A Structure Property of Optimal Policies for Maintenance Problems With Safety-Critical Components

Li Xia, *Student Member, IEEE*, Qianchuan Zhao, *Member, IEEE*, and Qing-Shan Jia, *Member, IEEE*

***Abstract*—The maintenance problem with safety-critical components is significant for the economical benefit of companies. Motivated by a practical asset maintenance project, a new joint replacement maintenance problem is introduced in this paper. The dynamics of the problem are modelled as a Markov decision process, whose action space increases exponentially with the number of safety-critical components in the asset. To deal with the curse of dimensionality, we identify a key property of the optimal solution: the optimal performance can always be achieved in a class of policies which satisfy the so-called shortest-remaining-lifetime-first (SRLF) rule. It reduces the action space from $\mathrm{O}(2^n)$ to $\mathrm{O}(n)$, where n is the number of safety-critical components. To further speed up the optimization procedure, some interesting properties of the optimal policy are derived. Combining the SRLF rule and the neuro-dynamic programming (NDP) methodology, we develop an efficient on-line algorithm to optimize this maintenance problem. This algorithm can handle the difficulties of large state space and large action space. Besides the theoretical proof, the optimality and efficiency of the SRLF rule and the properties of the optimal policy are also illustrated by numerical examples. This work can shed some insights to the maintenance problems in a more general situation.**

***Note to Practitioners*—Motivated by a practical asset maintenance problem, we introduce a new joint replacement maintenance model in this paper. This problem can be extended to other maintenance problems with safety-critical components and has important economical benefit for companies. During the optimization of joint replacement problems, the action space will grow exponentially with the system size. It makes the action selection very inefficient. This large action space problem is little addressed in the literature. We identify the SRLF rule which can reduce the action space to the linear size of the number of components. This rule can preserve the optimal actions in the reduced action space and make the action selection efficient. Based on this rule, we derive some properties of the optimal policy. These properties can be used to improve the efficiency of optimization algorithms. The SRLF rule is just a heuristics for the action selection. In the practical application, it should be combined with other optimization algorithms. This paper discusses its combination with NDP algorithm, which is usually used to deal with the large-scale stochastic optimization problems. This combined algorithm can handle the problems of large state space and large action space. The simulation experiments demonstrate its efficiency and scalability, but, when the practitioners use the NDP algorithm, it should be noted that the selection of basis functions (features) is quite empirical. It heavily depends on the practitioners' experiences on the studied problem.**

***Index Terms*—Joint replacement, maintenance actions, Markov decision processes, neuro-dynamic programming.**

NOMENCLATURE

MDP	Markov decision processes.
NDP	Neuro-dynamic programming.
SRLF	Shortest-remaining-lifetime-first.
TD(λ)	Temporal difference algorithm with parameter λ.

I. INTRODUCTION

IN THE past several decades, the optimization of maintenance problems has been extensively studied in the literature [6], [7], [19], [23]. This optimization problem is significant to improve the operation efficiency of systems and is very important in the area of industry, military, and so on. The maintenance problem for assets (e.g., jet engines or generators) is one kind of important maintenance problem. It pursues the optimal policy such that the maintenance cost is minimized, while keeping the asset in a good working condition. In this paper, we study an asset maintenance model with safety-critical components, which is illustrated in Fig. 1. This model is derived from a practical maintenance project reported previously in [20] and [22]. Since the total cost of a typical contract between the customer and the asset providers may exceed billions of dollars [3], this optimization problem has great economical importance.

As illustrated by Fig. 1, there are many safety-critical components with different new lifetimes and prices in an asset. When the asset is working, the remaining lifetimes of safety-critical components decrease linearly with time. The asset stops working when there is one expired component (i.e., its remaining lifetime is zero) or emergent failure (take the jet engine as an example: birds hit the engine, the engine gets stuck, and we should stop the engine to clean the mess). The asset is then sent to the workshop for maintenance, which causes a shop visit. During the shop visit period, the asset is disassembled into components. A maintenance policy determines

Manuscript received May 30, 2006; revised December 26, 2006 and April 16, 2007. This paper was recommended for publication by Associate Editor L. Shi and Editor N. Viswanadham upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grants NSFC (60274011, 60574064, 60574067, 60704008, 60721003, 60736027), in part by the NCET Program of China NCET-04-0094, in part by the National New Faculty Funding for Universities with Doctoral Program (20070003110), and in part by the National 111 International Collaboration Project.

L. Xia is with the Center for Intelligent and Networked Systems (CFINS), Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: xiali98@mails.tsinghua.edu.cn).

Q. Zhao and Q.-S. Jia are with CFINS, Department of Automation and TNLIST Laboratory, Tsinghua University, Beijing 100084, China (e-mail: zhaoqc@tsinghua.edu.cn; jiaqs@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2007.910763

1545-5955/$25.00 © 2008 IEEE

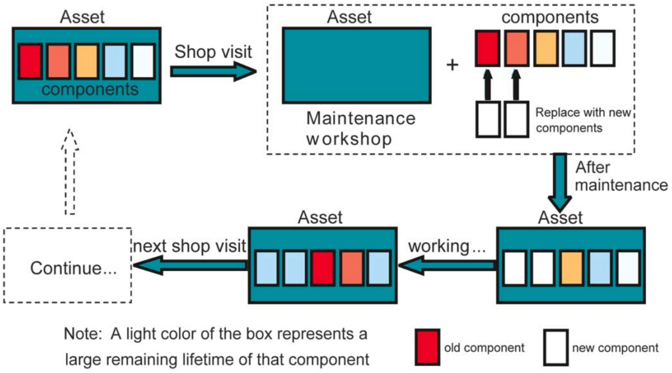


Fig. 1. Flowchart of the asset maintenance procedure.

which components should be replaced by new ones. Then the asset is assembled and this shop visit is finished. After the asset is installed again and works for some time, some components may expire or there is an emergent failure again, which causes another shop visit. The following procedures are similar to those described as above. This is the asset maintenance procedure.

At each shop visit, the maintenance cost includes the prices of new components for replacement and a fixed cost for shop visit. The assets such as jet engines or generators are very expensive. Therefore, the manufacturer usually signs a contract with the customer, such as electric utilities, airline companies, etc., to cover the maintenance cost of the asset for several years. Then, as an optimization problem to the maintenance service provider, the goal is to find the optimal maintenance policy that minimizes the total cost during the whole contract period. Since the asset is safety-critical, we assume that the probability of random failure of components is small enough to be negligible. It is different from the maintenance problems discussed in some literature. In our model, the random failure of the asset is only caused by the emergency whose probability is independent of the states of components. For further explanations on the reasonability of these assumptions, please refer to Section II-A and [20] and [22].

This problem is a multicomponents maintenance problem which is known as a difficult optimization problem in the literature [8], [11], [12], [18]. The difficulty is mainly due to the correlation among the replacement decisions of different components. In the literature, the problem with two identical components is first studied and the derived policy has a control-limit structure. When the number of components increases, the problem becomes very complex and has no simple policy structure. It is difficult to obtain the optimal policy analytically and some research efforts are given to find the heuristic rules [12]. For the optimization of our problem, [20] and [22] discuss it with two kinds of methods. Reference [20] introduces a “one-stage analysis” method for this optimization problem. It uses a rollout simulation to evaluate the average cost until to the next stage shop visit and chooses a good action from the simulation results. This method can be viewed as a feasible approach for large-scale maintenance problems, but it just can achieve a suboptimal performance. Our method introduced in this paper can be combined with this method to reduce the computation of evaluation for the feasible actions. It can improve the efficiency of this “one-stage analysis” method. Reference [22] uses the Lagrangian relaxation method to decompose the original problem into subproblems, and optimizes these subproblems to reduce the global cost. This approach is also suboptimal and its performance is not very satisfactory when the problem scale increases. This approach can also combine our method to improve the efficiency of the construction of feasible solutions.

As we know, the maintenance problem can be modelled as an MDP problem [9], [13]. We present an MDP model for this asset maintenance problem in Section II-B. The MDP problem is extremely challenging due to the large state space and the large action space. It is interesting to note that in the literature of MDP, many efforts (for example, NDP [2], [16]) focus on the problem of large state space, while the problem of large action space is little addressed. The main contribution of this paper is that we find a rule named SRLF to cut down the action space from $O(2^n)$ to $O(n)$, and combine the NDP algorithm with SRLF rule to handle this large-scale maintenance problem. We show that the optimal value to the maintenance problem can always be achieved by the policies obeying this rule. Since the NDP methodology has a theoretical guarantee for its convergence, this combined algorithm is also expected to have the global optimal performance. Moreover, this rule can also be combined with other algorithms, such as the approaches in [20], [22], to improve the efficiency of the algorithms. The proof of the optimality of the SRLF rule is given in Section III. To further speed up the optimization procedure, we derive some other properties of the optimal policy in Section IV-A. One of these properties is that the optimal value function is nonincreasing with respect to the remaining lifetimes of components. Another property is that the optimal policy is monotone when we only change the remaining lifetime of one of the components. In Section IV-B, we also give some general conditions where the SRLF rule is still effective. It brings the rule to a more general application situation. Furthermore, we discuss the limitation and future work of the SRLF rule in Section IV-C. In Section V, we develop an on-line algorithm which combines the advantages of NDP techniques and the SRLF rule. This algorithm is promising to solve the large-scale maintenance problem. To demonstrate the efficiency of the SRLF rule, we give some simulation experiments in Section VI. Finally, we conclude this paper with Section VII.

II. Model Description

In this section, we first introduce the asset maintenance problem in more detail, especially the difference from the maintenance models addressed in some of the literature. Then we present the MDP model of this problem.

