A Bivariate Maintenance Policy for Multi-State Repairable Systems With Monotone Process

Mimi Zhang, Min Xie, Fellow, IEEE, and Olivier Gaudoin

Abstract—This paper proposes a sequential failure limit maintenance policy for a repairable system. The objective system is assumed to have $k + 1$ states, including one working state and $k$ failure states, and the multiple failure states are classified potentially by features such as failure severity or failure cause. The system deteriorates over time and will be replaced upon the $N$th failure. Corrective maintenance is performed immediately upon each of the first $(N - 1)$ failures. To avoid the costly failure, preventive maintenance actions will be performed as soon as the system’s reliability drops to a critical threshold $R$. Both preventive maintenance and corrective maintenance are assumed to be imperfect. Increasing and decreasing geometric processes are introduced to characterize the efficiency of preventive maintenance and corrective maintenance. The objective is to derive an optimal maintenance policy $(R^*, N^*)$ such that the long-run expected cost per unit time is minimized. The analytical expression of the cost rate function is derived, and the corresponding optimal maintenance policy can be determined numerically. A numerical example is given to illustrate the theoretical results and the maintaining procedure. The decision model shows its adaptability to different possible characteristics of the maintained system.

Index Terms—Geometric process, multiple failure states, quasi-renewal process, sequential failure limit policy.

ACRONYM

| PM  | preventive maintenance |
| CM  | corrective maintenance  |
| GP  | geometric process      |
| QRP | quasi-renewal process  |

NOTATION

- $F(t)$: working time distribution function of the system at the beginning of a renewal cycle
- $G(t)$: corrective maintenance duration distribution function of a newly installed system
- $u$: the mean of the corrective maintenance duration distribution
- $l_n$: the length of the $n$th repair cycle, $n = 1, \ldots, N$
- $I_n$: the length of the $n$th renewal cycle
- $N$: the number of corrective maintenance activities in a renewal cycle
- $C_p$: cost of each preventive maintenance
- $C_f$: corrective maintenance cost rate due to the down-time
- $c_s$: failure damage of failure state $s$, $s = 1, \ldots, k$
- $W$: the length of a renewal cycle
- $a$: the impact factor of the preventive maintenance on the system’s lifetime distribution
- $b$: the impact factor of the preventive maintenance on the duration of corrective maintenance
- $a_s$: the impact factor on the system’s lifetime distribution by each occurrence of failure state $s$, $s = 1, \ldots, k$
- $b_s$: the impact factor on the duration of corrective maintenance by each occurrence of failure state $s$, $s = 1, \ldots, k$
- $p_s$: the probability of the occurrence of the $s$th-type failure, $s = 1, \ldots, k$
I. INTRODUCTION

For a repairable deteriorating system, the most advanced maintenance strategies rely on the monitoring of a measurable degradation process of the system and making maintenance decisions according to the monitoring information. Generally, such a condition-based maintenance policy is more efficient than a maintenance policy based on the system age and on the knowledge of the underlying lifetime distribution. However, building a degradation database is very expensive, and most systems lack the capability of acquiring sensor-based information due to the non-existence of suitable quality characteristics. This is, however, not the case with failure time data which can be easily retrieved from historical maintenance records (Gebraeel et al. [1]). In view of this result, this paper investigates a maintenance strategy which is applicable to systems which are inaccessible to the condition-based maintenance program. Note that the preliminarily used lifetime distributions can be updated by the subsequent real-time sensory signals.

It is well-known that in most cases maintenance is imperfect, restoring a system’s operating state to somewhere between as good as new and as bad as old ([2]–[4]). For a repairable system, a common phenomenon is that the successive operating durations after repairs demonstrate a decreasing trend, whereas the durations of the successive repairs show an increasing trend. To model the maintenance efficiency with this characteristic, Lam [5] introduced a geometric process (GP) repair model. It should be noted that Wang and Pham [6] proposed a quasi-renewal process (QRP) to characterize maintenance efficiency. Conceptually different as they are, the GP and the QRP share the same theoretical basis. As Lam put this idea out first, we use the notation “geometric process” in the following. The GP and the QRP models have been studied by many researchers. The readers are referred to [7]–[10] for more details.

The GP and the QRP were originally introduced to characterize the effectiveness of corrective maintenance (CM). Characterizing the effectiveness of preventive maintenance (PM) using the geometric process, however, has received surprisingly little attention until very recently. Wang and Pham [11] approached the assessing of PM efficiency by a quasi-renewal process in which the degradation critical threshold is raised proportionally after each PM. Doyen and Gaudoin [12] also pointed out that the quasi-renewal process can be generalized to imperfect PM by introducing a positive parameter characterizing the PM efficiency. Yet, this article just presents a general framework and does not give much detailed discussion on this generalized model.

In our paper, we propose a maintenance strategy, including both imperfect CM and imperfect PM, and assess the efficiency of these two maintenance actions both in the context of GP. As the PM and the CM are only different in the sense that the PM is planned whereas the CM is unplanned, it is reasonable to characterize PM and CM efficiency both using the GP, with different factors.

