A risk-aware maintenance model based on constrained Markov decision process

Jianyu Xu [a], Xiujie Zhao [b] and Bin Liu [c]
[a] International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China
[b] College of Management and Economics, Tianjin University, Tianjin, China
[c] Department of Management Science, University of Strathclyde, Glasgow, UK

Abstract

The Markov decision process (MDP) model has been widely studied and used in sequential decision making problems. In particular, it has been proved to be effective in maintenance policy optimization problems where the system state is supposed to continuously evolve under sequential maintenance policies. In traditional MDP models for maintenance, the long-run expected total discounted cost is concerned as the objective function. The maintenance manager's target is to evaluate an optimal policy that incurs the minimum expected total discounted cost through the corresponding MDP model. However, a significant drawback of these existing MDP-based maintenance strategies is that they fail to incorporate and characterize the safety issues of the system during the maintenance process. Therefore, in some applications that are sensitive to functional risks, such strategies fail to accommodate the requirement of risk awareness. In this study, we apply the concept of risk-aversion in the MDP maintenance model to develop risk-aware maintenance policies. Specifically, we use risk functions to measure some indexes of the system that reflect the safety level and formulate a safety constraint. Then, we summarize the problem as a constrained MDP model and use the linear programming approach to evaluate the proposed risk-aware optimal maintenance policy under concern.

Keywords: Risk aversion; condition-based maintenance; safety constraint; Markov decision process; imperfect repair.

1 Introduction

Maintenance has played an important role in preventing system failure and sustaining a safe operation for various industries. With the advances of sensor technology, system condition data can be captured in a much easier way, prompting the shift of maintenance strategies from age-based maintenance to condition-based maintenance (CBM). CBM tends to intervene the system based on the observed system state. Compared with the traditional

time-based maintenance, CBM is able to prevent system failure in a "timely" manner. Superiority of CBM has been recognized from both the academic and practical point of view (Ahmad and Kamaruddin, 2012; Zhao et al., 2021).

Renewal process theorem and Markov decision process (MDP) are the most commonly used approaches to model CBM processes. Renewal process tends to describe the maintenance process by a series of renewal cycles, which occur whenever the system is restored to the as-good-as-new state upon maintenance. Compared with renewal process, MDP is more flexible in CBM modeling in the sense that it can easily characterize multiple maintenance actions that does not constitute a renewal cycle. In addition, MDP is able to model maintenance process for a finite horizon, while renewal cycle theorem will find it tedious. In literature, a large body of studies have appeared on CBM modeling using MDP and its variants Zhu and Xiang (2021). Papakonstantinou and Shinozuka (2014) conducted a survey on application of the MDP models in structural inspection and maintenance policies. Zhang and Revie (2017) developed a partially observable Markov decision process (POMDP) to model the decision-making in maintenance, with application to rapid gravity filters of a water utility. Chen et al. (2015) developed a monotone control-limit CBM policy considering the update of degradation parameters. Elwany et al. (2011) formulated the maintenance problem into a MDP model for systems subject to continuous monitoring. A monotone control-limit policy was devised considering measurement noise. Liu et al. (2017) developed a MDP model for CBM considering age-state-dependent operating cost. Junca and Sanchez-Silva (2013) presented a maintenance model for systems deteriorating as a result of shocks and the optimal decision is obtained based on MDP. Byon et al. (2010) examined the optimal maintenance strategy for wind turbines considering stochastic weather conditions. A POMDP model was established to describe the maintenance process. Havinga and de Jonge (2020) formulated a MDP model for the cyclic patrolling repairman problem considering condition-based preventive maintenance. Lagos et al. (2020) developed a MDP model for airline maintenance operations. Flory et al. (2015) addressed the maintenance problem for a continuously degrading system that operates in a partially observable environment and formulated the problem as a POMDP.

While one can witness the popularity of MDP in maintenance modeling, a limitation is that almost all the studies focus on the single criterion of maintenance cost (e.g., aver-

age maintenance cost, discounted total maintenance cost, etc). However, the criterion of maintenance cost overlooks the risks associated with occasional high-cost events that occur during system operation (Gosavi, 2006). In particular, the expectation of the long-term cost which is extensively adopted as the criterion fails to capture the fluctuation of the observation sequence, thus fails to evaluate the potential risks associated with such fluctuation. In many practical scenarios, risk awareness is concerned with a high priority during system operation (Geibel and Wysotzki, 2005; Ngo and Krishnamurthy, 2009; Garcıa and Fernández, 2015; Tamar et al., 2016; Xu et al., 2021; Meraklı and Küçükyavuz, 2020). Similarly, in many cases where successive maintenance actions are conducted, risk-averse criterion is necessary to accommodate the maintenance managers who retain sensitivity to risky system conditions. For example, safety-critical systems have to keep a large margin from the failure threshold due to industry standards or safety concerns, and only concerning the average cost is not enough in the long run. In such situations, classical maintenance policies evaluated from MDP models aiming at minimizing expected total cost is no longer applicable. What a maintenance manager expects is a more sophisticated model that can quantify the risk sensitivity and determine the optimal maintenance strategy incorporating this factor (Gosavi, 2006; Wu et al., 2017). Actually, risk-aware MDP has been applied in various industries, such as portfolio selection (Huo and Fu, 2017; Zhang et al., 2020), option pricing (Chow and Ghavamzadeh, 2014), and energy storage systems (Xia, 2020).

One motivating example of the study is maintenance on bridge joints. Bridge joints are used to accommodate the necessary movements of bridge decks, withstand the traffic load, and protect bearings from corrosion (Liu et al., 2017). Asphalt Plug Joint (APJ) is one of the most common bridge joints due to its advantages in waterproof and noise reduction. APJs have an expected lifetime of 5 to 15 years based on the operating environment. Based on the knowledge from local maintenance experts, APJs deteriorate as a result of aging and environmental factors such as accumulated debris, corrosion and traffic load. Malfuction of an APJ will increase the failure risk of bridge deck and bearing, and may even lead to bridge failure. To mitigate the risk of APJ failure, general inspection is regulated to evaluate the health condition of APJ joints. Upon inspection, the maintenance engineers are eager to know whether they should repair or replace the APJ. They want a maintenance policy that minimizes the maintenance cost while can sustain a high safety

level. Another example is the maintenance of water treatment systems that are widely deployed in the petrochemical industry. The treatment vessel in the system is subject to periodic inspections and restoration of performance on demand. The shutdown of the water treatment system caused by the vessel failure can cause costly capacity loss and thus decision makers wish to control the risk of shutdown at a lower level. Meanwhile, to restore the performance of water treatment vessels costs considerable materials and manpower, and it is of great interest to the asset owner to reduce the restoration cost under a constrained risk level of shutdown.

To this end, this study aims to develop a risk-aware maintenance model considering safety constraint. The system under investigation is assumed to deteriorate according to a Markov chain and a safety constraint is imposed upon the system to sustain a reliable operation. Several approaches have been investigated for such problems in existing research, such as constructing objective functions based on risk-aware functions (Ruszczyński, 2010; Xu et al., 2021) or using robustness formulation for the model construction against adversarial environment (Kim and Lim, 2015). In our work, we choose to incorporate the safety requirements as risk-aware constraint functions in the maintenance model and formulate a constraint MDP model. In particular, we firstly introduce a constraint function that measures the safety level of the system during maintenance process and subsequently develop a risk-aware maintenance model. Then, we choose a metric that measures the safety level of the system according to the system state. Under maintenance policies, the system state continuously evolves and generates a sequence of stochastic states. Therefore, the safety level of the system, as a function of the system state, is also a stochastic sequence during the maintenance process. Following the paradigm of risk management, we use the risk measures (Artzner et al., 1999; Ruszczyński and Shapiro, 2006) to measure the risk of the (stochastic) safety level of the system through the maintenance process. Moreover, a risk threshold is imposed on the system and the risk measure of the safety level shall stay above the threshold to sustain a reliable operation. Based on these discussions, the maintenance problem is formulated as a constrained MDP. We restrict the risk measures concerned in this work to two popular risk measures, namely, value-at-risk (VaR) (Jorion, 2007) and conditional value-at-risk (CVaR) (Rockafellar and Uryasev, 2000). VaR calculates the quantile of the safety level, thus using VaR in the constraint function means that the safety level is

supposed to stay above some preset lower limit with a high probability. Constraint function using VaR is "hard" in the sense that it is too strict for any maintenance policy to be feasible in some problem settings. CVaR is an alternative to VaR in situations where soft safety constraint is needed. Different from VaR, CVaR involves the quantile of the safety level in a cumulative way. Thus, the constraint function using CVaR requires the average system safety level stays above a preset value and is more relaxed than the VaR constraint. Finally we prove that the underlying constrained MDP problem can be solved using linear programming approach and evaluate the corresponding optimal policy. Numerical properties of the approach are investigated through a series of simulation studies. Based on the discussion above, we summarize the contributions of this work as follows.

