Adaptive service rate control of an $M/M/1$ queue with server breakdowns

YI ZHENG*, JUXIHONG JULAITI*, AND GUODONG PANG†

Abstract. We study service rate control problems for an $M/M/1$ queue with server breakdowns in which the breakdown rate is assumed to be a function of the service rate. Assuming that the queue has infinite capacity, we first establish the optimality equations for the discounted cost problem and characterize the optimal rate control policies. Then, we characterize the ergodicity of the controlled queue and establish the optimality conditions for the average-cost (ergodic) control problem using the vanishing discounted method. We next study the ergodic control problem when the queue has a finite capacity and establish a verification theorem by directly involving the stationary distribution of the controlled Markov process.

For practical applications, we consider the adaptive service rate control problem for the model with finite capacity. Studying this problem is useful because the relationship between the server breakdown rate and the service rate is costly to observe in practice. We propose an adaptive (self-tuning) control algorithm, assuming that the relationship between the server breakdown rate and the service rate is linear with unknown parameters. We prove that the regret vanishes under the algorithm and the proposed policies are asymptotically optimal. In addition, numerical experiments are conducted to validate the algorithm.

1. Introduction

In this paper, we study service rate control problems for a single-server queue with Poisson arrivals, exponential service times, and server breakdowns. The server availability is modeled by a random process with 'up' and 'down' states. The system functions normally in the 'up' state, while the server stop serving customers in the 'down' state. The sojourn time that the server stays in the 'down' state follows an exponential distribution with a constant parameter. The time for the server from 'up' to 'down' obeys an exponential distribution with a parameter depending on the service rate. The service rate control problems for the system with infinite or finite capacity will be addressed in this paper. The controller is allowed to choose a service rate at each state during the 'up' times. The cost function includes an effort cost that increases with the service rate, a holding/delay cost associated with the system congestion, a maintenance cost that occurs during the down state, and a rejection cost for the system with finite capacity.

It is well known that any effort to reduce unplanned downtime can create considerable savings in industries. For example, in manufacturing [31], certain materials will be wasted when machines suddenly break during the production process, disruptions and delays can be caused to the production of the downstream items, and/or unnecessary energy consumptions can be incurred during these breakdown times. In data centers, certain computing tasks may be lost when servers break down and will need to be restarted. Many queueing models that have been developed and analyzed include interruptions, such as the models in [4,7,28,49,50,58]. In many applications, the breakdown

*The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, College of Engineering, Pennsylvania State University, University Park, PA 16802

†Department of Computational Applied Mathematics and Operations Research, George R. Brown School of Engineering, Rice University, Houston, TX 77005

E-mail addresses: yxz282,jpj5196@psu.edu, gdpang@rice.edu.

Key words and phrases. $M/M/1$ queue, finite or infinite capacity, server breakdowns, service rate control, adaptive (self-tuning) control, Markov decision process, discounted and long-run average (ergodic) cost criteria.

rate of a system may be closely related to its service rate. In particular, in some manufacturing facilities, it is often a common practice to adjust service speeds in order to reduce energy costs, and it is also observed that some machines tend to break down more frequently when they run at a higher speed due to the resulting overheating/higher temperature. To the best of our knowledge, the queueing model with a service breakdown rate depending on the service rate has not been developed previously.

The service rate control problems for a single-server queue with infinite capacity have been extensively addressed. The closely related papers are [2, 6, 8, 29, 30, 41, 42, 44, 45]. In [29], the authors study service rate control problem of the long-run average cost in an $M/M/1$ queue. In [2], the joint admission and service rate control problem is studied for an $M/M/1$ queue. In [6], control of both the arrival and service rates is considered for an $M/M/1$ queue. In [30], the optimal buffer size and arrival/service rate control problem is studied using a diffusion approximation for a single server queue in heavy traffic. The service rate control problem of a single-server queue with a Markov-modulated Poisson arrival process has been studied in [41]. The control problem in which the controller is allowed to remove the server and adjust the service rate when the server is turned on is addressed in [8] for $M/M/1$ queue under the average cost criterion. We also like to mention the papers [37,38,43,57] for the admission and service rate control problems in multi-server queues.

We first consider the service rate control problem under the discounted cost and long-run average cost criteria for the queueing model with infinite capacity. The controlled queueing process is identified as a Markov decision process. For the discounted cost problem, we show the optimality equations for the optimal value function and the existence of optimal controls in Theorem 2.1. Since the breakdown rate for the server depends on the service rate, the characterization equation (2.4) for optimal controls contains an additional term corresponding to the breakdown rate, and the dynamic equations for the optimal values have two different forms for 'up' and 'down' states, respectively. Furthermore, under the assumptions of the cost functions and assuming the convexity and monotonicity of the breakdown rate function, we provide a representation of the optimal service rate control in Theorem 2.2.

For the service rate control problem under the long-run average cost, we apply the vanishing discount approach in the spirit of [52]. The results for the long-run average cost problem are given in Theorem 2.3. Its proof relies on the stability condition of the joint Markov process consisting of the queueing state and server availability processes. A necessary and sufficient condition related to the effective service rate for the stability of the joint Markov process is given in Proposition 2.3. The transition rate matrix results for the quasi-birth-and-death process play an important role in the proof of the stability.

In [29], the optimal value under the long-run average cost criterion for an $M/M/1$ queue with infinite capacity is approximated by solving a sequence of problems with the truncated holding cost function. They show that the optimal policies of the approximating problems converge monotonically to the optimal policy of the original problem. Their proof crucially relies on the monotonicity property of the optimal controls, that is, there exists an optimal control policy in which the service rate increases as a function of queue length. Another natural approach for the approximation of the service rate control problem with infinite capacity is to use a sequence of problems with the truncated state space [2]. It is shown that the limiting policy of the sequence of constructed approximation policies exists and is optimal for the control problem with infinite capacity. Their proof for the convergence of approximating problems also relies on the monotonicity of the optimal controls in the number of jobs in the system. In these works, the monotonicity properties hold naturally, since in the objective of minimizing the holding cost and effort costs, the controller will necessarily increase the service rate to reduce the holding cost when the system is more congested.

However, in our model setup, when the service rate is high, the server may tend to break down more frequently, that is, the breakdown rate is nondecreasing in the service rate. When the system is highly congested, the controller must take into account both the positive and negative effects of increasing the service rate: on one hand, it may reduce the congestion level (hence the holding cost), and on the other hand, it may cause the server to break down more frequently, which not only incurs repair costs, but also causes more jobs accumulating in the queue (breakdowns do not affect the arrival process), consequently increasing the congestion level and incurs more holding cost. Therefore, the well celebrated monotone property ([29, 45]) no longer holds for our model in general. In Remark 2.1, we illustrate why the monotone property does not hold when the breakdown rate is a linear increasing function of the service rate. This is further illustrated in the numerical examples in Section 5.1. Moreover, in Remark 2.2, we identify a sufficient condition for the monotone property to hold for our model, which unfortunately would require unrealistic condition on the parameters (that is, the repair rate is higher than the sum of the maximum service rate and the increment of the breakdown rate). As a consequence, due to the lack of the monotonicity property, we must develop new methods in order to prove the results in Theorems 2.1–2.3.

We next consider the service rate control problem under the ergodic cost for the queueing model with server breakdowns and finite capacity. The relation between the breakdown rate and the service rate is assumed to be a general continuous function and the action space is assumed to be a compact set. The results of the characterization for optimal controls are stated in Theorem 3.1. We consider a joint Markov process associated with the queueing state and the background process related to 'up' and 'down' states. Because the joint Markov process has finite states, we prove Theorem 3.1 by using the properties of the stationary distribution of the process. This approach was previously used in [2, 5], where the service rate control problem under the ergodic cost was studied for a single-server queue with finite capacity. However, in this paper, the construction of the verification equations, which depend on the background process, is different from those in [2, 5]. Then, because we assume a compact action space, the existence and uniqueness of the solution to the system of equations is presented in Lemma 3.2. Its proof, which uses contraction, is much similar in spirit to the method in [44]. The result in this lemma is used in the study of the adaptive control problem.

In practice, the relationship between the service rate and the breakdown rate of the system may not be known and may change over time. In an online learning framework, the aforementioned optimal control problem with finite capacity and an unknown relationship becomes an adaptive control problem. The adaptive control problems of Markov chains have been extensively studied in [12–15, 24, 25, 40, 55, 59]. They consider controlled Markov chains whose transition probabilities depend on an unknown parameter. For queueing systems, adaptive control of service rates is studied in [33], adaptive priority assignment in [34], and adaptive scheduling and routing in [54, 59].

As the third goal of the paper, we consider the adaptive control problem for an $M/M/1$ queue with breakdowns and finite capacity, where the relationship between the service rate and the breakdown rate depends on unknown parameters taking values in a compact set. This implies that the transition rate matrix of the Markov process is specified up to the unknown parameters. We assume that the breakdown rate is a linear function of the service rate with two unknown parameters. We apply a self-tuning approach to the adaptive control problem. The self-tuning scheme is introduced in [47]. Mandl [47] provides several models for controlled Markov processes with unknown parameters. The self-tuning approach is identified as a procedure in which the controlled policy is continuously modified based on the estimation of unknown parameters to approach the optimal policy for the problem with true parameters. We estimate the unknown parameters in the relation between the server breakdown and service rates based on the historical data at each jump time of the Markov process, and then the control implemented is characterized by an optimality equation with the current parameter estimate. The optimality equation used is the same as that in the aforementioned optimal control problem when the true parameters are known. Since the linear relation between the service and breakdown rates leads to a nonlinear relation between the mean service times and mean 'up' times of the system, the quasi-maximum likelihood estimates are used for the estimation of unknown parameters. This method has been studied in [23, 35] and references therein. Asymptotic optimality under work-conserving Markov rate control policies, which results in uniform ergodicity for the joint Markov process of the queueing state and background processes, is established in Theorem 4.1.

We also evaluate the performance of the method by conducting numerical experiments. For the system dynamics, we consider three scenarios, in which the proportions of up times and the rejection probabilities in the long run are different. We plot the functions for optimal service rates of congestion levels in these scenarios. Cost parameters are chosen to cover different cases in practice. We observe that the optimal service rates may not be monotone in the number of jobs in the system. This is different from the monotonicity of optimal service rates for an $M/M/1$ queue in [2, 29]. For the adaptive service rate control problem, we conduct simulation experiments and show that the proposed policies converge to the optimal service rate control policy.

It is worth mentioning that online learning problems are also related to adaptive control problems, since estimation methods for unknown parameters are used in both types of problems. When the unknown parameters are fitted statistically, the regret of the algorithm is commonly used to measure its performance [56]. In this paper, we show that the regret of the proposed algorithm vanishes and the proposed policies are asymptotically optimal. Blackwell's Approachability Theorem [11] provides conditions to analyze the regret of online learning algorithms [1, 16]. Online problems for demand models have been addressed extensively; see, for example, [9, 10, 22, 36]. In [36], the authors consider the unknown demand model, where the demand for products is assumed to be linear with unknown parameters. They provide modified greedy iterated least-squares policies to achieve asymptotic optimality. However, studies on online learning problems for queueing systems are scarce. Recently, Chen et al. [20] study the dynamic pricing and capacity sizing problem in a $GI/GI/1$ queue. They develop an online stochastic gradient descent method and show the regret bound for the convergence. Applications of reinforcement learning to queueing systems are not new; see, for example, [21, 46, 51]. In a recent work, Liu et al. [46] propose a model-based learning algorithm to a server allocation and routing problem in a queueing system. They show that the proposed algorithm can obtain the optimal policy using a Lyapunov analysis. However, the convergence of reinforcement learning is not well understood [56] and there are limited studies on service rate control with an unreliable server. In this paper, we consider adaptive service rate control problems for an $M/M/1$ queue with breakdowns, where the optimal controls do not have closed-form representations and are characterized by a system of Bellman equations.

1.1. **Organization of the paper.** The paper is organized as follows. Section 2.1 contains a detailed description of the $M/M/1$ queueing model with service breakdowns. After defining control policies considered in this study, we state the assumptions and describe the system dynamics. In Section 2.2, we establish optimality equations for the discounted cost problem. The characteristics of the value function and cost functions are then presented. The optimality equations of the ergodic problems are stated in Section 2.3, followed by the properties of the optimal controls. In Section 3, we define an ergodic control problem when the queue has a finite capacity and show the optimality of the controls. In Section 4, we present the results of the adaptive control problem for a system with finite capacity when the relationship between the server breakdown rate and the service rate is a linear function and the parameters for the function are unknown. Finally, numerical examples of the adaptive control problem are presented in Section 5.

2. $M/M/1$ queue with infinite capacity and server breakdowns

2.1. **The model description.** We consider an $M/M/1$ queue with adjustable service rate and server breakdowns. Let $\{A(t)\}_{t\geq 0}$ denote the Poisson arrival process with arrival rate $\lambda > 0$. The queueing system is in an up-down environment. In the 'up' state, the system functions normally and the controller chooses a dynamic service rate μ from the compact set $\mathbb{U} := [0, \bar{\mu}]$ with $0 < \bar{\mu} < \infty$.

In the 'down' state, the server stops, while jobs keep joining the queue and any job in service will wait for the system to resume.

