Industrial Engineering and Applications
L.-C. Tang (Ed.)
© 2023 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE230032

Integrated Production/Inventory and Condition-Based Maintenance Control of a Multi-Item Production System Under Stochastic Demand

Alp DARENDELILER [a,1], Dieter CLAEYS [b] and El-Houssaine AGHEZZAF [b]

[a] *Department of Industrial Systems Engineering and Product Design, Ghent University, Ghent, Belgium*

[b] *Department of Industrial Systems Engineering and Product Design, Ghent University, Ghent, Belgium*

Industrial Systems Engineering (ISyE) Flanders Make, Kortrijk, Belgium

ORCiD ID: Alp Darendeliler https://orcid.org/0000-0002-5217-7444, Dieter Claeys https://orcid.org/0000-0002-7666-2479, El-Houssaine Aghezzaf https://orcid.org/0000-0003-3849-2218

Abstract. This paper studies the joint production/inventory and condition-based maintenance control for a multi-product manufacturing system with setup and maintenance times under stochastic product demands. The problem is modelled as a semi-Markov decision process (SMDP). The objective is to find a joint production and maintenance policy that minimizes the long run expected discounted cost including setup, holding, lost sales, preventive and corrective maintenance costs. A Q-learning method with state aggregation (QLA) is proposed to find near-optimal policies for large-scale problems that cannot be solved to optimality due to the curse of dimensionality. The numerical results show that QLA provides well-performing policies in a reasonable computational time.

Keywords. condition-based maintenance, lot-sizing, scheduling, multi-product, inventory, reinforcement learning

1. Introduction

Physical assets, no matter how well designed, are subject to deterioration over time since they are operating under certain stress or load in the real environment [1]. This deterioration eventually results in equipment failures, leading to high corrective maintenance costs, unplanned downtimes, and production losses. Recent advances in sensor technologies enable manufacturers to monitor the current health status of the production systems. Based on this real-time information, condition-based maintenance (CBM) recommends when to proactively perform maintenance operations to reduce the possibility of equipment failures while avoiding unnecessarily early preventive maintenance operations.

[1] Corresponding Author: Alp Darendeliler, Alp.Darendeliler@Ugent.be

Since production decisions influence the degradation pattern of manufacturing systems, equipment condition needs to be considered when deciding on production quantities and which product to produce. Also, an effective maintenance policy should also take the inventory levels into account to avoid unmet demand when the system is down due to maintenance. To address this interdependency, in the literature, integrated production lot-sizing and CBM models have been developed with the aim of minimizing the total production/inventory and maintenance costs. Most studies assume that a single product is produced, and the demand rate is deterministic and constant. However, in many practical situations, multiple items are produced by a single machine, and the demand is uncertain. Developing efficient production and maintenance control policies for such systems is a challenging problem due to the stochasticity in the machine degradation and product demands together with the setup times.

This paper covers the joint control of production/inventory and CBM for a single-machine, multi-product, make-to-stock production system with setup and maintenance times under stochastic demand studied by Darendeliler et al. [2]. The system stochastically degrades while producing products and the degradation behaviour differs depending on the item being produced. Based on the state of the system, production and maintenance decisions are dynamically made.

The described problem is modelled as a semi-Markov decision process (SMDP). Upon observing the degradation level, machine setup status and inventory levels of the products, one of the following actions is taken: keeping the machine idle, performing preventive maintenance, performing corrective maintenance in case of failure, produce an item for which the machine has been set up, or performing setup to initiate production for another item. The objective is to minimize the long run total expected discounted cost including corrective maintenance, preventive maintenance, setup, holding, and lost sales costs. Since the problem is subject to the curse of dimensionality, only small problem instances can be solved to optimality. To find near-optimal solutions to large-scale problems, in this paper, a model-free reinforcement learning algorithm (Q-learning) with state aggregation proposed in [3] is adapted for the present problem.

The remainder of the paper is organized as follows. The relevant literature is reviewed in Section 2. Then, in Section 3, the problem is formulated. Section 4 presents the proposed solution method. The numerical results are then presented in Section 5. Finally, Section 6 concludes the paper.

2. Literature Review

In literature, the integration of time-based maintenance and lot-sizing has been extensively studied under the assumptions of single-item and constant demand rate [4]-[8]. Aghezzaf et al. [9] and Shamsaei and Van Vyve [10] develop integrated models for the multi-item lot-sizing and preventive maintenance problem. However, these studies do not take the current equipment condition into account in making production and maintenance decisions.