(i) We introduce risk-aversion in maintenance problems by using risk measures to construct a metric of safety level and formulate the problem as a MDP model with safety constraints.

(ii) We use occupation measure method and transfer the non-linear safety constraint into linear functions in terms of occupation measure. Then, we transform the model into a linear programming problem.

(iii) Based on the theory of MDP with linear constraints, we prove that the existence of the optimal policy and provide its specific form.

The rest of the paper is organized as follows. Section 2 presents the model formulation of classical maintenance problem using the MDP models. Section 3 proposes the constraint function using risk measures and incorporates the safety constraint in the MDP model to formulate a constrained MDP model for the risk-aware maintenance problem. Then, two specific risk measures are introduced in our model. Section 4 develops an approach to evaluate the optimal maintenance policy for our model using the linear programming method. Section 5 investigates the numerical properties through examples and simulation studies. Section 6 summarizes the work and gives some final concluding remarks.

Notation list

$\mathcal{X}$	Set of system states
R	Transition probability of system state under natural deterioration
$\mathcal{A}$	Set of maintenance actions
$\mathcal{A}^x$	Set of feasible maintenance actions when system state is x.
M^a	Transient transition probability of system state under maintenance action a
P^a	Transition probability of system state under maintenance action a
c	Maintenance cost function
λ	Discount factor
π	A generic maintenance policy
V^π	Value function of policy π
π^*	The optimal maintenance policy
V^*	Value function of the optimal policy
q	Safety metric of system state
D_λ^π	Discounted stationary probability distribution under maintenance policy π
ρ	Risk measure
α	Confidence level
VaR_α	Value-at-risk measure
CVaR_α	Conditional value-at-risk measure

2 Problem formulation

Consider a system whose health condition is discretized into a finite number of states. Let $\mathcal{X} \triangleq \{1, 2, ..., N\}$ be the set of all possible system states, where states are in descending order in terms of the health condition of the system. In particular, state 1 indicates perfect (new) state and state N represents the failure state. Assume that the system is under periodic inspections at time $i\Delta t$, where $i = 0, 1, \cdots$, denoting the number of inspection, and Δt is the inspection interval. We assume that the system state is observed accurately each time. Let X_i be the state of the system at the ith inspection. When no maintenance action is taken, we assume that system state evolves according to a transition probability

matrix $R \triangleq \{R(x,\, y);\, x,\, y \in \mathcal{X}\}$, where

$$R(x,\, y) = P\{X_{i+1} = y | X_i = x\},\ \forall x,\, y \in \mathcal{X}.$$

We conclude that R can be induced by the physical model of the dynamics of system state in practice, e.g., the degradation of some critical quality characteristic of the system.

Let $\mathcal{A} \triangleq \{0, 1, ..., M\}$ be the set of all allowable maintenance actions before failure, and denote by $a \in \mathcal{A}$ a generic maintenance action in $\mathcal{A}$. The maintenance actions are labeled so that the effectiveness is increasing in the index number, e.g., $a = 0$ indicates that no maintenance action is taken and $a = M$ implies a replacement. At the ith inspection time, the decision-maker is supposed to observe the system state and then select a maintenance action a_i in $\mathcal{A}$.

We assume that the time on maintenance is negligible. If the system is found failed when the state is observed, the decision maker is supposed to choose only replacement as the maintenance action. To incorporate these requirements in the model, we define $\mathcal{A}^x \subset \mathcal{A}$ to be the set of all feasible maintenance actions for each state $x \in \mathcal{X}$, specifically,

$$\mathcal{A}^x = \begin{cases} \{M\}, & x = N \\ \mathcal{A}, & x \neq N \end{cases}.$$

Moreover, we define $\Theta \triangleq \{(x,\, a) \in \mathcal{X} \times \mathcal{A} :\ a \in \mathcal{A}^x\}$ as the set of all feasible state action pairs, each of which represents a combination of a possible system state and a feasible action under this state.

For all $a \in \mathcal{A}$, if a is taken, the system state makes an instant transition from the state X^- before maintenance to the state X^+ after maintenance, according to the transition probability matrix $M^a \triangleq \{M^a(x,\, y);\, x,\, y \in \mathcal{X}\}$, where

$$M^a(x,\, y) = P\{X^+ = y | X^- = x,\, a_i = a\},\ \forall x,\, y \in \mathcal{X}.$$

Based on our discussions about the maintenance actions, we conclude that Q^a satisfies the following conditions:

(C1). Maintenance on a worse initial state leads to a stochastically worse state, i.e.,

$$\sum_{y=s}^{N} M^a(x_1,\, y) \geq \sum_{y=s}^{N} M^a(x_2,\, y)\ \ \forall\ \ x_1 > x_2,\ s \in S,$$

(C2). Less efficient action leads to a stochastically worse state, i.e.,

$$\sum_{y=1}^{s} M^{a_1}(x,\, y) \geq \sum_{y=1}^{s} M^{a_2}(x,\, y) \;\; \forall\, a_1 > a_2,\; s \in S.$$

(C3). The transition probabilities under no maintenance action ($a = 0$) and replacement ($a = M$) are given as

$$M^0(x,\, y) = \begin{cases} 1,\; y = x \\ 0,\; y \neq x \end{cases}, \forall x,\, y \in \mathcal{X},$$

and

$$M^M(x,\, y) = \begin{cases} 1,\; y = 1 \\ 0,\; y \neq 1 \end{cases} \quad \forall x,\, y \in \mathcal{X}.$$

Now we incorporate maintenance actions in modeling the evolution of the system state. Denote the transition probability matrix of system state by $P^{\mathcal{A}}$, for all $a \in \mathcal{A}$. Define $P^a \triangleq \{P^a(x,\, y);\; x,\, y \in \mathcal{X}\}$, which is given by

$$P^a(x,\, y) = P\left\{X_{i+1} = y \,|\, X_i = x,\, a_i = a\right\} = \sum_{l=1}^{N} M^a(x,\, l) R(l,\, y),\; \forall x,\, y \in \mathcal{X},\; \forall i \geq 0.$$

The logic of $P^a(x,\, y)$ goes as follows. Since a maintenance action improves system health condition before transiting to the state at the $(i+1)$th inspection time , the transition probability is formulated by enumerating all the states that the system would reach by maintenance.

For all $(x,\, a) \in \Theta$, let $c(x, a)$ denote the maintenance cost of action a given the system state x at any time. For all $x \in \mathcal{X}$, $c(x,\, a)$, as a function defined on $\mathcal{A}^x$, is supposed to be monotonically increasing in a, which indicates that a more effective maintenance action incurs a higher cost. In the paradigm of sequential decision problems with infinite time horizon, all costs are discounted to the initial time with a discount factor $\lambda \in (0,\, 1)$. Based on the above discussions, we may summarize our proposed maintenance model as a Markov decision process (MDP), represented by a five-tuple $\left(\mathcal{X},\, \mathcal{A},\, c,\, P^{\mathcal{A}},\, \lambda\right)$.