Let $\{X(t)\}_{t\geq 0}$ and $\{K(t)\}_{t\geq 0}$ denote the number of jobs in the system (including those in queue and in service, either interrupted or not) and the server availability process, respectively. At time $t \geq 0$, $K(t) = 1$ if the system is in the "up" state, and $K(t) = 0$ otherwise. We say that a rate control policy U is admissible if it is non-anticipative, takes values in $\mathbb{U}$, and satisfies that $U(t) = 0$ if $K(t) = 0$ for $t \geq 0$, such that the system is work-conserving (no idling when there are jobs waiting in queue in the 'up' state). The set of admissible rate control policies is denoted by $\mathfrak{U}$. Let $\mathbb{Z}_+$ denote the set of nonnegative integers. An admissible rate control policy U is called stationary Markov if

$$U(t) \,=\, \nu(X(t), K(t)) \qquad t \geq 0\,,$$

for some $\nu\colon \mathbb{Z}_+ \times \{0,1\} \mapsto \mathbb{U}$ satisfying $\nu(0,1) \equiv 0$ and $\nu(x,0) \equiv 0$ for $x \in \mathbb{Z}_+$. The work-conserving condition means that $\nu(x,1) > 0$ for $x \geq 1$. The set of stationary Markov control policies is denoted by $\mathfrak{U}_{\mathrm{sm}}$.

We assume that given $U \in \mathfrak{U}$, $\{K(t)\}_{t\geq 0}$ is a continuous-time Markov process with state space $\{0,1\}$, and its transition rate matrix is given by

$$\begin{bmatrix} -\beta_d & \beta_d \\ \beta_u(U(t)) & -\beta_u(U(t)) \end{bmatrix}$$

for $t \geq 0$, where $\beta_u\colon \mathbb{U} \mapsto \mathbb{R}_+$ is a measurable function ($\mathbb{R}_+$ denotes the set of nonnegative real numbers), and β_d is a positive constant. $\beta_u(\cdot)$ and β_d represent the breakdown (from "up" to "down") and repair (from "down" to "up") rates, respectively. This implies that the repair of the server is started immediately when the server stops. The repair rate is usually assumed to be a constant, see, for example, Section 2 of [26]. In practice, the functioning time of a server, for example, a machine, depends on its effort. Thus, the breakdown rate is assumed to be a function of service rate. In addition, we assume that $\beta_u(\cdot)$ is strictly positive and continuously differentiable. Given $\nu \in \mathfrak{U}_{\mathrm{sm}}$ and $X(0)$, the state process $\{X(t)\}_{t\geq 0}$ evolves as the following

$$X(t) \,=\, X(0) + A(t) - S\biggl(\int_0^t \nu\bigl(X(s), K(s)\bigr)(X(s) \wedge 1)\,\mathrm{d}s\biggr) \qquad \forall\, t \geq 0\,, \tag{2.1}$$

where $\{S(t)\}_{t\geq 0}$ is a unit rate Poisson process, independent of the arrival process $\{A(t)\}_{t\geq 0}$. Note that provided $\nu \in \mathfrak{U}_{\mathrm{sm}}$, $\{(X(t), K(t))\}_{t\geq 0}$ is a well-defined Markov process.

The costs of our optimization problems consist of the effort, holding/delay and repair costs. The effort cost function is denoted by $\mathcal{R}(\cdot)$, that is, the cost rate is $\mathcal{R}(\mu)$ per unit time when a service rate $\mu \in \mathbb{U}$ is selected. We assume that $\mathcal{R}(\cdot)$ is strictly increasing, continuously differentiable, and such that $\mathcal{R}(0) = 0$. The holding cost function is defined by $H(x)$ for $x \in \mathbb{Z}_+$, which is assumed to be convex and nondecreasing. We also assume that during the down times, the system incurs a cost at a positive constant rate C_m. The total cost function is defined by

$$f(x,k,\mu) \,:=\, \mathcal{R}(\mu) + H(x) + C_m(1-k) \tag{2.2}$$

for $(x,k,\mu) \in \mathbb{Z}_+ \times \{0,1\} \times \mathbb{U}$. In the next two subsections, we consider the discounted and long-run average (ergodic) cost minimization problems.

2.2. **The discounted cost problem.** In this subsection, we present the results for the service rate control problem under the discounted cost criterion. We study the optimality equations and show the properties of the optimal controls.

For $U \in \mathfrak{U}$, the α-discounted cost criterion is given by

$$J_\alpha^U(x,k) \,:=\, \mathbb{E}_{x,k}^U\biggl[\int_0^\infty \mathrm{e}^{-\alpha s} f\bigl(X(s), K(s), U(s)\bigr)\,\mathrm{d}s\biggr] \qquad \forall\, \alpha > 0\,.$$

The optimal α-discounted value function is denoted by

$$V_\alpha(x,k) := \min_{U\in\mathfrak{U}} J_\alpha^U(x,k) \qquad \forall\, \alpha > 0\,. \tag{2.3}$$

We say that a control $U^* \in \mathfrak{U}$ is optimal if $J_\alpha^{U^*} = V_\alpha$.

In the next theorem, we show that a stationary optimal control for the α-discounted problem exists, and that V_α is the solution of optimality equations. We first define

$$\phi(w,y) := \max_{\mu\in\mathbb{U}}\bigl\{w\mu - \beta_u(\mu)y - \mathcal{R}(\mu)\bigr\} \quad \forall\,(w,y)\in\mathbb{R}\times\mathbb{R}\,. \tag{2.4}$$

Theorem 2.1. *There exists an optimal control policy $\nu_\alpha^* \in \mathfrak{U}_{\mathrm{sm}}$ for the α-discounted problem (2.3). The value function V_α satisfies the following discounted cost optimality equations*

$$\begin{aligned} V_\alpha(0,1) &= \frac{1}{\alpha+M}\Bigl(H(0) + \lambda V_\alpha(1,1) + \beta_u(0)V_\alpha(0,0) + \bigl(M-\lambda-\beta_u(0)\bigr)V_\alpha(0,1)\Bigr)\,,\\ V_\alpha(x,1) &= \frac{1}{\alpha+M}\Bigl(H(x) - \phi\bigl(W_\alpha(x,1), Y_\alpha(x)\bigr) + \lambda V_\alpha(x+1,1) + \bigl(M-\lambda\bigr)V_\alpha(x,1)\Bigr) \end{aligned} \tag{2.5}$$

for $x\in\mathbb{N}$, and

$$V_\alpha(x,0) = \frac{1}{\alpha+M}\Bigl(H(x) + C_m + \lambda V_\alpha(x+1,0) + \beta_d V_\alpha(x,1) + (M-\lambda-\beta_d)V_\alpha(x,0)\Bigr)$$

for $x\in\mathbb{Z}_+$, where $\mathbb{N}$ denotes the set of natural numbers, $M := \bar\mu + \lambda + \beta_d + \beta_u(\bar\mu)$,

$$W_\alpha(x,1) := V_\alpha(x,1) - V_\alpha(x-1,1) \quad \textit{and} \quad Y_\alpha(x) := V_\alpha(x,0) - V_\alpha(x,1)\,. \tag{2.6}$$

Proof. We use the uniformization technique to prove this theorem (see, for example, [41, 45]). Let $V_{n,\alpha}(x,k)$ be the optimal α-discounted expected value obtained during the last n transitions starting from the state (x,k). We assume that $V_{0,\alpha}(x,0) = V_{0,\alpha}(x,1) = 0$ for all $x\in\mathbb{Z}_+$.

The recursive formula for $V_{n,\alpha}$ is given by

$$\begin{aligned} V_{n+1,\alpha}(0,1) = \frac{1}{\alpha+M}\Bigl(&H(x) + \lambda V_{n,\alpha}(1,1) + \beta_u(0)V_{n,\alpha}(0,0)\\ &+ \bigl(M-\lambda-\beta_u(0)\bigr)V_{n,\alpha}(0,1)\Bigr)\,, \end{aligned} \tag{2.7}$$

$$\begin{aligned} V_{n+1,\alpha}(x,1) = \frac{1}{\alpha+M}\min_{\mu\in\mathbb{U}}\bigl\{&\mathcal{R}(u) + H(x) + \mu V_{n,\alpha}(x-1,1) + \lambda V_{n,\alpha}(x+1,1)\\ &+ \beta_u(\mu)V_{n,\alpha}(x,0) + \bigl(M-\lambda-\mu-\beta_u(\mu)\bigr)V_{n,\alpha}(x,1)\bigr\} \end{aligned} \tag{2.8}$$

for $x\in\mathbb{N}$, and

$$V_{n+1,\alpha}(x,0) = \frac{1}{\alpha+M}\Bigl(H(x) + C_m + \lambda V_{n,\alpha}(x+1,0) + \beta_d V_{n,\alpha}(x,1) + (M-\lambda-\beta_d)V_{n,\alpha}(x,0)\Bigr) \tag{2.9}$$

for $x\in\mathbb{Z}_+$. Let

$$W_{n,\alpha}(\cdot,k) := V_{n,\alpha}(\cdot,k) - V_{n,\alpha}(\cdot-1,k) \quad \text{and} \quad Y_{n,\alpha}(\cdot) := V_{n,\alpha}(\cdot,0) - V_{n,\alpha}(\cdot,1) \tag{2.10}$$

for $k\in\{0,1\}$. Then, (2.9) takes the form

$$V_{n+1,\alpha}(x,1) = \Bigl(H(x) - \phi\bigl(W_{n,\alpha}(x,1), Y_{n,\alpha}(x)\bigr) + \lambda V_{n,\alpha}(x+1,1) + \bigl(M-\lambda\bigr)V_{n,\alpha}(x,1)\Bigr)\,. \tag{2.11}$$

Applying [27, Proposition 3.1 (iii)], taking $n\to\infty$ in (2.7), (2.9), and (2.11) yields the equations in the statement. By part (ii) of Proposition 3.1 of [27], there exists an optimal control $\nu_\alpha^* \in \mathfrak{U}_{\mathrm{sm}}$. This completes the proof. □

The next two propositions are used to derive the characterization of the optimal controls. In the following proposition, we show that the α-discounted value functions are nondecreasing in the number of jobs in the system.

Proposition 2.1. *For $k\in\{0,1\}$, the value function $V_\alpha(\cdot,k)$ is a nondecreasing function.*

Proof. We prove this result by induction. Recall $V_n(x,1)$ and $V_n(x,0)$ given in (2.8) and (2.9), respectively. This result trivially holds when $n = 0$. Suppose that for $k \in \{0,1\}$, $V_{n,\alpha}(\cdot,k)$ is a nondecreasing function. We let ν_α^{n+1} be an optimal stationary Markov control of $(n+1)$-stage problem in (2.8). For the notational convenience, we denote

$$\mu_x := \nu_\alpha^{n+1}(x,1)\,. \tag{2.12}$$

Without loss of generality, we assume $\mu_0 = 0$. It follows by the inductive hypotheses and equations (2.7)–(2.9) that for $x \in \mathbb{N}$,

$$\begin{aligned} V_{n+1,\alpha}(x-1,1) \;&\le\; \frac{1}{\alpha+M}\Big(H(x-1) + \mathcal{R}(\mu_x) + \mu_x V_{n,\alpha}((x-2)^+,1) + \lambda V_{n,\alpha}(x,1) \\ &\qquad + \beta_u(\mu_x)V_{n,\alpha}(x-1,0) + \big(M-\lambda-\mu_x-\beta_u(\mu_x)\big)V_{n,\alpha}(x-1,1)\Big) \\ &\le\; V_{n+1,\alpha}(x,1)\,, \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} V_{n+1,\alpha}(x-1,0) \;&\le\; \frac{1}{\alpha+M}\Big(C_m + H(x-1) + \lambda V_{n,\alpha}(x,0) \\ &\qquad + \beta_d V_{n,\alpha}(x-1,1) + (M-\lambda-\beta_d)V_{n,\alpha}(x-1,0)\Big) \\ &\le\; V_{n+1,\alpha}(x,0)\,. \end{aligned}$$

The result follows by taking $n \to \infty$. □

In the next proposition, we show that if the repair cost is higher than the effort cost, then the α-discounted cost starting from 'down' state is higher than that starting from 'up' state.

Proposition 2.2. *Assume $C_m \ge \mathcal{R}(\bar\mu)$. Then, $V_\alpha(x,0) \ge V_\alpha(x,1)$, for $x \in \mathbb{Z}_+$.*

Proof. We use induction. It is evident that the result holds for $V_{0,\alpha}$. Suppose $V_{n,\alpha}(\cdot,0) \ge V_{n,\alpha}(\cdot,1)$. By taking the difference of (2.7) and (2.9), and using the inductive hypothesis, the result trivially holds when $x = 0$ for $(n+1)$-stage problem. Recall μ_x in (2.12), and $W_{n,\alpha}$ and $Y_{n,\alpha}$ in (2.10). By using (2.8) and (2.9), we obtain

$$\begin{aligned} (\alpha+M)Y_{n+1,\alpha}(x+1) \;&=\; C_m - \mathcal{R}(\mu_{x+1}) + \mu_{x+1}W_{n,\alpha}(x+1,1) \\ &\qquad + \lambda Y_{n,\alpha}(x+2) + \big(M-\lambda-\beta_u(\mu_{x+1})-\beta_d\big)Y_{n,\alpha}(x+1) \\ &\ge\; 0\,, \end{aligned} \tag{2.14}$$

where the inequality follows by (2.13), the inductive hypothesis, and $C_m \ge \mathcal{R}(\bar\mu)$. By taking $n \to \infty$, the result follows. □

Assumption 2.1. In addition to the assumptions stated in Section 2.1, we assume that the function $\beta_u(\cdot)$ is nondecreasing and convex, $\mathcal{R}(\cdot)$ is strictly convex, and $C_m \ge \mathcal{R}(\bar\mu)$.