In recent years, there has been growing interest on multi-state reliability models and optimization algorithms. Derman [13] is among the first to investigate a multi-state model, where the deterioration of the system was described as the movement from state to state by a Markov chain. Ding et al. [14] and Li and Kapur [15] investigated a new perspective to model the states of the components and systems using fuzzy set theory. More works on multi-state systems can be found in [16]–[19]. A special multi-state system, which we will deliberate in this paper, is a degenerative system having $k + 1$ states with one working state and $k$ different failure states (Zhang and Wang [20]). The occurrences of these $k$ failure states are stochastic and mutually exclusive. The $k$-failure-states model can be visualized in many contexts. For example, in the context of gracefully degrading systems, a failure can be classified by its severity, and thus the cost related to each failure type is different. Another example is that a failure can be classified by its cause, and thus the treatment related to each failure type is different. Lam et al. [21] studied a monotone process repair model for such systems with $k$ failure modes. They applied a replacement policy based on the number of failures of the system in which the system state after repair cannot be statistically as good as new. They showed that the repair model for the multi-state degenerative system forms a general monotone process repair model which includes the GP repair model as a special case. Several extensions of this multi-state model have been investigated in Zhang et al. [22] and Zhang and Wang [20]. However, the multi-state failure system subjected to imperfect CM and imperfect PM, both in the form of GP, is a blank topic of interest.

Conventional PM is scheduled periodically, and it often holds the same time interval for all PM actions (Sheu et al. [23]). However, because the PM is generally imperfect and cannot restore the system to as good as new, the age-dependent or periodic PM policy is unavoidably ineffective. On the other hand, as Wang [24] pointed out, an optimal maintenance policy must be based not only on cost measures but also on reliability measures. For multi-component systems, minimizing system maintenance cost may not imply maximizing system reliability because various components in the system may have different maintenance costs and different reliability importance. Sometimes when the maintenance cost is minimized the system reliability may be so low that the optimal maintenance policy is not acceptable in practice. This problem leads to the sequential failure limit PM policy presented herein. Under the sequential failure limit PM policy, the PM actions are scheduled such that the system reliability does not drop below a critical threshold. Specifically, after installation, the objective system is preventively maintained whenever the system reliability reaches a predetermined threshold, and the imperfect PM restores the system condition to somewhere between as good as new and as bad as old. Many literature works focus on reliability centered PM policies. Zhou et al. [25] integrated a sequential imperfect maintenance policy into a condition based predictive maintenance policy, and an imperfect PM is performed whenever the system reliability reaches a predetermined threshold. In Liao et al. [26], it is assumed that the system’s reliability could be monitored continuously, and an imperfect PM is performed whenever the system’s reliability reaches a predetermined threshold. Doyen and Gaudoin [12] also mentioned a sequential failure limit PM policy in which the system is preventively maintained as soon as a reliability indicator exceeds a predetermined threshold.

The maintenance strategy proposed in this paper is a combination of the failure limit policy and the repair number counting policy (Wang [24]) and includes them as special cases. Upon
each of the first \((N-1)\) failures, CM is immediately performed to keep the system operating. The system will be replaced by a new identical one upon its \(N\)th failure. PM is performed whenever the system’s reliability drops to a predetermined critical threshold \(R\). Both the CM and the PM are assumed to be imperfect; that is, the successive working durations of the system after repairs form a stochastically decreasing process. We further assume that the consecutive corrective repair durations form a stochastically increasing process. Because PM actions are pre-scheduled, the PM times are assumed to be negligible. The policy decision variables are \(R\) and \(N\). The purpose is to determine the optimal policy parameters \((R^*, N^*)\) such that the long-run expected cost per unit time is minimized.

The rest of the paper is organized as follows. In Section II, we introduce a monotone process model for a deteriorating system, which has one working state and \(k\) failure states. In Section III, we derive an analytical expression of the cost rate function. The optimal policy \((R^*, N^*)\) minimizing the cost rate function can be determined numerically or analytically. Section IV illustrates the proposed procedure via a simulation study. The last section concludes the paper.

II. MODEL ASSUMPTIONS

For easy reference, we give the definitions of stochastic order and geometric process as follows (see Ross [27]).

**Definition 1:** Given two real-valued random variables \(Z_1\) and \(Z_2\), \(Z_1\) is said to be stochastically larger than \(Z_2\), which is denoted by \(Z_1 \geq_{st} Z_2\) or \(Z_2 \leq_{st} Z_1\), if

\[
P(Z_1 > z) \geq P(Z_2 > z), \quad \text{for all real } z.
\]

We say that a stochastic process \(\{Z_n, n = 1, 2, \ldots\}\) is stochastically decreasing if \(Z_n \geq_{st} Z_{n+1}\), and stochastically increasing if \(Z_n \leq_{st} Z_{n+1}\), for all \(n = 1, 2, \ldots\).

**Definition 2:** Assume that \(\{Z_n, n = 1, 2, \ldots\}\) is a stochastic process composed of statistically independent non-negative random variables. If we have relationship \(P(Z_n \leq z) = P(Z_1 \leq \xi^{n-1} z)\) for some factor \(\xi > 0\), then \(\{Z_n, n = 1, 2, \ldots\}\) is called a geometric process. Furthermore, if \(\xi > 1\), \(\{Z_n, n = 1, 2, \ldots\}\) is called a stochastically decreasing geometric process; that is,

\[
Z_n \geq_{st} Z_{n+1}, \quad \text{for all } n = 1, 2, \ldots.
\]

If \(0 < \xi < 1\), \(\{Z_n, n = 1, 2, \ldots\}\) is called a stochastically increasing geometric process; that is,

\[
Z_n \leq_{st} Z_{n+1}, \quad \text{for all } n = 1, 2, \ldots.
\]

If \(\xi = 1\), the geometric process reduces to a renewal process.

The model assumptions are listed below.

**Assumption 1:** At the beginning, a new system is installed and will be replaced upon its \(N\)th failure by a new physically and statistically identical one. The duration of replacement is negligible.