An admissible maintenance policy for the decision maker is defined as $\pi = (\pi_1,\, \pi_2, \cdots)$. For all $x \in \mathcal{X}$, each $\pi_t : x \to \mathcal{P}(\mathcal{A}^x)$ is a mapping from x to the set of all probability distributions $\mathcal{P}(\mathcal{A}^x)$ on $\mathcal{A}^x$. In particular, if all π_t's are the same, we call π a stationary

policy. Let Π be the set of all admissible maintenance policies. In reality, a policy implies the maintenance strategy of the decision maker facing an observed system state each time. Given any policy $\pi \in \Pi$ and initial system state $x \in \mathcal{A}$, the value function of policy π is defined in terms of the expected total discounted cost as

$$V^{\pi}(x) \triangleq \mathbf{E}\left[(1-\lambda)\cdot\sum_{i=0}^{\infty}\lambda^{i}c(X_t,\,\pi_t(X_t))\,\Big|\,X_0=x\right]. \tag{1}$$

For ease of discussion, we assume that the system is new when it starts functioning, which implies that the initial state of the underlying process is $X_0 = 1$ from a practical perspective. The decision maker's target is to decide a maintenance policy $\pi \in \Pi$ to minimize the expected total discounted cost given initial state $X_0 = 1$, namely, find the following optimal policy

$$\pi^{*} = \arg\min_{\pi\in\Pi} V^{\pi}(1). \tag{2}$$

Along with the assumption that maintenance actions are instantaneous, it can be shown that the value function of π^*, denoted by V^*, is the unique solution of the following dynamic programming equation

$$V^{*}(x) = (1-\lambda)\cdot\min_{a\in\mathcal{A}^{x}}\left\{c(x,\,a)+\lambda\mathbf{E}_{X\sim P^{a}(x,\,\cdot)}\left[V^{*}(X)\right]\right\},\ \forall x\in\mathcal{X}. \tag{3}$$

where $\mathbf{E}_{X\sim P^{a}(x,\,\cdot)}\left[V\left(X\right)\right] = \sum_{j=1}^{N} P^{a}(x,\,j)V(j)$. Moreover, the optimal policy π^* is given as

$$\pi^{*}(x) = \arg\min_{a\in\mathcal{A}^{x}}\left\{c(x,\,a)+\lambda\sum_{j=1}^{N}P^{a}(x,\,j)V^{*}(j)\right\}.$$

The above formulation of the problem and the optimal policy directly follow from classical results of MDP models. However, we note that the value function V^{π} in (1) only concerns the expected cost and somehow ignores the tracking of the system states. As a result, the corresponding optimal policy π^* fails to incorporate risk-aversion, and thus may lead to risky situations. For example, using only the expected cost as the criterion, there is no constraint on the total time when the system arrives at some inferior states, or even failure. Thereby, under the optimal policy π^*, the total maintenance cost may be averagely low, yet the system may stay in unhealthy situations in a significantly large proportion of time. When the application environment retains a low tolerance for the risk that is caused by unhealthy system conditions, π^* above possibly fails to accommodate

risk-aware agents. To cope with this gap, we introduce an additional constraint function that measures the safety level of any policy in terms of the health condition of the system throughout the maintenance process in the next section. Then, we formulate a risk-aware sequential maintenance model.

3 Safety constraint on maintenance actions

In situations where safety guarantees are required, researchers impose safety constraint upon the system to sustain a stable operation. In this section, we introduce a general risk function to measure the safety level of any given policy and formulate a safety constrained maintenance model. Then, we give two specific and extensively used examples of risk functions. To formulate the safety constraint, we firstly need to choose a proper index that measures the system's safety level according to its state. Various metrics can be applied to measure the safety, such as reliability value, remaining useful life, etc. In this study, we employ a more general metric $q : \mathcal{X} \to (0, 1)$ to measure the system safety level. In particular, $q(x)$ quantifies the system's safety level when system is in state $x \in \mathcal{X}$. Meanwhile, $q(x)$ is supposed to be monotonically decreasing in $x \in \mathcal{X}$, implying that a "worse" state is considered to be more risky. The exact form of q is not specified, as q is a general mapping from system state to a normalized metric. An exact form highly depends on the scenario under investigation. For example, we can use the probability of system survival until the next inspection to construct $q(x)$, indicating that the decision maker tends to keep the system operating at a reasonable probability between consequent inspections. Also, q could be a distance function reflecting the distance from system state X to the failure threshold. Under a given policy, the system continuously evolves and generates a sequence of states, causing different distributions of states each time. So, the challenge is how to use q to measure the safety level when the system state is stochastic with non-stationary distribution. We use the same idea of discounting the future observation and define the following discounted limit probability function (Mattila et al., 2017) on $\mathcal{X}$ induced by π as

$$D_\lambda^\pi(x) = (1-\lambda) \cdot \sum_{i=1}^{\infty} \lambda^t P\{X_i = x \,|\, X_0 = 1,\, \pi\},\ \ \forall x \in \mathcal{X}.$$

D^π is quite similar to the traditional stationary probability distribution of a controlled Markov chain, namely, $\lim_{T\to\infty} 1/T \sum_{i=1}^{T} P\{X_i = x \,|\, X_0 = 1,\, \pi\},\ \forall x \in \mathcal{X}$. We conclude two differences between them, because of which we use the former. The first difference is that the stationary distribution take average of the distribution each time, while the discounted distribution uses the exponential-weighted average, which is more consistent with our model setting that any future feedback of the maintenance policy is less important than the current feedback and is exponentially discounted. The second difference is that the stationary distribution does not necessarily exist under an arbitrary policy, and additional assumptions such as the uni-chain assumption is needed to ensure the existence. However, the discounted stationary distribution always exists in the sense that the set of partial sum sequences $\left\{(1-\lambda) \cdot \sum_{i=1}^{T} \lambda^{i-1} P\{X_i = x \,|\, X_0 = 1,\, \pi\}\right\}_{T=1}^{\infty}$ is a non-negative and monotonic Cauchy sequence.

Based on D^π, we define a random variable, Q_λ^π, denoting the index of system safety level under π. In particular, we let $P\{Q_\lambda^\pi = q(x)\} = D_\lambda^\pi(x)$. Namely, we use q as the safety metric and consider the discounted stationary probability $D_\lambda^\pi(\cdot)$ as some reference probability on X generated by π. It might be controversial in some situations to discount the safety metric Q_λ^π in the same way as the maintenance cost. However, we would argue the rationality from the following two perspectives. On one hand, it is somewhat practical to discount the safety level in the future. For example, with the operation and maintenance process of the system going, more knowledge about the system degradation will accumulate and the corresponding safety restriction on the system operation can be less conservative. Therefore, discounting the safety level indicates that the decision maker is more confident and imposes less risk on the system state with time going. On the other hand, to use the discounted stationary distribution consistently for both the maintenance cost and safety level, the mathematical tractability of the model can be guaranteed and the optimal policy can be solved through linear programming, as discussed in later sections.

Note that Q_λ^π is a random variable, one may consider to use the expectation $\mathbf{E}(Q_\lambda^\pi)$ to measure the system safety level under π. However, to accommodate risk-aware decision

makers, only considering $\mathbf{E}(Q_\lambda^\pi)$ is not enough. Because $\mathbf{E}(Q_\lambda^\pi)$ contains no information about the fluctuation of Q_λ^π and thus fails to detect some potential risky situations that appear with non-negligible probabilities. Following the paradigm of risk management, we introduce a risk function of Q_λ^π as the metric for safety. Broadly speaking, a (standardized) risk function in the cost-concerning context, denoted by ρ, is a mapping defined on all the space $\mathcal{L}$ of random variables taking values on [0, 1] that satisfies the following axioms (Ruszczyński and Shapiro, 2006)

(i) *Normalization*: $\rho(0) = 0$, where the "0" in $\rho(0)$ represents the random variable that always equals to 0;

(ii) *Translation invariance*: for all $r \in \mathbb{R}$ and $L \in \mathcal{L}$, $\rho(L + r) = \rho(L) + r$;

(iii) *Monotonicity*: for all L_1, $L_2 \in \mathcal{L}$ and $L_1 \geq L_2$ a.s., $\rho(L_1) \geq \rho(L_2)$.

Risk measures are widely used as a metric to judge if the risk of a specific performance index (e.g., outcome, asset and cost) is acceptable. To control the risks and maintain a high safety level, a safety-concerned maintenance policy is supposed to keep $\rho(Q_\lambda^\pi)$ at a low level. Given a preset threshold of safety value $\tau \in [0, 1]$, we claim the following safety constraint

$$\rho(Q_\lambda^\pi) \geq \tau, \tag{4}$$

in particular, $\tau = 0$ means that there is no safety constraint. By incorporating the safety constraint in the original optimization problem (2), the resulting safety constrained problem is

$$\begin{aligned} &\min_{\pi \in \Pi} V^\pi(1), \\ &\text{s.t. } \rho(Q_\lambda^\pi) \geq \tau. \end{aligned} \tag{5}$$

Remark. Since $q(x) \leq 1$ for all $x \in \mathcal{X}$, we have $Q_\lambda^\pi \leq 1$. According to the normalization and translation invariance of ρ, for all $r \in \mathbb{R}$, $\rho(r) = \rho(0 + r) = \rho(0) + r = r$. Moreover, because of the monotonicity of ρ, $Q_\lambda^\pi \leq 1$ implies that $\rho(Q_\lambda^\pi) \leq \rho(1) = 1$. Thus, to ensure that problem (5) is feasible, we always set $\tau \in [0, 1]$.