The assumption of $\beta_u(\cdot)$ in Assumption 2.1 implies that the system is more likely to breakdown when the service rate is at high level.

Theorem 2.2. *Grant Assumption 2.1. Then, the results in Theorem 2.1 hold, and there exists an optimal control $\nu_\alpha^* \in \mathfrak{U}_{\mathrm{sm}}$ for the α-discounted problem such that*

$$\nu_\alpha^*(x,1) \;=\; \psi\big(W_\alpha(x,1), Y_\alpha(x)\big)\,, \tag{2.15}$$

where the function ψ is the maximizer of ϕ in (2.4) and satisfies

$$\psi(w,y) \;=\; \begin{cases} 0\,, & \text{for } w \;\le\; \mathcal{R}'(0) + \beta_u'(0)y\,, \\ (y\beta_u' + \mathcal{R}')^{-1}(w) & \text{for } \mathcal{R}'(0) + \beta_u'(0)y \;<\; w \;\le\; \mathcal{R}'(\bar\mu) + \beta_u'(\bar\mu)y\,, \\ \bar\mu & \text{for } w \;>\; \mathcal{R}'(\bar\mu) + \beta_u'(\bar\mu)y\,. \end{cases} \tag{2.16}$$

Proof. Since $\mathcal{R}(\cdot)$ is a strictly convex function, then by the assumption that $\beta_u(\cdot)$ is convex, the set of maximizers of ϕ becomes a singleton if $y \geq 0$. But it follows by Proposition 2.2 that $Y_\alpha(x) \geq 0$ for any $x \in \mathbb{Z}_+$. Using Assumption 2.1, $(y\beta'_u + \mathcal{R}')^{-1}$ is continuous and strictly increasing. Thus, (2.15) holds, and the rest of the proof is the same as that of Theorem 2.1. □

Remark 2.1. The following special case is frequently used in the rest of paper. If

$$\beta_u(\mu) \,=\, \kappa_1 + \kappa_2\mu \tag{2.17}$$

for some positive constants κ_1 and κ_2, then

$$\psi(w,y) \,=\, \begin{cases} 0\,, & \text{for } w-\kappa_2 y \,\leq\, \mathcal{R}'(0)\,, \\ (\mathcal{R}')^{-1}(w-\kappa_2 y) & \text{for } \mathcal{R}'(0) \,<\, w-\kappa_2 y \,\leq\, \mathcal{R}'(\bar{\mu})\,, \\ \bar{\mu} & \text{for } w-\kappa_2 y \,>\, \mathcal{R}'(\bar{\mu})\,. \end{cases} \tag{2.18}$$

In this special case, it can be easily seen that the optimal control policy may not be monotone in the queue length (recalling the expressions of $W_\alpha(x,1)$ and $Y_\alpha(x)$ in (2.6)). This is also very intuitive: since the breakdown rate is increasing in the service rate, even if the system is at a high congestion level, the controller may not risk to increase the speed since a breakdown can possibly increase congestion more severely and thus induce more delay costs (and lost jobs in the case of finite capacity). See also the numerical examples for illustration in Section 5.1.

Remark 2.2. One may ask under what conditions a monotonically optimal control exists. We provide a sufficient condition for this to hold: (2.17) holds, $\mathcal{R}(\cdot)$ is strictly convex, $C_m \geq \mathcal{R}(\bar{\mu})$ and

$$\beta_d \,\geq\, (\kappa_2+1)\bar{\mu}\,. \tag{2.19}$$

The inequality (2.19) implies that the repair rate is higher than the sum of the maximum service rate and the increment of the breakdown rate. It is clear that this condition may be unrealistic. So for practical problems, there is a lack of monotonicity property as in the existing literature on rate control of single server queues [6, 29]. Nevertheless, it can be shown that under these conditions, the following structural properties of the optimal policies hold:

(i) For $k \in \{0,1\}$, $V_\alpha(\cdot,k)$ is a convex function.
(ii) Y_α is a nondecreasing function.
(iii) There exists an optimal Markov control policy $\nu^*_\alpha \in \mathfrak{U}_{\mathrm{sm}}$ such that $\nu^*_\alpha(\cdot,1)$ is a nondecreasing function.

We omit the details to prove these properties for brevity.

2.3. **The long-run average cost problem.** In this subsection, we first establish a necessary and sufficient condition for the stability of the joint Markov process. Under the stability condition, we show the existence and characterization of optimal controls for the long-run average expected cost problem by utilizing the vanishing discounted method.

Recall f in (2.2). Under $U \in \mathfrak{U}$, the expected long-run average cost criterion is given by

$$\varrho^U(x,k) \,:=\, \limsup_{T\to\infty} \frac{1}{T}\,\mathbb{E}^U_{x,k}\left[\int_0^T f\big(X(s),K(s),U(s)\big)\,\mathrm{d}s\right].$$

The optimal expected long-run average cost is defined by

$$\varrho_*(x,k) \,:=\, \inf_{U\in\mathfrak{U}} \varrho^U(x,k)\,. \tag{2.20}$$

We say that a policy ν_* is optimal for long-run average cost if $\varrho_*(x,k) = \varrho^{\nu_*}(x,k)$. In Proposition 2.3, we show that ϱ_* is finite and independent of (x,k) under the following assumption on the holding cost function (recall that we have assumed that $H(\cdot)$ is convex, and note that the subgeometric condition below holds in the linear case). The finiteness of ϱ_* implies that the average cost problem is well-posed. Similar assumptions on the holding cost function were also used in [29, 41].

Assumption 2.2. The holding cost function $H(\cdot)$ is subgeometric, that is,

$$\sum_{n=0}^{\infty} H(n)\gamma^n < \infty\,, \quad \forall\, \gamma \in [0,1)\,. \tag{2.21}$$

In the next proposition, we provide a necessary and sufficient condition for the long-run average cost to be finite. We say that a policy $\nu \in \mathfrak{U}_{\mathrm{ssm}}$ is stable if (X,K) is positive recurrent under ν. Here the set of stationary and stable Markov control policies is denoted by $\mathfrak{U}_{\mathrm{ssm}}$.

Proposition 2.3. *There exists a stable policy $\nu \in \mathfrak{U}_{\mathrm{ssm}}$ if and only if*

$$\min_{\mu\in\mathbb{U}} \left\{ \frac{\lambda\bigl(\beta_u(\mu)+\beta_d\bigr)}{\beta_d \mu} \right\} \;<\; 1\,. \tag{2.22}$$

Moreover, under Assumption 2.2, there exists a stable policy $\nu \in \mathfrak{U}_{\mathrm{ssm}}$ such that ϱ^ν is finite and independent of (x,k).

Proof. Under $\nu \in \mathfrak{U}_{\mathrm{sm}}$, the infinitesimal generator of (X,K) is given by

$$\mathcal{Q} \;:=\; \begin{bmatrix} Q_0 & A & & & & \\ B_1 & Q_1 & A & & & \\ & B_2 & Q_2 & A & & \\ & & B_3 & Q_3 & A & \\ & & & \ddots & \ddots & \ddots \end{bmatrix},$$

where

$$A \;:=\; \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \qquad Q_0 \;:=\; \begin{bmatrix} -\bigl(\beta_d+\lambda\bigr) & \beta_d \\ \beta_u(0) & -\bigl(\beta_u(0)+\lambda\bigr) \end{bmatrix},$$

$$B_x \;:=\; \begin{bmatrix} 0 & 0 \\ 0 & \nu(x,1) \end{bmatrix}, \quad \text{and} \quad Q_x \;:=\; \begin{bmatrix} -\bigl(\beta_d+\lambda\bigr) & \beta_d \\ \beta_u(\nu(x,1)) & -\bigl(\beta_u(\nu(x,1))+\nu(x,1)+\lambda\bigr) \end{bmatrix},$$

with $x \in \mathbb{N}$. We first prove the sufficiency. It is evident that under $\nu \in \mathfrak{U}_{\mathrm{sm}}$, (X,K) is irreducible when $\inf_{x\in\mathbb{N}} \nu(x,1) > 0$. Let $\mu_\circ$ be a minimizer of (2.22), and $\nu(x,1) \equiv \mu_\circ$ for $x \in \mathbb{N}$. Then, the embedded Markov chain of the joint process (X,K) becomes a discrete time level-independent quasi-birth-and-death process, see [32] for the detailed definition. We define

$$Q \;:=\; A + B_1 + Q_1\,.$$

For $x \in \mathbb{R}^d$, x^{T} denotes the transpose of x. It is straightforward to verify that

$$\eta \;:=\; \left(\frac{\beta_u(\mu_\circ)}{\beta_u(\mu_\circ)+\beta_d}, \frac{\beta_d}{\beta_u(\mu_\circ)+\beta_d} \right)^{\mathsf{T}}$$

solves the system of equations $\eta^{\mathsf{T}} Q = 0$ and $e^{\mathsf{T}}\eta = 1$. Then, it follows by [32, Theorem 3.2.1] that under $\nu \in \mathfrak{U}_{\mathrm{sm}}$, (X,K) is recurrent if

$$\eta^{\mathsf{T}} A e < \eta^{\mathsf{T}} B_1 e\,. \tag{2.23}$$

It is evident that (2.23) is equivalent to $\lambda\bigl(\beta_u\bigl(\mu_\circ\bigr)+\beta_d\bigr) < \beta_u(\mu_\circ)\mu_\circ$. Since the spectral radius of $\mathcal{Q}$ is bounded, the stationary distribution of (X,K) exists and thus $\nu \in \mathfrak{U}_{\mathrm{ssm}}$.

For the necessity, it follows by Theorem 3.1.1 of [48] that (X,K) is positive recurrent under $\mu_\circ$ only if $\eta^{\mathsf{T}} A e < \eta^{\mathsf{T}} B_1 e$.

Similarly as above, we choose $\nu(x,1) \equiv \mu_\circ$. We use R_ν to denote the rate matrix that satisfies

$$(\pi_\nu(n-1,0), \pi_\nu(n-1,1))\, R_\nu \;=\; (\pi_\nu(n,0), \pi_\nu(n,1)) \quad \text{for } n \geq 1\,,$$

where π_ν denotes the stationary distribution of (X,K) governed by ν. Applying Theorem 3.1.1 of [48] again, R_ν is the minimal nonnegative solution of the quadratic matrix equation

$$(R_\nu)^2\, B_x + R_\nu\, Q_x + A \;=\; 0\,, \tag{2.24}$$

and the spectral radius of R_ν is less than 1. We refer the reader to Lemma 3.1 for an explicit representation of R_ν. Thus, the spectral radius of $(R_\nu)^m$ converges to 0 at a geometric decay as $m \to \infty$. Therefore, by Assumption 2.2, we have that ϱ^ν is finite. This completes the proof. □

Remark 2.3. We rewrite (2.22) as

$$\lambda \,<\, \max_{\mu\in\mathbb{U}}\biggl\{\frac{\mu\beta_d}{\beta_u(\mu)+\beta_d}\biggr\}\,. \tag{2.25}$$

Note that $\frac{\beta_d}{\beta_d+\beta_u(\mu)}$ can be viewed as the fraction of time that the server functions normally, and thus, the expression in the maximum on the right-hand side of (2.25) represents the effective service rate. Therefore, (2.25) means that there exists a service rate such that the arrival rate is less than the effective service rate, that is, the stability condition for (X, K).

In the next theorem, we present the existence of the solution to the average cost optimality inequalities (ACOI) and the optimal policy for the long-run average expected cost. To prove this theorem, we apply the vanishing discounted method.

Theorem 2.3. *Suppose that $C_m \ge \mathcal{R}(\bar\mu)$, and Assumption 2.2 and (2.22) hold. Then, the following items hold:*

(i) *As $\alpha \searrow 0$, $V_\alpha(\cdot,k) - V_\alpha(0,1)$ converges, along a subsequence, to a function $V(\cdot,k)$, for $k \in \{0,1\}$, and $\varrho_* = \lim_{\alpha\to 0}\alpha V_\alpha(x,k)$ for every $(x,k)\in\mathbb{Z}_+\times\{0,1\}$. Moreover, $\{V(\cdot,k)\colon k\in\{0,1\}\}$ and ϱ_* satisfy the ACOI:*

$$\begin{cases} V(x,1) \,\ge\, \frac{1}{M}\bigl(H(x)+\lambda V(x+1,1)-\phi(W(x,1),Y(x))-\varrho_*+(M-\lambda)V(x,1)\bigr)\,,\\ V(x,0) \,\ge\, \frac{1}{M}\bigl(H(x)+\lambda V(x+1,0)+C_m-\beta_d Y(x)-\varrho_*+(M-\lambda)V(x,0)\bigr)\,,\end{cases} \tag{2.26}$$

for $x\in\mathbb{N}$, and

$$0 \,\ge\, \frac{1}{M}\bigl(H(0)+\mathcal{R}(0)+\lambda V(1,1)+\beta_u(0)Y(0)-\varrho_*\bigr)\,, \tag{2.27}$$

where $W(x,1) := V(x+1,1)-V(x,1)$ and $Y(x) := V(x,0)-V(x,1)$ for $x\in\mathbb{Z}_+$.

