time. A longer planning time therefore has a greater impact on the policy with planned corrective maintenance. This also emerges from Fig. 3 in which η^*_{pcm} increases faster than η^*_{er}, and M^*_{pcm} decreases faster than M^*_{er}. Planned preventive maintenance is preferred if the planning time is relatively short. The policy with immediate emergency repairs upon failure becomes optimal if the planning time exceeds some threshold. This threshold depends on the downtime cost and the emergency repair cost, relative to the preventive and corrective maintenance costs. A higher downtime cost c_{d} makes emergency repairs more attractive, resulting in a lower threshold. A higher emergency repair cost, on the other hand, favors planned corrective maintenance, resulting in a higher threshold.

We continue to consider the effect of the volatility of the deterioration process. We choose combinations of values for the parameters a and b of the gamma process such that the mean time to failure (MTTF) equals 1. Determining such combinations is not straightforward, we refer to [5] for details on this procedure. We let $\sigma = \sqrt{a} \cdot b$ denote the standard deviation of the level of deterioration at time 1, this can be seen as a measure for the amount of volatility in the deterioration process.

Fig. 4 shows how the optimal cost rates η^*_{pcm} and η^*_{er} respond to a varying level of volatility in the deterioration process. The figure is based on a preventive maintenance cost $c_{\text{pm}} = 1$ and a corrective maintenance cost $c_{\text{cm}} = 3$. Furthermore, in part (a), the downtime cost is $c_{\text{d}} = 10$ and the emergency repair cost is $c_{\text{er}} = 3.5$, whereas these respectively are $c_{\text{d}} = 5$ and $c_{\text{er}} = 4$ in part (b). The planning time is $s = 0.2$.

If the amount of volatility in the deterioration process is low, i.e., if the deterioration behavior is very stable, failures are very well predictable. This means that preventive maintenance can be scheduled very effectively, namely just before failure, resulting in optimal cost rates at comparatively low levels for both policies. For higher levels of volatility, failures become less predictable, resulting in increasing cost rates for both policies. Furthermore, a clear preference for one of the two policies is clearly visible, independent of the exact amount of volatility. If the downtime cost c_{d} is relatively high, failures are relatively expensive under the policy with planned corrective maintenance, and the policy with emergency repairs is preferred, see Fig. 4 (a). If, on the other hand, the emergency repair cost c_{er} is relatively high, the policy with planned corrective maintenance is preferred, see Fig. 4 (b). Both cost rates converge for high levels of volatility σ, meaning that the difference between the optimal cost rates of the two policies stabilizes. If σ is very high, failures are caused by sudden very large deterioration increments. Such failures cannot be avoided by preventive maintenance, implying that maintenance is only carried out reactively (either corrective maintenance or emergency repairs). For the policy with planned corrective maintenance the cost per cycle is then $s \cdot c_{\text{d}} + c_{\text{cm}}$, and the mean cycle length $MTTF + s$, implying that the cost rate η^*_{pcm} converges to $(s \cdot c_{\text{d}} + c_{\text{cm}})/(MTTF + s)$ when σ increases. For the policy with emergency repairs the cost per cycle is c_{er}, and the mean cycle length is $MTTF$, resulting in a cost rate η^*_{er} that converges to