3.1 *Value at risk*

Value at risk (VaR) is a widely utilized risk function for risk-aversion in different areas. VaR calculates the quantile of a random variable. Given a confidence level $\alpha \in (0, 1)$, the VaR of a random variable W is defined as

$$\mathrm{VaR}_\alpha(W) \triangleq \inf\{r \in \mathbb{R} : \ F_W(r) \geq \alpha\} = F_W^{-1}(\alpha), \tag{6}$$

where $F_W(\cdot)$ is the cumulative distribution function (CDF) of W. $\mathrm{VaR}_\alpha(Q_\lambda^\pi)$ reveals the upper bound on Q_λ^π with a high probability α. Compared with expectation, $\mathrm{VaR}_\alpha(Q_\lambda^\pi)$ can better captures the fluctuation of Q_λ^π. Using VaR_α as the constraint, the agent can restrict the value of Q_λ^π at a high level with a high probability. Choosing VaR as ρ in (5), the safety constraint problem can be written as

$$\begin{aligned} &\min_{\pi \in \Pi} V^\pi(1), \\ &\text{s.t. } \mathrm{VaR}_\alpha(Q_\lambda^\pi) \geq \tau. \end{aligned} \tag{7}$$

3.2 *Conditional value at risk*

A widely used alternative to VaR is the conditional value at risk (CVaR). As a risk function, CVaR not only considers the quantile of a distribution, but the whole tail distribution beyond the quantile by taking average on all quantiles whose confidence level is above some preset value. Given a confidence level $\alpha \in (0, 1)$, the CVaR of a random variable W is defined as

$$\mathrm{CVaR}_\alpha(W) \triangleq (1-\alpha)^{-1} \int_\alpha^1 \mathrm{VaR}_\beta(W) d\beta. \tag{8}$$

Instead of making a hard constraint on the quantile as VaR, the CVaR constraint makes a restriction on the quantile in a cumulative way. It makes the CVaR constraint more flexible and practical in scenarios where the demand for safety is less strict. Using CVaR_α, the safety constraint problem (5) can be written as

$$\begin{aligned} &\min_{\pi \in \Pi} V^\pi(1), \\ &\text{s.t. } \mathrm{CVaR}_\alpha(Q_\lambda^\pi) \geq \tau. \end{aligned} \tag{9}$$

Since the confidence level α is decided by the decision maker, here and thereafter, we fix the confidence level α for both VaR and CVaR. The challenge of solving (7) and (9) lies in the non-linear constraint function. Due to the existence of constraint, classical dynamic programming method is no longer feasible. Alternatively, we choose to use the linear programming method based on occupation measure which is specified in the following context. However, the safety constraint is non-linear so that existing linear programming methods for MDP with linear constraints can not be directly used. In the next section, we will show that the safety constraint can be transferred to an equivalent linear function in terms of the occupation measure and the optimal maintenance policy can be solved using conventional numerical methods for linear programming.

4 Optimal policy

In this section, we achieve the optimal policy for Problem (7) and (9) using the linear programming method for MDP models. To proceed, we introduce the approach of the discounted occupation measure as below, which formulates the fundamentals of linear programming for MDP. Given the discount factor λ, for all $\pi \in \Pi$ and all feasible state action pair $(x,\, a) \in \Theta$, we define

$$f_\lambda^\pi(x,\, a) = (1-\lambda) \sum_{i=0}^{\infty} \lambda^i P\left\{X_t = x,\, \pi_i(x) = a \,|\, \pi\right\}.$$

As a function defined on Θ, it is easy to verify that $f_\lambda^\pi(\cdot,\, \cdot)$ is a probability measure on Θ, which can be interpreted as the proportion of the expected number of visits to each feasible state action pair under policy π. By Theorem 3.2 in Altman (1999), given initial state $x_0 = 1$, the set of probability measures induced by all policies in Π can be equivalently presented by the set of all vectors $\varphi \in [0,\, 1]^{|\Theta|}$ satisfying

$$\begin{cases} \sum_{a \in \mathcal{A}^x} \varphi(x,\, a) - (1-\lambda)\mathbf{I}\left\{x = 1\right\} = \lambda \sum_{y \in \mathcal{X}} \sum_{a \in \mathcal{A}^y} \varphi(y,\, a) P^a(y,\, x), \;\; \forall x \in \mathcal{X} \\ \varphi(x,\, a) \geq 0, \;\; \forall (x,\, a) \in \Theta \end{cases}. \quad (10)$$

where $\mathbf{I}\left\{\cdot\right\}$ is the indicator function, i.e.,

$$\mathbf{I}\left\{x = 1\right\} = \begin{cases} 1, & x = 1 \\ 0, & \text{otherwise} \end{cases}.$$

An important benefit by introducing the discounted occupation measure is that we can represent any discounted total cost in a linear way. In particular, if we denote by φ^π the discounted occupation measure induced by policy $\pi \in \Pi$, then we may rewrite the total expected discounted cost (2) as

$$V^\pi(x) = \sum_{(x,\, a)\in\Theta} \varphi^\pi(x,\, a)c(x,\, a). \tag{11}$$

Meanwhile, according to the definition of Q_λ^π and φ, we conclude that Q_λ^π takes the value of $q(x)$ with probability $\sum_{a\in\mathcal{A}^x} \varphi^\pi(x,\, a)$.

In the following sections, we will show that using (10) and (11), we may transfer problem (7) and (9) into the linear programming programs and thus use traditional numerical methods to solve the problem.

4.1 *Optimal policy for VaR constraint*

According to previous discussions, to use the linear programming method to solve Problem (7), we need to transfer the VaR constraint $\mathrm{VaR}_\alpha(Q_\lambda^\pi) \geq \tau$ into a constraint function using φ. Note that we have,

$$P\{Q_\lambda^\pi = q(x)\} = D_\lambda^\pi(x) = \sum_{a\in\mathcal{A}^x} \varphi(x,\, a).$$

Thus, an equivalent condition to the VaR constraint $\mathrm{VaR}_\alpha(Q_\lambda^\pi) \geq \tau$ is given in the lemma below.

Lemma 1. *Given $\pi \in \Pi$ and the occupation measure φ^π induced by π, an equivalent condition to the constraint function $\mathrm{VaR}_\alpha(Q_\lambda^\pi) \geq \tau$ can be given using φ^π by*

$$\sum_{q(x)\leq\tau} \sum_{a\in\mathcal{A}^x} \varphi^\pi(x,\, a) \leq \alpha.$$

The proof of Lemma 1 is comparatively trivial and detailed in the Supplemental Online Materials. According to Lemma 1, if φ is an occupation measure induced by some policy that makes $\mathrm{VaR}_\alpha(Q_\lambda^\pi) \geq \tau$ holds, then combining with (11), Problem (7) can be rewritten

as

$$\begin{aligned}
\min_{\varphi} \quad & \sum_{(x,\,a)\in\Theta} \varphi(x,\,a)c(x,\,a) \\
\text{s.t.} \quad & \sum_{a\in\mathcal{A}^x} \varphi(x,\,a) - \lambda \sum_{y\in\mathcal{X}} \sum_{a\in\mathcal{A}^y} \varphi(y,\,a)P^a(y,\,x) = (1-\lambda)\mathbf{I}\left\{x=1\right\},\ \ \forall x\in\mathcal{X}, \\
& \sum_{q(x)\le\tau} \sum_{a\in\mathcal{A}^x} \varphi(x,\,a) \le \alpha, \\
& \varphi(x,\,a) \ge 0,\ \ \forall (x,\,a)\in\Theta.
\end{aligned} \tag{12}$$

In the following theorem, we conclude that the original problem (7) is feasible if and only if problem (12) is feasible. Moreover, problem (7) admits a stationary optimal policy that can be induced by the optimal solution of linear programming problem (12).