(ii) *There exists a long-run average cost optimal control $\nu^*\in\mathfrak{U}_{\mathrm{ssm}}$, which is a limit of a sequence of optimal controls for the discounted cost problem.*

Proof. To prove this theorem, we verify Assumptions 1-8 in [52]. Since the uniformization rate M is positive and finite, then Assumptions 1 and 2 of [52] are satisfied. To verify Assumptions 3 and 4 of [52], it suffices to show that Lemma 2.1 (i) of [52] holds. But it follows directly by Proposition 2.3, which implies that there exists a stable policy such that the long-run average cost is finite and (X, K) is an ergodic Markov process. Let $V_\alpha(0, 1)$ be the distinguishing point. Applying Propositions 2.1 and 2.2, we obtain

$$V_\alpha(x,k)-V_\alpha(0,1) \,\ge\, 0 \quad \forall\,(x,k)\in\mathbb{Z}_+\times\{0,1\}\,.$$

Thus, Assumption 5 of [52] is satisfied. Let

$$C_{x,0}(\alpha,0) \,:=\, \frac{C_m+H(x)}{\alpha+M} \quad\text{and}\quad C_{x,1}(\alpha,\mu) \,:=\, \frac{\mathcal{R}(\mu)+H(x)}{\alpha+M}\,.$$

It is evident that for each $(x,k)\in\mathbb{Z}_+\times\{0,1\}$, $C_{x,0}$ and $C_{x,1}$ are continuous functions on $[0,\infty)\times\mathbb{U}$. Hence, Assumption 6 of [52] is verified. To verify Assumption 7 of [52], we define

$$L_{(x,0),(x',k')}(\alpha,\mu) = \begin{cases} \frac{\lambda}{\alpha+M} & \text{if } x'=x+1 \text{ and } k'=0\,,\\ \frac{\beta_d}{\alpha+M} & \text{if } x'=x \text{ and } k'=1\,,\\ 0 & \text{otherwise}\,,\end{cases}$$

for $x \in \mathbb{Z}_+$,

$$L_{(0,1),(x',k')}(\alpha,\mu) = \begin{cases} \frac{\lambda}{\alpha+M} & \text{if } x'=1 \text{ and } k'=1\,, \\ \frac{\beta_u(0)}{\alpha+M} & \text{if } x'=0 \text{ and } k'=0\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

and

$$L_{(x,1),(x',k')}(\alpha,\mu) = \begin{cases} \frac{\lambda}{\alpha+M} & \text{if } x'=x+1 \text{ and } k'=1\,, \\ \frac{\mu}{\alpha+M} & \text{if } x'=x-1 \text{ and } k'=1\,, \\ \frac{\beta(\mu)}{\alpha+M} & \text{if } x'=x \text{ and } k'=0\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

for $x \in \mathbb{N}$. Then, it is evident that for all $(x,k) \in \mathbb{Z}_+ \times \mathbb{U}$, $L_{(x,k),(x',k')}(\alpha,\mu)$ is a continuous function on $[0,\infty) \times \mathbb{U}$, and Assumption 7 of [52] is satisfied. Since the expected sojourn times for (X,K) under any ν are equal to $1/M$, then Assumption 8 of [52] holds. Thus, by Theorem 12 of [52], we have shown (i) and (ii). □

Remark 2.4. It can be also proved that under the hypotheses in Theorem 2.3 and the sufficient conditions as stated in Remark 2.2, there exists an optimal policy $\nu^* \in \mathfrak{U}_{\mathrm{ssm}}$ such that $\nu^*(\cdot,1)$ is nondecreasing. However, the sufficient condition (2.19) may be unrealistic, and the optimal policy in Theorem 2.3 does not have monotone property in general.

Remark 2.5. In practice, to obtain the optimal policies, people usually study the approximation of the original problem, since (2.26) is almost impossible to solve directly for the infinite state space. In the literature, there are two approaches for the approximation. One approach is to use a sequence of control problems with truncated holding cost functions to produce the approximations. In [29], an asymptotic method that uses a truncated holding cost function to compute the optimal policy is developed for the average-cost problem of the $M/M/1$ queue. They show that the optimal policies of the approximating problems converge monotonically to the optimal policy of the original problem, and the optimal controls are nondecreasing functions in the congestion level of the system. The other approach is to use a sequence of optimal objective values for the finite state space to approximate the original optimal value. This approach has been studied in [2,5] for the $M/M/1$ queue with rejection cost (see also [38] in a multi-server setting). For the approximation with the truncated holding cost function, it is crucial to first obtain a monotone sequence of approximation for optimal policies. However, for the $M/M/1$ queue with server breakdowns, since the breakdown rate depends on the service rate, the optimal controls may not be monotone in the number of jobs. Therefore, we consider an $M/M/1$ queue with finite capacity for practical problems and establish the results for the system in next section.

3. $M/M/1$ QUEUE WITH FINITE CAPACITY AND SERVER BREAKDOWNS

In this section, we consider the same queueing system in the previous section except that the capacity is finite, denoted as N, and focus on the ergodic control problem. When the system is full, new arrivals are rejected. We let $p > 0$ denote a fixed penalty to reject a customer. For notation convenience, we let $\nu_x \equiv \nu(x,1)$ for $x \in \mathbb{Z}_+$. The ergodic cost criterion is given by

$$\varrho^\nu := \sum_{x=0}^{N} \Big(\pi_\nu(x,1)\big(H(x) + \mathcal{R}(\nu_x)\big) + \pi_\nu(x,0)\big(H(x) + C_m\big) \Big) + \lambda p\big(\pi_\nu(N,1) + \pi_\nu(N,0)\big)\,. \tag{3.1}$$

Here $\{\pi_\nu(x,k)\colon 0 \le x \le N, k \in \{0,1\}\}$ denotes the stationary distribution of the irreducible joint Markov process (X,K) under a policy $\nu \in \mathfrak{U}_{\mathrm{sm}}$ (for this model $\mathfrak{U}_{\mathrm{sm}} = \mathfrak{U}_{\mathrm{ssm}}$). The optimal ergodic cost is defined by

$$\varrho_* := \inf_{\nu \in \mathfrak{U}_{\mathrm{sm}}} \varrho^\nu\,. \tag{3.2}$$

We replace the assumptions on the action space, the cost functions and the server breakdown rate function $\beta_u(\cdot)$ in the previous section with the following relaxed assumptions.

Assumption 3.1. The following conditions hold.

(i) The action space $\mathbb{U}$ is a compact subset of $[0,\infty)$ satisfying $0 \in \mathbb{U}$.
(ii) The holding cost function $H(\cdot)$ is nondecreasing.
(iii) The effort cost function $\mathcal{R}(\cdot)$ is nondecreasing and continuous satisfying $\mathcal{R}(0) = 0$.
(iv) The function $\beta_u(\cdot)$ is strictly positive and continuous.

Recall $\bar{\mu} := \max\{\mu : \mu \in \mathbb{U}\}$ and $M := \bar{\mu} + \lambda + \beta_u(\bar{\mu}) + \beta_d$. By applying the uniformization technique, the optimality equations for the ergodic control problem are given by

$$V(0,1) \;=\; \frac{1}{M}\big(\mathcal{R}(0) + H(0) + \lambda V(1,1) + \beta_u(0)V(0,0) - \varrho + (M - \lambda - \beta_u(0))V(0,1)\big)\,, \tag{3.3}$$

and

$$\begin{aligned} V(x,1) \;=\; \frac{1}{M}\min_{\mu\in\mathbb{U}}\big\{&\mathcal{R}(\mu) + H(x) + \lambda V(x+1,1) + \mu V(x-1,1) \\ &+ \beta_u(\mu)V(x,0) - \varrho + (M - \lambda - \mu - \beta_u(\mu))V(x,1)\big\} \end{aligned} \tag{3.4}$$

for $1 \le x \le N-1$, and

$$V(x,0) \;=\; \frac{1}{M}\big(H(x) + C_m + \lambda V(x+1,0) + \beta_d V(x,1) - \varrho + (M - \lambda - \beta_d)V(x,0)\big) \tag{3.5}$$

for $0 \le x \le N-1$, and

$$\begin{aligned} V(N,1) \;=\; \frac{1}{M}\min_{\mu\in\mathbb{U}}\big\{&\mathcal{R}(\mu) + H(N) + \lambda p + \mu V(N-1,1) \\ &+ \beta_u(\mu)V(N,0) - \varrho + (M - \mu - \beta_u(\mu))V(N,1)\big\}\,, \end{aligned} \tag{3.6}$$

and

$$V(N,0) \;=\; \frac{1}{M}\big(H(N) + C_m + \lambda p + \beta_d V(N,1) - \varrho + (M - \beta_d)V(N,0)\big)\,. \tag{3.7}$$

To simplify the notation, we define the relative cost differences

$$\begin{aligned} &W_1 := V(0,1), \quad W_x := V(x,1) - V(x-1,1), \quad \text{for} \quad 1 \le x \le N, \\ &Y_x := V(x,0) - V(x,1), \quad \text{for} \quad 0 \le x \le N. \end{aligned}$$

By using the relative cost differences, the optimality equations take the form

$$\begin{cases} \lambda W_1 \;=\; -\beta_u(0)Y_0 - H(0) + \varrho\,, \\ \lambda(W_1 + Y_1) \;=\; (\lambda + \beta_d)Y_0 - C_m - H(0) + \varrho\,, \end{cases} \tag{3.8}$$

$$\begin{cases} \lambda W_{x+1} \;=\; \phi\big(W_x, Y_x\big) - H(x) + \varrho\,, \\ \lambda(W_{x+1} + Y_{x+1}) \;=\; (\lambda + \beta_d)Y_x - C_m - H(x) + \varrho\,, \end{cases} \tag{3.9}$$

for $1 \le x \le N-1$, and

$$\begin{cases} \lambda p \;=\; \phi\big(W_N, Y_N\big) - H(N) + \varrho\,, \\ \lambda p \;=\; \beta_d Y_N - C_m - H(N) + \varrho\,, \end{cases} \tag{3.10}$$

where ϕ is defined in (2.4). Note that the system of equations (3.3)–(3.7) is equivalent to that of equations (3.8)–(3.10).

In the following lemma, we provide the recursive formula for the stationary distribution of (X, K) when the system is not full.

Lemma 3.1. *Let $\nu \in \mathfrak{U}_{\mathrm{sm}}$ be such that the stationary distribution of (X, K) exists under ν. Then, for $1 \le x \le N-1$,*

$$\bigl(\pi_\nu(x-1,0), \pi_\nu(x-1,1)\bigr)\, R_\nu(x) \;=\; \bigl(\pi_\nu(x,0), \pi_\nu(x,1)\bigr)\,, \tag{3.11}$$

where

$$R_\nu(x) \;=\; \begin{pmatrix} \frac{\lambda(\beta_u(\nu_x)+\nu_x)}{\nu_x(\lambda+\beta_d)} & \frac{\lambda}{\nu_x} \\ \frac{\lambda\beta_u(\nu_x)}{\nu_x(\lambda+\beta_d)} & \frac{\lambda}{\nu_x} \end{pmatrix}\,, \quad \text{and} \quad \bigl(R_\nu(x)\bigr)^{-1} \;=\; \begin{pmatrix} \frac{\lambda+\beta_d}{\lambda} & -\frac{\beta_u(\nu_x)}{\lambda} \\ -\frac{\lambda+\beta_d}{\lambda} & \frac{\nu_x+\beta_u(\nu_x)}{\lambda} \end{pmatrix}.$$

Proof. It is evident that the balance equations for (X, K) take the form

$$\begin{cases} \bigl(\lambda+\beta_u(0)\bigr)\pi_\nu(0,1) \;=\; \beta_d\pi_\nu(0,0)+\nu_1\pi(1,1)\,, \\ \bigl(\lambda+\beta_d\bigr)\pi_\nu(0,0) \;=\; \beta_u(0)\pi_\nu(0,1)\,, \end{cases} \tag{3.12}$$

and

$$\begin{cases} \bigl(\lambda+\beta_u(\nu_x)+\nu_x\bigr)\pi_\nu(x,1) \;=\; \beta_d\pi_\nu(x,0)+\nu_{x+1}\pi_\nu(x+1,1)+\lambda\pi_\nu(x-1,1)\,, \\ \bigl(\lambda+\beta_d\bigr)\pi_\nu(x,0) \;=\; \beta_u(\nu_x)\pi_\nu(x,1)+\lambda\pi_\nu(x-1,0)\,, \end{cases} \tag{3.13}$$

for $1 \le x \le N-1$. We sum the equations in (3.12) and get

$$\lambda\bigl(\pi_\nu(0,0)+\pi_\nu(0,1)\bigr) \;=\; \nu_1\pi(1,1)\,. \tag{3.14}$$

By applying (3.14) and adding the equations in (3.13), we have that for $1 \le x \le N-1$,

$$\lambda\bigl(\pi_\nu(x,1)+\pi_\nu(x,0)\bigr) = \nu_{x+1}\pi_\nu(x+1,1)\,. \tag{3.15}$$

Thus, by (3.13) and (3.15), we obtain

$$\begin{cases} \lambda\pi_\nu(x-1,0) & = \; \bigl(\lambda+\beta_d\bigr)\pi_\nu(x,0)-\beta_u(\nu_x)\pi_\nu(x,1)\,, \\ \lambda\pi_\nu(x-1,1) & = \; -(\lambda+\beta_d)\pi_\nu(x,0)+\bigl(\beta_u(\nu_x)+\nu_x\bigr)\pi_\nu(x,1)\,, \end{cases}$$

for $1 \le x \le N-1$. Thus, equation (3.11) holds, and this completes the proof. □

By Lemma 3.1, it follows that $\pi_\nu(x,k)$ is continuous with respect to ν (in the total variation norm) for each $1 \le x \le N$ and $k \in \{0,1\}$. It is evident that π_ν, for each $\nu \in \mathfrak{U}_{\mathrm{sm}}$, takes values in a compact set. Then the minimum of (3.2) exists.