Recent studies investigate the joint optimization of economic production quantity (EPQ) and CBM. Jafari and Makis [11] develop a joint optimization model of the EPQ and preventive maintenance policy for a production system under condition monitoring. They propose a proportional hazards model which takes the age and the condition

of the equipment into account. Zheng et al. [12] extend this work by incorporating a CBM policy with multiple actions and dynamic control limits into their model. For a two-unit production system, Jafari and Makis [13] develop a model to find the optimal EPQ and preventive maintenance policy, where the condition of the first unit is partially observable and only the age of the second unit is known. Peng and van Houtum [14] propose an integrated lot-sizing and CBM model in which the degradation is modelled as a continuous-time continuous state stochastic process. The objective is to find the optimal preventive maintenance threshold and EPQ to minimize average long run cost rate. Khatab et al. [15] consider an integrated model to determine the optimal inspection period and the maintenance threshold to minimize the production, inspection, maintenance and the cost of producing nonconforming items. Cheng et al. [16] study an integrated problem of lot-sizing, quality control and CBM, where product quality decreases with degradation level. For a multi-component system, Cheng et al. [17] propose a joint optimization model of lot-sizing and CBM. They use the Birnbaum importance measure to determine the threshold level for each component.

In literature, fewer works propose joint production and CBM policies for systems with stochastic demand [18]-[21]. These studies develop MDP models to determine the optimal policy for single-product systems.

Darendeliler et al. [3] study the joint optimal control of production and CBM for a single-machine multi-product system under stochastic demand, assuming that the system is reviewed at equidistant time points. Hence, the durations of the production lots and maintenance operations are assumed to be equal. Also, production setup times are not considered. Then, the previous work is extended in [2] by relaxing these assumptions, and a SMDP model is proposed in which the system is reviewed at nonequidistant time epochs corresponding to the completion of a unit production, setup, maintenance or arrival of a demand when the machine is idle. The tabular Q-learning is used to obtain the optimal policy. However, due to its lack of scalability, tabular Q-learning becomes inapplicable as the state space exponentially increases with the number of products. To overcome this, in this paper, we apply a Q-learning algorithm with state aggregation to obtain near-optimal policies for the large-scale problems in a reasonable computational time.

Scheduling the production of multiple items on a single machine with limited capacity under random demand, possibly random setup times and possibly random production times is defined as the stochastic economic lot scheduling problem (SELSP) [22]. Finite production capacity has to be dynamically distributed among the products to be responsive to the stochastic demand, processing and setup times. The presence of the setup times in addition to the stochastic environment add complexity to the problem [22]. Winands et al. [22] categorize the SELSP by sequencing and lot-sizing decisions. In this study, the production policy has the characteristics of dynamic sequence (sequencing decisions are made based on the system state rather than a fixed sequence) and global lot-sizing (the production amount depends on the other products' stock levels in addition to the stock level of the current product). In addition, equipment deterioration and maintenance decision-making are integrated into the SELSP, while they are not considered in the existing SELSP literature.

In the literature, few studies consider dynamic sequencing and global lot-sizing for the SELSP. Qui and Loulou [23] model the problem as a SMDP. Only limited sized problems can be solved by their procedure due to the curse of dimensionality. Wang et al.

[24] propose the Q-learning for SMDP and the Q-learning with a learning improvement heuristic to solve the SELSP. However, the proposed methods are not scalable to large-scale problems. Löhndrorf and Minner [25] apply approximate value iteration technique to the SELSP. Although, it performs well for small instances, as the problem size increases, it is outperformed by the globally optimized based-stock and fixed cycle policies.

3. Problem Formulation

Consider a single-machine make-to-stock manufacturing system that produces N products one at a time and deteriorates with production according to known probabilities. The machine deterioration level is described by the states $\{1,\ldots,F\}$, where 1 is the as-new state, and F is the failure state. In case of failure, corrective maintenance is performed which takes CT time units and costs c_c. At any operational state $i \in \{1,\ldots,F-1\}$, preventive maintenance can be carried out, which takes PT time units and costs c_p. After preventive or corrective maintenance, the equipment becomes as good as new. Production of product $n \in \{1,\ldots,N\} = M$ takes ρ_n time per unit. If product n is decided to be produced, the machine must have been set up for product n. The setup status of the machine is denoted by $k \in \{0,1\ldots n,\ldots,N\}$, where $k = n$ indicates that the machine is setup for production of item n, and the machine is idle when $k = 0$. It is assumed that the setup status cannot be maintained over an idle period. The setup time and cost of product n are ST_n and cs_n respectively. During production, setup and maintenance, the system cannot be interrupted. For each product $n = 1\ldots,N$, the inventory level x_n cannot exceed the maximum allowable stock level $\bar{x}_n$. The inventory holding cost ch_n is incurred per item per unit time for product n. The demand for each product n, $D_n(t)$, follows a Poisson process with rate λ_n. Unmet demand is lost and a lost sales cost cl_n is charged per unit for product n.