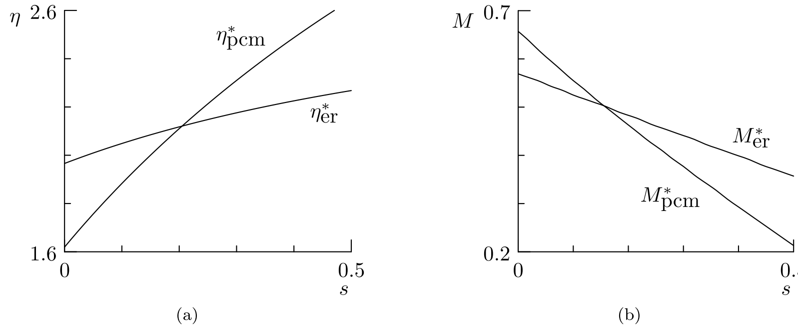

Fig. 3. Optimal cost rates for the policies with planned corrective maintenance $\eta^*_{\rm pcm}$ and with emergency repairs $\eta^*_{\rm er}$ (a) and corresponding optimal maintenance thresholds $M^*_{\rm pcm}$ and $M^*_{\rm er}$ (b), as a function of the planning time s; for a gamma deterioration process with $a = 1$, $b = 1$, $L = 1$, and maintenance costs $c_{\rm pm} = 1$, $c_{\rm cm} = 3$, $c_{\rm d} = 8$, $c_{\rm er} = 4$.

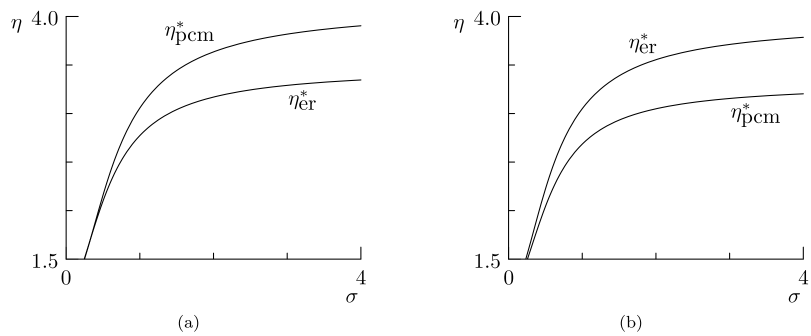

Fig. 4. Optimal cost rates for the policies with planned corrective maintenance $\eta^*_{\rm pcm}$ and with emergency repairs $\eta^*_{\rm er}$ as a function of the level of volatility in the deterioration process σ for maintenance costs $c_{\rm pm} = 1$, $c_{\rm cm} = 3$, $c_{\rm d} = 10$, $c_{\rm er} = 3.5$ (a) and for maintenance costs $c_{\rm pm} = 1$, $c_{\rm cm} = 3$, $c_{\rm d} = 5$, $c_{\rm er} = 4$ (b); for a gamma deterioration process with $MTTF = 1$ and $L = 1$, and a planning time $s = 0.2$.

$c_{\rm er}/MTTF$ when σ increases. These converging values can be verified based on Fig. 4.

6. Concluding remarks

The matrix algebra approach proposed in this paper may be used more broadly in the area of maintenance optimization, leading to a number of suggestions for future research. Instead of a fixed planning time, one could also consider a stochastic duration. This is realistic if maintenance is carried out as quickly as possible after its initiation, or if the delivery time for spare components is unpredictable. The extension with a stochastic planning time is relatively simple if an exponential planning time is adopted, but becomes more complicated if another distribution is chosen. Also, if the main reason for having a planning time is that spare components need to be ordered, one could extend the analysis based on matrix algebra to a model with two deterioration thresholds; one for ordering a spare component, and a second one for carrying out maintenance after the arrival of the spare components [10,19]. The rationale is that spare components can wait for maintenance much easier than hired personnel.

Another possible extension is to consider imperfect maintenance [14]. The effect of maintenance could be stable over time, but one could also assume that the effect of maintenance becomes smaller in the number of maintenance interventions [13]. After a while, replacement of the unit would then be required. Regarding the failure behavior of the unit, instead of assuming a fixed deterioration level at which failure occurs, stochasticity in this failure level could be considered, for instance by incorporating a failure rate that depends on the deterioration level [5,22]. Also environmental factors that influence the deterioration behavior could be taken into account [11,18]. The final suitable extension that we would like to mention is to consider a single policy that combines planned corrective maintenance and emergency repairs. Emergency repairs are likely to be preferred in case of failure either before or in an early phase of the planning time, whereas it is expected that waiting for the already planned maintenance intervention is better if there is not that much time left.