The time interval between the \((m-1)\)th replacement and the \(m\)th replacement of the system is called the \(m\)th renewal cycle, \(m = 1, 2, \ldots\). Let \(T_m(m = 1, 2, \ldots)\) denote the time interval between the \((m-1)\)th replacement and the \(m\)th replacement of the system. Obviously, according to Assumption 1, \(\{T_1, T_2, \ldots\}\) forms a renewal process.

**Assumption 2:** The system is a multi-state system with one working state and \(k\) failure states. We use \(0, 1, 2, \ldots, k\) to denote the working state, the first-type failure state, the second-type failure state, \(\ldots\), and the \(k\)th-type failure state. If the system fails, the system will be in the \(s\)th-type failure state with probability \(p_s\), where we have \(p_s \geq 0\) for \(s = 1, 2, \ldots, k\), and \(\sum_{s=1}^{k} p_s = 1\).

An implication behind Assumption 2 is that the probability of the occurrence of each failure state does not change with time. This scenario can be visualized in the context of, for example, a multi-component system under competing risks. Specifically, for a multi-component system under competing risks, the failure of each component will lead directly to the failure of the system, and the failed component will be replaced by a new identical component. The failure state is classified by the failed component. For repairable multi-component systems under competing risks, most of the existing research assumes statistical independency of component failure [28]. Under the assumption of statistical independency of component failure and the assumption of perfect replacement, the probability of the occurrence of each failure state can be treated as constant.

Because we can always arrange the \(k\) failure states in order of failure severity, we might assume that failure state \((s + 1)\) is more serious than failure state \(s\) for \(s = 1, 2, \ldots, k - 1\). Thus, state 1 is a failure state with the lowest severity while state \(k\) is a failure state with the highest severity.

**Assumption 3:** The CM is conducted immediately upon the system’s failure. Between two consecutive CMs, the PM is conducted as soon as the system’s reliability drops to the critical threshold \(R\). Each PM is assumed to take negligible time while the duration of each CM cannot be neglected. The system will be replaced by a new identical one upon its \(N\)th failure. Both PM and CM are assumed to be imperfect. Each CM is assumed to have a factor \(a\) on the working time distribution, and a factor \(b\) on the CM duration distribution. Each CM is assumed to have a factor \(a_s\) on the working time distribution, and a factor \(b_s\) on the CM duration distribution, where \(s = 1, 2, \ldots, k\) indicates the failure state.

In a renewal cycle, the time interval between the completion of the \((n-1)\)th corrective maintenance and the completion of the \(n\)th corrective maintenance is called the \(n\)th repair cycle of the system, \(n = 1, 2, \ldots, N\). Let random variable \(V_n\) denote the number of PM actions performed in the \(n\)th repair cycle. In the \(n\)th repair cycle, \(n = 1, 2, \ldots, N\), let \(x_1, x_2, \ldots, x_N\) denote, respectively, the working time between the \((j-1)\)th PM and the \(j\)th PM \((j = 1, \ldots, N)\), the working time between the \(V_n\)th PM and the failure, the duration of the CM, and the length of the \(n\)th repair cycle. A possible course of the \(n\)th repair cycle is shown in Fig. 1, and a schematic of the \(n\)th renewal cycle is shown in Fig. 2.

**Assumption 4:** \(\{X_1, X_2, \ldots, X_N\}\) and \(\{Y_1, Y_2, \ldots, Y_N\}\) are two statistically independent processes.

**Assumption 5:** Each PM contributes a factor \(a \geq 1\) on the remaining useful life distribution, and a factor \(b \leq 1\) on the CM duration distribution. Each CM, with failure state \(s \in \{1, 2, \ldots, k\}\), contributes a factor \(a_s \geq 1\) on the remaining
useful life distribution, and a factor $b_s < 1$ on the CM duration distribution.

Assume that the lifetime of the system at the beginning of a renewal cycle has distribution function $F(t), t \geq 0$. Hence, the reliability of the system is supposed to drop to the critical threshold $R$ at epoch $F^{-1}(1 - R)$, at which time a potential PM is scheduled. Here and throughout the paper, by the superscript $-1$, we mean the reciprocal of a function. If the system survives beyond the time point $F^{-1}(1 - R)$, the potential PM will be released; namely, we have $x_1^n = F^{-1}(1 - R)$. The remaining useful life of the system, right after the first PM in the first repair cycle, has distribution function $F(at), t > 0$. Parameter $a \geq 1$ implies that, as the system deteriorates, the maintained system, compared with a newly installed system, is more prone to fail. Suppose that the system survives beyond the $v_1$th PM point yet fails before the $(v_1 + 1)$th PM point; that is, $V_1 = v_1$. Then we have $x_1^n = F^{-1}(1 - R) + a^{-1}(j = 1, \ldots, v_1)$, and $X_1 = F(a^{-1}t)$. Right after the first CM, that is at the beginning of the second repair cycle, the remaining useful life of the system has distribution function $F(a^{a_s} t)$, where $s_1 \in \{1, 2, \ldots, k\}$ is the failure state in the first repair cycle with $1 \leq a_1 \leq a_2 \leq \ldots \leq a_k$. Parameter $a_s$ represents the impact on the system’s remaining useful life by each occurrence of failure state $s \in \{1, 2, \ldots, k\}$. By analogy, given $(V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_{n - 1} = s_{n - 1})$, in the $n$th repair cycle we have

\[
x_n^n = \frac{F^{-1}(1 - R)}{a^{\sum_{i=1}^{n-1} v_i + j - 1} \prod_{i=1}^{n-1} a_s},
\]

\[
P(X_n < t|V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_{n - 1} = s_{n - 1}) = \frac{F\left(a^{\sum_{i=1}^{n} v_i} \prod_{i=1}^{n-1} a_s t \right)}{1 - R},
\]

where $j = 1, \ldots, v_n, 0 < t < \tau_n, s_1, s_2, \ldots, s_{n-1} \in \{1, 2, \ldots, k\}$, and

\[
\tau_n = \frac{F^{-1}(1 - R)}{a^{\sum_{i=1}^{n} v_i} \prod_{i=1}^{n-1} a_s},
\]