A. Description of the Asset Maintenance Problem

In our practical asset maintenance project, an asset consists of many components. We focus on the replacement of so-called safety-critical components. These components are critical for the asset safety. In order to guarantee the safety of the asset, each new component i has a maximum working lifetime $s^0(i)$, which is also called the new lifetime of component i. $s^0(i)$ is specified according to the statistics of the asset, so we can view it as a known parameter. We assume that the components are not interdependent and the remaining lifetimes of components decrease linearly with time. This assumption is acceptable for the safety-critical components since these components are very

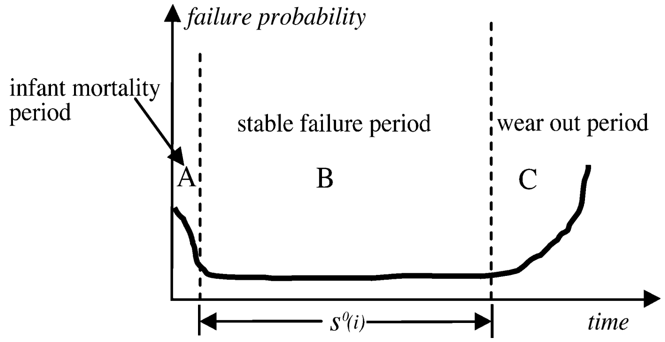


Fig. 2. Bathtub-shaped failure probability of components.

stable during the normal operation process and the interdependency among components can be neglected. If the component i has worked for time $s^0(i)$, its remaining lifetime reaches zero. It should be replaced immediately by a new component i. When the asset is working, it is assumed that the components do not encounter any random failure. This assumption is different from that in some literature of maintenance problems where the exponentially distributed random failure is assumed for each component. In fact, the practical failure probability of components usually has a bathtub-shaped pattern, which is illustrated in Fig. 2. In this figure, section A is the infant mortality period and its failure probability is quite high. Section B is the stable failure period and its failure probability is a very small constant. Section C is the wear out period and its failure probability increases rapidly with time. In our practical project, our industry partners guarantee that during the working period $s^0(i)$ of the component i, its failure probability is small enough that we can neglect it. This assumption is quite reasonable for the high-safety required asset.

However, the whole asset may fail due to the random emergent events. For example, when a bird is inhaled into the jet engine, we should disassemble the engine to clean the mess. But it does not affect the remaining lifetimes of components, because the asset has a special design to protect its safety-critical components. The probability of these random emergent events is assumed to be a constant. It is independent of the remaining lifetimes of components. This assumption is reasonable because the working state of the asset is quite stable during its normal operation period. When the remaining lifetimes of components are all positive, these components do not encounter any failure. Therefore, the random failure of the asset is caused only by the uniformly distributed emergent events.

The remaining lifetime of each component decreases linearly with time when the asset is working. When the remaining lifetime of one of the components reaches zero or the asset encounters a random failure, the asset should be sent to the workshop for maintenance. During the maintenance period, we need to decide which components in the asset should be replaced. Obviously, it is required that the expired components should be replaced. Moreover, considering the cost for shop visit, it is economical to replace some components whose remaining lifetimes are quite small. It can prevent the next shop visit from occurring soon and reduce the number of shop visits. We call this a joint replacement. The lifetime and price of each new component may be different. We assume that the capacity of the maintenance workshop is adequate and the maintenance time is negligible.

In each maintenance period the maintenance cost has two terms: the cost of shop visit and the cost of new components which are used to replace the old ones. The cost of shop visit is a fixed value, which consists of the disassembling cost of the asset (e.g., cost to take a jet engine from the aircraft), the transportation cost from the working site to the workshop, and the cost for the asset being out of service. The cost of new components is the sum of prices of the new components for replacement.

In this model, the action variable is the replacement decision of each component when a shop visit occurs. The objective function is the total maintenance cost during the contract period of the asset. Our optimization problem is how to make replacement decision at each shop visit time in order to minimize the total maintenance cost during the contract period of the asset.

B. Markov Decision Model

A Markov decision model is used to formulate this asset maintenance problem. First, we give the following notations which are used in the rest of the paper.

Notations

n Number of components in the asset.

T Period of the contract.

t Time unit in the total maintenance procedure, $t = 0, 1, 2, \ldots, T$.

p Random failure probability of the asset in each time unit.

w_t Indicator of the asset being in the workshop for maintenance at time t. If the asset encounters a random failure or one of its components is expired, we let $w_t = 1$. Otherwise, $w_t = 0$.

T_u Time when the uth shop visit happens and the maintenance action has not been executed, $u = 1, 2, \ldots$. It is easy to know that if $w_t = 1$, then $t \in \{T_u, u = 1, 2, \ldots\}$.

T_u^+ Time just after the uth shop visit happens and the maintenance action has been executed. The difference between T_u^+ and T_u is infinitely small.

c_0 Fixed cost for each shop visit.

$\mathbf{c}$ Price vector of new components. It is an n-by-1 vector whose element $c(i)$ is the price of component $i, i = 1, 2, \ldots, n$.

$\mathbf{s}^0$ New lifetime vector of all components. It is an n-by-1 vector whose element $s^0(i)$ is a constant representing the new lifetime of component $i, 0 < s^0(i) \le T, i = 1, 2, \ldots, n$.

$\mathbf{s}_t$ Remaining lifetime vector of all components at time t. It is an n-by-1 vector whose element $s_t(i)$ is the remaining lifetime of component i at time $t, i = 1, 2, \ldots, n$. Particularly, $\mathbf{s}_0$ is the remaining lifetimes at time $t = 0$, which is also called the initial remaining lifetimes.

$\mathbf{d}_{T_u}$ Decision vector of all components at the shop visit time T_u. It is an n-by-1 vector whose element $d_{T_u}(i)$ is the replacement decision of component i at time $T_u, i = 1, 2, \ldots, n$. For each i, $d_{T_u}(i)$ can choose from only two values; $d_{T_u}(i) = 1$ means to replace component i and $d_{T_u}(i) = 0$ means not to replace component i.

a^* Virtual action when there is no shop visit, i.e., the action at time $t, t \notin \{T_u, u = 1, 2, \ldots\}$. This action has no effect on the asset. It is introduced just for the completeness of the MDP model.

X_t State vector of the Markov system at time t. We define $X_t = (\mathbf{s}_t, w_t)$.

$\mathcal{S}$ State space of the Markov system. $\mathcal{S} := \{\text{all } (\mathbf{s}_t, w_t) : 0 \leq s_t(i) < s^0(i), w_t = 0 \text{ or } 1, i = 1, 2, \ldots, n\}$.

$\mathcal{A}$ Action space of the Markov system. $\mathcal{A} := \{\{0,1\}^n, a^*\}$, where $\{0,1\}^n$ is an n-dimension vector space consisting of elements 0 or 1. Obviously, $\mathcal{A}(X_t) = \{a^*\}$, if $w_t = 0$; $\mathcal{A}(X_t) = \{0,1\}^n$, if $w_t = 1$; and $\mathbf{d}_{T_u} \in \{0,1\}^n$.

π Replacement policy of the asset maintenance problem. We only consider the deterministic Markov policies which can also attain the optimal cost of randomized history-dependent policies for this problem [15]. We denote $\pi = \{\pi_0, \pi_1, \ldots, \pi_t, \ldots, \pi_T\}$, where π_t is the decision rule at time t, i.e., π_t is a mapping from state space $\mathcal{S}$ to action space $\mathcal{A}$. In fact, for $t \in \{T_u, u = 1, 2, \ldots\}$, i.e., $w_t = 1$, we know $\pi_t(X_t) = \mathbf{d}_t \in \{0,1\}^n$; for $t \notin \{T_u, u = 1, 2, \ldots\}$, i.e., $w_t = 0$, we know $\pi_t(X_t) = a^*$. Therefore, in this paper, we only need to decide how to choose $\pi_t(X_t)$ when $w_t = 1$.

$r(X_t, \pi_t(X_t))$ Cost function of the MDP model. It is easy to know that $r(X_t, a^*) = 0, r(X_t, \mathbf{d}_t) = c_0 + \mathbf{c}^T \cdot \mathbf{d}_t$, where the superscript T denotes the vector transpose.

ξ A sample path during the maintenance procedure which is obtained by simulation technique. To simulate this maintenance problem, at each time unit t, we need to generate a random number ξ_t which is uniformly distributed in [0,1]. If $\xi_t < p$, it means that the asset encounters a random failure at this time unit. Otherwise, there is no random failure at this time unit. Therefore, the total simulation process depends on the sequence of random numbers $\xi = \{\xi_0, \xi_1, \xi_2, \ldots, \xi_T\}$. We can consider that ξ represents all the randomness in the maintenance problem.

$(\xi, \pi, \mathbf{s}_0)$ System trajectory in the sample path ξ with initial remaining lifetimes $\mathbf{s}_0$ and policy π. It can completely determine the total states trajectory of the system in the current sample path.

$f^{\xi,\pi,\mathbf{s}_0}_{(t_1,t_2)}$ Total maintenance costs from time t_1 to t_2 in the trajectory $(\xi, \pi, \mathbf{s}_0)$.

From the description of this asset maintenance problem, we know that the maintenance process has Markovian property: If we know the current system state, the future system states are independent of the historical states. We can model this problem as a finite-horizon discrete-time Markov decision process. The state vector of the Markov system is $X_t = (\mathbf{s}_t, w_t), t = 0, 1, 2, \ldots, T$. It is obvious that the size of the state space $\mathcal{S}$ is $|\mathcal{S}| = 2\prod_{i=1}^{n} s^0(i)$, which grows exponentially with the number of components n. As we know, if there is no shop visit at time t, i.e., $w_t = 0$, the asset is in a good working condition and we need not to make replacement decision. Therefore, there is no action choosing when $w_t = 0$. It is different from the standard Markov decision processes where the action choosing is required at each state transition epoch. For the completeness of the MDP model, we define a special action a^* which represents the virtual action when $w_t = 0$. Thus, the available action space is $\mathcal{A}(X_t) = \{a^*\}$ when $w_t = 0$, and $\mathcal{A}(X_t) = \{0,1\}^n$ when $w_t = 1$. The deterministic Markov policy π is a mapping from $\mathcal{S}$ to $\mathcal{A}$ at each time t, i.e., $\pi = \{\pi_0, \pi_1, \ldots, \pi_t, \ldots, \pi_T\}$. The decision rule at time t, π_t, is defined as $\pi_t(X_t) = \mathbf{d}_t \in \{0,1\}^n$ when $w_t = 1, \pi_t(X_t) = a^*$ when $w_t = 0$. The cost functions are $r(X_t, a^*) = 0$ and $r(X_t, \mathbf{d}_t) = c_0 + \mathbf{c}^T \cdot \mathbf{d}_t$. The state transition probability at time t under decision rule $\pi_t, P_t\{X_{t+1}|X_t, \pi_t\}$, can be written as