In the following verification theorem, we characterize the optimal controls. Its proof relies on the results of Lemma 3.1. Recall that the function ψ represents the minimizer of (2.4).

Theorem 3.1. *Let $\varrho < \infty$ and $(W_1, W_2, \dots, W_N, Y_0, Y_1, \dots, Y_N)$ be a solution to (3.8)–(3.10). Under Assumption 3.1, if $Y_x \ge 0$ for $0 \le x \le N$, $W_x \ge 0$ and $\nu_x^* := \psi(W_x, Y_x) > 0$ for $1 \le x \le N$, then $\{\nu_x^*\colon 1 \le x \le N\}$ is optimal and $\varrho^{\nu^*} = \varrho = \varrho_*$.*

Proof. Let ν be an optimal rate control policy of (3.2). Multiplying both sides of the first equation in (3.9) by $\pi_\nu(x,1)$ and both sides of the second equation in (3.9) by $\pi_\nu(x,0)$, we obtain

$$\begin{cases} \bigl(H(x)+\mathcal{R}(\nu_x)-\varrho\bigr)\pi_\nu(x,1) \;\ge\; W_x\nu_x\pi_\nu(x,1)-Y_x\beta_u(\nu_x)\pi_\nu(x,1)-\lambda W_{x+1}\pi_\nu(x,1)\,, \\ \bigl(H(x)+C_m-\varrho\bigr)\pi_\nu(x,0) \;=\; (\lambda+\beta_d)Y_x\pi_\nu(x,0)-\lambda\bigl(W_{x+1}+Y_{x+1}\bigr)\pi_\nu(x,0)\,, \end{cases} \tag{3.16}$$

for $1 \le x \le N-1$. It follows by (3.11) and (3.15) that

$$\begin{aligned} &W_x\nu_x\pi_\nu(x,1)+Y_x\bigl((\lambda+\beta_d)\pi_\nu(x,0)-\beta_u(\nu_x)\pi_\nu(x,1)\bigr) \\ &\qquad\qquad = \; W_x\lambda\bigl(\pi_\nu(x-1,1)+\pi_\nu(x-1,0)\bigr)+Y_x\lambda\pi_\nu(x-1,0)\,. \end{aligned} \tag{3.17}$$

Then, summing the equations in (3.16) and applying (3.17), we obtain that for $1 \le x \le N-1$,

$$\begin{aligned} &H(x)\bigl(\pi_\nu(x,1)+\pi_\nu(x,0)\bigr)+\mathcal{R}(\nu_x)\pi_\nu(x,1)+C_m\pi_\nu(x,0)-\varrho\bigl(\pi_\nu(x,1)+\pi_\nu(x,0)\bigr) \\ &\qquad\qquad \ge W_x\lambda\bigl(\pi_\nu(x-1,1)+\pi_\nu(x-1,0)\bigr)+Y_x\lambda\pi_\nu(x-1,0) \\ &\qquad\qquad\quad -\bigl(W_{x+1}\lambda(\pi_\nu(x,1)+\pi_\nu(x,0))+Y_{x+1}\lambda\pi_\nu(x,0)\bigr)\,. \end{aligned} \tag{3.18}$$

It follows by the balance equation

$$\begin{cases} \big(\lambda + \beta_u(\nu_N) + \nu_N\big)\pi_\nu(N,1) \;=\; \beta_d \pi_\nu(N,0) + \lambda \pi_\nu(N-1,1)\,, \\ \big(\lambda + \beta_d\big)\pi_\nu(N,0) \;=\; \beta_u(\nu_N)\pi_\nu(N,1) + \lambda \pi_\nu(N-1,0)\,, \end{cases}$$

that

$$\beta_d \pi_\nu(N,0) - \beta_u(\nu_N)\pi_\nu(N,1) \;=\; \lambda \pi_\nu(N-1,0)\,. \tag{3.19}$$

In analogy to (3.18), applying (3.10), (3.15), and (3.19), we obtain

$$\begin{aligned} (\lambda p + H(N) - \varrho)\big(\pi_\nu(N,1) + \pi_\nu(N,0)\big) + \mathcal{R}(\nu_N)\pi_\nu(N,1) + C_m \pi_\nu(N,0) \\ \geq\; W_N \lambda(\pi_\nu(N-1,1) + \pi_\nu(N-1,0)) + Y_N \lambda \pi_\nu(N-1,0)\,. \end{aligned} \tag{3.20}$$

Adding all the equations in (3.18) for $1 \leq x \leq N-1$ together with (3.20), and applying (3.1), we have

$$\begin{aligned} \varrho_* - \varrho - \big(H(0) - \varrho\big)\big(\pi_\nu(0,0) + \pi_\nu(0,1)\big) - C_m \pi_\nu(0,0) \\ \geq\; W_1 \lambda(\pi_\nu(0,1) + \pi_\nu(0,0)) + Y_1 \lambda \pi_\nu(0,0)\,. \end{aligned} \tag{3.21}$$

It follows by (3.12) that

$$(\lambda + \beta_d)\pi_\nu(0,0) - \beta_u(0)\pi_\nu(0,1) \;=\; 0\,. \tag{3.22}$$

Then, by (3.8) and (3.22), we have

$$\lambda W_1 \pi_\nu(0,1) + \lambda(W_1 + Y_1)\pi_\nu(0,0) \;=\; -C_m \pi_\nu(0,0) - \big(H(0) - \varrho\big)\big(\pi_\nu(0,0) + \pi_\nu(0,1)\big)\,. \tag{3.23}$$

Thus, applying (3.21) and (3.23), we obtain $\varrho_* \geq \varrho$. It implies $\varrho_* = \varrho$, and then the proof is completed. □

In the next lemma, we show the existence and uniqueness of the value functions. The result of the lemma is also used in the study of the adaptive control problem in the next section.

Lemma 3.2. *There exists a unique solution $\{V(x,k)\}_{(x,k)\in\mathcal{S}}$ to the set of the optimality equations (3.3)–(3.7).*

Proof. We prove this lemma by contraction. It follows by Theorem 3.1 that $\varrho = \varrho_*$ for any solutions to (3.3)–(3.7). Let $\mathcal{S}$ denote the state space of (X,K). We define the operator $\mathcal{T}_V\colon \mathcal{S} \mapsto \mathbb{R}$ by

$$\mathcal{T}_V(0,1) \;:=\; \frac{1}{M}\big(\mathcal{R}(0) + H(0) + \lambda V(1,1) + \beta_u(0)V(0,0) - \varrho_* + (M - \lambda - \beta_u(0))V(0,1)\big)\,, \tag{3.24}$$

$$\begin{aligned} \mathcal{T}_V(x,1) \;:=\; \frac{1}{M}\Big(H(x) - \varrho_* - \phi\big(V(x,1) - V(x-1,1), V(x,0) - V(x,1)\big) \\ + \lambda V(x+1,1) + (M-\lambda)V(x,1)\Big) \end{aligned} \tag{3.25}$$

for $1 \leq x \leq N-1$,

$$\mathcal{T}_V(N,1) \;:=\; \frac{1}{M}\Big(H(N) - \varrho_* - \phi\big(V(N,1) - V(N-1,1), V(N,0) - V(N,1)\big) + \lambda p\Big)\,, \tag{3.26}$$

$$\mathcal{T}_V(x,0) \;:=\; \frac{1}{M}\big(H(x) + C_m + \lambda V(x+1,0) + \beta_d V(x,1) - \varrho_* + (M - \lambda - \beta_d)V(x,0)\big) \tag{3.27}$$

for $0 \leq x \leq N-1$, and

$$\mathcal{T}_V(N,0) \;:=\; \frac{1}{M}\big(H(N) + C_m + \lambda p + \beta_d V(N,1) - \varrho_* + (M - \beta_d)V(N,0)\big)\,. \tag{3.28}$$

Let V_1 and V_2 be any functions satisfying (3.3)–(3.7). It suffices to show that there exists some positive constant $C < 1$ such that

$$\max_{(x,k)\in\mathcal{S}} |\mathcal{T}_{V_1}(x,k) - \mathcal{T}_{V_2}(x,k)| \leq C \max_{(x,k)\in\mathcal{S}} |V_1(x,k) - V_2(x,k)|\,. \tag{3.29}$$

We show that (3.25) satisfies (3.29). Note that for any functions g and f,

$$|\max_x g(x) - \max_x f(x)| \leq \max_x |g(x) - f(x)| \,.$$

Thus, for any $0 \leq x \leq N-1$,

$$\begin{aligned}
&|\mathcal{T}_{V_1}(x,1) - \mathcal{T}_{V_2}(x,1)| \\
&\leq \frac{1}{M}\Big(\bar{\mu}\big(|V_1(x,1) - V_2(x,1)| + |V_1(x-1,1) - V_2(x-1,1)|\big) \\
&\quad + \beta(\bar{\mu})\big(|V_1(x,1) - V_2(x,1)| + |V_1(x,0) - V_2(x,0)|\big) + \lambda|V_1(x+1,1) - V_2(x+1,1)| \\
&\quad + (M-\lambda)|V_1(x,1) - V_2(x,1)|\Big) \\
&\leq C \max_{(x,k)\in\mathcal{S}} |V_1(x,k) - V_2(x,k)| \,,
\end{aligned}$$

where C is some constant such that

$$\max\left\{\frac{\bar{\mu}}{M}, \frac{\beta_u(\bar{\mu})}{M}, \frac{\lambda}{M}, \frac{M-\lambda}{M}\right\} \leq C < 1 \,.$$

By repeating the procedure described above and applying (3.24) and (3.26)–(3.28), we obtain (3.28). It follows that $\mathcal{T}_V$ is a contraction, and it has a unique fixed point. This completes the proof. □

Remark 3.1. For the service rate control problem of an $M/M/1$ queue with server breakdowns and finite capacity under the discounted cost criterion, one may apply the same approach, which uses the uniformization technique, as in Section 2.2 to obtain the existence and characterization of optimal controls. The properties in Propositions 2.1 and 2.2 also hold for the discounted problem in the model with finite capacity. In this section, we have focused on the service rate control problem under the ergodic cost criterion. We apply the approach by involving the stationary distribution of the joint Markov process (X, K). This approach is different from the vanishing discounted method used for the problem with infinite capacity as described in Section 2.3. Since we consider a queue with finite capacity, that is, the state process has finite states, the stationary distribution for the joint process (X, K) can be expressed as a finite-dimensional vector. Therefore, it is natural to consider the approach with a stationary distribution as shown in the proof of Theorem 3.1.

4. The adaptive control problem

In practice, the function $\beta_u(\cdot)$ in (3.3)–(3.7) may be unobservable from data. At each state x_i, $i = 1, 2, \cdots$, of the state process X, the controller must choose the service rate $\nu_i(x_i)$ for the server. The function $\beta_u(\cdot)$ can be inferred from the history of $\{\nu_i(x_i)\}_{i\in\mathbb{N}}$ and the sojourn times when the system is in the up state.

To simplify the notation, throughout this section, we assume that (2.17) holds with κ_1 and κ_2 taking values in compact sets $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively. We define $\kappa := (\kappa_1, \kappa_2)^{\mathsf{T}}$, and assume that κ is initially unknown and $\mathcal{K}_1, \mathcal{K}_2 \subset \mathbb{R}_+$. We consider a queueing system with finite capacity and the objective function as in Section 3. Note that in Remark 4.1, we provide some approaches to relax the assumption in (2.17).

We let the sequence $\{\nu_i(x_i)\}_{i\in\mathbb{N}}$ with $\nu_i \in \mathfrak{U}_{\mathrm{sm}}$ denote the design variables corresponding to the service rates during the "up" times, that is, when the process $K = 1$. We use $T(\nu_i(x_i))$ to denote the sojourn time of the joint Markov process governed by $\nu_i \in \mathfrak{U}_{\mathrm{sm}}$ when the processes $X = x_i$ and $K = 1$. Given $\{\nu_i(x_i)\}_{i\in\mathbb{N}}$, it is evident that each $T(\nu_i(x_i))$ is exponentially distributed with the parameter $\lambda + \kappa_1 + (\kappa_2 + 1)\nu_i(x_i)$. Let $\{t_i\}_{i\in\mathbb{N}}$ be a sequence of realizations of random variables $\{T(\nu_i(x_i))\}_{i\in\mathbb{N}}$. Then, given a sample of design variables and responses with size n, we let $\hat{\kappa}^n := (\hat{\kappa}^n_1, \hat{\kappa}^n_2)^{\mathsf{T}} = (\kappa^n_1 + \lambda, \kappa^n_2 + 1)^{\mathsf{T}}$ denote a solution of the quasi-likelihood equations taking the form

$$\begin{cases} \sum_{i=1}^{n} \nu_i(x_i)\Big(t_i - \big(\hat{\kappa}_1^n + \hat{\kappa}_2^n \nu_i(x_i)\big)^{-1}\Big) \;=\; 0\,, \\ \sum_{i=1}^{n} \Big(t_i - \big(\hat{\kappa}_1^n + \hat{\kappa}_2^n \nu_i(x_i)\big)^{-1}\Big) \;=\; 0\,. \end{cases} \tag{4.1}$$

The solution of (4.1) may not be unique. However, we may choose the "correct" root with the lowest mean least square error, see, for example, [35, Chapter 13.3]. Such a root of (4.1) is a quasi maximum likelihood estimate of $\hat{\kappa} := (\kappa_1 + \lambda, \kappa_2 + 1)^{\mathsf{T}}$. Given the true parameter $\kappa = (\kappa_1, \kappa_2)^{\mathsf{T}}$, we define the error terms $\{\varepsilon_i\}_{i\in\mathbb{N}}$ by

$$\varepsilon_i \;:=\; t_i - \big(\lambda + \kappa_1 + (\kappa_2 + 1)\nu_i(x_i)\big)^{-1}\,.$$

Note that there is no decision to make during the down times, and if decisions are made based on current parameter estimations during the up times, then $\{\varepsilon_i\}_{i\in\mathbb{N}}$ forms a martingale with respect to its natural filtration. Therefore, we consider the adaptive design case in this section.