Suppose that the CM duration of a newly installed system has cumulative distribution function $G(t), t \geq 0$, namely $G(t)$ is a generic cumulative distribution function of the CM duration. In the first repair cycle, given $V_1 = v_1$ and failure state $s_1$, the CM duration $Y_1$ has distribution function $G(b_{s_1}^{b_{s_1}} t)$, where $s_1 \in \{1, 2, \ldots, k\}$, and $0 < b_1 < \ldots < b_k < 1$. Parameter $b_s \leq 1$ implies that, as the system deteriorates, the CM durations form a stochastically increasing process. Parameter $b_s$ represents the impact on the CM duration by each occurrence of failure state $s \in \{1, 2, \ldots, k\}$. By analogy, given $(V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_n = s_n)$, in the $n$th renewal cycle we have

\[
P(Y_n < t|V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_n = s_n) = G\left(b^{\sum_{i=1}^{n} s_i} \prod_{i=1}^{n} b_s t \right)
\]

where $t > 0$, and $s_1, s_2, \ldots, s_n \in \{1, 2, \ldots, k\}$.

The following theorem shows that the CM duration of a newly installed system is a stochastically decreasing process and that the $\{Y_1, Y_2, \ldots, Y_n\}$ is a stochastically increasing process.

**Theorem 1:** For $n = 1, 2, \ldots, t > 0$, we have

\[
P(X_n > t) \geq P(X_{n+1} > t),
\]

and

\[
P(Y_n > t) \leq P(Y_{n+1} > t).
\]

The proof of this theorem is given in the Appendix.

Denote the cost of each occurrence of failure by $C_n (n = 1, 2, \ldots)$. We assume that $C_n$ are statistically independent and identically distributed random variables with distribution $P(C_n = c_s) = p_s$, where $p_s (s = 1, \ldots, k)$ is the occurring probability of the $s$th-type failure defined in Assumption 2. The corrective maintenance cost rate, due to down-time, is denoted by $c_f$. The replacement cost in each renewal cycle is $C$, and the cost for each PM is $C_p$. 
III. COST RATE FUNCTION UNDER POLICY \((R, N)\)

In this section, we give the expression of the expected cost rate function under the bivariate policy \((R, N)\) with critical threshold \(R\) and failure number \(N\). Our objective is to determine the optimal policy \((R^*, N^*)\) such that the long-run expected cost per unit time is minimized.

Let \(C(R, N)\) denote the expected cost rate under policy \((R, N)\). According to the renewal reward theorem (see Ross [29]), we have

\[
C(R, N) = \frac{\text{expected cost incurred in a renewal cycle}}{\text{expected length of the renewal cycle}} = \frac{EU}{EW}.
\]

According to the assumptions, we have

\[
W = \left[ \sum_{n=1}^{N} \sum_{j=1}^{N_Y} x_{n,j}^n + \sum_{n=1}^{N} X_n \right] + \sum_{n=1}^{N-1} Y_n.
\]

The first part of the right-hand side denotes the overall working time while the second part denotes the overall repair time in a renewal cycle. The total cost of a renewal cycle is given by

\[
U = C + \sum_{n=1}^{N} \sum_{j=1}^{N_Y} C_{n,j} + c_f \sum_{n=1}^{N-1} Y_n + \sum_{n=1}^{N} C_n.
\]

Thus, the expected cost rate \(C(R, N)\) can be written as

\[
C(R, N) = \frac{C + \sum_{n=1}^{N} \sum_{j=1}^{N_Y} (EY_n + \sum_{n=1}^{N-1} EC_n)}{\sum_{n=1}^{N} \sum_{j=1}^{N_Y} x_{n,j}^n + \sum_{n=1}^{N} E X_n + \sum_{n=1}^{N-1} E Y_n}.
\]

All the expectations in function \(C(R, N)\) are stated below, and their proofs are given in the Appendix.

\[
E \left( \sum_{n=1}^{N} C_n \right) = N \sum_{k=1}^{K} c_k p_k;
\]

\[
EV_n = \frac{R}{1 - R};
\]

\[
E[X_n] = \left( \frac{p_1}{a_1} + \ldots + \frac{p_k}{a_k} \right)^{n-1} (1 - R)^{n-1} \frac{\lambda(R) a^n}{(a - R)^n};
\]

\[
\lambda(R) = \int_0^a t dF(t), \quad a > R;
\]

\[
E(Y_n) = (1 - R)^{n-1} \frac{u b^n}{(b - R)^n} \left( \frac{p_1}{b_1} + \ldots + \frac{p_k}{b_k} \right)^n;
\]

\[
u = \int b dG(t), \quad b > R;
\]

\[
E \left( x_{n,j}^2 \right) = \left( \frac{p_1}{a_1} + \ldots + \frac{p_k}{a_k} \right)^{n-1} (1 - R)^{n-1} \frac{a^n}{(a - R)^n} \times \frac{(R^j)^{n-1}(1 - R)}{a^{j-1}}.
\]

For simplicity of reference, we define \(A\) and \(B\) as

\[
A = \frac{p_1}{a_1} + \ldots + \frac{p_k}{a_k}, \quad B = \frac{p_1}{b_1} + \ldots + \frac{p_k}{b_k}.
\]