$$
\begin{aligned}
&P_t\{X_{t+1} \mid X_t, \pi_t\} \\
&= P_t\{(\mathbf{s}_{t+1}, w_{t+1}) \mid (\mathbf{s}_t, w_t), \pi_t\} \\
&= \begin{cases}
(1-p)I(\mathbf{s}_{t+1}), & \mathbf{s}_{t+1} = \mathbf{s}_t - \mathbf{e} + w_t[(\mathbf{s}^0 - \mathbf{s}_t) \otimes \mathbf{d}_t] \\
& w_{t+1} = 0 \\
(p-1)I(\mathbf{s}_{t+1}) + 1, & \mathbf{s}_{t+1} = \mathbf{s}_t - \mathbf{e} + w_t[(\mathbf{s}^0 - \mathbf{s}_t) \otimes \mathbf{d}_t] \\
& w_{t+1} = 1 \\
0, & \text{otherwise}
\end{cases}
\end{aligned}
$$

where $I(\mathbf{s}_{t+1})$ is an indicator function which is defined as $I(\mathbf{s}_{t+1}) = 0$ (or 1) if $\min_{i=1,2,\ldots,n}\{s_{t+1}(i)\} = 0$ (or > 0), $\mathbf{e}$ is the unitary vector defined as $\mathbf{e} := (1, 1, \ldots, 1)^T_{1\times n}, \mathbf{a} \otimes \mathbf{b} := (a_1 b_1, a_2 b_2, \ldots, a_n b_n)^T$ means the respective elements of the n-by-1 column vectors $\mathbf{a}$ and $\mathbf{b}$ are multiplied. The optimization objective is to choose the optimal policy π^* which minimizes the total maintenance costs in the contract period. This optimization problem can be written as

$$
\begin{aligned}
\pi^* &= \arg\min_{\pi} E_\xi \left\{ f^{\xi,\pi,\mathbf{s}_0}_{(0,T)} \right\} \\
&= \arg\min_{\pi} E_\xi \left\{ \sum_{t=0}^{T} r(X_t, \pi_t(X_t)) \right\} \qquad (1)
\end{aligned}
$$

where $E_\xi\{\cdot\}$ denotes the expectation over all the sample paths ξ.

In this Markov decision model, since $\pi_t(X_t) = a^*$ when $w_t = 0$, we only need to choose the action $\pi_t(X_t) = \mathbf{d}_t$ when $w_t = 1$. For simplicity, $\mathbf{d}_t$ is also referred to as $\mathbf{d}_{T_u}$ since $t \in$

$\{T_u, u = 1, 2, \ldots\}$ when $w_t = 1$. As we know, the decision variable $\mathbf{d}_{T_u}$ is a vector with n dimensions. Its element $d_{T_u}(i)$ has two alternative values, 1 and 0, which respectively represent to whether replace this component or not. Therefore, the number of total actions at each shop visit time are $|\mathcal{A}| = 2^n$. In the practical problem, the number of components n is usually about several dozen, so the action space is very large. This is a curse of dimensionality if we want to find the optimal actions in the whole action space through enumeration.

There are some classical algorithms to solve the MDP problems, such as the policy iteration and value iteration for the infinite-horizon MDP, and the backward induction for the finite-horizon MDP. But these algorithms always suffer the curse of dimensionality when the system size increases. In the literature of MDP, many research interests are focused on the problem of large state space. For example, NDP [2], [16] attempts to use the neural network to approximate all the value functions. Other approaches for the problem of large state space include state aggregation [1], time aggregation [4], randomization [17], and so on. But little literature addresses the problem of large action space. When the action space is large, these approaches for large state space are even more difficult to apply. In this paper, we find the SRLF rule for our maintenance problem. It reduces the size of the action space from $O(2^n)$ to $O(n)$. Thus, the large action space problem is solved for this model. The details of the SRLF rule are presented in Section III.

III. SRLF Rule and Its Optimality

The SRLF rule is found originally from the asset maintenance problem. It can reduce the action space greatly. In this section, we first give the motivation which leads to the SRLF rule and then introduce this rule. Furthermore, we prove the optimality of this rule which means that it can preserve the optimal solutions in the reduced action space.

A. Motivation

In order to tackle the curse of dimensionality in MDP, methods are usually introduced using the special properties of a certain type of problems. Similarly, in our asset maintenance problem, we find the SRLF rule according to the property of this type of problems. First, we give some simple examples which motivate the SRLF rule.

For simplicity, we consider an infinite-horizon optimization problem for the asset maintenance where there are only two components. The fixed cost of a shop visit is 5 and the random failure probability of the asset is 0.1. When the asset enters the workshop for maintenance, the remaining lifetimes of components are 21 and 4, respectively. Below we discuss the optimal replacement decision in several cases.

In the first case, we assume that the two components are identical, e.g., their new lifetimes are both 30 and their prices are both 2. In this case, it is easy to understand that replacing component 2 is the most urgent, and then component 1. It is reasonable to replace the old component first since all these components are identical. So in this situation, we only need to decide how many old components should be replaced. The decision procedure is quite simple.

In the second case, we assume that the prices of components are 1 and 3, respectively, and their new lifetimes are both 30. Since component 2 is more expensive than component 1, it is suspicious whether we should replace the expensive component 2 first. How should we make the replacement decision in this situation?

Furthermore, in the third case, we suppose the prices of components are 1 and 3, respectively, and their new lifetimes are 30 and 7, respectively. Since the remaining lifetime of component 2 is quite close to its new lifetime, should we first replace this component which seems quite new?

From these cases, we find that the replacement decision is quite complex when the components are different. When the number of components increases, the action space increases exponentially. This adds on the difficulty of the decision procedure. Fortunately, we find that the optimal decisions in these three cases have a special property. The optimal decisions in the first and third case are both (0, 1), and the optimal decision in the second case is (1, 1). In fact, the structure of the optimal policy is similar to that in Fig. 3. These optimal decisions are accordant with our guess in the first case, i.e., we should replace the old component first. Furthermore, after we carefully investigate the characteristics of this type of maintenance problems, we find that the replacement rule in the first case is also correct in all the other situations. It is named the SRLF rule and it can simplify the decision procedure greatly.

B. Description of the Rule

The SRLF rule is described as follows.

When an asset enters the workshop for maintenance, we sort the remaining lifetimes of components in ascending order. If a component with large remaining lifetime is replaced, it is required that the components whose remaining lifetimes are smaller than this one should also be replaced.

As indicated by the name, the SRLF rule requires that components with short remaining lifetime should be replaced first. In the previous examples of Section III-A, the SRLF rule requires that if we decide to replace component 1, we should also replace component 2.

This rule looks somewhat straightforward, but if we consider that the new lifetimes and prices of all components are different, it is not clear whether the optimal solutions can be reserved after using this rule. Moreover, in some situations, this rule is counterintuitive. For example, in the practical maintenance procedure, the workers are usually inclined to replace the cheap components first while not replace the expensive components. In another example, when the remaining lifetimes of two components are identical, it is also inclined to replace the component whose new lifetime is large while keeping the one whose new lifetime is short. These empirical methods are conflicted with the SRLF rule. It is necessary to prove the optimality of the rule rigorously before we use it.

C. Optimality of the SRLF Rule

In this section, we use a sample path based analysis to prove the optimality of the SRLF rule. Some notations which are used in the proof can be referred to in Section II-B. The detailed proof of Lemma 1 and Propositions 1 and 2 can be found in the Appendix.

Lemma 1: If $\mathbf{s}'_0 \geq \mathbf{s}_0$ (component-wise), then for any policy π there exists a policy π' such that $f^{\xi,\pi',\mathbf{s}'_0}_{(0,T'^{+}_1)} \leq f^{\xi,\pi,\mathbf{s}_0}_{(0,T'^{+}_1)}$ and $\mathbf{s}'_{T'^{+}_1} \geq \mathbf{s}_{T'^{+}_1}$, for all ξ.

This lemma means that if the remaining lifetime at time 0 is larger, we can find a better policy whose cost is smaller at the first shop visit. By induction, we can extend Lemma 1 to a more general result.

Proposition 1: If $\mathbf{s}_0{}' \geq \mathbf{s}_0$, then for any policy π there exists a policy π' which has $f^{\xi,\pi',\mathbf{s}_0'}_{(0,t)} \leq f^{\xi,\pi,\mathbf{s}_0}_{(0,t)}$, for all ξ and t. This proposition indicates that the cost in a sample path is monotone with respect to the initial remaining lifetime. Based on Proposition 1, we have Proposition 2 as follows.

Proposition 2: For any trajectory $(\xi, \pi, \mathbf{s}_0)$, if at time $T_u, s_{T_u}(i) < s_{T_u}(j), d_{T_u}(i) = 0, d_{T_u}(j) = 1$, then a non-worse policy π' which is accordant with the SRLF rule can be constructed, i.e., $f^{\xi,\pi',\mathbf{s}_0}_{(0,T)} \leq f^{\xi,\pi,\mathbf{s}_0}_{(0,T)}$.

Inspired by Proposition 2, we have the following theorem about the optimality of the SRLF rule.

Theorem 1 (The optimality of the SRLF rule): The SRLF rule can preserve the optimal solutions in the reduced action space.

Proof: This theorem can be proved with Proposition 2 by iteration. From Proposition 2, it is known that for the actions of any two components not accordant with the SRLF rule at time T_u, a non-worse trajectory $(\xi, \pi', \mathbf{s}_0)$ that adopts this rule at time T_u for these two components can be constructed. This procedure can be iterated until at time T_u all the actions obey this rule. At the next shop visit time T_{u+1}, this procedure can be continued until T, so, for any time between 0 and T, if the actions of some components do not accord with the SRLF rule, we can use Proposition 2 to construct a non-worse trajectory $(\xi, \pi'', \mathbf{s}_0)$ which obeys this rule. This proves Theorem 1. □

We can use this rule to simplify the decision procedure greatly. At each shop visit time, instead of deciding what collection of components should be replaced, we only need to decide how many old components should be replaced, which are sorted ascendingly with their remaining lifetimes. Therefore, the action space is reduced from 2^n to $n + 1$. In our practical asset maintenance problem, the number of components in an asset is typically about 30. Then, the action space is reduced from 2^{30} (about 10^9) to 31. It is a great saving of the computation resource.

From the proof of the SRLF rule, it is shown that when the remaining lifetimes of some components are identical, their actions should also be identical. In other words, when some components have the same remaining lifetimes at the maintenance time, we should either replace all these components or not replace any of them, in spite of their different prices.