It is well known that the estimations from (4.1) may not be consistent (convergence in probability of the estimators to the true parameter values). Therefore, we provide a family of rate control policies under which the estimations from (4.1) are consistent. First, we present a sufficient condition for the strong consistency of $\hat{\kappa}^n$, that is, the almost sure convergence of the estimators to the true parameter values.

Let $\underline{\zeta}^n$ and $\bar{\zeta}^n$ be the smallest and the largest eigenvalues of the design matrix

$$\sum_{i=1}^{n} (1, \nu_i(x_i))^{\mathsf{T}} (1, \nu_i(x_i)) \;=\; \begin{bmatrix} n & \sum_{i=1}^{n} \nu_i(x_i) \\ \sum_{i=1}^{n} \nu_i(x_i) & \sum_{i=1}^{n} \big(\nu_i(x_i)\big)^2 \end{bmatrix}\,,$$

respectively. The following lemma directly follows by Theorem 2.1 of [23].

Lemma 4.1. *Assume that* $\underline{\zeta}^n \to \infty$ *as* $n \to \infty$ *a.s., and*

$$\liminf_{n\to\infty} \frac{\underline{\zeta}^n}{(\bar{\zeta}^n \log \bar{\zeta}^n)^{1/2} (\log\log \bar{\zeta}^n)^{1/2+\delta}} > 0 \quad a.s.$$

for some $\delta > 0$. *Then,* (4.1) *has a solution* $\hat{\kappa}^n$ *such that*

$$|\hat{\kappa}^n - \hat{\kappa}| = \mathfrak{o}\left(\frac{(\bar{\zeta}^n \log \bar{\zeta}^n)^{1/2} (\log\log \bar{\zeta}^n)^{1/2+\delta}}{\underline{\zeta}^n} \right) \quad a.s. \tag{4.2}$$

In the following lemma, we verify the conditions in Lemma 4.1 for a class of rate control policies, and then show that there exists a sequence of estimators from (4.1) that are strongly consistent.

Lemma 4.2. *Under any sequence of work-conserving Markov rate control policies, the conditions in Lemma 4.1 are satisfied. Moreover,* (4.1) *has a solution* $\hat{\kappa}^n$ *such that*

$$|\hat{\kappa}^n - \hat{\kappa}| = \mathfrak{o}\left(\frac{(\log n)^{1/2} (\log\log n)^{1/2+\delta}}{\sqrt{n}} \right) \quad a.s. \tag{4.3}$$

for any $\delta > 0$.

Proof. Since $\{\nu_i(x_i)\}_{i\in\mathbb{N}}$ is uniformly bounded, it is evident that $\bar{\zeta}^n = \mathcal{O}(n)$. We show the lower bound of $\underline{\zeta}^n$ for all large n. Recall that for any constant rate policy $\nu \equiv \mu \in \mathbb{U}$, the Markov process (X, K) is ergodic. Since the service rate for work-conserving Markov policy is bounded away from 0 by assumption, it follows that the mean return time of the embedded Markov chain of (X, K) to the state $(0, 1)$ is uniformly bounded under any sequence of work-conserving Markov policies. Recall that $\{\nu_i(x_i)\}_{i\in\mathbb{N}}$ denote the design variables during the up times. Let $\bar{\nu}^n := n^{-1} \sum_{i=1}^{n} \nu_i(x_i)$ and $\underline{\nu}$ denote the policy satisfying $\underline{\nu}(0, 1) = 0$ and

$$\underline{\nu}(x, 1) \;\equiv\; \underline{\nu} \;=\; \arg\min_{\mu\in\mathbb{U}} \left\{ \frac{\mu^2 \big(1 - \pi_\mu(0, 1)\big)^2 \pi_\mu(0, 1)}{\big(\sum_{j=0}^{N} \pi_\mu(j, 1)\big)^3} \right\} \quad \forall x \in \mathbb{N}\,. \tag{4.4}$$

The minimum of (4.4) can be attained because $\{\pi_\mu(x,k)\colon (x,k)\in\mathbb{Z}_+\times\{0,1\}\}$ are continuous functions of μ on $\mathbb{U}$ by Lemma 3.1, (3.12) and (3.19). Let $\mathbb{1}(\cdot)$ denote the indicator function. Thus, based on the ergodic theory,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bigl(\nu_i(x_i)-\bar{\nu}^n\bigr)^2 \;&\ge\; \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bigl(\nu_i(x_i)-\bar{\nu}^n\bigr)^2\mathbb{1}(x_i=0)\\ &=\;\liminf_{n\to\infty}(\bar{\nu}^n)^2\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(x_i=0)\;\ge\;\frac{\underline{\nu}^2\bigl(1-\pi_{\underline{\nu}}(0,1)\bigr)^2\pi_{\underline{\nu}}(0,1)}{\bigl(\sum_{j=0}^{N}\pi_{\underline{\nu}}(j,1)\bigr)^3}\end{aligned}\tag{4.5}$$

a.s., where the last inequality follows by (4.4) and the fact $\nu_i(0)=0$ for any i. Applying Lemma 2 of [36], we have

$$\underline{\zeta}^n\;\ge\;C_1\sum_{i=1}^{n}\bigl(\nu_i(x_i)-\bar{\nu}^n\bigr)^2\tag{4.6}$$

where $C_1={}^{2}/{(1+2\underline{\mu}-\bar{\mu})^2}$. Then, by (4.5) and (4.6), it follows that

$$\underline{\zeta}^n\;\ge\;C_2(1+n)$$

for some positive constant C_2. Therefore, both $\underline{\zeta}^n$ and $\bar{\zeta}^n$ are at the order of n, and we have verified the conditions in Lemma 4.1. It follows by (4.2) that (4.3) holds, and this completes the proof. □

Let $\hat{\varrho}^n$ be the optimal value of (3.2) with the parameters κ_1 and κ_2 replaced by $\hat{\kappa}_1^n-\lambda$ and $\hat{\kappa}_2^n-1$, respectively. Recall that ϱ_* denotes the true optimal ergodic cost in (3.2). In the following lemma, we show the convergence of optimal values.

Lemma 4.3. *If $\hat{\kappa}^n\to\hat{\kappa}$ as $n\to\infty$ a.s., then $\hat{\varrho}^n\to\varrho_*$ as $n\to\infty$ a.s.*

The result of Lemma 4.3 directly follows by the expression in (3.1) together with the Lemma 3.1, and applying Theorem 2.3 of [3], and we omit its proof.

Lemma 4.4. *Assume that (2.17) holds and $\mathcal{R}(\cdot)$ is strongly convex and continuously differentiable. Let ν^i be the optimal control obtained by solving (3.8)–(3.10) under the estimator $\hat{\kappa}^i$. Then, it follows that*

$$\nu^i(x,1)\;\to\;\nu^*(x,1)\qquad a.s.\tag{4.7}$$

as $i\to\infty$, for $1\le x\le N$.

Proof. It is evident that $\phi(w,y)$ is a continuous function of w and y. For any (w,y), $\phi(w,y)$ is also continuous with respect to the parameter κ. Let $\{W_x^i,Y_x^i\colon 1\le i\le N\}$ be the solution of equations (3.8)–(3.10) under the estimate $\hat{\kappa}^i$. Let $\hat{\varrho}^i$ denote the optimal value in (3.2) under the estimate $\hat{\kappa}^i$. Note that $\{W_x,Y_x\colon 0\le x\le N\}$ are continuous functions of κ and ϱ. Since $\hat{\kappa}^i\to\hat{\kappa}$ and $\hat{\varrho}^i\to\varrho_*$ a.s. as $i\to\infty$ by Lemma 4.3, then it follows by the continuous mapping theorem that

$$W_x^i\to W_x^*\,,\quad\text{and}\quad Y_x^i\to Y_x^*\quad a.s.$$

as $i\to\infty$ for $1\le x\le N$. By the strong convexity and continuity of ϕ, the convergence of optimal values $\phi(W_x^i,Y_x^i)$, $i\in\mathbb{N}$, implies the convergence of maximizers of $\{\phi(W_x^i,Y_x^i)\colon i\in\mathbb{N}\}$. We refer the readers to (2.18) for the representation of maximizers. We have shown (4.7). □

For the queue with a finite capacity, the dynamics in (2.1) becomes

$$X(t)\;=\;\biggl(X(0)+A(t)-S\biggl(\int_0^t\nu\bigl(X(s),K(s)\bigr)(X(s)\wedge1)\,\mathrm{d}s\biggr)\biggr)\wedge N\qquad\forall\,t\ge0\,.$$

Recall the cost function $f(x,k,\mu)$ in (2.2). To simplify the notation, we let $f_\nu(x,k) = f(x,k,\nu(x,k))$ for $\nu \in \mathfrak{U}_{\mathrm{sm}}$. Let m_t denote the number of jumps for the process (X,K) before time t, $\tau_0 = 0$, and τ_i, $1 \le i \le m_t$, denote the i-th jump time of (X,K). We define the cumulative cost function F by

$$\begin{aligned} F(t) \;:=\; &\sum_{i=1}^{m_t} \int_{\tau_{i-1}}^{\tau_i} f_{\nu^{i-1}}\big(X(s),K(s)\big)\,\mathrm{d}s + \int_{\tau_{m_t}}^{t} f_{\nu^{m_t}}\big(X(s),K(s)\big)\,\mathrm{d}s \\ &+ p\sum_{i=1}^{m_t}\big(A(\tau_i)-A(\tau_{i-1})\big)\int_{\tau_{i-1}}^{\tau_i} \mathbb{1}(X(s)=N)\,\mathrm{d}s \\ &+ p\big(A(t)-A(\tau_{m_t})\big)\int_{\tau_{m_t}}^{t} \mathbb{1}(X(s)=N)\,\mathrm{d}s \end{aligned} \tag{4.8}$$

for $t \ge 0$, where ν^i denotes the policy updated by solving (3.8)–(3.10) under the estimator $\hat{\kappa}^i$, and third and fourth terms on the left hand side (LHS) correspond to the penalty of rejections.

In the next theorem, we present the main result of this section. The theorem implies that if we estimate the unknown parameters κ_1 and κ_2 under work-conserving rate controls at each state and update the rate controls by solving (3.8)–(3.10) and under estimated parameters, then the long-run average cost converges to the optimal cost. Because the transition rate matrix is updated over time due to the change of parameters, the joint Markov process is time-varying, and the proof of the theorem relies on Kruglov strong law of large numbers; see Theorem 2 in [39].

Theorem 4.1. *Assume that (4.7) holds. Then,*

$$\lim_{t\to\infty} \frac{1}{t}\,\mathbb{E}\big[F(t)\big] \;=\; \varrho_*. \tag{4.9}$$

Proof. It is evident that

$$\frac{1}{t}\int_0^t f_{\mu(s)}(X(s),K(s))\,\mathrm{d}s \;=\; \sum_{x=0}^{N}\sum_{k=0}^{1}\frac{1}{t}\int_0^t \mathbb{1}(X(s)=x,K(s)=k) f_{\mu(s)}(x,k)\,\mathrm{d}s\,, \tag{4.10}$$

where $\mu(s) = \nu^i(X(s),K(s))$ for $\tau_i \le s < \tau_{i+1}$. For $(x,k) \in \mathcal{S}$, let $\tau_n(x,k)$ denote the n-th time at which the Markov process jumps into state (x,k) with $\tau_0(x,k) = 0$, and let $h_n(x,k)$ denote the n-th holding time in the state (x,k). Define $T_n(x,k) := \tau_n(x,k) - \tau_{n-1}(x,k)$ for $n \ge 1$. We use $N_{x,k}(t)$ to denote the number of transitions of the Markov process into state (x,k) before time t. Then, we have

$$\begin{aligned} &\int_0^t \mathbb{1}(X(s)=x,K(s)=k) f_{\mu(s)}(x,k)\,\mathrm{d}s \\ &\quad= \sum_{i=1}^{N_{x,k}(t)-1} h_i(x,k) f_{\mu_i}(x,k) + \big((t-\tau_{N_{x,k}(t)}) \wedge h_{N_{x,k}(t)}(x,k)\big) f_{\mu_{N_{x,k}(t)}}(x,k)\,, \end{aligned} \tag{4.11}$$

where μ_i, for $i \in \mathbb{N}$, denotes the service rate during the i-th holding time in the state (x,k). Applying the strong Markov property, given $\{\nu^i \colon i \in \mathbb{N}\}$, $\{h_i(x,k) \colon i \in \mathbb{N}\}$ are independent. Let $h_*(x,k)$ denote the holding time in the state (x,k) of the Markov process under the optimal service rate control ν_*. For $k = 0$, we have $\mu_i \equiv 0$ for $i \in \mathbb{N}$, and then $h_i(x,0)$ are i.i.d., distributed as $h_*(x,0)$. For $k = 1$, applying (4.7), it follows that

$$\lim_{i\to\infty} \mathbb{E}[h_i(x,1) f_{\mu_i}(x,k)] \;=\; \mathbb{E}[h_*(x,1) f_{\nu_*}(x,k)]\,. \tag{4.12}$$