Theorem 2: Under policy \((R, N)\), the explicit expression of \(C(R, N)\) is

\[
C(R, N) = \frac{C + NC_p R/(1 - R) + c_f \Psi_3 + N \sum_{n=1}^{N} c_n p_n a^n}{\Psi_1 + \Psi_2 + \Psi_3},
\]

where

\[
\Psi_1 = \sum_{n=1}^{N} \sum_{j=1}^{N_Y} \left( \sum_{i=1}^{N} x_{n,j}^i \right)^n - R^{-1}(1 - R) \times a R \times \frac{1 - \left[ A(1 - R) \frac{a}{a - R} \right]^N}{1 - A(1 - R) \frac{a}{a - R}},
\]

\[
\Psi_2 = \sum_{n=1}^{N} E X_n - \frac{\lambda(R) a}{a - R} \times \frac{1 - \left[ A(1 - R) \frac{a}{a - R} \right]^N}{1 - A(1 - R) \frac{a}{a - R}};
\]

and

\[
\Psi_3 = \sum_{n=1}^{N} E Y_n - B(1 - R) \times \frac{ub}{b - R} \times \frac{1 - \left[ H(1 - R) \frac{b}{b - R} \right]^N}{1 - B(1 - R) \frac{b}{b - R}}.
\]

For a proof, see the Appendix.

From the expression of the cost rate function, we can see that the formation of the generic distribution function of the CM duration, \(G(\cdot)\), makes no difference as long as the mean \(u\) remains unchanged. The closed-form expression of the cost rate function makes the ensuing minimizing procedure easy to achieve, showing another advantage of the proposed maintenance strategy. There is a tremendous amount of literature on proposing various kinds of maintenance strategies. However, a formidable problem faced by most of them is the complexity of the optimization problem, and thus the arduous process in finding the optimal solution. Note that the bivariate optimization problem, \(\min_{N, R} C(R, N)\), is an integer optimization problem in \(N\), and is equivalent to the optimization problem \(\min_{N, R} \min_{I, J} C(R, N)\). Because \(N\) is an integer-valued variable, the latter optimization problem can be easily solved by solving two univariate minimization problems. Specifically, by fixing \(N\), the optimization problem \(\min_{R} C(R, N)\) can be solved by analytical or numerical methods; the optimal reliability is thus a function of \(N\), denoted as \(R^*_N\). By choosing \(N\) to be 1, 2, 3, we can find \(R^*_1, R^*_2, R^*_3, \ldots\) respectively such that the corresponding \(C(R^*_1, 1), C(R^*_2, 2), C(R^*_3, 3), \ldots\) are minimized. Because of the degeneration, the length of a renewal cycle is limited (see Wang and Pham [7]). Therefore, we can determine the global minimum of the cost rate function by comparing the values of \(C(R^*_1, 1), C(R^*_2, 2), C(R^*_3, 3), \ldots\).

Remark 1: Given the deteriorating trajectory \((V_1 = v_1, \ldots, V_{n-1} = v_{n-1}, S_1 = s_1, \ldots, S_{n-1} = s_{n-1})\), it is of interest to derive the survival function \(P(l_n > t) = \text{Pr}(l_n > t|V_1 = v_1, \ldots, V_{n-1} = v_{n-1}, S_1 = s_1, \ldots, S_{n-1} = s_{n-1})\), where \(l_n = \sum_{j=1}^{V_n} X_j + X_n\) denotes the time to failure of the \(n\)th repair cycle. Because

\[
x_n^j = \frac{F(1 - R)}{\sum_{i=1}^{N} a_i^{j-1}} \prod_{i=1}^{N} a_i^{-1}, \quad j = 1, 2, \ldots.
\]
for $0 \leq t < x_n^*$, we have

$$P(l_n > t) = 1 - F\left(\sum_{i=1}^{n-1} x_i a_i t\right);$$

and for $\sum_{i=1}^{n-1} x_i a_i t \leq t < \sum_{i=1}^{n-1} x_i a_i t$, we have

$$P(l_n > t) = \sum_{i=1}^{n-1} P(l_n > t | V_n = v_n) \times R^{v_n} (1 - R)\left(1 - R\right) = \left[1 - F\left(\sum_{i=1}^{n-1} x_i a_i t\right)\right] \times R^{v_n} (1 - R) + \sum_{i=1}^{n-1} R^{v_n} (1 - R) = \left[1 - F\left(\sum_{i=1}^{n-1} x_i a_i t\right)\right] \times R^{v_n} (1 - R) + R^2.$$

Another interesting measurement of system reliability is the mean time to failure (MTTF). By invoking the law of total expectation, the expression for the MTTF is given as

$$\text{MTTF} = E[V_n | V_1 = v_1, \ldots, V_{n-1} = v_{n-1}, V_n = v_n] = F \left(\sum_{i=1}^{n-1} x_i a_i t\right)\frac{(R(1-R))^{v_n}}{a \sum_{i=1}^{n-1} (1 - (1 - R)^{v_n})}.$$

Remark 2: The estimation of the unknown parameters in the distribution function $F(t)$ and the unknown factors can be carried out by the maximum likelihood method. The mean of the CM duration distribution, namely $\text{MTTF}$, can be assessed using experts’ judgment. One may argue that the assumption of constant factors is unrealistic and that factors changing with time might be more practical. However, on one hand, with the development of modern manufacturing technologies, each maintenance activity could restore a system to a good condition. In other words, the reliability of the system will not drop much between two consecutive maintenance activities. On the other hand, because a power $\zeta^n$ decreases ($\zeta < 1$) or increases ($\zeta > 1$) exponentially with $n$, the values of the factors $a$ and $b_a$ should be close to 1. Otherwise, the estimated system reliability will decrease rapidly, which discords with the stated fact about the maintenance. Likewise, the values of the factors $b$ and $b_a$ should be close to 1. From the above, we say that the factors are insensitive to time and that the assumption of constant factors makes sense. Moreover, the introduction of time-varying factors will make the parameter-estimation problem rather burdensome, which, on the contrary, deteriorates the model performance. Applications of the GP model to real-life problems can be found in Lam [30] and Lam and Chan [31].