Theorem 1. *Let φ^* be the optimal solution of problem (12). Then, problem (7) admits a stationary admissible optimal policy $\pi^*\in\Pi$ given as*

$$\pi^*(x)(a) = \frac{\varphi^*(x,\,a)}{\sum_{k\in\mathcal{A}^x}\varphi^*(x,\,k)}, \tag{13}$$

for $\sum_{k\in\mathcal{A}^x}\varphi^*(x,\,k) > 0$ and arbitrarily decided otherwise. Conversely, if problem (7) admits an optimal admissible policy, then it also admits a stationary optimal policy, which induces an optimal solution to problem (12).

Proof. We define a complementary cost function $d^{\mathrm{VaR}}:\ \Theta\to\mathbb{R}$ as

$$d^{\mathrm{VaR}}(x,\,a) = \mathbf{I}\left\{q(x)\le\tau\right\},$$

then, the constraint $\sum_{q(x)\le\tau}\sum_{a\in\mathcal{A}^x}\varphi(x,\,a)\le\alpha$ can be rewritten as

$$\sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}^x}\varphi(x,\,a)d^{\mathrm{VaR}}(x,\,a)\le\alpha.$$

The l.h.s of the above inequality can be reversely transformed as

$$\min_{\pi\in\Pi}\mathbf{E}\left[(1-\lambda)\cdot\sum_{i=0}^{\infty}\lambda^i d^{\mathrm{VaR}}(x_i,\,\pi_t(X_i))\,\middle|\,X_0=1\right].$$

Therefore, we can transfer problem (12) reversely into an equivalent original problem as

$$\begin{aligned} &\min_{\pi\in\Pi} V^{\pi}(1), \\ &\text{s.t. } \mathbf{E}\left[(1-\lambda)\cdot\sum_{i=0}^{\infty}\lambda^i d^{\mathrm{VaR}}(x_i,\,\pi_i(X_i))\,\Big|\,X_0=1\right]\le\alpha. \end{aligned} \tag{14}$$

Since problem (14), problem (12) and problem (9) are equivalent, we can apply the Theorem 3.3 of Altman (1999) to problem (14) and conclude the proof. □

There are situations where one is only interested in the optimal value function, instead of the corresponding optimal policy. For computational convenience, we introduce the dual problem to problem (12) in the following corollary, which directly calculates the optimal value function.

Corollary 1. *The dual to problem (12) is*

$$\begin{aligned} &\max_{V,\,\varphi}\ [V(1)-\alpha\psi] \\ &\text{s.t. } V(x)\le c(x,\,a)+\psi\cdot\mathbf{I}\{q(x)\le\tau\}+\sum_{(y,\,a)\in\Theta}P^a(y,\,x)V(x),\ \forall(x,\,a)\in\Theta, \\ &\quad V(x)\ge 0,\ \forall x\in\mathcal{X},\ \psi\ge 0. \end{aligned} \tag{15}$$

Moreover, strong duality holds between problem (12) and problem (15).

Corollary 1 can be directly concluded by applying the Theorem 3.7 of Altman (1999) to problem (14), we omit the proof for brevity.

Unlike the classical MDP problems with no constraint, our proposed constraint model does not admit an optimal value function or policy that can be directly calculated using dynamic programming method. However, we can still provide a recursive representation for the optimal value function using Corollary 1. Let ψ^* be the optimal solution of problem (15). We note that due to the theory of superharmonic functions, the dual problem (15) immediately induces the following dynamic programming equation for the optimal value $V^{\mathrm{VaR}}(x)$ in the case of VaR constraint as

$$V^{\mathrm{VaR}}(x)=\min_{a\in\mathcal{A}^x}\left\{c(x,\,a)+\psi^*\cdot\mathbf{I}\{q(x)\le\tau\}+\lambda\mathbf{E}_{P^a(x,\,\cdot)}[V(X)]\right\}.$$

Since there are unknown parameters in this dynamic equation, it is not feasible for numerical calculation. However, we may have some insight into the optimal value function under safety constraint. The role of $\psi^* \cdot \mathbf{I}\{q(x) \leq \tau\}$ in the dynamic programming function above is intuitive. We may consider the occupation measure of each state action pair as an additional cost multiplied by a control variable ψ^*. Whenever the system x is risky such that $q(x) \leq \tau$, the cost is counted. This is consistent with our motivation of introducing VaR constraint, namely, we would like the safety level metric Q_λ^π to be above τ with a high probability by means that we add a cost of $\mathbf{I}\{q(x) \leq \tau\}$ for punishment each time $q(x) \leq \tau$.

4.2 *Optimal policy for CVaR constraint*

We firstly introduce a useful representation of the CVaR function that is critical for the construction of a linear programming framework. The CVaR of a random variable W, as defined in the previous section, can be represented as (Ruszczyński and Shapiro, 2006)

$$\mathrm{CVaR}_\alpha(W) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + (1-\alpha)^{-1} \mathbf{E}(W-\eta)_+ \right\}, \tag{16}$$

where

$$(W-\eta)_+ = \begin{cases} W-\eta, & W \geq \eta \\ 0, & \text{Otherwise} \end{cases}.$$

Using (16) and the discounted occupation measure, we provide an equivalent condition to the CVaR constraint in the following lemma.

Lemma 2. *Given $\pi \in \Pi$ and the occupation measure φ^π induced by π, an equivalent condition to the constraint function $\mathrm{CVaR}_\alpha(Q_\lambda^\pi) \geq \tau$ can be given using φ^π by*

$$\eta + (1-\alpha)^{-1} \sum_{(x,\,a) \in \Theta} \varphi(x,\,a)\,[q(x)-\eta]_+ \geq \tau, \ \ \forall \eta \in \mathcal{Q},$$

where $\mathcal{Q}$ is the set of all possible values of $q(x)$, $x \in \mathcal{X}$.

The proof of Lemma 2 is similar to the proof of Lemma 1 and given in the Supplemental Online Materials for details.

Now, we incorporate this representation in (1) from the Supplemental Online Materials

in problem (9) and rewrite the problem as

$$\begin{aligned}
&\min_{\varphi}\ (1-\lambda)^{-1}\sum_{(x,\,a)\in\Theta}\varphi(x,\,a)c(x,\,a)\\
&\text{s.t.}\ \sum_{a\in\mathcal{A}^x}\varphi(x,\,a)-\sum_{y\in\mathcal{X}}\sum_{a\in\mathcal{A}^y}\varphi(y,\,a)P^a(y,\,x)=(1-\lambda)\mathbf{I}\left\{x=1\right\},\ \forall x\in\mathcal{X},\\
&\quad\eta+(1-\alpha)^{-1}\sum_{(x,\,a)\in\Theta}\varphi(x,\,a)\left[q(x)-\eta\right]_+\geq\tau,\ \forall\eta\in\mathcal{Q},\\
&\quad\varphi(x,\,a)\geq 0,\ \forall(x,\,a)\in\Theta.
\end{aligned}\tag{17}$$

In the following theorem, we conclude that problem (9) is feasible if and only if problem (17) is feasible. Similar to the case of VaR constraint, the optimal policy for problem (9) is shown to be stationary and induced by the optimal solution of problem (17).

Theorem 2. *Let φ^* be an optimal solution of problem (17). Then, problem (9) admits a stationary optimal policy given as*

$$\pi^*(x)(a)=\frac{\varphi^*(x,\,a)}{\sum_{k\in\mathcal{A}^x}\varphi^*(x,\,k)},\tag{18}$$

for $\sum_{k\in\mathcal{A}^x}\varphi^*(x,\,k)>0$ and arbitrarily decided otherwise. Conversely, if problem (9) admits an optimal admissible policy, then it also admits a stationary optimal policy, which induces an optimal solution to problem (17).

