KYBERNETIKA — VOLUME 57 (2021), NUMBER 3, PAGES 493–512

PARTIALLY OBSERVABLE QUEUEING SYSTEMS WITH CONTROLLED SERVICE RATES UNDER A DISCOUNTED OPTIMALITY CRITERION

Yofre H. García, Saul Diaz-Infante and J. Adolfo Minjárez-Sosa

We are concerned with a class of $GI/GI/1$ queueing systems with controlled service rates, in which the waiting times are only observed when they take zero value. Applying a suitable filtering process, we show the existence of optimal control policies under a discounted optimality criterion.

Keywords: queueing models, partially observable systems, discounted criterion, optimal policies

Classification: 90C39, 90B22

1. INTRODUCTION

In this paper we consider $GI/GI/1$ queueing systems with controlled service rate and partially observable (PO) waiting times, described as follows. Customers are enumerated according to the order in which they arrive at the queueing system. Thus, let x_t and u_t be the waiting time and the service rate of the tth customer, respectively. The process $\{x_t\}$ evolves according to the stochastic difference equation

$$x_{t+1} = \max\{x_t + a_t\eta_t - \xi_t, 0\}, \quad t = 1, 2, \ldots, \tag{1}$$

where η_t represents a random "base" service time of the tth customer and ξ_t denotes the interarrival time between the tth and $(t+1)$th customers, while a_t is the control variable defined as the reciprocal of the service rate u_t, i. e., $a_t = 1/u_t$. In addition, the waiting time is PO in the sense that the controller only observe when $x_t = 0$. That is, the controller only realizes when a customer arrives directly to the server, and in such a case it is possible to record the occurrence of the event $[x_t = 0]$. In contrast, the controller is unable to measure the possible waiting time of customers when the server is busy, which means that the variable x_t is not observable when it takes positive values. Hence, we can define the observation process $\{y_t\}$ as

$$y_t := I_{[x_t=0]}, \quad t = 1, 2, \ldots \tag{2}$$

DOI: 10.14736/kyb-2021-3-0493

The actions or controls at each stage are selected by the controller taking into account the observed history, in order to minimize an infinite horizon total expected discounted cost. In this sense, the control problem we are dealing with is finding a discounted optimal policy within this partially observable scenario.

We follow the standard approach (see, e. g., [2, 3, 8, 15, 17, 18] and references therein) to solve the PO control problem, which essentially consists of to transform it into a completely observable (CO) control problem whose dynamic is given by a recursive equation of the form

$$z_{t+1} = \Psi(z_t, a_t, y_{t+1}), \quad t = 1, 2, \ldots, \tag{3}$$

where Ψ is a suitable function and z_t is the conditional distribution of x_t given the observed history up to time t. Defining properly the corresponding total discounted cost, this new CO control problem is equivalent to the PO control problem in the sense that an optimal policy for one is optimal for the other. Hence, the analysis of the PO problem focuses on guaranteeing the existence of a solution for the CO problem, by applying one of the several well-known procedures in the field of Markov decision processes (see, e. g., [2, 3, 7, 12, 13, 14, 15, 17]). It is worth observing that the good behavior of such procedures depends on knowing the function Ψ in (3) as well as its properties (e. g., continuity in the pair (z, a)), which can be obtained according to the specific problem that is being analyzed.

Our main objective in this paper is precisely to present an explicit form of the dynamic of the CO problem corresponding to the PO queueing system (1) – (2). In general terms, assuming that the distribution z_t of the non observed waiting time x_t has a conditional density κ_t given the observed history up to time t and applying a convenient filtering technique, we obtain a recursive equation $\kappa_{t+1} = \Psi(\kappa_t)$ generating the process $\{\kappa_t\}$. Then, imposing suitable continuity and compactness condition, we prove the existence of optimal control policies for the CO problem, which are optimal for the PO queueing system (1) – (2).