IV. A NUMERICAL EXAMPLE

To demonstrate the maintenance program and the methodology developed in this paper, a numerical example is discussed in this section. In the following analysis, we assume that the lifetime distribution function $F(t)$ is known, and its parameters are given (estimated). For illustrative purpose, we consider a repairable deteriorating system with only three states: two failure states, and one working state $k = 2$.

Assume that, at the beginning of a renewal cycle, the lifetime of a new system has a Weibull distribution

$$F(t) = 1 - \exp \left(\frac{t}{\beta}\right)^{\alpha}$$

with parameters $\alpha = 1.5$ and $\beta = 2000$. By setting $p_1 = 0.45$, $p_2 = 0.55$, $a_1 = 1.03$, $a_2 = 1.1$, $a_2 = 1.2$, $b = 0.98$, $b_1 = 0.9$, $b_2 = 0.8$, we have

$$A = \frac{p_1}{a_1} + \frac{p_2}{a_2} = 0.7992,$$  

$$B = \frac{p_1}{b_1} + \frac{p_2}{b_2} = 1.3125.$$  

We assume that the other parameters are obtained as $C_N = 5000$, $c_f = 100$, $c_1 = 100$, $c_2 = 10000$, $C = 500000$, $y = 240$. For each value of $N$, we search for an optimal reliability $R_N^*$ which minimizes the cost rate function. Fig. 3 shows the results. The $x$-axis denotes the number of failures (repair cycles) in a renewal cycle, and the $y$-axis denotes the corresponding minimized cost rate. It is obvious that the optimal policy can be found uniquely from the cost rate surface.

Under the optimal maintenance policy $(R^*, N^*) = (0.6488, 6)$, in each renewal cycle the observations of

$$\{x_1, x_1^*, \ldots, x_1^{V_1}, x_2, x_2^*, \ldots, x_2^{V_2}, x_3, \ldots, x_6, x_6^*, \ldots, x_6^{V_6}, X_6\}$$

can be generated as follows. For the nth repair cycle ($n = 1, 2, \ldots, 6$), we randomly simulate a failure state $s_n$ with probabilities $P(s_n = 1) = p_1 = 0.45$ and $P(s_n = 2) = p_2 = 0.55$. Then, for $n = 1, 2, \ldots, 6$,
TABLE I
THE OPTIMAL RELIABILITY AND THE CORRESPONDING MINIMIZED EXPECTED LONG-RUN COST RATE FOR EACH VALUE OF N

<table>
<thead>
<tr>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_N</td>
<td>0.91</td>
<td>0.85</td>
<td>0.79</td>
<td>0.74</td>
<td>0.69</td>
<td>0.65</td>
<td>0.61</td>
<td>0.57</td>
</tr>
<tr>
<td>C(R_N, N)</td>
<td>163.57</td>
<td>106.53</td>
<td>89.01</td>
<td>81.75</td>
<td>78.86</td>
<td>78.31</td>
<td>79.15</td>
<td>80.84</td>
</tr>
</tbody>
</table>

We randomly simulate an observation of the remaining useful life from distribution function \( F(a \sum_{i=1}^{n-1} v_i + j - 1 \prod_{i=1}^{n-1} a_{s_i}, t) \). If the observation is larger than \( x_n^j = F^{-1}(1 - R^*)/(a \sum_{i=1}^{n-1} v_i + j - 1 \prod_{i=1}^{n-1} a_{s_i}) \), implying that the system survives beyond the point at which a potential PM is scheduled, then the PM is released, and the remaining useful life distribution function changes into \( F(a \sum_{i=1}^{n-1} v_i + j \prod_{i=1}^{n-1} a_{s_i}, t) \). If the observation is smaller than \( x_n^j \), implying that the system fails before the PM point, then we have \( V_n = j - 1 \), and the remaining useful life distribution function at the beginning of \((n + 1)\)th repair cycle is \( F(a \sum_{i=1}^{n-1} v_i \prod_{i=1}^{n-1} a_{s_i}, t) \).

Table II gives the observations of \{ \( \{X_n, V_n\}, n = 1, 2, \ldots, 6 \} \), in which there are 10 renewal cycles.

Define a cumulative statistic \( \Lambda_n = \sum_{i=1}^{n} X_i, n = 1, 2, \ldots \). Theorem 1 states that \( \{ X_1, X_2, \ldots, X_n \} \) is a stochastically decreasing process, namely \( P(X_n > t) \geq P(X_{n+1} > t), t > 0 \). Therefore, the increment of \( \Lambda_n - \Lambda_{n-1} \), that is \( X_n \), should present a decreasing trend. Note that \( X_n \) can be treated as the slope of \( \Lambda_n \), because we can re-write \( X_n \) as \( X_n = (\Lambda_n - \Lambda_{n-1})/(n - (n - 1)) \). Hence, the trajectory of \( \Lambda_n \) versus the number of repair cycles should by and large present a concave shape. To visually demonstrate the stochastically decreasing property of the process \{ \( \{X_1, X_2, \ldots, X_n \} \) \}, we plot the trajectory of the cumulative statistic in Fig. 5 with the sample size being 30. In Fig. 5, the \( x \)-axis represents the repair cycle, and the \( y \)-axis represents the cumulative statistic \( \Lambda_n \). The concave curves in Fig. 5 attest that the process \{ \( \{X_1, X_2, \ldots, X_n \} \) \} is stochastically decreasing.