References

[1] R. Ahmad, S. Kamaruddin, An overview of time-based and condition-based maintenance in industrial applications, Comput. Ind. Eng. 63 (2012) 135–149.
[2] C. Bérenguer, A. Grall, L. Dieulle, M. Roussignol, Maintenance policy for a continuously monitored deteriorating system, Probab. Eng. Inf. Sci. 17 (2003) 235–250.
[3] B. De Jonge, Discretizing continuous-time continuous-state deterioration processes, with an application to condition-based maintenance optimization, Reliab. Eng. Syst. Saf. 188 (2019) 1–5.
[4] B. De Jonge, P.A. Scarf, A review on maintenance optimization, Eur. J. Oper. Res. 285 (2020) 805–824.
[5] B. De Jonge, R.H. Teunter, T. Tinga, The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance, Reliab. Eng. Syst. Saf. 158 (2017) 21–30.
[6] Q. Feng, J.G. Shanthikumar, How research in production and operations management may evolve in the era of big data, Prod. Oper. Manag. 27 (2018) 1670–1684.
[7] M.J.A. Havinga, B. De Jonge, Condition-based maintenance in the cyclic patrolling repairman problem, Int. J. Prod. Econ. 222 (2020).
[8] S.M. Iravani, V. Krishnamurthy, Workforce agility in repair and maintenance environments, Manuf. Serv. Oper. Manag. 9 (2007) 168–184.
[9] R. Jiang, Optimization of alarm threshold and sequential inspection scheme, Reliab. Eng. Syst. Saf. 95 (2010) 208–215.
[10] G.P. Kiesmüller, F.E. Sachs, Spare parts or buffer? How to design a transfer line with unreliable machines, Eur. J. Oper. Res. 284 (2020) 121–134.
[11] M. Kurt, J.P. Kharoufeh, Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment, Oper. Res. Lett. 38 (2010) 273–279.
[12] R. Li, J.K. Ryan, A Bayesian inventory model using real-time condition monitoring information, Prod. Oper. Manag. 20 (2011) 754–771.
[13] H. Liao, E.A. Elsayed, L.Y. Chan, Maintenance of continuously monitored degrading systems, Eur. J. Oper. Res. 175 (2006) 821–835.
[14] Y. Liu, Y. Chen, T. Jiang, Dynamic selective maintenance optimization for multi-state systems over a finite horizon: a deep reinforcement learning approach, Eur. J. Oper. Res. 283 (2020) 166–181.

[15] S. Mercier, I.T. Castro, On the modelling of imperfect repairs for a continuously monitored gamma wear process through age reduction, J. Appl. Probab. 50 (2013) 1057–1076.
[16] S.I. Resnick, Adventures in Stochastic Processes, Springer, 2013.
[17] B. Saassouh, L. Dieulle, A. Grall, Online maintenance policy for a deteriorating system with random change of mode, Reliab. Eng. Syst. Saf. 92 (2007) 1677–1685.
[18] I.B. Sidibé, A. Khatab, C. Diallo, K.H. Adjallah, Kernal estimator of maintenance optimization model for a stochastically degrading system under different operating environments, Reliab. Eng. Syst. Saf. 147 (2016) 109–116.
[19] E. Topan, A.S. Eruguz, W. Ma, M.C. Van Der Heijden, R. Dekker, A review of operational spare parts service logistics in service control towers, Eur. J. Oper. Res. 282 (2020) 401–414.
[20] M.A.J. Uit het Broek, R.H. Teunter, B. De Jonge, J. Veldman, Joint condition-based maintenance and condition-based production optimization, Reliab. Eng. Syst. Saf. 214 (2021) 107743.
[21] M.A.J. Uit het Broek, R.H. Teunter, B. De Jonge, J. Veldman, N.D. Van Foreest, Condition-based production planning: adjusting production rates to balance output and failure risk, Manuf. Serv. Oper. Manag. 22 (2020) 792–811.
[22] A. Van Horenbeek, L. Pintelon, A dynamic predictive maintenance policy for complex multi-component systems, Reliab. Eng. Syst. Saf. 120 (2013) 39–50.
[23] J.M. Van Noortwijk, A survey of the application of gamma processes in maintenance, Reliab. Eng. Syst. Saf. 94 (2009) 2–21.