IV. Extensions of the SRLF Rule

In this section, we discuss some extensions of the SRLF rule. Based on the SRLF rule, we derive some important theorems. These theorems describe the characteristics of the optimal policy in this maintenance problem. They are helpful for us to understand how to choose the optimal decisions. Furthermore, we generalize this particular asset maintenance problem to a class of maintenance problems. Some generalized conditions are introduced. The SRLF rule and the related theorems are still correct in this class of maintenance problems. Finally, we discuss the limitations of the SRLF rule and its future work.

A. Theorems Derived From the SRLF Rule

First, we derive some theorems based on the SRLF rule. These theorems are quite helpful for us to study the special properties of the optimal policy in the maintenance problem. They also can be used as guidelines for us to choose the optimal replacement decisions. Some of these theorems are similar to the related results in [14], but their problem model is different from ours.

Theorem 2: The optimal value function in this model is nonincreasing with respect to the remaining lifetime vector. In other words, if $X = (\mathbf{s}_0, w), X' = (\mathbf{s}_0', w)$, and $\mathbf{s}_0 \leq \mathbf{s}_0'$, then $v_T^*(X) \geq v_T^*(X')$.

Proof: In the finite-horizon MDP model, the optimal value function v_T^* is defined as [15]

$$v_T^*(X) := E\left\{\sum_{t=0}^{T} r\left(X_t, \pi_t^*(X_t)\right) \mid X_0 = X\right\} \tag{2}$$

where T is the total time stage, $\pi_t^*(X_t)$ is the optimal actions according to the optimal policy π^* at time t. From Section II-B, it is known that $\sum_{t=0}^{T} r(X_t, \pi_t^*(X_t)) \mid X_0 = (\mathbf{s}_0, w)$ is equal to $f^{\xi,\pi^*,\mathbf{s}_0}_{(0,T)}$, where ξ is a random sequence generated for the initial state $(\mathbf{s}_0, w)$, so the optimal value function can be written as $v_T^*(X) = E\{\sum_{t=0}^{T} r(X_t, \pi_t^*(X_t)) \mid X_0 = (\mathbf{s}_0, w)\} = E_\xi\{f^{\xi,\pi^*,\mathbf{s}_0}_{(0,T)}\}$. With Proposition 1, it is known that if two initial remaining lifetimes $\mathbf{s}_0 \leq \mathbf{s}_0'$, there exists a policy π' which has $f^{\xi,\pi',\mathbf{s}_0'}_{(0,T)} \leq f^{\xi,\pi^*,\mathbf{s}_0}_{(0,T)}$ for any sample path ξ. Since π^* is the optimal policy, it has $E_\xi\{f^{\xi,\pi^*,\mathbf{s}_0'}_{(0,T)}\} \leq E_\xi\{f^{\xi,\pi',\mathbf{s}_0'}_{(0,T)}\}$. Therefore, we can derive that $E_\xi\{f^{\xi,\pi^*,\mathbf{s}_0'}_{(0,T)}\} \leq E_\xi\{f^{\xi,\pi^*,\mathbf{s}_0}_{(0,T)}\}$. Therefore, we know that $v_T^*(X) \geq v_T^*(X')$ and $v_T^*(X)$ is nonincreasing with respect to $\mathbf{s}_0$. □

The nonincreasing property of the optimal value function is very helpful for us to analyze the optimal policy in this maintenance problem. Obviously, this property can be further extended to the infinite-horizon MDP model.

Theorem 3: For any sample path ξ and any shop visit time T_u, if the optimal decision for $\mathbf{s}_{T_u}$ is $\mathbf{d}_{T_u}^*$, then for any $\mathbf{s}_{T_u}'$ satisfying $s_{T_u}'(i) \leq s_{T_u}(i)$, if $d_{T_u}^*(i) = 1$; $s_{T_u}'(j) = s_{T_u}(j)$, if $d_{T_u}^*(j) = 0$, the optimal decision for $\mathbf{s}_{T_u}'$ is also $\mathbf{d}_{T_u}^*$.

Proof: Because $\mathbf{s}_{T_u}' \leq \mathbf{s}_{T_u}$, in order to use Proposition 1, we can just suppose that the time starts from T_u instead of 0, i.e., ξ and π both start from time T_u, and $\mathbf{s}_{T_u}$ is the initial remaining lifetimes of the sample path ξ. Therefore, it is easy to know that for any policy π' there exists policy π such that $f^{\xi,\pi,\mathbf{s}_{T_u}}_{(T_u,T)} \leq f^{\xi,\pi',\mathbf{s}_{T_u}'}_{(T_u,T)}$ for all ξ, so we can derive $E_\xi\{f^{\xi,\pi,\mathbf{s}_{T_u}}_{(T_u,T)}\} \leq E_\xi\{f^{\xi,\pi',\mathbf{s}_{T_u}'}_{(T_u,T)}\}$. Suppose the optimal policy of this problem is π^*. From the definition of the optimal policy, we know $E_\xi\{f^{\xi,\pi^*,\mathbf{s}_{T_u}}_{(T_u,T)}\} \leq E_\xi\{f^{\xi,\pi,\mathbf{s}_{T_u}}_{(T_u,T)}\}$. If we choose $\mathbf{d}_{T_u}^*$ as the decision for $\mathbf{s}_{T_u}'$, it is obvious that after the replacement action $\mathbf{d}_{T_u}^*$, it has $\mathbf{s}_{T_u+1}' = \mathbf{s}_{T_u+1}$. Therefore, the two trajectories, $(\xi, \pi^*, \mathbf{s}_{T_u}')$ and $(\xi, \pi^*, \mathbf{s}_{T_u})$, are the same after time T_u, i.e., $f^{\xi,\pi^*,\mathbf{s}_{T_u}'}_{(T_u+1,T)} = f^{\xi,\pi^*,\mathbf{s}_{T_u}}_{(T_u+1,T)}$. It is easy to know that $r((\mathbf{s}_{T_u}', 1), \mathbf{d}_{T_u}^*) + f^{\xi,\pi^*,\mathbf{s}_{T_u}'}_{(T_u+1,T)} = r((\mathbf{s}_{T_u}, 1), \mathbf{d}_{T_u}^*) + f^{\xi,\pi^*,\mathbf{s}_{T_u}}_{(T_u+1,T)} = f^{\xi,\pi^*,\mathbf{s}_{T_u}}_{(T_u,T)}$. Therefore, we can derive $E_\xi\{r((\mathbf{s}_{T_u}', 1), \mathbf{d}_{T_u}^*) + f^{\xi,\pi^*,\mathbf{s}_{T_u}'}_{(T_u+1,T)}\} = E_\xi\{f^{\xi,\pi^*,\mathbf{s}_{T_u}}_{(T_u,T)}\} \leq E_\xi\{f^{\xi,\pi,\mathbf{s}_{T_u}}_{(T_u,T)}\} \leq E_\xi\{f^{\xi,\pi',\mathbf{s}_{T_u}'}_{(T_u,T)}\}$, i.e., $E_\xi\{r((\mathbf{s}_{T_u}', 1), \mathbf{d}_{T_u}^*) + f^{\xi,\pi^*,\mathbf{s}_{T_u}'}_{(T_u+1,T)}\} \leq E_\xi\{f^{\xi,\pi',\mathbf{s}_{T_u}'}_{(T_u,T)}\}$ for any policy π'. Therefore, the optimal decision for $\mathbf{s}_{T_u}'$ is also $\mathbf{d}_{T_u}^*$ and Theorem 3 is proved. □

With Theorem 3, we know that the optimal decisions for some remaining lifetime vectors are identical, so we can only focus on the replacement optimization of parts of remaining lifetime vectors. It simplifies the asset maintenance problem and speeds up the optimization procedure.

Theorem 4: At any shop visit time T_u, if two remaining lifetime vectors, $\mathbf{s}_{T_u}$ and $\mathbf{s}'_{T_u}$, have the relationship: $s_{T_u}(i) \geq s'_{T_u}(i)$ for any given i, and $s_{T_u}(j) = s'_{T_u}(j)$ for all the other $j \neq i$, then the optimal decisions have $d^*_{T_u}(i) \leq d^{*\prime}_{T_u}(i)$.

Proof: Theorem 4 says that at shop visit time T_u for any component i, if we fix the remaining lifetimes of other $n-1$ components and change the remaining lifetime of component $i, s_{T_u}(i)$, then the optimal decision of component i is nonincreasing with respect to $s_{T_u}(i)$. It can be easily derived from Theorem 3. We prove it from two cases according to the value of $d^*_{T_u}(i)$.

1) If $d^*_{T_u}(i) = 1$:
   In this case, we find that the two remaining lifetime vectors $\mathbf{s}_{T_u}$ and $\mathbf{s}'_{T_u}$ are just accordant with the condition in Theorem 3. From Theorem 3, we know that the optimal decisions of these two remaining lifetime vectors are identical, i.e., $\mathbf{d}^*_{T_u} = \mathbf{d}^{*\prime}_{T_u}$. Therefore, it is obvious that $d^*_{T_u}(i) = d^{*\prime}_{T_u}(i)$.
2) If $d^*_{T_u}(i) = 0$:

The optimal decision $d^{*\prime}_{T_u}(i)$ is either 1 or 0, both satisfying $d^*_{T_u}(i) \leq d^{*\prime}_{T_u}(i)$.

So it is proved that the optimal decision of component i is nonincreasing with respect to its remaining lifetime when the remaining lifetimes of other components are fixed. □

In fact, Theorem 4 can be viewed as a special case of Theorem 3, but Theorem 4 emphasizes the monotone property of the optimal policy. It is an important property of this kind of maintenance problem.

These three theorems can be viewed as the derivations of Proposition 1. They are very easy to use and quite helpful for us to choose the optimal replacement decisions in the maintenance problem. They also give some interesting insights into the characteristics of the optimal policy. The examples in Section IV illustrate how to use these theorems to speed up the optimization procedure of the maintenance problem.

B. Generalized Conditions of Maintenance Problems

As we know, the proof of the SRLF rule is based on our particular asset maintenance problem. In fact, we can generalize some conditions of this problem and keep the SRLF rule still effective under these conditions, so we can extend the rule to a wider application situation. The generalized conditions mainly include the following aspects.

Condition 1: The random failure probability of the asset can be any distribution.

In our particular maintenance problem, we assume that the random failure probability of the asset is a constant. In fact, we can generalize it to an arbitrary random failure distribution, but we still require that the random failure distribution should be independent of the remaining lifetimes of components.