To analyze the sensitivity of the optimal maintenance policy to the cost configurations, we vary one of the maintenance costs, fix the other costs, and investigate the evolution of the optimal long-run expected cost per unit time. The results show that, as the PM cost increases, the optimal critical threshold \( R^* \) decreases, leaving the system operating under an increasing risk of failure. When the CM cost rate increases, the CM action will be more expensive because of, e.g., the unavailability, production losses, and unplanned intervention. In this case, the optimal maintenance policy sets a high critical threshold, and involves more PM actions to avoid system unavailability. The replacement cost has its influence mainly on the optimal number of corrective maintenance actions. The more expensive the replacement is, the larger the optimal number of corrective maintenance actions will be.
TABLE II
THE WORKING DURATIONS \( \{X_1, X_2, \ldots, X_N\} \) AND THE PM NUMBERS \( \{V_1, V_2, \ldots, V_N\} \) WITH TEN RENEWAL CYCLES

<table>
<thead>
<tr>
<th>Renewal Cycle</th>
<th>((X_1, V_1))</th>
<th>((X_2, V_2))</th>
<th>((X_3, V_3))</th>
<th>((X_4, V_4))</th>
<th>((X_5, V_5))</th>
<th>((X_6, V_6))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(3845, 0)</td>
<td>(1348, 3)</td>
<td>(1427, 2)</td>
<td>(319, 0)</td>
<td>(472, 0)</td>
<td>(767, 4)</td>
</tr>
<tr>
<td>2</td>
<td>(591, 0)</td>
<td>(2016, 2)</td>
<td>(675, 6)</td>
<td>(683, 0)</td>
<td>(627, 9)</td>
<td>(296, 1)</td>
</tr>
<tr>
<td>3</td>
<td>(2715, 1)</td>
<td>(1274, 6)</td>
<td>(110, 7)</td>
<td>(83, 3)</td>
<td>(258, 1)</td>
<td>(491, 2)</td>
</tr>
<tr>
<td>4</td>
<td>(2481, 1)</td>
<td>(2251, 2)</td>
<td>(1515, 0)</td>
<td>(1078, 0)</td>
<td>(783, 4)</td>
<td>(498, 3)</td>
</tr>
<tr>
<td>5</td>
<td>(867, 3)</td>
<td>(1605, 4)</td>
<td>(1798, 0)</td>
<td>(289, 5)</td>
<td>(546, 5)</td>
<td>(716, 10)</td>
</tr>
<tr>
<td>6</td>
<td>(2297, 0)</td>
<td>(441, 4)</td>
<td>(1077, 6)</td>
<td>(4, 2)</td>
<td>(309, 1)</td>
<td>(73, 1)</td>
</tr>
<tr>
<td>7</td>
<td>(1651, 4)</td>
<td>(1553, 3)</td>
<td>(463, 10)</td>
<td>(939, 0)</td>
<td>(445, 3)</td>
<td>(351, 0)</td>
</tr>
<tr>
<td>8</td>
<td>(2141, 3)</td>
<td>(616, 0)</td>
<td>(1183, 4)</td>
<td>(512, 1)</td>
<td>(391, 0)</td>
<td>(508, 0)</td>
</tr>
<tr>
<td>9</td>
<td>(2328, 1)</td>
<td>(937, 2)</td>
<td>(904, 1)</td>
<td>(1685, 1)</td>
<td>(384, 1)</td>
<td>(757, 0)</td>
</tr>
<tr>
<td>10</td>
<td>(2503, 0)</td>
<td>(1561, 1)</td>
<td>(1396, 0)</td>
<td>(389, 3)</td>
<td>(918, 5)</td>
<td>(440, 1)</td>
</tr>
</tbody>
</table>

Fig. 5. Cumulative working time in each renewal cycle.

V. CONCLUDING REMARKS

In this paper, a multi-state deteriorating system is studied, and a bivariate maintenance policy, namely \((R, N)\), is developed. This multi-state system has \(k\) failure states and one working state. This is a general formulation, and many systems can be described under this framework with a suitable failure and state specification. Under some commonly used assumptions, we have derived an analytical expression of the cost rate function, and thus the optimal solutions can be readily obtained.

The maintenance strategy proposed herein is very flexible, including many maintenance strategies as special cases. The PM can be assumed to be perfect, in which case we have \(a = b = 1\). If the critical threshold is set to be zero, meaning no preventive maintenance will be performed, then our model reduces to the repair number counting policy. If we do not take into consider the failure states when we are making maintenance schemes, then we can set \(k = 1\). Because these maintenance policies are particular cases, it is obvious that the proposed maintenance strategy is more flexible and versatile. Possible extensions of the research are several.

1) We could analyze systems with multiple components and more than one working state.
2) The probabilities of occurrences of these \(k\) failure states could depend on the working time.
3) We could combine other imperfect maintenance treatment methods with a monotone process model, such as the virtual age method, the improvement factor method, the multiple \((p, q)\) rule, and so on.