Controlled queueing systems for the CO case have been studied in several scenarios (see, e. g., [6, 10, 16]). In particular, in [6] is analyzed a queueing system of the form (1). On the other hand, for PO systems, a special case of (1) is when $\eta_t = 1$, which models an inventory system where the only possible observation is when the stock is zero. This kind of PO inventory systems were studied in [1] considering unnormalized probabilities to analyze the dynamic of the corresponding CO problem and supposing finite expectation respect to the conditional densities κ_t. In addition, our work is also related with non controlled queueing systems modeled by the Lindley equation [11]

$$x_{t+1} = \max\{x_t + \eta_t - \xi_t, 0\}, \quad t = 1, 2, \ldots, \tag{4}$$

when the waiting time x_t is only observed if it takes zero value, being η_t a random variable that represents the service time of the customer that arrives at position t. Thus, our results show a recursive filtering process allowing to study the PO system (4) in a CO environment. Within the field of the stochastic process filtering theory (see,

e. g., [5] and references therein), the latter constitutes the first step to obtain important mathematical properties of the original PO queueing system (4).

The paper is organized as follows. In Section 2 we present a review on general PO control systems and delineate the approach applied to analyze the PO queueing system described in Section 3. Next, in Section 4 we introduce the filtering process which define the dynamic of the densities, whereas in Section 5 the CO control problem is defined. Next, in Section 6, we prove the existence of optimal policies. Finally, Section 7 contains some concluding remarks.

2. REVIEW ON GENERAL PO SYSTEMS

We assume that all stochastic processes considered throughout the paper are defined on an underlying probability space $(\Omega, \mathcal{F}, P)$. In addition we shall use the following notation.

For a *Borel space* Y — that is, a Borel subset of a complete separable metric space — $\mathcal{B}(Y)$ denotes the Borel σ–algebra and "measurability" always means measurability with respect to $\mathcal{B}(Y)$. Given two Borel spaces Y and Y', a *stochastic kernel* $q(\cdot|\cdot)$ on Y given Y' is a function such that $q(\cdot|y')$ is a probability measure in Y for each $y' \in Y'$, and $q(B|\cdot)$ is a measurable function on Y' for each $B \in \mathcal{B}(Y)$. The set of all probability measures on Y is denoted by $\mathbb{P}(Y)$, endowed with the usual weak topology. That is, a sequence $\{\theta_t\}$ in $\mathbb{P}(Y)$ converges weakly to $\theta \in \mathbb{P}(Y)$ if

$$\int_Y \varphi \mathrm{d}\theta_t \to \int_Y \varphi \, \mathrm{d}\theta,$$

for all continuous and bounded function φ on Y. The sets $\mathbb{R}^+$ and $\mathbb{N}$ stand for the nonnegative real numbers set and positive integers subset, respectively. We denote by $\mathcal{L}(Y)$ the class of lower semicontinuous functions on Y bounded below and by $\mathcal{L}_+(Y) \subset \mathcal{L}(Y)$ the subclass of nonnegative functions in $\mathcal{L}(Y)$. Moreover, $L_1(\mathbb{R}^+)$ is the space of integrable functions on $\mathbb{R}^+$. In addition, denote by $\mathbb{D}$ the set of all density functions φ on $\mathbb{R}^+$. Observe that $\mathbb{D}$ is a closed subset of $L_1(\mathbb{R}^+)$ respect to the L_1–norm, that is, if $\{\varphi_t\}$ is a sequence in $\mathbb{D}$ such that

$$\|\varphi_t - \varphi\| := \int_0^\infty |\varphi_t(\omega) - \varphi(\omega)| \, \mathrm{d}\omega \to 0, \quad \text{as} \quad t \to \infty, \tag{5}$$

then $\varphi \in \mathbb{D}$. Finally, $I_B(\cdot)$ stands for the indicator function of the set B and $(x)^+ := \max\{0, x\}$.