Proof. The proof of Theorem 2 is similar to the proof of Theorem 1. Let $|\mathcal{S}|=K$ and $\mathcal{Q}=\{q_1,\cdots,q_K\}$. We define a set of complementary costs $\left\{d_k^{\mathrm{CVaR}}\right\}_{k=1}^K$ as

$$d_k^{\mathrm{CVaR}}=-(\alpha-1)\left[q(x)-q_k\right]_+,$$

and the constraint $\eta+(1-\alpha)^{-1}\sum_{(x,\,a)\in\Theta}\varphi(x,\,a)\left[q(x)-\eta\right]_+\geq\tau$, $\forall\eta\in\mathcal{Q}$ can be transferred as

$$\sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}^x}\varphi(x,\,a)d^{\mathrm{CVaR}}(x,\,a)\leq q_k-\tau,\ k=1\cdots,K.\tag{19}$$

The rest of the proof is completely the same as the proof of Theorem 1. □

Same as the previous discussions for VaR constraint, we also provide the dual linear programming problem to problem (17) as below.

Corollary 2. *Let $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_K) \in \mathbb{R}^K$ be a K-dimensional non-negative vector, the dual problem to (17) is*

$$\begin{aligned}
&\max_{V,\,\varepsilon}\ V(1) - \sum_{k=1}^{K} \varepsilon_k (q_k - \tau) \\
&\text{s.t. } V(x) \le c(x,\, a) + (\alpha - 1)^{-1} \sum_{k=1}^{K} \varepsilon_k [q(x) - q_k]_+ + \sum_{(y,\,a) \in \Theta} P^a(y,\, x) V(x),\ \ \forall (x,\, a) \in \Theta, \\
&\qquad V(x) \ge 0,\ \ \forall x \in \mathcal{X}, \\
&\qquad \varepsilon_k \ge 0,\ \ \forall k = 1, \cdots, K.
\end{aligned} \tag{20}$$

Moreover, strong duality holds between problem (17) and problem (20).

The proof of Corollary 2 can also be directly concluded using the Theorem 3.7 of Altman (1999). Details are omitted for brevity. Same as the VaR case, given the optimal solution ε^* for problem (20), we conclude the following dynamic programming equation for the optimal value function $V^{\mathrm{CVaR}}(x)$ for the CVaR case

$$V^{\mathrm{CVaR}}(x) = \min_{a \in \mathcal{A}^x} \left\{ c(x,\, a) + (\alpha - 1)^{-1} \sum_{k=1}^{K} \varepsilon_k^* [q(x) - \eta]_+ + \lambda \mathbf{E}_{P^a(x,\,\cdot)} \left[V(X) \right] \right\}.$$

Similar to the VaR constraint case, there is an additional cost of $(\alpha - 1)^{-1} \sum_{k=1}^{K} \varepsilon_k^* [q(x) - \eta]_+$ in the dynamic programming equation. However, in the CVaR case, the difference is that the additional cost is an weighted average among residual terms $[q(x) - q_k]_+$, $k = 1, \cdots, K$. This meets with the idea of CVaR that takes average among quantiles.

Remark. The optimal policy under safety constraint, as derived in this section, is not necessarily deterministic as the optimal policy under no constraint. In practice, engineers may not prefer random policies in some scenarios even they admit mathematical optimality. We propose an alternative here without rigorous interpretation. When deterministic policies are preferred, the previous random optimal policies can be transferred to deterministic policies according to practical safety requirements. Specifically, if the cost saving is more important and safety restriction is relaxed, then maintenance actions with low levels can be chosen with probability 1 for those system states that admit random optimal maintenance actions. In the contrast, if safety requirement is high and strict, then maintenance actions with relatively high level can be chosen with probability 1 for the underlying states. We will make more discussions about this in Section 5 through a numerical example.

5 Numerical study

We demonstrate the foregoing elaborations with a case study of condition-based maintenance on systems subject to performance degradation. A water treatment system in a petrochemical plant generally degrades over time and usage due to the corrosion of treatment vessel. The treatment vessel is inspected periodically and necessary maintenance actions are carried out upon the inspection. Apart from minimizing the operational and maintenance cost, an important issue to which decision makers pay great attention is the risk of shutdown of the system. Shutdowns not only induce more cost to the plant, but also lead to safety risks. Similar problems also exist in manufacturing systems. Manufacturers oftentimes utilize the machine as long as possible, while unexpected system shutdowns may cause unfulfilled orders and other negative market effects (Yang et al., 2017). Therefore, decision makers seek for a maintenance policy for such systems to minimize the maintenance cost under a constrained risk of shutdown. Consider such a system of which the degradation levels can be discretized by $\Delta d, 2\Delta d, \ldots, N\Delta d$. In this manner, system state space can be fully characterized by $\mathcal{X} \triangleq \{1, 2, ..., N\}$. Note that Δd is preferably set as a small value to achieve higher accuracy. For the purpose of illustration, we assume that the system degradation can be generally modeled by a drifted Wiener process. We denote the state of the system at time t by $Y(t)$, which is given by

$$Y(t) = \mu t + \sigma \mathcal{B}(t), \tag{21}$$

where μ and σ are the drift parameter and diffusion parameter, respectively, and $\mathcal{B}(\cdot)$ is a standard Brownian motion. Let the decision interval be Δt. We can approximately give the probability matrix R. Specifically, we have $R(x, y) = \mathbf{I}\{y = N\}$ for $x = N$. For $x < N$, $R(x, y)$ is given as follows:

$$\begin{aligned} R(x,y) =& \mathbf{I}\{y = N\} \left[1 - \Phi\left(\frac{N\Delta d - x\Delta d - \mu\Delta\tau}{\sigma\sqrt{\Delta t}}\right)\right] \\ &+ \mathbf{I}\{N > y > x\} \left[\Phi\left(\frac{y\Delta d - x\Delta d + \Delta d - \mu\Delta\tau}{\sigma\sqrt{\Delta t}}\right) - \Phi\left(\frac{y\Delta d - x\Delta d - \mu\Delta\tau}{\sigma\sqrt{\Delta t}}\right)\right] \\ &+ \mathbf{I}\{y = x\}\Phi\left(\frac{\Delta d - \mu\Delta\tau}{\sigma\sqrt{\Delta t}}\right). \end{aligned} \tag{22}$$

Further, the elements in Q^a are given in a form of the truncated normal distribution.

Specifically, for $1 \leq a \leq N-2$, the maintenance action is imperfect repair, thus we have

$$\begin{aligned} M^a(x,y) =& \mathbf{I}\{y=1\}\left[1 - F_{\mathcal{TN}}\left(x; a, \sigma_R, 0, x-1\right)\right] \\ &+ \mathbf{I}\{y \neq 1\}\left[F_{\mathcal{TN}}\left(x-y+1; a, \sigma_R, 0, x-1\right) - F_{\mathcal{TN}}\left(x-y; a, \sigma_R, 0, x-1\right)\right], \end{aligned} \tag{23}$$

where $F_{\mathcal{TN}}$ is the CDF of Truncated Normal distribution with the scale parameter σ_R describing the uncertainty in the effect of the imperfect repair. Specifically, for no maintenance action ($a = 0$) and replacements ($a = N-1$), the following equations hold:

$$M^0(x,y) = \mathbf{I}\{y=x\},\ M^{N-1}(x,y) = \mathbf{I}\{y=1\}. \tag{24}$$

Next, we set the single-time maintenance cost $c(x,a)$ in the following form:

$$c(x,a) = \begin{cases} c_r, & \text{if } a = N-1 \\ c_m a, & \text{otherwise} \end{cases} \tag{25}$$

In this sense, we assume that the replacement cost is constant while the repair cost is proportional to a. In addition, if the system fails and needs a corrective replacement, an addition cost of c_f is incurred to the replacement cost. The setting of q can be rather flexible under various circumstances. As an example, we let $q(x)$ be the probability that the system survives until the next decision epoch. Specifically,

$$q(x) = \sum_{x'=x}^{N} R(x,x'). \tag{26}$$

It is noteworthy that the probability that the system survives until the next decision epoch is an important criterion for industrial asset management. Generally, industrial systems are expected to operate uninterruptedly between inspections to guarantee the required performance. For water treatment vessels, unexpected shutdowns are deemed as risky events of which the probability of occurrence between inspections should be controlled strictly (Zhao et al., 2021). Therefore, manufacturers or asset managers usually control the probability of shutdown, which is equivalent to $q(x)$ that we have defined, to a satisfactory level. Other risk measures that satisfy the conditions listed in Section 3 can also be adopted to optimize the maintenance policies, for example, the expected system uptime and expected cumulative performance (uit het Broek et al., 2020; Chen et al., 2021).