The system is reviewed at epochs when the production or setup of an item has just been completed, or demand of any item arrives in case the machine is at the idle state. At those epochs, the system state $s = (i,k,x_1,\ldots,x_N)$ is observed, with the degradation level i, setup status k, and the product inventories $x_1,\ldots,x_N$. Upon observing the state, an action $a \in A(s)$ is taken, which can be one of the following: (1) keeping the machine idle $(a = 0)$; (2) producing item $n\,(a = n)$; (3) performing preventive maintenance $(a = -1)$; (4) performing corrective maintenance $(a = -2)$ if and only if $i = F$. Note that if $a = n \neq k$, then the machine is going to be setup for item n.

For state $s = (i,k,x_1,\ldots,x_N)$ and action a, the transition probabilities for the degradation can be expressed as

$$P_{ij}^{k,a} = \begin{cases} 1 & if\ a = 0,\ i = j \\ 1 & if\ a = -2,\ i = F,\ j = 1 \\ 1 & if\ a = -1,\ j = 1 \\ p_{ij}(n) & if\ a = n = k,\ j \in \{i,\ldots,F\} \\ 1 & if\ a = \ n \neq k,\ j = i \\ 0 & otherwise, \end{cases} \tag{1}$$

where the degradation remains at the same level until the next epoch if no production occurs, and it will be at state j in case of production.

Let,

$\tau(s,a)$: time until the next decision epoch in state s under action a,
$c(s,a)$: cost incurred (discounted) until the next decision epoch in state s under action a,
d_n: demand for product n between the current and next decision epoch; note that given the time interval $\tau(s,a)$,
$d_n \sim Poisson(\tau(s,a)\lambda_n)$,
S_n^r: time at which r^{th} demand occurs $(r = 1,\ldots,d_n)$ for product $n = 1,\ldots,N$,
$s' = (i', k', x_1{}', \ldots, x_N{}')$: state at the next epoch,
$\mathbf{1}_{\{A\}}$: indicator function that equals 1 if condition A holds, otherwise it equals 0,
γ: discount factor.

Then the following equations describe the evolution of the system:

1. If $i = F$ and $a = -2$ (corrective maintenance is the only option), then

$$\tau(s,a) = CT,\ \ c(s,a) = c_c + \sum_{n=1}^{N}\sum_{r=1}^{d_n}\left[\mathbf{1}_{\{x_n \geq r\}}\frac{\left(1-e^{-\gamma S_n^r}\right)}{\gamma}ch_n + \ \mathbf{1}_{\{r>x_n\}}e^{-\gamma S_n^r}cl_n\right]$$
$$+\frac{\left(1-e^{-\gamma\tau(s,a)}\right)}{\gamma}\sum_{n=1}^{N}ch_n\left(x_n-d_n\right)^+,$$
$$s' = (i'=1, k'=0, x_1{}' = (x_1-d_1)^+, \ldots, x_N{}' = (x_N-d_N)^+).$$

2. If $i < F$ and $a = -1$ (preventive maintenance), then

$$\tau(s,a) = PT,\ \ c(s,a) = c_p + \sum_{n=1}^{N}\sum_{r=1}^{d_n}\left[\mathbf{1}_{\{x_n \geq r\}}\frac{\left(1-e^{-\gamma S_n^r}\right)}{\gamma}ch_n + \ \mathbf{1}_{\{r>x_n\}}e^{-\gamma S_n^r}cl_n\right]$$
$$+\frac{\left(1-e^{-\gamma\tau(s,a)}\right)}{\gamma}\sum_{n=1}^{N}ch_n\left(x_n-d_n\right)^+,$$
$$s' = (i'=1, k'=0, x_1{}' = (x_1-d_1)^+, \ldots, x_N{}' = (x_N-d_N)^+).$$

3. If $i < F$ and $a = 0$ (keep the machine idle), then, let $S^* = \min\left\{S_1^1, \ldots, S_N^1\right\}$, and $S_r^1 = S^*$ *for some* $r \in \{1,\ldots,N\}$,

$$\tau(s,a) = S^*,\ \ c(s,a) = \frac{\left(1-e^{-\gamma\tau(s,a)}\right)}{\gamma}\sum_{n=1}^{N}ch_n x_n + e^{-\gamma\tau(s,a)}cl_r\left(1-x_r\right)^+,$$
$$s' = (i'=i, k'=0, x_r{}' = (x_r-1)^+,\ x_m{}' = x_m\ \ for\ m \in \{1,\ldots,N\}\setminus r).$$