Since the total randomness in this maintenance problem is the random failure of the asset, we know that the arbitrary random failure distribution only affects the maintenance time which is caused by the random failure. Other system characteristics are not affected by this changed random failure distribution. From the simulation methodology, we know that the system randomness is all generated from a sequence of uniformly distributed random numbers $\xi = \{\xi_0, \xi_1, \xi_2, \ldots\}$. With ξ and the random failure distribution function $F(t)$, the time sequence of random failure $\{T_1^f, T_2^f, T_3^f, \ldots\}$ can be determined by the inverse-transform method, where T_i^f is the time of the ith random failure. In detail, they are determined as $T_1^f = 0 + F^{-1}(\xi_0), T_2^f = T_1^f + F^{-1}(\xi_1), T_3^f = T_2^f + F^{-1}(\xi_2), \ldots$. Therefore, we can still use ξ to represent the system sample path, and the system trajectory is also determined by $(\xi, \pi, \mathbf{s}_0)$. Therefore, the change of the random failure distribution does not affect the proof of the SRLF rule. This rule is still correct for any distribution of random failure probability.

Condition 2: The criterion of MDP model can be the infinite-horizon average cost criterion.

Our original problem is modelled as a finite-horizon Markov decision process. If we extend the contract time T to infinite, this problem becomes an infinite-horizon Markov decision problem. It is obvious that this modification does not affect any proof of the SRLF rule except that we let T go to ∞. Thus, it is reasonable to expect that the SRLF rule is still effective with the infinite-horizon average cost criterion.

Furthermore, we have a preliminary consideration about a general condition where the random failure probability of components cannot be neglected. In this situation, we consider that each component has a random failure distribution and the random failure of the entire asset can be omitted. When one of the components encounters a random failure or its remaining lifetime is zero, the asset should be sent to the workshop for maintenance. The following schemes are similar to those of the original model. We consider a simple situation, where the random failure distribution of each component i is an exponential distribution with parameter $1/s_t(i)$. When $s_t(i)$ is larger, the expected time for the random failure occurring will be larger. For this new model, it is natural to expect that the SRLF rule is still correct. But the theoretical proof is not straightforward because it is difficult to determine the order of the random failure of each component occurring. This problem is quite interesting and deserves further investigation.

With these generalized conditions, the SRLF rule has a much wider application situation. The related theorems are also correct in these new situations. These properties are significant for us to study the optimization of this kind of maintenance problems.

C. Limitations of the SRLF Rule

Although the SRLF rule is quite helpful for us to optimize the maintenance problems, it also has some limitations. In order to help the practitioners to use it in the practical applications, we give a discussion about the related limitations and its future work.

The first limitation is about the interdependency among components. In our model, we assume the components are independent and their remaining lifetimes decrease linearly with time. As we discussed in Section II-A, this assumption is reasonable for the safety-critical components since their working status is required to be very stable. But, if we consider the maintenance of less critical components, this assumption may be not satisfied. In this situation, the components can be interdependent and the remaining lifetime of one component can affect the decreasing rate of the remaining lifetimes of other components.

This model is much more complicated. The correctness of SRLF rule depends on the details of the interdependency among components. In some situations, the SRLF rule will be not correct. For example, we consider three components with remaining lifetimes (5,7,60). The remaining lifetimes of components 1 and 2 are small and are going to zero, which means they are in bad working status. Component 3 is in good working status. Since the components have interdependency, the bad working status of components 1 and 2 will affect the decreasing rate of component 3. The current bad status of component 2 makes the remaining lifetime of component 3 decrease with time at rate 2, while the current bad status of component 1 makes the remaining lifetime of component 3 decrease with time at rate 1. After components 1 and 2 are replaced and in good working status, the remaining lifetime of component 3 will decrease with time at rate 1. In this example, it is reasonable for us to replace component 2 first, because it affects component 3 more heavily. Thus, the SRLF rule is not satisfied in this situation. It demonstrates that the interdependency among components will affect the correctness of the SRLF rule. This problem is quite interesting and deserves further investigation.

Another limitation is that we assume the random failure of components can be neglected. If the random failure of components cannot be neglected, we need further study about the SRLF rule. Concerning this problem, we have given a preliminary discussion in Section IV-B. This is another future work which can extend the SRLF rule.

These limitations clarify the application situations of the SRLF rule. In Section V, we discuss how to use the SRLF rule and NDP techniques to handle a large-scale maintenance problem.

V. Combination With NDP Techniques

As we know, NDP is an approach to solve the large-scale stochastic optimization problems [2], [10], [16]. It is also referred to as reinforcement learning in the field of artificial intelligence [21]. NDP uses the architecture approximation techniques to solve the large state space problem. But it is difficult for NDP to handle the large action space problem. Since the NDP can tackle the large state space problem and the SRLF rule can cut down the action space, it is natural for us to consider the combination of NDP techniques and the SRLF rule. In this section, we develop an on-line algorithm which combines the temporal difference (TD(λ)) learning algorithm and the SRLF rule. This algorithm is efficient to solve the large-scale maintenance problem. Since the SRLF rule has been generalized to the infinite-horizon MDP model, we will use the infinite-horizon average cost criterion in the following sections. Thus, we only need to consider the stationary policies, i.e., $\pi_t \equiv \pi$ for all time $t = 0, 1, 2, \ldots$. This can simplify the optimization analysis of maintenance problems. On the other hand, we can view the optimization of stationary policies as an approximation for the finite-horizon MDP problem.

As we know, in an ergodic MDP problem with infinite-horizon average cost criterion and stationary policy, we can define the performance potential as below [5] (which is also called relative value function or bias)

$$g(X) = \lim_{T\to\infty} E\left\{\sum_{t=0}^{T}[r(X_t, d_t) - \eta] \,|\, X_0 = X\right\} \tag{3}$$

where d_t is the action at time t with a little notation abuse, and η is the average cost as

$$\eta = \lim_{T\to\infty} \frac{1}{T+1} E\left\{\sum_{t=0}^{T} r(X_t, d_t)\right\}. \tag{4}$$

From the policy iteration of MDP, we know that the policy can be improved through the following step:

$$d_t = \arg_{d\in\mathcal{A}(X_t)} \min E\{r(X_t, d) - \eta + g(X_{t+1})\}. \tag{5}$$

When the state space increases exponentially with the system size, it is difficult to estimate all the performance potentials at every state. NDP is proposed to solve this problem using the architecture approximation techniques. It uses some architecture functions, e.g., the artificial neural networks, to approximate the performance potentials of systems. Below, we discuss one of the NDP algorithms, the TD(λ) learning algorithm, to solve this maintenance problem.

In the TD(λ) algorithm, we use the linear summation of basis functions to approximate the performance potentials

$$g(X) = \sum_{k=1}^{K} \psi_k(X) \cdot v_k \tag{6}$$

where $\psi_k(\,\cdot\,)$ is the kth basis function and v_k is its coefficient. The selection of basis functions, which are also called features, is heavily dependent on the user's experience. With these approximate performance potentials, we can make our decision according to (5) at each decision time. In order to strengthen the exploration ability of the algorithm, we use the ϵ-greedy method to adopt the decision, i.e., we adopt the actions of (5) with probability $1 - \epsilon$ and adopt other random actions from the whole $\mathcal{A}(X_t)$ with probability ϵ, where ϵ is a small probability which may change with time. After adopting action d_t, we can obtain the cost $r(X_t, d_t)$ and the next system state X_{t+1}. The temporal difference δ_t can be calculated as

$$\begin{aligned}\delta_t &= r(X_t, d_t) - \eta_t + g(X_{t+1}) - g(X_t)\\ &= r(X_t, d_t) - \eta_t + \sum_{k=1}^{K} \psi_k(X_{t+1}) \cdot v_k\\ &\quad - \sum_{k=1}^{K} \psi_k(X_t) \cdot v_k\end{aligned} \tag{7}$$

where η_t is the average cost which can be estimated as

$$\begin{aligned}\eta_t &= \frac{1}{t+1}\sum_{\tau=0}^{t} r(X_\tau, d_\tau)\\ &= \eta_{t-1} + \frac{1}{t+1}[r(X_t, d_t) - \eta_{t-1}].\end{aligned} \tag{8}$$

With (7), we get the following TD(λ) learning formula to update the coefficient vector $\mathbf{v} = (v_1, \ldots, v_K)$:

$$\mathbf{v} = \mathbf{v} + \gamma_t \mathbf{Z}_t \delta_t \tag{9}$$

where γ_t is a series of learning step-size, $\mathbf{Z}_t$ is called the eligibility traces which are defined as

$$\begin{aligned}\mathbf{Z}_t &= \sum_{\tau=0}^{t} \lambda^{t-\tau} \nabla_{\mathbf{v}} g(X_\tau) \\ &= \lambda \mathbf{Z}_{t-1} + \nabla_{\mathbf{v}} g(X_t) \\ &= \lambda \mathbf{Z}_{t-1} + \boldsymbol{\psi}(X_t) \end{aligned} \tag{10}$$

where λ is a parameter in [0,1], $\nabla_{\mathbf{v}}$ denotes the gradients with respect to $\mathbf{v}$, $\boldsymbol{\psi}(\cdot)$ is the vector consisting of basis functions, i.e., $\boldsymbol{\psi}(\cdot) = (\psi_1(\cdot), \ldots, \psi_K(\cdot))$.

Therefore, we give the following TD(λ) algorithm to optimize this asset maintenance problem.

Algorithm 1. TD(λ) Algorithm With the SRLF Rule:

```
Step 1) Select the proper basis functions $\boldsymbol{\psi}(\cdot)$, $\gamma_t$, and $\lambda$. Set
        the initial parameters $\mathbf{v}$ and $X_0$.
Step 2) At each decision time $t$, use (5) and the $\epsilon$-greedy
        method to determine the action $d_t$.
Step 3) With (7), (8), (9), and (10), update the coefficient
        vector $\mathbf{v}$.
Step 4) Continue step 2 at the next decision time $t+1$, to the
        end of the simulation period.
```

In the above algorithm, we should combine the SRLF rule in the action selection formula (5). It can reduce the size of action space $\mathcal{A}(X_t)$ to $n+1$. Without the SRLF rule, the size of $\mathcal{A}(X_t)$ is 2^n and the action selection in (5) is infeasible for the practical problems. Since the NDP algorithm can converge to the global optimum theoretically and the optimality of SRLF rule has been proved, it is natural to know that Algorithm 1 can also converge to the global optimum theoretically. Therefore, this algorithm combines the advantages of NDP techniques and SRLF rule. It can handle the large state space problem and the large action space problem. Moreover, this algorithm can be implemented with an on-line manner. It is important for the application in practice. The efficiency of this algorithm is demonstrated in Section VI.