5.1 *VaR constraint model*

We firstly present an example for the VaR constraint model. We choose a value of 10 for N, which is common in problems with moderate sizes. The parameter setting for the model is given in details in Table 1.

Table 1: Parameter setting for the illustrative example

Parameters	Value
N	10
μ	2.5
σ	4
Δt	1
σ_R	3
c_r	20
c_m	1
c_f	20
τ	0.7
λ	0.9
α	0.3

The optimal maintenance policy can be achieved by solving the linear programming problem in (12), and is summarized in (27). The minimum discounted cost is 74.1273 under the optimal policy. We notice that the optimal decision is deterministic for all system states except for $x = 7$. To be specific, when the system state is 1 or 2, the optimal decision is doing nothing, i.e., $\pi^*(1) = \pi^*(2) = 0$; it is optimal to imperfectly repair the system when $3 \leq x \leq 6$. When $x = 7$, the optimal decision is randomly chosen from {8, 9} with probabilities 0.0396 and 0.9604, respectively; finally, replacement is the optimal decision when $x \geq 8$. Specifically, we investigate how the optimal discounted cost changes when the random policy is altered to a deterministic one. On the one hand, under conservative considerations, we can set $\pi(7, 9)$=1, that is, when the system state is 7, the optimal action 9 is chosen with probability 1. Optimal decisions for other system states stay the same as those in (27). The conservative policy results in a discounted cost of 74.43, which is slightly larger than that of the random optimal policy. On the other hand, we can choose action 8

when the system state is 7, which thus forms a less conservative policy. The policy results in a discounted cost of 66.79, which is considerably lower than the original optimal cost. However, the less conservative policy cannot satisfy the constraints for the problem.

$$\pi^*(x,a) = \begin{cases} 1, & \text{for } (x,a) \in \{(1,0),(2,0),(3,1),(4,3),(5,6),(6,8),(8,9),(9,9),(10,9)\}, \\ 0.0396, & \text{for } (x,a) = (7,8), \\ 0.9604, & \text{for } (x,a) = (7,9), \\ 0, & \text{otherwise}, \end{cases} \tag{27}$$

To demonstrate the influence on the optimal policy and the total discounted cost caused by introducing the safety constraint, we denote by $\pi_0^*(x,a)$ the optimal policy under $\alpha = 1$, where no safety constraint is imposed to the model, and $\pi_0^*(x,a)$ is specified in (28). The minimum discounted cost is 58.34, which is considerable lower than that under the safety constraint characterized by $\tau = 0.7$ and $\alpha = 0.3$.

$$\pi_0^*(x,a) = \begin{cases} 1, & \text{for } (x,a) \in \{(1,0),(2,0),(3,1),(4,1),(5,1),(6,4),(7,6), \\ & \qquad (8,7),(9,8),(10,9)\}, \\ 0, & \text{otherwise}, \end{cases} \tag{28}$$

From (28), we note that the optimal decision is deterministic for all states, which is consistent with the results for non-constrained MDP model. Moreover, replacement is the optimal decision only when the system has failed ($x = 10$), meaning that without risk-aversion, the optimal policy retains a high tolerance for "bad" states. In light of the comparison between optimal policies π^* and π_0^*, the safety constraint leads to more conservative maintenance policies by repairing or replacing the system earlier, yet this inevitably induces a larger maintenance cost.

Since both τ and α determine the risk-aware constraint and thus influence the optimal decisions, we are interested in how the change in τ and α affects the value of the objective function. To study the sensitivity of the objective function to τ and α, the discounted total cost under different values of τ and α are exhibited in Figure 1. We can observe that when τ and $1-\alpha$ are both small (the blue area in the figure), the safety constraint does not affect the problem, therefore the optimal policy is the same as that in the non-constrained

case. In contrast, in the cases where τ and $1-\alpha$ are both large or either of them is very large (the yellow area), the constraint can never be satisfied, leading to infeasibility of the problem. In general, as τ or $1-\alpha$ increases, the safety constraint exerts more influences on the optimal decisions and discounted cost.

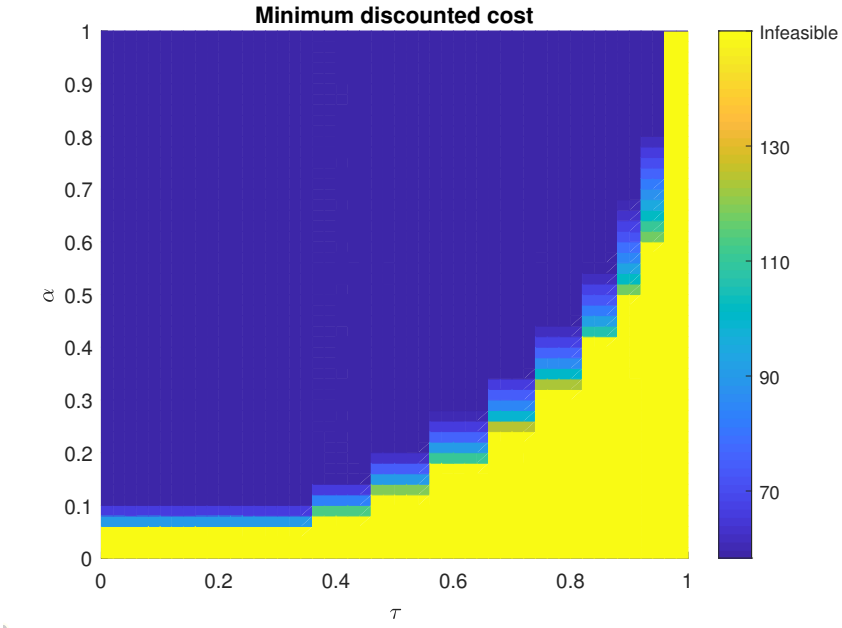

Figure 1: Optimal discounted cost with different τ and α under the VaR constraint

5.2 *CVaR safety constraint*

Next, we demonstrate the problem under the CVaR constraint in Section 4.2 by a similar example. For the sake of consistency, we keep all the model inputs the same as in Table 1. By numerical solution, we find that the optimal maintenance policy is exactly the same as the optimal policy π_0^* in (28). It seems that under the same model parameter setting, CVaR constraint exerts more relaxed risk-aversion on the model than VaR constraint. We may conclude the reason as follows. According to the form of VaR and CVaR in (6) and (8), VaR constraint makes a restriction on the absolute value of the quantile of Q_λ^π, yet CVaR constraint makes a restriction on the cumulative average of the quantile Q_λ^π. Even the upper bound of the quantile may reach some high value, its average value may still stay at a low level. Thus, the VaR constraint is more restrictive than the CVaR constraint, and

thus leads to a more conservative optimal policy.

To study the sensitivity to τ and α under CVaR constraint, a plot of the discounted total cost against τ and α that is analogous to Figure 1 is shown in Figure 2. In comparison to the case where VaR constraint is employed, the discounted total cost under CVaR cost is uniformly lower. In particular, the CVaR constraint yields a much smaller set of τ and α that lead to infeasibility. The results are consistent with our interpretation that CVaR constraint is softer than VaR constraint.

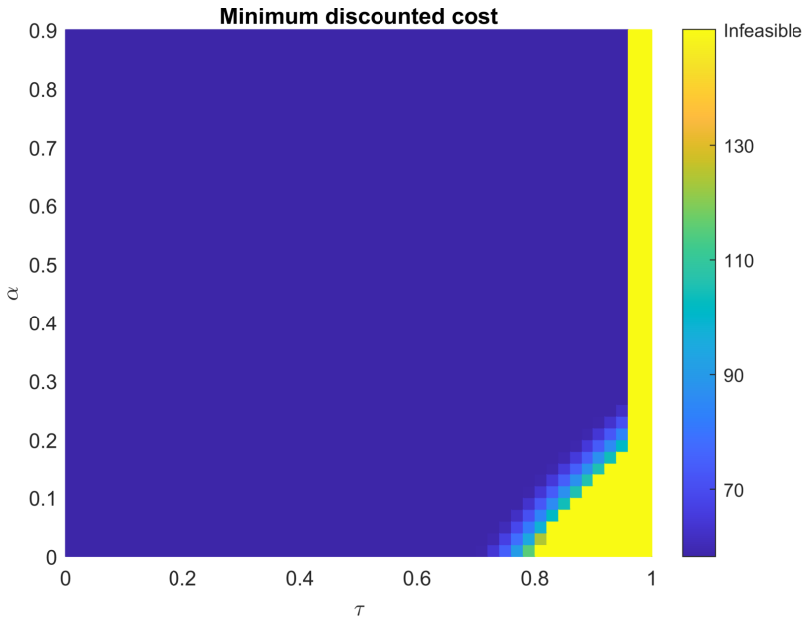

Figure 2: Optimal discounted cost with different τ and α under the CVaR constraint

The sensitivity analysis, as part of the numerical example, is exhibited in the Supplemental Online Materials.