4. If $i < F$ and $a = n$ (produce item n)

4.1. If $a = n = k$

$$\tau(s,a) = \rho_n, \;\; c(s,a) = \sum_{n=1}^{N}\sum_{r=1}^{d_n}\left[\mathbf{1}_{\{x_n \geq r\}}\frac{\left(1-e^{-\gamma S_n^r}\right)}{\gamma}ch_n + \mathbf{1}_{\{r>x_n\}}e^{-\gamma S_n^r}cl_n\right] + \frac{\left(1-e^{-\gamma\tau(s,a)}\right)}{\gamma}\sum_{n=1}^{N}ch_n\left(x_n-d_n\right)^+,$$

$$s' = (i' = j \geq i,\; k' = n,\; x_n' = \left((x_n - d_n)^+ + 1\right), x_m{}' = (x_m - d_m)^+ \; for\; m \in \{1,\ldots,N\} \setminus n).$$

4.2. If $a = n \neq k$

$$\tau(s,a) = ST_n, \;\; c(s,a) = cs_n + \sum_{n=1}^{N}\sum_{r=1}^{d_n}\left[\mathbf{1}_{\{x_n \geq r\}}\frac{\left(1-e^{-\gamma S_n^r}\right)}{\gamma}ch_n + \mathbf{1}_{\{r>x_n\}}e^{-\gamma S_n^r}cl_n\right] + \frac{\left(1-e^{-\gamma\tau(s,a)}\right)}{\gamma}\sum_{n=1}^{N}ch_n\left(x_n-d_n\right)^+,$$

$$s' = (i' = i, k' = n, x_1{}' = (x_1 - d_1)^+, \ldots, x_N{}' = (x_N - d_N)^+).$$

The objective is to find the policy that minimizes the long run total expected discounted cost. Hence, the optimization problem can be expressed as

$$\min_{\mu} E\left[\sum_{t=0}^{\infty} e^{-\gamma\sum_{k=0}^{t-1}\tau(s_k,\mu(s_k))}c\left(s_t,\mu\left(s_t\right)\right)\right] \tag{2}$$

where s_t is the state the system visits at epoch t, $\mu(s_t)$ is the action taken according to the policy μ at state s_t, $c(s_t,a_t)$ is the discounted cost incurred between the epochs t and $t+1$, and $\tau(s_k,\mu(s_k))$ is the time length between these epochs.

4. Solution Method

Q-learning is a well-known model-free reinforcement learning algorithm that can find the optimal policy under certain conditions [26]. Instead of using complete transition probabilities as in the model-based methods, Q-learning estimates the optimal state-action values (Q-values) readily from the sampled values.

The tabular Q-learning for SMDP, proposed by [27] was applied to the present problem by [2] to obtain the optimal policy. The steps of the algorithm are shown in Figure 1. The Q-values are updated by 2.3 (Figure 1), where $c(s,a)$ is the accumulated discounted cost in period of length $\tau(s,a)$, α is the learning rate, and γ is the discount factor.

1. Initialize a starting state s, and $Q(s,a)$ for all $s \in S$ and action $a \in A(s)$
2. For $t = 1, 2, \ldots, T$
   - 2.1. Choose action $a \in A(s)$ for state s $(\varepsilon - greedy)$
   - 2.2. Compute the cost $c(s,a)$, the sojourn time $\tau(s,a)$ and the next state s' based on state s and action a
   - 2.3. Update the Q-values by the equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(c(s,a) + e^{-\gamma \tau(s,a)} \min_{a' \in A(s')} Q(s',a') - Q(s,a) \right)$$

   - 2.4. Update $s \leftarrow s'$
3. Return $\mu(s) = \arg\min_{a \in A(s)} Q(s,a)$ for all $s \in S$

Figure 1. Steps of the Q-learning algorithm

The tabular Q-learning method stores each state-action separately in a table or an array. For large-scale problems, it is not feasible to store each state-action pair separately due to the memory limitations. Another problem with the tabular methods is the time and data needed to fill the large tables accurately [26]. Our problem is subject to the curse of dimensionality due to the exponentially growing state space with the number of products. Thus, it is not possible to solve large-scale problems by the tabular Q-learning. To tackle this, we adapt the state aggregation method from [3] to the present problem.