VI. Numerical Examples

The SRLF rule can be implemented within the framework of the traditional MDP algorithm or the NDP algorithm. The action space can be reduced dramatically without loss of the policy optimality. In this section, first we use the value iteration algorithm to demonstrate the optimality of the SRLF rule and the related theorems. Then we give an experiment of Algorithm 1 to demonstrate the efficiency of NDP algorithm with the SRLF rule. Please note, the optimization criterion in this section is the infinite-horizon average cost criterion.

A. *Experiment 1*

In order to visually describe the structure of the policy space, the number of components is chosen as 2. The new lifetimes of the two components are 10 and 15, respectively. The fixed cost of a shop visit is 5. The prices of these two components are 1 and 2, respectively. The random failure probability of the asset at each time unit is 0.1. The objective is to minimize the average maintenance cost in infinite horizon. The stopping criterion of the value iteration algorithm is that the maximum error of the value functions between two iterations is smaller than $\varepsilon = 0.001$. Under this condition, we get the ε-optimal policy. The detailed value iteration algorithm for the infinite-horizon average cost MDP model can be referred to in [15].

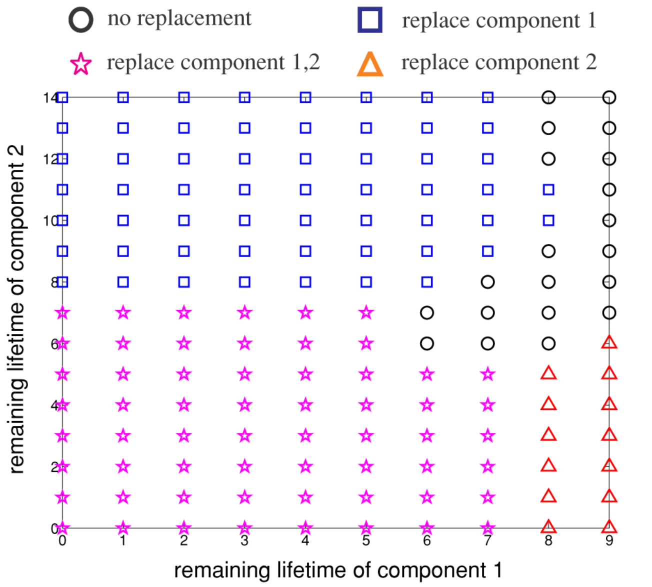


Fig. 3. ε-optimal policy obtained by value iteration.

Two different implementations of the value iteration algorithm are compared. One uses the SRLF rule when choosing actions in the action space. The other does not use the SRLF rule and enumerates all the possible actions in the whole action space. These two implementations get the same optimal policy which is demonstrated in Fig. 3, so it is verified that the SRLF rule can preserve the optimality of actions and reduce the action space greatly at the same time.

As illustrated in Fig. 3, the optimal policy is not a threshold-type policy, which means that if the remaining lifetime of component is smaller than a threshold, this component should be replaced. Below, we demonstrate that the optimal policy obtained agrees with the theorems in Section IV-A.

From Fig. 3, we see that for the state (8, 5), the optimal decision is (0, 1); for the state (8, 9), the optimal decision is (0, 0); for the state (8, 11), the optimal decision is (1, 0); for the state (8, 12), the optimal decision is (0, 0). It means that when the remaining lifetime of component 1 is fixed and the remaining lifetime of component 2 is increased, the optimal decision of component 2 is nonincreasing. This agrees with Theorem 4.

From Fig. 3, we also see that for the state (6, 8), the optimal decision is (1, 0). For the left parts of the horizontal line across (6, 8), their optimal decisions are all (1, 0). On the other hand, for the state (8, 5), the optimal decision is (0, 1). For the down parts of the vertical line across (8, 5), their optimal decisions are all (0, 1). These agree with Theorem 3.

Moreover, we give an example which is somewhat counterintuitive but still accordant with the SRLF rule, thus shows the SRLF rule is a nontrivial rule. For the state (8, 10), the optimal decision is (1, 0). Since the state (6, 6) is worse than (8, 10), we may empirically think that it needs more replacement. But the optimal decision for (6, 6) is (0, 0). It is counterintuitive, but still accordant with the SRLF rule. It can be partially explained as follows. After we replace component 1 for state (8, 10), the state will become (10, 10), where the remaining lifetimes of two components are the same and we call these two components "synchronized." The synchronized components can be utilized more efficiently, since probably we can replace the two components when they are both exhausted after ten time units. Thus, the remaining lifetimes of these two components are not wasted in this situation. It is also easy to understand that the optimal decision for state (6, 6) is (0, 0), since the components are synchronized and we can utilize them more efficiently when they are expired after six time units. This is an interesting phenomenon in the multicomponents maintenance problem and deserves further investigation.

The usage of Theorems 3 and 4 can also be demonstrated in this figure. For example, if we know the optimal decision for the state (5, 7) is (1, 1), with Theorem 3 we can infer that the optimal decisions for the rectangle area located at the left-down of the point (5, 7) are all (1, 1). For another example, if we know the optimal decision for the state (7, 6) is (0, 0), with Theorem 4 we can infer that the optimal decisions for the points located at the up parts of the vertical line across (7, 6) are all $(*, 0)$. We can also infer that the optimal decisions for the points located at the right parts of the horizontal line across (7, 6) are all $(0, *)$. With these theorems, the process of choosing the optimal decisions can be simplified greatly.

B. Experiment 2

To demonstrate the efficiency of the SRLF rule, we give a comparison of two simulation experiments of Algorithm 1. One is implemented with the SRLF rule and the other is implemented without this rule. The cost of each shop visit is $c_0 = 10$ and the random failure probability is $p = 0.015$. For simplicity, we assume that all the components are identical and the new lifetime is 100, the price is 2. The simulation length of the two experiments are both $T = 10^5$. The selection of basis functions is important for Algorithm 1. We choose the basis functions as follows. The first basis function indicates whether the asset is in maintenance status, i.e., $\psi_1(X_t) = w_t$. The other basis functions are defined as the number of components whose remaining lifetimes are between a particular zone, e.g., $\psi_k(X_t) = |\{i : (k-2)\Delta \le s_t(i) < (k-1)\Delta, i = 1, 2, \ldots, n\}|, k = 2, \ldots, K$, where Δ is the separate time period. In this experiment, we set $\Delta = 5$. The learning step-size is $\gamma_t = 0.0001, \epsilon = 0.01$, and $\lambda = 0.1$. The selection of these parameters are mainly based on our experiences of this project and trial-and-error. The initial state X_0 and coefficients $\mathbf{v}$ are chosen arbitrarily.

The simulation program is implemented with the MATLAB v7.1 and it is tested on a personal computer with an Intel Pentium D 2.8 GHz CPU, 2 GB RAM, and Windows XP Professional operating system. When increasing the number of components n from 2 to 11, we compare the computation time of the two simulations. The results are illustrated by Fig. 4. It is found that without the SRLF rule, the computation time of Algorithm 1 will grow exponentially with the system size, while the computation time of Algorithm 1 with SRLF rule will be approximately constant. Therefore, it is demonstrated that the SRLF rule cuts down the action space greatly and improves the efficiency of optimization algorithms.

Below, we give another experiment about using the TD(λ) algorithm with the SRLF rule to optimize a practical asset maintenance problem. The number of components is $n = 30$ and the new lifetimes of components are $\mathbf{s}^0 = (112, 225, 130, 152, 241, 280, 175, 192, 102, 233, 61, 201, 236, 247, 142, 138, 226, 92, 124, 269, 211, 119, 188, 195, 129, 243, 264, 193, 172, 148)^T$. The prices of components are $\mathbf{c} = (1, 2, 2, 3, 1, 3, 2, 1, 3, 1, 1, 1, 2, 2, 3, 1, 1, 1, 3, 1, 1, 2, 1, 1, 2, 2, 1, 1, 1, 2)^T$. The total simulation length is $T = 10^6$ and we choose an arbitrary initial state X_0. The initial coefficient vector $\mathbf{v}$ can also be chosen arbitrarily and here we set it as all zeros. All the other parameters are the same as those in the previous experiment. We consider the average cost criterion to minimize the average maintenance cost. The simulation results are illustrated by Fig. 5. We run the simulation for ten replications. The mean cost is 0.6551 and the standard deviation is 0.0036. As a comparison, we consider the "one-stage analysis" method used in [20]. The main idea of this method is to choose the action which minimizes the cost of current shop visit over the expectation of time period to the next shop visit. The simulation length is also $T = 10^6$ and we run the simulation for ten replications. The mean cost is 0.6683 and the standard deviation is 0.0025. From these two experiments, we can see that Algorithm 1 with the SRLF rule has a competitive performance compared with the "one-stage analysis" method. Moreover, in Algorithm 1, since quite a few parameters are selected by the user's experience, it is possible to find an even

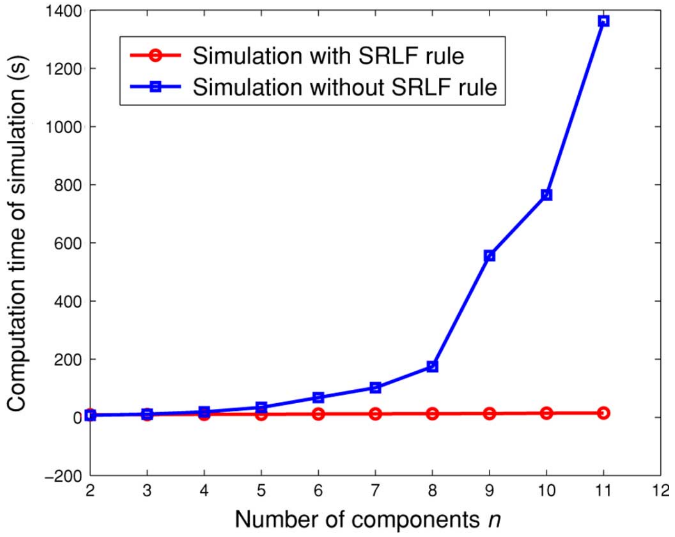


Fig. 4. Comparison of computation time in two simulations.