Since the service rate is bounded, it is straightforward to check that

$$\sup_{i\in\mathbb{N}} \mathbb{E}|h_i(x,k) f_{\mu_i}(x,k)| \;<\; \infty\,,$$

and

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(|h_i(x,k)f_{\mu_i}(x,k)| > y) \le \mathbb{P}(|f_0(1,0)\Psi| > y)$$

for all $y \ge 0$, where Ψ denotes a random variable having exponential distribution with parameter λ. Thus, by Kruglov strong law of large numbers, conditioning on $\{\nu^i : i \in \mathbb{N}\}$, we have

$$\frac{1}{n}\sum_{i=1}^{n} h_i(x,k)f_{\mu_i}(x,k) \to \mathbb{E}_{\nu_*}[h_*(x,k)f_{\nu_*}(x,k)] \quad \text{a.s.} \tag{4.13}$$

as $n \to \infty$. For the rejection cost in (4.8), given the state process $X = N$, the holding time $h_i(N,k)$ is independent of the arrival process A, and then

$$\lim_{i\to\infty} \mathbb{E}[A(h_i(N,k))] = \lim_{i\to\infty} \lambda\,\mathbb{E}[h_i(N,k)] = \lambda\,\mathbb{E}[h_*(N,k)]\,. \tag{4.14}$$

The similar result in (4.13) holds for the rejection cost. By repeating the procedure as above, we have that $\{T_i(x,k) : i \in \mathbb{N}\}$ are independent conditioning on $\{\nu^i : i \in \mathbb{N}\}$, and

$$\lim_{i\to\infty} \mathbb{E}[T_i(x,k)] = \mathbb{E}[T_*(x,k)]\,,$$

where $T_*(x,k)$ denotes the return time to the state (x,k) of the Markov process under the optimal service rate control policy ν_*. By applying Kruglov strong law of large numbers and repeating the proof for the elementary renewal theory, it follows that given $\{\nu^i : i \in \mathbb{N}\}$,

$$\frac{N_{x,k}(t)}{t} \to \frac{1}{\mathbb{E}_{\nu_*}[T_*(x,k)]} \quad \text{a.s.} \tag{4.15}$$

as $t \to \infty$. Thus, it follows by (4.13) and (4.15) that given $\{\nu^i : i \in \mathbb{N}\}$,

$$\begin{aligned}\frac{1}{t}\sum_{i=1}^{N_{x,k}(t)-1} h_i(x,k)f_{\mu_i}(x,k) &= \frac{N_{x,k}(t)-1}{t}\cdot\frac{1}{N_{x,k}(t)-1}\sum_{i=1}^{N_{x,k}(t)-1} h_i(x,k)f_{\mu_i}(x,k)\\ &\to \frac{\mathbb{E}_{\nu_*}[h_*(x,k)]f_{\nu_*}(x,k)}{\mathbb{E}_{\nu_*}[T_*(x,k)]} \quad \text{a.s.}\end{aligned} \tag{4.16}$$

as $t \to \infty$. Note that $\pi_{\nu_*}(x,k) = \frac{\mathbb{E}_{\nu_*}[h_*(x,k)]}{\mathbb{E}_{\nu_*}[T_*(x,k)]}$. Since the service rate is bounded, we have

$$\sup_{i\in\mathbb{N}} \frac{\mathbb{E}[h_i(x,k)f_{\mu_i}(x,k)]}{t} \to 0\,. \tag{4.17}$$

Then, by (4.11), (4.16), and (4.17), and applying the dominated convergence theorem, we have

$$\lim_{t\to\infty}\frac{1}{t}\,\mathbb{E}\left[\mathbb{E}\left[\int_0^t \mathbb{1}(X(s)=x, K(s)=k) f_{\mu(s)}(x,k)\,\mathrm{d}s \,\middle|\, \{\nu^i : i\in\mathbb{N}\}\right]\right] = \pi_{\nu_*}(x,k)f_{\nu_*}(x,k)\,. \tag{4.18}$$

Similarly, for the rejection cost in (4.8), by using (4.14), we obtain

$$\lim_{t\to\infty}\frac{1}{t}\,\mathbb{E}\left[p\sum_{i=1}^{m_t}\big(A(\tau_i)-A(\tau_{i-1})\big)\int_{\tau_{i-1}}^{\tau_i}\mathbb{1}(X(s)=N)\,\mathrm{d}s\right] = p\lambda\big(\pi_{\nu_*}(N,1)+\pi_{\nu_*}(N,0)\big)\,. \tag{4.19}$$

Note that the expectation of the fourth term on the LHS of (4.8) is bounded. Therefore, by using (4.8), (4.10), (4.18), and (4.19), we have shown (4.9). This completes the proof. □

Remark 4.1. To extend the results to the problem with a nonlinear relationship between the breakdown rate and service rate, one may apply the same analysis by replacing (2.17) with a polynomial function, where the coefficients may be initially unknown. One may change the likelihood equations in (4.1) accordingly and use the generalized least square estimation in [23]. On the other hand, instead of assuming the functional form of the relationship between the breakdown and service rates, one may treat the breakdown rate as a general function of the service rate. In this case, some non-parametetric approaches in the study of online problems for inventory models may be adopted to study the service rate control problem; see, for example, [18, 19, 53] and references therein.

5. Numerical examples

In this section, we show the numerical results for the queueing system with finite capacity as in Section 3 and the adaptive control problem as in Section 4. In Section 5.1, we provide the results for the optimal service rate controls under different parameters of the system dynamics and the cost functions. In Section 5.2, we present the simulation study for the convergence of regret under adaptive controls.

We first determine the cost functions and the queueing system for the numerical study. The holding/delay cost function in all examples satisfies $H(x) = C_h x^2$ for $x \geq 0$, where C_h is a positive constant. We set the effort cost function $\mathcal{R}(\mu) = C_r \mu^2$ for $\mu \in \{0\} \cup \mathbb{U}$, where C_r is a positive constant and $\mathbb{U} = [\underline{\mu}, \bar{\mu}]$ with $\underline{\mu}, \bar{\mu} > 0$. We consider the cases in which the number of jobs in the system is truncated at N. When there are N jobs in the system, new arrivals are rejected with the cost p for a single job. Recall that the arrival rate is denoted by λ and the breakdown rate is assumed to be a linear function of the service rate satisfying $\beta_u(\mu) = \kappa_1 + \kappa_2 \mu$ for $\mu \in \{0\} \cup \mathbb{U}$. β_d is used to represent the maintenance rate and is a positive constant. The maintenance cost is C_m per unit of time.

5.1. **The optimal service rate controls.** The parameters for the system dynamics are listed in Table 1. We consider a scenario under different cost settings and compare the results of the optimal service rate controls. As mentioned in Remark 2.3, the ratio $\beta_d/(\beta_d + \beta_u(\mu))$ represents the proportion of up times on average for the system under the service rate μ. The proportion of up times approximately ranges from 73.37% to 89.05% on average. As a result, the range of the effective service rate becomes $[3.56, 8.80]$.

In this scenario, we use days as the time unit for the parameters. Then, the server is likely to break down every 12 days on average if the server runs at its lowest rate, while the server is likely to break down every 4 days on average if the server runs at its highest rate. It takes about 1.5 days to repair the server on average. The setting of breakdown and maintenance rates is very close to the real data from [17], where the server corresponds to a coal unloader and the jobs correspond to trainloads waiting to be unloaded. Here we assume that the unloading rate is adjustable and the breakdown rate depends on the unloading rate.

Table 1. Parameter combinations for the system dynamics.

λ	$\underline{\mu}$	$\bar{\mu}$	β_d	κ_1	κ_2	N
8	4	12	$2/3$	$1/500$	$1/50$	25

In Table 2, we provide the parameters of the cost functions in the scenario. In total, there are 15 parameter combinations for the numerical study of the optimal service rate controls. As shown in Table 2, the cost parameters taking values in the set $\{1, 10\}$ are permuted to show the impact of costs on the optimal policy.

Table 2. Cost parameter settings.

Settings	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C_r (dollar/hour)	1	10	1	1	1	10	10	10	1	1	1	10	10	10	1
C_h (dollar/hour)	1	1	10	1	1	10	1	1	10	10	1	10	10	1	10
C_m (dollar/hour)	1	1	1	10	1	1	10	1	10	1	10	10	1	10	10
p (dollar/hour)	1	1	1	1	10	1	1	10	1	10	10	1	10	10	10

TABLE 3. System performances.

Settings	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Service Utilization(%)	67.5	89.3	60.8	66.8	66.6	66.6	88.5	89.5	61.0	60.8	67.2	66.7	67.3	88.4	60.6
Rejection Rate(%)	8.0	55.6	8.5	9.3	8.4	8.7	55.7	54.6	9.2	8.3	8.4	9.0	7.6	54.8	8.3

The rejection rate is the ratio between number of rejected jobs and the total number of arrivals. The rejection rate and the service utilization from implementing the optimal policy are listed in the Table 3, which are obtained via simulating 500,000 events.

For every set of parameters, we compute the optimal service rate policy by solving (3.8)–(3.10) Figure 1 shows the optimal policies for the scenario, where the x-axis represents the number of jobs in the system and the y-axis corresponds to the service rate under the optimal policy. We find that the optimal policies may not be monotone in the number of jobs in the system. Specifically, under the cost settings $2, 7, 8$ and 14, the holding cost parameters are relatively low whereas the effort cost parameter is relatively high. In these cost parameter settings, Figure 1 shows that the optimal service rate policies are non-monotone. For all the scenarios under the cost settings 3, 9 and 15, since the effort cost parameters are set as "low "while the holding cost parameters are set as "high", the optimal service rates are chosen at the highest value when the number of jobs in the system is large.

In addition, the rejection rates in Table 3 indicate that when the effort cost is high, the system tends to have a high rejection rate (see, for example, cost settings 2, 7, 8, and 14). This is reasonable because the optimal control tends to run the server at a lower service rate when the effort cost is high (see, for example, controls under cost settings 2, 7, 8, and 14 in Figure 1), and the system is likely to be at a high congestion level. Therefore, it is more likely to observe rejections. When the holding cost increases, the rejection rate decreases even when the effort cost is high (see, for example, cost settings 6, 12, and 13). This is justifiable as an increased holding cost could result in optimal control policies that prevent the system from entering a high congestion level.

However, increasing the rejection cost when the effort cost is high does not have a significant impact on the rejection rate (see, for instance, cost settings 8 and 14). From the optimal controls under cost settings 2 and 8 in Figure 1, we can see that an increased rejection penalty does not change the policy significantly when the effort cost is high. Therefore, the system has a high likelihood of entering a high congestion level. Hence, the rejection rate remains high.

In summary, when the effort cost is high, the system controller may choose a lower service rate to decrease the likelihood of unplanned downtime even when the system is near its capacity limit, which results in a non-monotone control policy. When the holding cost is high, the controller may run the server at a relatively high service rate to avoid congestion despite the risk of encountering breakdowns.

5.2. **The adaptive service rate control problem.** We first introduce the procedure used to solve the adaptive service rate control problem. Recall that the parameters κ_1 and κ_2 are initially unknown and take values in compact sets. At the beginning of the implementation, we randomly choose the initial parameters $\hat{\kappa}_1^0$ and $\hat{\kappa}_2^0$ from the sets $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively. In the numerical study, we assume $\beta_u(\mu) \leq \beta_d$ for $\underline{\mu} \leq \mu \leq \bar{\mu}$, and $\mathcal{K}_1$ and $\mathcal{K}_2$ are constructed to ensure that the inequality to hold.

We estimate the parameters for every 50 jumps of the process (X, K). The cycle of each estimation is identified as an iteration. At each iteration, we simulate the system under the control policy which is obtained by solving equations (3.8)–(3.10). During the calculation, we use the estimation of parameters κ_1 and κ_2 in the equations. The estimation of parameters is updated based on the solution of (4.1) in which we use the data collected from simulations. We project the estimated

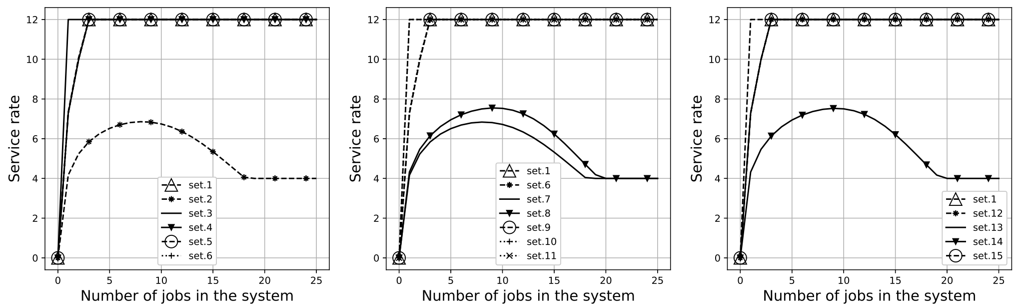

Figure 1. Optimal control policies under different cost settings.

parameters to the boundaries of $\mathcal{K}_1 \times \mathcal{K}_2$ if the estimation lies outside the domain $\mathcal{K}_1 \times \mathcal{K}_2$. We simulate the process (X, K) under the adaptive service rate controls in a finite time horizon. The performance of the algorithm is measured by the average regret

$$R(n) := \frac{1}{t_n} \mathbb{E}\big[F(t_n)\big] - \rho^* \tag{5.1}$$

for positive integer $n \leq L$, where the cost $F(t)$ is defined in (4.8) and L denotes the number of timestamps in the simulation study.