Since the product with the lowest expected runout time has the biggest impact on the optimal Q-values, only the index and the inventory level of the most urgent product is stored in the aggregated state while the stock levels of the other products are not included as separate variables; however, they are used in the calculation of the one-period costs and the next system state. The urgencies of the products are determined by the product runout times $x^n/\lambda_n - ST_n,\ n = 1, \ldots, N$; these are the average times until stock out, where λ_n is the demand rate for item n [28]. Hence, the most urgent product's index is $u = \arg\min_{n \in M} (x_n/\lambda_n - ST_n)$. For a complete state $s = (i,k,x_1,\ldots,x_N)$, the corresponding aggregate state is $\bar{s} = \left(i,k,u,x_u,\sum_{n=1}^{N} x_n\right)$, where $\sum_{n=1}^{N} x_n$ is total inventory level, which is stored to take the overall stock into account in making production and maintenance decisions. The tabular representation of the state-action values has exponential worst-case space complexity of $O\left(FN^2\bar{x}^N\right)$, where $\bar{x} = \max_{n \in N} \bar{x}_n$. However, the proposed aggregation scheme has pseudo-polynomial complexity of $O\left(FN^3\bar{x}^2\right)$.

Figure 2 shows the steps of the Q-learning with state aggregation (QLA). In the first step, the system state s is initialized, the corresponding aggregated state is determined, and all Q-values are set to zero. For the system state $s = (i,k,x_1,\ldots,x_N)$, the aggregate state is $\bar{s} = \left(i,k,u,x_u,\sum_{n=1}^{N} x_n\right)$. Step 2.1 chooses action according to the $\varepsilon - greedy$ policy in which with probability $1-\varepsilon$, the greedy action is selected, otherwise a random action is chosen. Based on the aggregated state $\bar{s}$, the chosen action $a \in A(\bar{s})$ is one of the following: (1) performing corrective maintenance if and only if a failure occurs; (2) performing preventive maintenance; (3) keeping the machine idle; (4) producing the most urgent item u; (5) producing the product for which the machine has already been setup. At step 2.2, product demands, demand arrival times, and the degradation path are sampled. Then, the current period length $\tau(s,a)$ and the next state $s' = (i',k',x_1{}',\ldots,x_N{}')$ are determined through the system equations given in Section 3. The discounted cost incurred until the next epoch $c(s,a)$, is calculated by the equations given in Section 3. Step 2.3. determines the next aggregated state. At step 2.4., aggregated Q-values are updated.

1. Initialize
   $s = (i, k, x_1, \ldots, x_N)$
   $u = \arg\min_{n \in M} (x_n/\lambda_n - ST_n)$
   $\bar{s} = \left(i, k, u, x_u, \sum_{n=1}^{N} x_n\right)$
   $Q(\bar{s}, a) = 0$ for all $\bar{s} \in \bar{S}$ and $a \in A(\bar{s})$
2. For $t = 1, 2, \ldots, T$
   2.1. Choose action $a \in A(\bar{s})$ $(\varepsilon - greedy)$
   2.2. Sample demands $d_1, \ldots, d_N$ and the next degradation state i'
   Compute $\tau(s,a)$, $c(s,a)$ and s' by the system equations
   2.3. Compute
   $u' = \arg\min_{n \in M} (x_n'/\lambda_n - ST_n)$,
   $\bar{s}' = \left(i', k', u', x_u', \sum_{n=1}^{N} x_n'\right)$
   2.4. Update the Q-value by the equation:
   $$Q(\bar{s}, a) \leftarrow Q(\bar{s}, a) + \alpha \left(c(s,a) + e^{-\gamma \tau(s,a)} \min_{a' \in A(\bar{s}')} Q(\bar{s}', a') - Q(\bar{s}, a) \right)$$
   2.5. Update $\bar{s} \leftarrow \bar{s}'$
3. Return $\mu(\bar{s}) = \arg\min_{a \in A(\bar{s})} Q(\bar{s}, a)$ for all $\bar{s} \in \bar{S}$

Figure 2. Steps of Q-learning with state aggregation

5. Numerical Results

In this section, we present results for three, five and ten-product examples. The machine degradation follows a Gamma Process with shape $\alpha = 0.5$ and scale parameter $\beta = 1$. The failure threshold L is 10. The gamma process is approximated by a discrete-time Markov chain with 21 states based on the method in [29], where 1 is the as-new state and 21 is the failure state F. The setup costs are set to zero since they primarily correspond to the time lost when the system in not operating, which is represented by the setup times.

In the Q-learning algorithms (QL and QLA), $\varepsilon - greedy$ policy is followed with $\varepsilon = 0.1$ (Step 2.1). A harmonic learning rate $\alpha = b/(b + N_t(\bar{s}, a) - 1)$, is used in updating the Q-values (Step 2.4), where $N_t(\bar{s}, a)$ is the number of times a state action pair $(\bar{s}, a)$ has been visited up to the time t, the parameter b is tuned based on the procedure suggested in [30]. In all calculations, the discount factor is $\gamma = 0.01$. The performance of the algorithms is evaluated based on the average cost per unit time (AVC). At every $100,000$ iterations, AVC is updated by running a simulation of $10,000$ iterations under the greedy policy that the algorithm has learned so far.