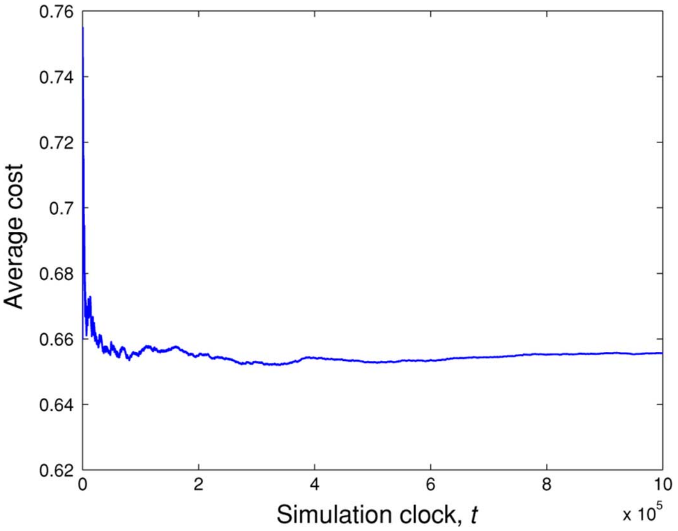


Fig. 5. Simulation results of TD(λ) algorithm with the SRLF rule.

better parameter setting and improve the algorithm performance. The selection of parameters is heavily dependent on the user's experience of the projects and trial-and-error. Therefore, it is demonstrated that Algorithm 1 is quite promising for such a large-scale stochastic optimization problem. However, if we do not use the SRLF rule, we find that the simulation can be hardly executed, because the action space, $|\mathcal{A}(X)| = 2^{30}$, is too large to search with formula (5).

VII. Conclusion

In this paper, a new joint replacement maintenance model with safety-critical components is introduced. This maintenance problem is modelled as an MDP problem. We find the SRLF rule to reduce the large action space. The optimality of the rule has been proved and some related theorems are also derived. With this rule, we find that the optimal policy has monotone property with respect to the remaining lifetimes of components. The rule and related theorems are quite helpful for us to choose the optimal replacement decisions in the maintenance problem. Furthermore, we give the on-line learning algorithm which combines the advantages of NDP techniques and the SRLF rule. This algorithm is promising to handle the large-scale maintenance problem. However, there are still some problems that need further investigation. For example, according to the property of this type of maintenance problems, we have the problem of how to choose the basis functions of the TD(λ) algorithm. It has important effects on the algorithm performance. Another future work is how to extend our results to a more complicated model which considers the random failure and interdependency among components. Our work gives some useful insights into these types of maintenance problems.

Appendix

Proof of Lemma 1: We show how to construct such a policy π' as follows. Define $T_{\min} = \min_{i=1,2,\ldots,n}\{s_0(i)\}$ and $T'_{\min} = \min_{i=1,2,\ldots,n}\{s'_0(i)\}$. Since $\mathrm{s}'_0 \geq \mathrm{s}_0$, we have $T_{\min} \leq T'_{\min}$.

Case 1: $T_1 < T_{\min}$.

In this case, the first shop visit in trajectory (ξ, π, s_0) is caused by a random failure, because at time T_1 the shortest remaining lifetime of components is $T_{\min} - T_1 > 0$. Since the randomness in both trajectories (ξ, π, s_0) and $(\xi, \pi', \mathrm{s}'_0)$ is identical, the first shop visit in trajectory $(\xi, \pi', \mathrm{s}'_0)$ is also caused by the random failure, and $T'_1 = T_1$. At time T'_1, we can construct the policy π' such that the action of $(\xi, \pi', \mathrm{s}'_0)$ is the same as that of (ξ, π, s_0) at time T_1, i.e., $d'_{T'_1}(i) = d_{T'_1}(i)$, for all i. As shown in Fig. 6, we have $\mathrm{s}'_{T_1'^+} \geq \mathrm{s}_{T_1'^+}$ and $f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_1'^+)} = f^{\xi,\pi,\mathrm{s}_0}_{(0,T_1'^+)}$.

Case 2: $T_1 = T_{\min}$.

In this case, the first shop visit in trajectory (ξ, π, s_0) is caused by the expiration of some components, and there is no random failure during period $(0, T_1)$. Because the randomness in both trajectories is identical and $T'_{\min} \geq T_{\min}$, the first shop visit in trajectory $(\xi, \pi', \mathrm{s}'_0)$ happens later due to either random failure or the expiration of some components, i.e., $T'_1 \geq T_1$. At time T'_1, if component i is replaced during period $[0, T'_1]$ in trajectory (ξ, π, s_0), then the policy π' replace this component too; otherwise, it is not replaced, i.e., $d'_{T'_1}(i) = \max_{0\leq t\leq T'_1} d_t(i)$, for all i. Then we have $\mathrm{s}'_{T_1'^+} \geq \mathrm{s}_{T_1'^+}$. For illustration purposes, we show a simple example in Fig. 7. Because the number of shop visits and the set of replaced components in trajectory $(\xi, \pi', \mathrm{s}'_0)$ are not larger than those in trajectory (ξ, π, s_0) during period $[0, T_1^+]$, we have $f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_1'^+)} \leq f^{\xi,\pi,\mathrm{s}_0}_{(0,T_1'^+)}$. This completes the proof. □

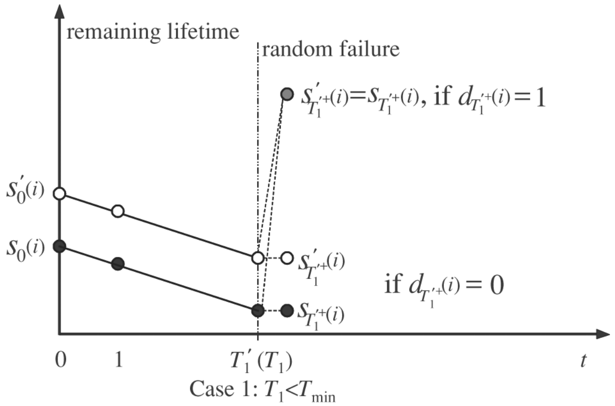


Fig. 6. Sample-path-based proof of case 1 in Lemma 1.

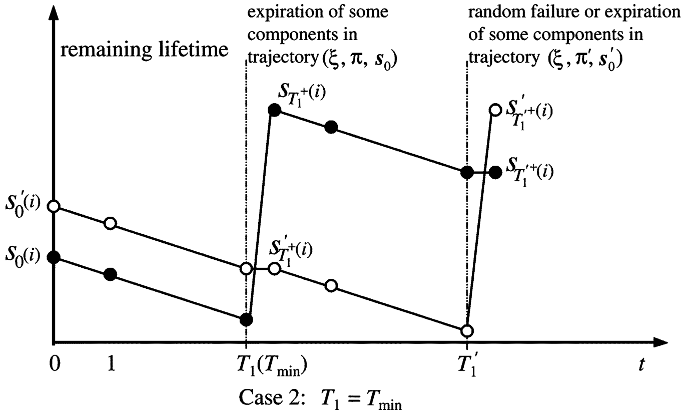


Fig. 7. Sample-path-based proof of case 2 in Lemma 1.

Proof of Proposition 1: First, we consider the case $t = T_1'^+, T_2'^+, \ldots$. Lemma 1 is the special case of Proposition 1, where $t = T_1'^+$. Using Lemma 1, we can construct a policy π' s.t. $\mathrm{s}_{T_1'^+} \geq \mathrm{s}_{T_1'^+}$ and $f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_1'^+)} \leq f^{\xi,\pi,\mathrm{s}_0}_{(0,T_1'^+)}$. If we regard $\mathrm{s}'_{T_1'^+}$ and $\mathrm{s}_{T_1'^+}$ as the initial conditions of period $(T_1'^+, T_2'^+)$ and apply Lemma 1 again, we can construct a policy π' s.t. $\mathrm{s}'_{T_2'^+} \geq \mathrm{s}_{T_2'^+}$ and $f^{\xi,\pi',\mathrm{s}'_0}_{(T_1'^+,T_2'^+)} \leq f^{\xi,\pi,\mathrm{s}_0}_{(T_1'^+,T_2'^+)}$. Following this idea and noting that $f^{\xi,\pi,\mathrm{s}_0}_{(0,T_u'^+)} = f^{\xi,\pi,\mathrm{s}_0}_{(0,T_1'^+)} + f^{\xi,\pi,\mathrm{s}_0}_{(T_1'^+,T_2'^+)} + \cdots + f^{\xi,\pi,\mathrm{s}_0}_{(T_{u-1}'^+,T_u'^+)}$ and $f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_u'^+)} = f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_1'^+)} + f^{\xi,\pi',\mathrm{s}'_0}_{(T_1'^+,T_2'^+)} + \cdots + f^{\xi,\pi',\mathrm{s}'_0}_{(T_{u-1}'^+,T_u'^+)}$, we can prove Proposition 1 for $t = T_1'^+, T_2'^+, \ldots$.

Second, we consider the case $T_u'^+ < t < T_{u+1}'^+$. By definition, there is no shop visit during period $(T_u'^+, T'_{u+1})$ and the maintenance cost happens only in shop visits. Therefore, we have $f^{\xi,\pi',\mathrm{s}'_0}_{(0,t)} = f^{\xi,\pi',\mathrm{s}'_0}_{(0,T_u'^+)}$. But in trajectory (ξ, π, s_0), there may be some shop visits during period $(T_u'^+, t)$, so $f^{\xi,\pi,\mathrm{s}_0}_{(0,t)} \geq f^{\xi,\pi,\mathrm{s}_0}_{(0,T_u'^+)}$. Then it is obvious that $f^{\xi,\pi',\mathrm{s}'_0}_{(0,t)} \leq f^{\xi,\pi,\mathrm{s}_0}_{(0,t)}$. This completes the proof. □

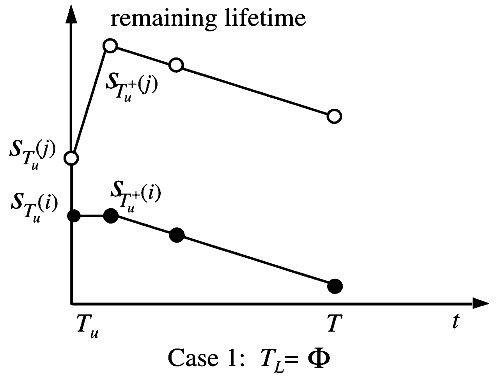


Fig. 8. Example for case 1 in the proof of Proposition 2.