We set L as 1000. The parameter t_L and the sets $\mathcal{K}_1$ and $\mathcal{K}_2$ for each setting are shown in Table 4. In each setting, t_L is chosen to be sufficiently large such that the average regret is near zero when n is large. In Figure 2, the x-axis represents the timestamps in the simulations with the difference between the timestamps equal to t_L/L, and the y-axis corresponds to the average regret at each timestamp. In each case, the expectation is approximated by the average over the values of 300 trajectories. We conduct experiments for the adaptive service rate control problem under the cost parameter settings 13 and 14 for the scenario shown in Table 1. As shown in Table 3, the rejection rate under cost setting 14 is significantly larger than the one under cost setting 13. The experiments under the other cost parameter settings are similar. Here we focus on two sets of cost parameters for simplicity. As shown in Figure 2, the average regrets converge to 0. This verifies the theoretical results in Theorem 4.1.

Table 4. The sets $\mathcal{K}_1$ and $\mathcal{K}_2$ and the parameter t_L.

Cost Setting	$\mathcal{K}_1$	$\mathcal{K}_2$	t_L
13	$[10^{-5}, 0.665]$	$[10^{-5}, 0.055]$	2011
14	$[10^{-5}, 0.665]$	$[10^{-5}, 0.055]$	2357

Acknowledgements

This work was supported in part by the US National Science Foundation grants DMS-1715875 and DMS-2108683/2216765, and Army Research Office grant W911NF-17-1-0019. The authors are grateful for the helpful comments from the associate editors and reviewers that have improved the exposition of the results in the paper.

References

[1] J. Abernethy, P. L Bartlett, and E. Hazan, *Blackwell approachability and no-regret learning are equivalent*, Proceedings of the 24th Annual Conference on Learning Theory, 2011, pp. 27–46.

[2] K. M. Adusumilli and J. J. Hasenbein, *Dynamic admission and service rate control of a queue*, Queueing Systems **66** (2010), no. 2, 131–154.

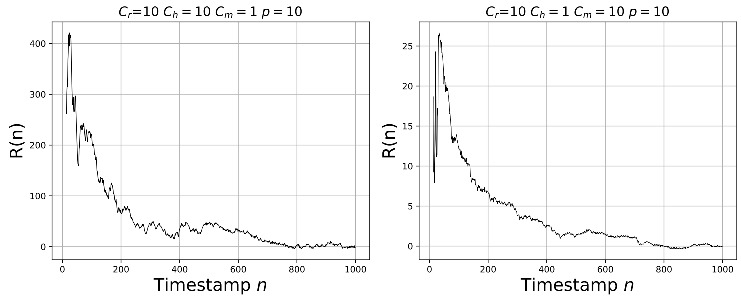

FIGURE 2. The average regret $R(n)$ under cost settings 13 and 14.

[3] J. Alvarez-Mena and O. Hernández-Lerma, *Convergence and approximation of optimization problems*, SIAM J. Optim. **15** (2004/05), no. 2, 527–539. MR2144179

[4] A. Arapostathis, G. Pang, and Y. Zheng, *Optimal scheduling of critically loaded multiclass GI/M/n+M queues in an alternating renewal environment*, Applied Mathematics and Optimization **84** (2021), no. 2, 1857–1901.

[5] B. Ata, *Dynamic power control in a wireless static channel subject to a quality-of-service constraint*, Oper. Res. **53** (2005), no. 5, 842–851. MR2171655

[6] B. Ata and S. Shneorson, *Dynamic control of an $M/M/1$ service system with adjustable arrival and service rates*, Management Science **52** (2006), no. 11, 1778–1791.

[7] B. Avi-Itzhak and P. Naor, *Some queuing problems with the service station subject to breakdown*, Operations Res. **11** (1963), 303–320. MR162293

[8] P. Badian-Pessot, M. E. Lewis, and D. G. Down, *Optimal control policies for an $M/M/1$ queue with a removable server and dynamic service rates*, Probab. Engrg. Inform. Sci. **35** (2021), no. 2, 189–209. MR4236494

[9] O. Besbes and A. Zeevi, *On the (surprising) sufficiency of linear models for dynamic pricing with demand learning*, Management Science **61** (2015), no. 4, 723–739.

[10] O. Besbes and A. Zeevi, *Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms*, Oper. Res. **57** (2009), no. 6, 1407–1420. MR2597918

[11] D. Blackwell, *An analog of the minimax theorem for vector payoffs*, Pacific Journal of Mathematics **6** (1956), no. 1, 1–8.

[12] V. Borkar and P. Varaiya, *Adaptive control of Markov chains. I. Finite parameter set*, IEEE Trans. Automat. Control **24** (1979), no. 6, 953–957. MR566454

[13] V. Borkar, *Recursive self-tuning control of finite markov chains*, Applicationes Mathematicae **24** (1997), no. 2, 169–188.

[14] V. Borkar and P. Varaiya, *Identification and adaptive control of Markov chains*, SIAM J. Control Optim. **20** (1982), no. 4, 470–489. MR661027

[15] V. Borkar, *The Kumar-Becker-Lin scheme revisited*, Journal of Optimization Theory and Applications **66** (1990), 289–309.

[16] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*, Cambridge university press, 2006.

[17] K. Chelst, A. Z. Tilles, and J. S. Pipis, *A coal unloader: A finite queueing system with breakdowns*, INFORMS Journal on Applied Analytics **11** (1981), no. 5, 12–25.

[18] B. Chen, X. Chao, and C. Shi, *Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand*, Math. Oper. Res. **46** (2021), no. 2, 726–756.

[19] Q. Chen, S. Jasin, and I. Duenyas, *Nonparametric self-adjusting control for joint learning and optimization of multiproduct pricing with finite resource capacity*, Math. Oper. Res. **44** (2019), no. 2, 601–631. MR3959086

[20] X. Chen, Yunan. Liu, and G. Hong, *An online learning approach to dynamic pricing and capacity sizing in service systems*, Operations Research **Forthcoming** (2023). https://doi.org/10.1287/opre.2020.612.

[21] R. H Crites and A. G Barto, *Improving elevator performance using reinforcement learning*, Advances in Neural Information Processing Systems (1996), 1017–1023.

[22] A. den Boer and B. Zwart, *Simultaneously learning and optimizing using controlled variance pricing*, Management Science **60** (2014), no. 3, 770–783.

[23] A. V. den Boer and B. Zwart, *Mean square convergence rates for maximum quasi-likelihood estimators*, Stoch. Syst. **4** (2014), no. 2, 375–403. MR3353222

[24] E. Drabik and Ł. Stettner, *On adaptive control of markov chains using nonparametric estimation*, Applicationes Mathematicae **27** (2000), no. 2, 143–152.

[25] T. Duncan, B Pasik-Duncan, and L Stettner, *Adaptive control of a partially observed discrete time markov process*, Applied Mathematics and Optimization **37** (1998), 269–293.

[26] A. Federgruen and K. C. So, *Optimal time to repair a broken server*, Adv. in Appl. Probab. **21** (1989), no. 2, 376–397. MR997729

[27] E. A. Feinberg and M. E. Lewis, *Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem*, Math. Oper. Res. **32** (2007), no. 4, 769–783. MR2363196

[28] D. P. Gaver Jr., *A waiting line with interrupted service, including priorities*, J. Roy. Statist. Soc. Ser. B **24** (1962), 73–90. MR141173

[29] J. M. George and J. M. Harrison, *Dynamic control of a queue with adjustable service rate*, Oper. Res. **49** (2001), no. 5, 720–731. MR1860424

[30] A. P Ghosh and A. P Weerasinghe, *Optimal buffer size and dynamic rate control for a queueing system with impatient customers in heavy traffic*, Stochastic Processes and Their Applications **120** (2010), no. 11, 2103–2141.

[31] M. P. Groover, *Fundamentals of modern manufacturing: materials, processes, and systems*, John Wiley & Sons., 2010.

[32] Q.-M. He, *Fundamentals of matrix-analytic methods*, Springer, New York, 2014. MR3112230

[33] O. Hernández-Lerma and S. I Marcus, *Adaptive control of service in queueing systems*, Systems & Control Letters **3** (1983), no. 5, 283–289.

[34] O. Hernàndez-Lerma and S. I Marcus, *Optimal adaptive control of priority assignment in queueing systems*, Systems & Control Letters **4** (1984), no. 2, 65–72.

[35] C. C. Heyde, *Quasi-likelihood and its application*, Springer Series in Statistics, Springer-Verlag, New York, 1997. A general approach to optimal parameter estimation. MR1461808

[36] N. B. Keskin and A. Zeevi, *Dynamic pricing with an unknown demand model: asymptotically optimal semi-myopic policies*, Oper. Res. **62** (2014), no. 5, 1142–1167. MR3269424

[37] Y. L. Koçağa, *An approximating diffusion control problem for dynamic admission and service rate control in a $G/M/N+G$ queue*, Operations Research Letters **45** (2017), no. 6, 538–542.

[38] Y. L. Koçağa and A. R Ward, *Admission control for a multi-server queue with abandonment*, Queueing Systems **65** (2010), 275–323.

[39] V. M. Kruglov, *Strong law of large numbers*, Vol. 3 Stability Problems for Stochastic Models: Proceedings of the Fifteenth Perm Seminar Perm, Russia, June 2–6, 1992, 2020, pp. 139–150.

[40] P. R. Kumar and W. Lin, *Optimal adaptive controllers for unknown Markov chains*, IEEE Trans. Automat. Control **27** (1982), no. 4, 765–774. MR680480

[41] R. Kumar, M. E. Lewis, and H. Topaloglu, *Dynamic service rate control for a single-server queue with Markov-modulated arrivals*, Naval Res. Logist. **60** (2013), no. 8, 661–677. MR3146992

[42] C.-C. Kuo, K.-H. Wang, and S.-L. Lee, *Optimal control of the $\langle p, N\rangle$-policy $M/G/1$ queue with server breakdowns and general startup times*, Internat. J. Inform. Management Sci. **20** (2009), no. 4, 565–577. MR2656192

[43] N. Lee and V. G Kulkarni, *Optimal arrival rate and service rate control of multi-server queues*, Queueing Systems **76** (2014), 37–50.

[44] S. A. Lippman, *Semi-Markov decision processes with unbounded rewards*, Management Sci. **19** (1972/73), 717–731. MR337340

[45] S. A. Lippman, *Applying a new device in the optimization of exponential queuing systems*, Operations Res. **23** (1975), no. 4, 687–710. MR443125

[46] B. Liu, Q. Xie, and E. Modiano, *Reinforcement learning for optimal control of queueing systems*, 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2019, pp. 663–670.

[47] P. Mandl, *On self-optimizing control of Markov processes*, Mathematical Control Theory, 1985, pp. 345–360. MR851236

[48] M. F. Neuts, *Matrix-geometric solutions in stochastic models*, Dover Publications, Inc., New York, 1994. An algorithmic approach, Corrected reprint of the 1981 original. MR1313503

[49] G. Pang and W. Whitt, *Heavy-traffic limits for many-server queues with service interruptions*, Queueing Syst. **61** (2009), no. 2-3, 167–202. MR2485887

[50] G. Pang and W. Whitt, *Service interruptions in large-scale service systems*, Management Science **55** (2009), no. 9, 1499–1512.

[51] M. Raeis, A. Tizghadam, and A. Leon-Garcia, *Reinforcement learning-based admission control in delay-sensitive service systems*, GLOBECOM 2020-2020 IEEE Global Communications Conference, 2020, pp. 1–6.

[52] L. I. Sennott, *Average cost semi-Markov decision processes and the control of queueing systems*, Probability in the Engineering and Informational Sciences **3** (1989), no. 2, 247–272.

[53] C. Shi, W. Chen, and I. Duenyas, *Technical note—nonparametric data-driven algorithms for multiproduct inventory systems with censored demand*, Oper. Res. **64** (2016), no. 2, 362–370. MR3500609

[54] A. Shwartz and A. M Makowski, *An optimal adaptive scheme for two competing queues with constraints*, Analysis and Optimization of Systems: Proceedings of the Seventh International Conference on Analysis and Optimization of Systems, Antibes, June 25-27, 1986, pp. 515–532.

[55] Ł. Stettner, *On nearly self-optimizing strategies for a discrete-time uniformly ergodic adaptive model*, Applied Mathematics and Optimization **27** (1993), 161–177.

[56] N. Walton and K. Xu, *Learning and information in stochastic networks and queues*, INFORMS TutORials in Operations Research (2021), 161–198.

[57] A. Weerasinghe, *Diffusion approximations for $G/M/n + GI$ queues with state-dependent service rates*, Mathematics of Operations Research **39** (2014), no. 1, 207–228.

[58] H. White and L. S. Christie, *Queuing with preemptive priorities or with breakdown*, Operations Res. **6** (1958), 79–95. MR92283

[59] L. Yeh, *A finite algorithm for ϵ-optimal solutions of adaptive queueing control*, Journal of Mathematical Analysis and Applications **125** (1987), no. 1, 218–233.