5.1. Three-product examples

In this part, the performance of QLA is compared to QL, which is proven to converge to the optimal policy under certain conditions, which are satisfied by QL [31]. The problem parameters are randomly selected from the ranges given in Table 1. Eight three-product problem instances ($N = 3$) are generated with these parameters. The sample mean of the percentage gaps between the average cost rates of QL and QLA over the generated instances is 1.387% ($Gap(\%) = 100\ (AVC^{QLA} - AVC^{QL})/AVC^{QL}$). This demonstrates that QLA provides well-performing policies that are close to optimal.

Table 1. Parameter ranges for the three-product problem

Parameter	Range
Demand rates	[0.1,0.25]
Production rates	[1,2]
Preventive maintenance time	[4,8]
Corrective maintenance time	[10,16]
Setup time	[0.5,2]
Preventive maintenance cost	[300,500]
Corrective maintenance cost	[800,1000]
Holding costs	[0.5,2]
Lost sales costs	[100,200]

Figure 3 shows the average cost rates of QL and QLA for an example generated from Table 1. There is not a considerable difference between their performances; the average cost rate for QLA converges to 36.986 while the average cost rate for QL converges to 36.649. However, QLA converges much faster than QL.

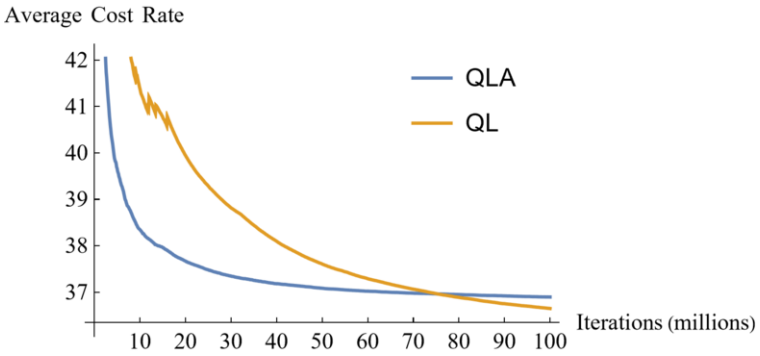

Figure 3. Average cost curves of QL and QLA for the three-product problem

5.2. Five and ten-product examples

In this part, we present computational results of the Q-learning with state aggregation (QLA) for the five ($N = 5$) and ten-product ($N = 10$) examples. Since, the state space grows exponentially with the number of products, it becomes computationally infeasible to solve problems with more than three products by the tabular Q-learning. Hence, it is not possible to compare QLA's performance with the optimal policy. The problems' parameters are randomly selected from the ranges given in Table 2. In both examples, the average cost rate converges towards the end of the simulation (Figure 4 and Figure 5). This demonstrates that QLA has reached a stable policy in a reasonable computational time.

Table 2. Parameter ranges for the five and ten-product problems

	Number of products	
	5	10
Demand rates	[0.1,0.2]	[0.01,0.1]
Production rates	[1,2]	[1,2]
Preventive maintenance time	[4,8]	[4,8]
Corrective maintenance time	[10,16]	[10,16]
Setup time	[0.5,2]	[0.5,2]
Preventive maintenance cost	[300,500]	[300,500]
Corrective maintenance cost	[800,1000]	[800,1000]
Holding costs	[0.5,2]	[0.5,2]
Lost sales costs	[100,200]	[100,200]