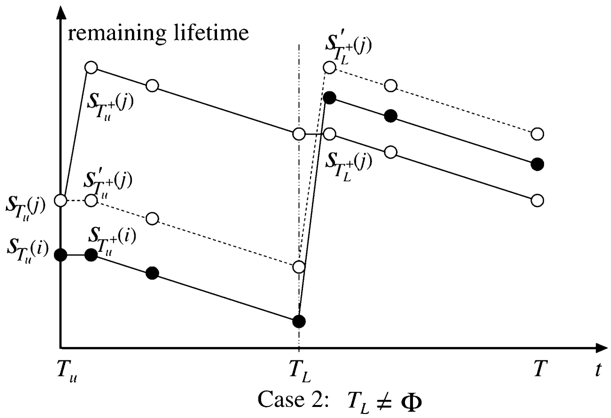


Fig. 9. Example for case 2 in the proof of Proposition 2.

Proof of Proposition 2: We construct such a policy π' as follows. Let $\pi'_t = \pi_t$ for $0 \le t < T_u$, so we have $f^{\xi,\pi',\mathrm{s}_0}_{(0,T_u)} = f^{\xi,\pi,\mathrm{s}_0}_{(0,T_u)}$ and $\mathrm{s}'_{T_u} = \mathrm{s}_{T_u}$. At time T_u, let policy π' not replace component j and take actions identical to policy π on the rest of the components, i.e., $d'_{T_u}(k) = d_{T_u}(k)$, for all $k \neq j$; $d'_{T_u}(j) = 0$. Define T_L as the first time that component i is replaced after T_u in trajectory (ξ, π, s_0), i.e., $T_L = \min\{t : d_t(i) = 1, T_u < t \le T\}$.

Case 1: $T_L = \emptyset$.

It means that component i is not replaced during period (T_u, T). Because $s_{T_u}(j) > s_{T_u}(i)$, the remaining lifetime of component j is always positive during period (T_u, T) in both trajectories. Therefore, no shop visit (if any) is caused by the expiration of component j. As an illustration, we show a simple case in Fig. 8. Let $\pi'_t = \pi_t$, for all $T_u < t \le T$. Then we can see that the numbers of shop visits in both trajectories are identical. Since component j is replaced at time T_u in (ξ, π, s_0) but not in $(\xi, \pi', \mathrm{s}_0)$, we have $f^{\xi,\pi',\mathrm{s}_0}_{(T_u,T)} < f^{\xi,\pi,\mathrm{s}_0}_{(T_u,T)}$.

Case 2: $T_L \neq \emptyset$.

Let $\pi'_t = \pi_t$ for $T_u < t < T_L$. Then we have $f^{\xi,\pi',\mathrm{s}_0}_{(T_u^+,T_L)} = f^{\xi,\pi,\mathrm{s}_0}_{(T_u^+,T_L)}$. At time T_L, let policy π' replace component j and take actions identical to policy π on the rest of the components, i.e., $d'_{T_L}(k) = d_{T_L}(k)$, for all $k \neq j$; $d'_{T_L}(j) = 1$. Then we have $\mathrm{s}'_{T_L^+} \ge \mathrm{s}_{T_L^+}$. As an example, we show a simple trajectory in Fig. 9. At time T_u and T_L, component j is replaced only once in trajectory $(\xi, \pi', \mathrm{s}_0)$, but no less than once in trajectory (ξ, π, s_0), so $f^{\xi,\pi',\mathrm{s}_0}_{(T_u,T_L^+)} \le f^{\xi,\pi,\mathrm{s}_0}_{(T_u,T_L^+)}$. For period (T_L^+, T), since $\mathrm{s}'_{T_L^+} \ge \mathrm{s}_{T_L^+}$, Proposition 1 ensures the existence of π', s.t. $f^{\xi,\pi',\mathrm{s}_0}_{(T_L^+,T)} \le f^{\xi,\pi,\mathrm{s}_0}_{(T_L^+,T)}$, so in this case, we also have $f^{\xi,\pi',\mathrm{s}_0}_{(T_u,T)} \le f^{\xi,\pi,\mathrm{s}_0}_{(T_u,T)}$.

Combining Case 1, Case 2, and $f^{\xi,\pi',\mathrm{s}_0}_{(0,T_u)} = f^{\xi,\pi,\mathrm{s}_0}_{(0,T_u)}$, we have constructed π' which is accordant with the SRLF rule and $f^{\xi,\pi',\mathrm{s}_0}_{(0,T)} \le f^{\xi,\pi,\mathrm{s}_0}_{(0,T)}$. This completes the proof. □

ACKNOWLEDGMENT

The authors would like to thank Prof. P. B. Luh and Prof. X. R. Cao for their helpful suggestions. The authors also wish to thank Prof. X. Chen, Prof. X. P. Guo, Y. Xu, T. Sun, and G. Tu for their discussions and comments. Particularly, the authors thank the anonymous referees and the Associate Editor for their constructive comments that helped improve the quality of this paper.

REFERENCES

[1] R. W. Aldhaheri and H. K. Khalil, "Aggregation of the policy iteration method for nearly completely decomposable Markov chains," *IEEE Trans. Autom. Control*, vol. 36, no. 2, pp. 178–187, Feb. 1991.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[3] R. A. Bowman and J. Schmee, "Pricing and managing a maintenance contract for a fleet of aircraft engines," *Simulation*, vol. 76, no. 2, pp. 69–77, 2001.

[4] X. R. Cao, Z. Ren, S. Bhatnagar, M. C. Fu, and S. I. Marcus, "A time aggregation approach to Markov decision processes," *Automatica*, vol. 38, pp. 929–943, 2002.

[5] X. R. Cao, "From perturbation analysis to Markov decision processes and reinforcement learning," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 13, pp. 9–39, 2003.

[6] D. I. Cho and M. Parlar, "A survey of maintenance models for multi-unit systems," *Eur. J. Operat. Res.*, vol. 51, pp. 1–23, 1991.

[7] R. Dekker, "Applications of maintenance optimization models: A review and analysis," *Reliability Eng. Syst. Safety*, vol. 51, pp. 229–240, 1996.

[8] R. Dekker, F. A. Van der Duyn Schouten, and R. E. Wildeman, "A review of multi-component maintenance models with economic dependence," *Math. Methods Operat. Res.*, vol. 45, pp. 411–435, 1997.

[9] R. Dekker, R. E. Wildeman, and R. van Egmond, "Joint replacement in an operational planning phase," *Eur. J. Operat. Res.*, vol. 91, pp. 74–88, 1996.

[10] E. A. Feinberg and A. Shwartz*, Handbook of Markov Decision Processes, Methods and Applications*. Norwell, MA: Kluwer, 2002.

[11] A. Haurie and P. Lecuyer, "A stochastic control approach to group preventive replacement in a multicomponent system," *IEEE Trans. Autom. Control*, vol. 27, no. 2, pp. 387–393, Apr. 1982.

[12] W. J. Hopp and Y. L. Kuo, "Heuristics for multicomponent joint replacement: Applications to aircraft engine maintenance," *Naval Res. Logistics*, vol. 45, pp. 435–458, 1998.

[13] R. A. Howard, *Dynamic Programming and Markov Processes*. New York: Wiley, 1960.

[14] S. Ozekici, "Optimal periodic replacement of multicomponent reliability systems," *Operat. Res.*, vol. 36, pp. 542–552, 1988.

[15] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ: Wiley, 1994.

[16] B. V. Roy, D. P. Bertsekas, Y. C. Lee, and J. N. Tsitsiklis, "A neuro-dynamics programming approach to retailer inventory management," in *Proc. 36th Conf. Decision and Control*, San Diego, CA, Dec. 1997.

[17] J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica*, vol. 65, p. 487C516, 1997.

[18] F. A. Van der Duyn Schouten and S. G. Vanneste, "Two simple control policies for a multi-component maintenance system," *Operat. Res.*, vol. 41, pp. 1125–1136, 1993.

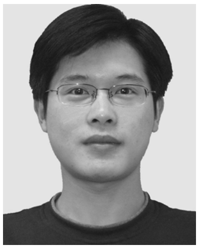
[19] Y. Sherif and M. Smith, "Optimal maintenance models for systems subject to failure—A review," *Naval Res. Logistics*, vol. 28, pp. 47–14, 1981.

[20] T. Sun, Q. C. Zhao, P. B. Luh, and R. N. Tomastik, "Joint replacement optimization for multi-part maintenance problems," in *Proc. 2004 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Sendai, Japan, Sep. 2004, pp. 1232–1237.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[22] G. Y. Tu, P. B. Luh, Q. C. Zhao, and R. N. Tomastik, "An optimization method for joint replacement decisions in maintenance," in *Proc. 43rd Conf. Decision and Control*, Atlantis, Paradise Island, Bahamas, Dec. 2004, pp. 3674–3679.

[23] H. Z. Wang, "A survey of maintenance policies of deteriorating systems," *Eur. J. Operat. Res.*, vol. 139, pp. 469–489, 2002.



Li Xia (S'02) received the B.E. degree in automation, in July 2002, and the Ph.D. degree in control science and engineering, in July 2007, both from Tsinghua University, Beijing, China.

He is currently with the research staff at IBM China Research Laboratory, Beijing, China. His research interests include discrete event dynamic systems (DEDS) theory and applications, simulation optimization techniques, and wireless sensor networks.

Qianchuan Zhao (M'06) received the B.E. degree in automatic control in July 1992, the B.S. degree in applied mathematics in July 1992, and the Ph.D. degree in control theory and its applicaitons in July 1996, all from Tsinghua University, Beijing, China.

He is currently a Professor and the Associate Director of the Center for Intelligent and Networked Systems (CFINS), Department of Automation, Tsinghua University, Beijing, China. He was a Visiting Scholar at Carnegie Mellon University and Harvard University in 2000 and 2002, repsectively. He was a Visiting Professor at Cornell University in 2006. His research interests include discrete event dynamic systems (DEDS) theory and applications, optimalization of complex systems, and wireless sensor networks. He is an Associate Editor for the *Journal of Optimization Theory and Applications* (JOTA).

Qing-Shan Jia (S'02–M'06) received the B.E. degree in automation in July 2002 and the Ph.D. degree in control science and engineering in July 2006, both from Tsinghua University, Beijing, China.

He is currently a Lecturer at the Center for Intelligent and Networked Systems (CFINS), Department of Automation, Tsinghua University, Beijing, China. He was a Visiting Scholar at Harvard University in 2006. His research interests include discrete event dynamic systems (DEDS) theory and applications and simulation-based performance evaluation and optimalization of complex systems.