APPENDIX

PROOF OF THEOREM 1

It is obvious that

\[ P(\{V_n = v_n\} = R^n(1 - R), \quad v_n = 0, 1, 2, \ldots) \]

and that

\[ P(\{X_n < t|V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_{n-1} = s_{n-1}\}) = \frac{F\left(\sum_{i=1}^{n} r_i \prod_{i=1}^{n-1} a_{s_i} t_i\right)}{1 - R} \]

Because the numbers of PM actions in each repair cycle are statistically independent random variables, we have

\[ P(\{V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_{n-1} = s_{n-1}\}) = \prod_{i=1}^{n} P(V_i = v_i) \prod_{i=1}^{n-1} P(S_i = s_i) \]

\[ = R^\sum_{i=1}^{n} r_i (1 - R)^{n-1} \prod_{i=1}^{n-1} p_{s_i} \]
For $0 < t < \tau_{n+1}$, we have

$$P(X_{n+1} \leq t) = \sum_{v_1=0}^{+\infty} \sum_{v_2=0}^{+\infty} \sum_{v_3=0}^{+\infty} \sum_{v_4=0}^{+\infty} \mathbb{G} \left( b^\sum_{i=1}^{n+1} v_i n^1 b_{s_i} t \right) \times R^\sum_{i=1}^{n+1} v_i (1 - R)^{n+1} \prod_{i=1}^{n+1} p_{s_i},$$

and conditional expectation

$$E[X_k|V_1 = v_1, \ldots, V_n = v_n, S_1 = s_1, \ldots, S_n = s_n] = \frac{\lambda(R)}{a^\sum_{i=1}^{n} v_i n^1 a_{s_i}},$$

where $\lambda(R) = \int_0^{R+1} t^1 \frac{dF(t)}{1 - R}$.

The constraint $a > R$ can be removed because $a > 1$, and $1 > R$. Similarly, we use the law of total expectation to derive the expected working time $E[X_n]$ and the expected CM duration $b_{s_i}(Y_n)$.

**DERIVATION OF THE EXPECTATIONS**

Expectations $E(\sum_{n=1}^{N} C_n)$ and $EV_n$ are easy to derive and are given below:

$$E\left( \sum_{n=1}^{N} C_n \right) = \sum_{n=1}^{N} E C_n = N^k \sum_{s=1}^{k} v_s p_s,$$

$$EV_n = \sum_{r=0}^{+\infty} r P(V_n = r) = \sum_{r=0}^{+\infty} r R^r (1 - R) = \frac{R}{1 - R}.$$
Because

\[ E[Y_1|V_1 - v_1, \ldots, V_n - v_n, S_1 = s_1, \ldots, S_n = s_n]\]

\[
= \int_0^{+\infty} t dG \left( b^{n-1} \prod_{i=1}^{n} b_i t \right)
= \frac{\int_0^{+\infty} t dG(t)}{b^{\sum_{i=1}^{n} s_i} \prod_{i=1}^{n} b_i}, \quad u = \int_0^{+\infty} t dG(t),
\]

we have

\[ E[Y_n] = \sum_{v_1 = 0}^{+\infty} \cdots \sum_{v_n = 0}^{+\infty} \prod_{i=1}^{n} b_i \]

\[
\times R\sum_{n=1}^{\infty} v_n (1 - R)^n \prod_{i=1}^{n} p_{s_i}
= (1 - R)^n \frac{n b^n}{(b - R)^n} \left( \frac{p_1}{b_1} + \cdots + \frac{p_k}{b_k} \right)^n, \quad b > R.
\]

**Proof of Theorem 2**

\[
\sum_{n=1}^{N} E X_n
= \sum_{n=1}^{N} A^{n-1} (1 - R)^{n-1} \frac{\lambda(R) a^n}{(a - R)^n}
= \frac{\lambda(R) a}{a - R} \times \frac{1 - \left[ A(1 - R) \frac{a}{a - R} \right]^N}{1 - A(1 - R) \frac{a}{a - R}};
\]

\[
\sum_{n=1}^{N} E Y_n
= \sum_{n=1}^{N-1} B^n (1 - R)^n \frac{a b^n}{(b - R)^n}
= B(1 - R) \frac{a b}{b - R} \times \frac{1 - \left[ B(1 - R) \frac{1}{b - R} \right]^{N-1}}{1 - B(1 - R) \frac{b}{b - R}};
\]

\[
\sum_{n=1}^{N} E \left( \sum_{j=1}^{V_n} x_j^2 \right)
= \sum_{n=1}^{N} \left( \sum_{j=1}^{V_n} x_j \right)^2
\]

**References**


Mimi Zhang received her B.S. (2011) in Statistics from University of Science and Technology of China. She is currently a Ph.D. student at the Department of Systems Engineering and Engineering Management, City University of Hong Kong. Her research interests include maintenance modeling, data completion, and Weibull.

Min Xie (M’91–SM’94–F’06) received his Ph.D. in Quality Technology in 1987 from Linkoping University in Sweden. He was awarded the prestigious LKY research fellowship in 1991, and currently he is a chair Professor at City University of Hong Kong. He has authored or co-authored numerous refereed journal papers, and some books on quality and reliability engineering, including Software Reliability Modelling by World Scientific Publisher, Weibull Models by John Wiley, Computing Systems Reliability by Kluwer Academic, and Advanced QFD Applications by ASQ Quality Press. He is an Editor of International Journal of Reliability, Quality and Safety Engineering, Department Editor of IIE Transactions, Associate Editor of IEEE TRANSACTIONS ON RELIABILITY, and on the editorial board of a number other international journals. He is a fellow of IEEE.

Olivier Gaudoin is a Professor at University Grenoble Alpes, France, where he teaches probability, statistics, and reliability. He received his Ph.D. in Applied Mathematics from Grenoble University in 1990. His research interests are probabilistic modelling and statistical analysis for the reliability of complex systems, including ageing and maintenance modelling, competing risks, software reliability, and random processes in reliability.