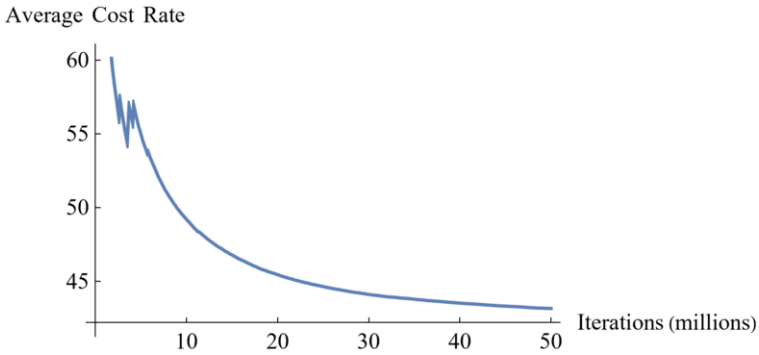

Figure 4. Average cost curve of QLA for the five-product problem

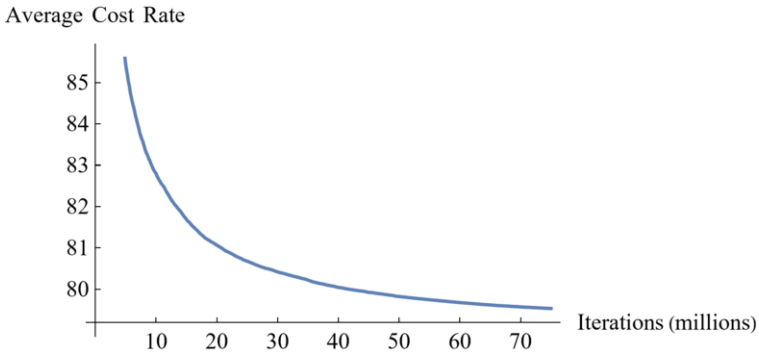

Figure 5. Average cost curve of QLA for the ten-product problem

6. Conclusions

In this paper, we have studied the joint production/inventory and condition-based maintenance (CBM) control problem for a multi-product manufacturing system with setup and maintenance times under stochastic demand. The problem has been modelled as a SMDP, in which production and maintenance decisions are made based on the setup status and

the degradation state of the machine as well as the stock levels of the products. Due to the curse of dimensionality, only limited-sized problems can be solved to optimality. Hence, a Q-learning method with state aggregation has been proposed to find near-optimal policies for larger problems. Its performance has been evaluated on problem instances with up to ten products. The results show that QLA converges in a reasonable computational time, and it is scalable to large-scale problems. In addition, it is compared to the tabular Q-learning (QL) that converges to the optimal policy. The results confirm that QLA performs very close to QL while converging considerably faster.

References

[1] Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing. 2006 Oct;20(7):1483–510, https://doi.org/10.1016/j.ymssp.2005.09.012.

[2] Darendeliler A, Claeys D, Afgezzaf E-H. Joint multi-item production and condition-based control of a system with setup times and stochastic demand. In: Liberatore F, Wesolkowski S, Parlier GH, editors. Proceedings of the 12th International Conference on Operations Research and Enterprise Systems: ICORES. Portugal. Science and Technology Publications (SciTePress): 2023. p. 185-192, 10.5220/0011741300003396.

[3] Darendeliler A, Claeys D, Aghezzaf E-H. Integrated condition-based maintenance and multi-item lot-sizing with stochastic demand. Journal of Industrial and Management Optimization. 2023 Sep 5;19(9):6908–47,10.5220/0011741300003396.

[4] Ben-Daya M, Makhdoum M. Integrated production and quality model under various preventive maintenance policies. Journal of the Operational Research Society. 1998;49(8):840–53, doi: https://doi.org/10.1038/sj.jors.2600586.

[5] Ben-Daya M. The economic production lot-sizing problem with imperfect production processes and imperfect maintenance. International Journal of Production Economics. 2002 Apr;76(3):257–64, Apr. 2002, doi: https://doi.org/10.1016/s0925-5273(01)00168-2.

[6] El-Ferik S. Economic production lot-sizing for an unreliable machine under imperfect age-based maintenance policy. European Journal of Operational Research. 2008 Apr;186(1):150–63, doi: https://doi.org/10.1016/j.ejor.2007.01.035.

[7] Liao GL, Sheu SH. Economic production quantity model for randomly failing production process with minimal repair and imperfect maintenance. International Journal of Production Economics. 2011 Mar;130(1):118–24, doi: https://doi.org/10.1016/j.ijpe.2010.12.004.

[8] Suliman SMA, Jawad SH. Optimization of preventive maintenance schedule and production lot size. International Journal of Production Economics. 2012 May;137(1):19–28, doi: https://doi.org/10.1016/j.ijpe.2012.01.006.

[9] Aghezzaf EH, Jamali MA, Ait-Kadi D. An integrated production and preventive maintenance planning model. European Journal of Operational Research. 2007 Sep;181(2):679–85, doi: https://doi.org/10.1016/j.ejor.2006.06.032.

[10] Shamsaei F, Van Vyve M. Solving integrated production and condition-based maintenance planning problems by MIP modeling. Flexible Services and Manufacturing Journal. 2016 May 3;29(2):184–202, doi: https://doi.org/10.1007/s10696-016-9244-8.

[11] Jafari L, Makis V. Joint optimal lot sizing and preventive maintenance policy for a production facility subject to condition monitoring. International Journal of Production Economics. 2015 Nov;169:156–68, doi: https://doi.org/10.1016/j.ijpe.2015.07.034.