6 Conclusions

In this work, we propose a maintenance modeling approach to accommodate risk-aware practitioners and engineers. Specifically, we develop a constrained MDP model to fit the proposed problem, where the risk-aversion is incorporated in the constraint function. A safety metric is defined as a metric of the system's safety level in the long run. Subsequently, two popular risk measures are used to measure the fluctuation of the safety metric,

and a constraint is imposed that the risk measures are not allowed to exceed some preset thresholds. Moreover, we construct the optimal policy of the proposed constrained MDP model that leads to the lowest total discounted cost.

We conclude that our model framework is not limited to the model assumptions and risk measures used in this work. Potential extension of this study can be considered as follows. Alternative risk measures can be used other than the VaR and CVaR risks, as long as the risk measure induces a linear constraint using the occupation measure. In addition, the definition of safety metric q and the transition probability are flexible and adapt to many existing popular models, though we choose specific models, i.e., the probability of survival for q and the Wiener process for the transitions, in our numerical study.

Meanwhile, we note that the incorporation of safety constraint in the maintenance process may significantly increase the maintenance cost in some situations. Therefore, the decision maker needs to balance between cost control and risk-aversion. From a more conceptual perspective, other feasible ways to introduce risk-aversion in the maintenance model are worth investigating in the future research.

Acknowledgements

The authors are grateful to the editors and the anonymous reviewers for the valuable comments and suggestions which contributed to a significant improvement of the original version of this paper. This work was supported in part by the National Natural Science Foundation of China under grant numbers 72002149, 72032005 and 71971181 and in part by Royal Society International Exchange Cost Share Project (IEC\NSFC\201401), and Guangdong Technology International Cooperation Project (2020A0505100024).

References

Ahmad, R. and S. Kamaruddin (2012). An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering 63*(1), 135–149.

Altman, E. (1999). *Constrained Markov decision processes*, Volume 7. CRC Press.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*(3), 203–228.

Byon, E., L. Ntaimo, and Y. Ding (2010). Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Transactions on Reliability 59*(2), 393–404.

Chen, L., J. Wang, and W. Yang (2021). A single machine scheduling problem with machine availability constraints and preventive maintenance. *International Journal of Production Research 59*(9), 2708–2721.

Chen, N., Z.-S. Ye, Y. Xiang, and L. Zhang (2015). Condition-based maintenance using the inverse gaussian degradation model. *European Journal of Operational Research 243*(1), 190–199.

Chow, Y. and M. Ghavamzadeh (2014). Algorithms for cvar optimization in mdps. In *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2*, pp. 3509–3517.

Elwany, A. H., N. Z. Gebraeel, and L. M. Maillart (2011). Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations Research 59*(3), 684–695.

Flory, J. A., J. P. Kharoufeh, and D. T. Abdul-Malak (2015). Optimal replacement of continuously degrading systems in partially observed environments. *Naval Research Logistics (NRL) 62*(5), 395–415.

Garcıa, J. and F. Fernández (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research 16*(1), 1437–1480.

Geibel, P. and F. Wysotzki (2005). Risk-sensitive reinforcement learning applied to control under constraints. *Journal of Artificial Intelligence Research 24*, 81–108.

Gosavi, A. (2006). A risk-sensitive approach to total productive maintenance. *Automatica 42*(8), 1321–1330.

Havinga, M. J. and B. de Jonge (2020). Condition-based maintenance in the cyclic patrolling repairman problem. *International Journal of Production Economics 222*, 107497.

Huo, X. and F. Fu (2017). Risk-aware multi-armed bandit problem with application to portfolio selection. *Royal Society Open Science 4*(11), 171377.

Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk.* The McGraw-Hill Companies, Inc.

Junca, M. and M. Sanchez-Silva (2013). Optimal maintenance policy for a compound poisson shock model. *IEEE Transactions on Reliability 62*(1), 66–72.

Kim, M. J. and A. E. Lim (2015). Robust multiarmed bandit problems. *Management Science 62*(1), 264–285.

Lagos, C., F. Delgado, and M. A. Klapp (2020). Dynamic optimization for airline maintenance operations. *Transportation Science 54*(4), 998–1015.

Liu, B., Z. Liang, A. K. Parlikad, M. Xie, and W. Kuo (2017). Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model. *Reliability Engineering & System Safety 168*, 200–209.

Liu, B., S. Wu, M. Xie, and W. Kuo (2017). A condition-based maintenance policy for degrading systems with age-and state-dependent operating cost. *European Journal of Operational Research 263*(3), 879–887.

Mattila, R., C. R. Rojas, V. Krishnamurthy, and B. Wahlberg (2017). Computing monotone policies for markov decision processes: a nearly-isotonic penalty approach. *IFAC-PapersOnLine 50*(1), 8429–8434.

Meraklı, M. and S. Küçükyavuz (2020). Risk aversion to parameter uncertainty in Markov decision processes with an application to slow-onset disaster relief. *IISE Transactions 52*(8), 811–831.

Ngo, M. H. and V. Krishnamurthy (2009). Monotonicity of constrained optimal transmission policies in correlated fading channels with arq. *IEEE Transactions on Signal Processing 58*(1), 438–451.

Papakonstantinou, K. G. and M. Shinozuka (2014). Planning structural inspection and maintenance policies via dynamic programming and markov processes. part i: Theory. *Reliability Engineering & System Safety 130*, 202–213.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk 2*, 21–42.

Ruszczyński, A. (2010). Risk-averse dynamic programming for markov decision processes. *Mathematical programming 125*(2), 235–261.

Ruszczyński, A. and A. Shapiro (2006). Optimization of convex risk functions. *Mathematics of Operations Research 31*(3), 433–452.

Tamar, A., Y. Chow, M. Ghavamzadeh, and S. Mannor (2016). Sequential decision making with coherent risk. *IEEE Transactions on Automatic Control 62*(7), 3323–3338.

uit het Broek, M. A. J., R. H. Teunter, B. de Jonge, J. Veldman, and N. D. Van Foreest (2020). Condition-based production planning: adjusting production rates to balance output and failure risk. *Manufacturing & Service Operations Management 22*(4), 792–811.

Wu, S., F. P. Coolen, and B. Liu (2017). Optimization of maintenance policy under parameter uncertainty using portfolio theory. *IISE Transactions 49*(7), 711–721.

Xia, L. (2020). Risk-sensitive markov decision processes with combined metrics of mean and variance. *Production and Operations Management 29*(12), 2808–2827.

Xu, J., L. Chen, and O. Tang (2021). An online algorithm for the risk-aware restless bandit. *European Journal of Operational Research 290*(2), 622–639.

Yang, L., X. Ma, and Y. Zhao (2017). A condition-based maintenance model for a three-state system subject to degradation and environmental shocks. *Computers and Industrial Engineering 105*, 210–222.

Zhang, M. and M. Revie (2017). Continuous-observation partially observable semi-markov decision processes for machine maintenance. *IEEE Transactions on Reliability 66*(1), 202–218.

Zhang, Y., P. Zhao, B. Li, Q. Wu, J. Huang, and M. Tan (2020). Cost-sensitive portfolio selection via deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*.

Zhao, X., Z. Liang, A. K. Parlikad, and M. Xie (2021). Performance-oriented risk evaluation and maintenance for multi-asset systems: A Bayesian perspective. *IISE Transactions*, in press, doi:10.1080/24725854.2020.1869871.

Zhu, Z. and Y. Xiang (2021). Condition-based maintenance for multi-component systems: Modeling, structural properties, and algorithms. *IISE Transactions 53*(1), 88–100.