[12] Zheng R, Zhou Y, Gu L, Zhang Z. Joint optimization of lot sizing and condition-based maintenance for a production system using the proportional hazards model. Computers and Industrial Engineering. 2021 Apr;154:107157, doi: https://doi.org/10.1016/j.cie.2021.107157.

[13] Jafari L, Makis V. Joint optimization of lot-sizing and maintenance policy for a partially observable two-unit system. The International Journal of Advanced Manufacturing Technology. 2016 Mar 9;87(5-8):1621–39, doi: https://doi.org/10.1007/s00170-016-8556-x.

[14] Peng H, van Houtum GJ. Joint optimization of condition-based maintenance and production lot-sizing. European Journal of Operational Research. 2016 Aug;253(1):94–107, doi: https://doi.org/10.1016/j.ejor.2016.02.027.
[15] Khatab A, Diallo C, Aghezzaf EH, Venkatadri U. Integrated production quality and condition-based maintenance optimisation for a stochastically deteriorating manufacturing system. International Journal of Production Research. 2018 Sep 14;57(8):2480–97, doi: https://doi.org/10.1080/00207543.2018.1521021.
[16] Cheng GQ, Zhou BH, Li L. Integrated production, quality control and condition-based maintenance for imperfect production systems. Reliability Engineering and System Safety. 2018 Jul;175:251–64, doi: https://doi.org/10.1016/j.ress.2018.03.025.
[17] Cheng GQ, Zhou BH, Li L. Joint optimization of lot sizing and condition-based maintenance for multi-component production systems. Computers and Industrial Engineering. 2017 Aug;110:538–49, doi: https://doi.org/10.1016/j.cie.2017.06.033.
[18] Iravani SMR, Duenyas I. Integrated maintenance and production control of a deteriorating production system. IIE Transactions. 2002 May;34(5):423–35, doi: https://doi.org/10.1080/07408170208928880.
[19] Sloan TW. A periodic review production and maintenance model with random demand, deteriorating equipment, and binomial yield. Journal of the Operational Research Society. 2004 Jun;55(6):647–56, doi: https://doi.org/10.1057/palgrave.jors.2601725.
[20] Xiang Y, Cassady CR, Jin T, Zhang CW. Joint production and maintenance planning with machine deterioration and random yield. International Journal of Production Research. 2013 Oct 17;52(6):1644–57, doi: https://doi.org/10.1080/00207543.2013.843037.
[21] Jafari L, Makis V. Optimal production and maintenance policy for a partially observable production system with stochastic demand. International Journal of Industrial and Systems Engineering. 2019;13(7).
[22] Winands EMM, Adan IJBF, van Houtum GJ. The stochastic economic lot scheduling problem: A survey. European Journal of Operational Research. 2011 Apr;210(1):1–9, doi: https://doi.org/10.1016/j.ejor.2010.06.011.
[23] Jin Qiu, Loulou R. Multiproduct production/inventory control under random demands. IEEE Transactions on Automatic Control. 1995;40(2):350–6, doi: https://doi.org/10.1109/9.341809.
[24] Wang J, Li X, Zhu X. Intelligent dynamic control of stochastic economic lot scheduling by agent-based reinforcement learning. International Journal of Production Research. 2012 Aug 15;50(16):4381–95, doi: https://doi.org/10.1080/00207543.2011.592158.
[25] Löhndorf N, Minner S. Simulation optimization for the stochastic economic lot scheduling problem. IIE Transactions. 2013 Jul;45(7):796–810, doi: https://doi.org/10.1080/0740817x.2012.662310.
[26] Sutton RS, Barto A. Reinforcement learning : an introduction. Cambridge, Ma ; London: The Mit Press; 2018.
[27] Bradtke S, Duff M. Reinforcement Learning Methods for Continuous-Time Markov Decision Problems. In: Tesauro G, Touretzky D, Leen T, editors. Proceedings of the Advances in Neural Information Processing Systems: MIT Press; 1994.
[28] Gascon A, Leachman RC, Lefrancois P. Multi-item, single-machine scheduling problem with stochastic demands: a comparison of heuristics. International Journal of Production Research. 1994 Mar;32(3):583–96, doi: https://doi.org/10.1080/00207549408956954.
[29] de Jonge B. Discretizing continuous-time continuous-state deterioration processes, with an application to condition-based maintenance optimization. Reliability Engineering and System Safety. 2019 Aug;188:1–5, doi: https://doi.org/10.1016/j.ress.2019.03.006.
[30] Powell WB. Approximate Dynamic Programming. John Wiley and Sons; 2011.
[31] Parr RE. Hierarchical control and learning for Markov decision processes [PhD Thesis]. [University of California at Berkeley]; 1998.