We consider a general PO stochastic system of the form

$$x_{t+1} = F\left(x_t, a_t, w_t^{(1)}\right), \quad t \in \mathbb{N}, \tag{6}$$

$$y_t = G\left(x_t, w_t^{(2)}\right), \quad t \in \mathbb{N}, \tag{7}$$

where F and G are known functions, x_t, a_t, and y_t represent the state, action, and observation, at time t, with values in X, A, and Y, respectively; $\left\{w_t^{(1)}\right\}$ and $\left\{w_t^{(2)}\right\}$ are independent sequences of independent and identically distributed (i.i.d.) random variables with values in S_1 and S_2 and distributions $\theta_1 \in \mathbb{P}(S_1)$ and $\theta_2 \in \mathbb{P}(S_2)$, respectively. We assume that the state space X, the observation space Y, and the disturbance spaces S_1 and S_2 are all Borel spaces, while the control space A is a compact metric space. Moreover, the one-stage cost is a nonnegative and continuous function $c : X \times A \to \mathbb{R}$.

The evolution of the PO stochastic system (6) – (7) is as follows. Suppose that the initial state $x_1 \in X$ has a given distribution $\nu \in \mathbb{P}(X)$. If $x_1 = x$, then the initial observation $y_1 \in Y$ is generated according to the stochastic kernel

$$K(C|x) := \int_{S_2} I_C\left[G(x,s)\right]\theta_2(\mathrm{d}s), \quad C \in \mathcal{B}(Y). \tag{8}$$

Taking into account the observation y_1 the controller selects an action $a_1 = a \in A$. Next, a cost $c(x,a)$ is incurred and the system moves to a new state $x_2 = x'$ according to the transition law

$$Q(B|x,a) := \int_{S_1} I_B\left[F(x,a,s)\right]\theta_1(\mathrm{d}s), \quad B \in \mathcal{B}(X). \tag{9}$$

Once the system is in state x' a new observation y_2 is generated according to the stochastic kernel $K(\cdot|x')$, and the process is repeated over and over again.

The actions are selected by means of control policies which are sequences $\pi = \{a_t\}$ of $A-$valued random variables such that a_t is $\mathcal{Y}_t-$measurable for each $t \in \mathbb{N}$, where $\mathcal{Y}_t := \sigma(y_1, y_2, \ldots, y_t)$ is the $\sigma-$algebra generated by the observation process. We denote by Π the set of all control policies. In addition, the costs are accumulated in an infinite horizon under the following discounted optimality criterion. For each policy $\pi \in \Pi$ and initial distribution $\nu \in \mathbb{P}(X)$ we define

$$V(\pi,\nu) := E_\nu^\pi \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, a_t), \tag{10}$$

where $\alpha \in (0,1)$ is the discount factor and E_ν^π is the expectation operator with respect to the probability measure P_ν^π induced by π and ν (see, e. g., [3, 4] for construction of P_ν^π). Hence, the PO optimal control problem is to find a policy $\pi^* \in \Pi$ such that

$$V^*(\nu) := \inf_{\pi \in \Pi} V(\pi,\nu) = V(\pi^*,\nu), \quad \nu \in \mathbb{P}(X). \tag{11}$$

The analysis of the problem (11) is based on the standard approach which consists in to transform it into a new CO control problem by using a filtering process. To fix ideas, for each $\pi \in \Pi$ and $\nu \in \mathbb{P}(X)$, we consider the filtering process $\{z_t\} \subset \mathbb{P}(X)$ defined, for $B \in \mathcal{B}(X)$, as

$$z_1(B) := P_\nu^\pi\left(x_1 \in B\right) = \nu(B) \tag{12}$$

and

$$z_t(B) := P^\pi_\nu\left(x_t \in B|\mathcal{Y}_{t-1}\right), \quad t > 1. \tag{13}$$

It is well-known (see, e. g., [2, 3, 5, 8, 15, 17, 18]) that there exists a measurable function $\Psi : \mathbb{P}(X) \times A \times Y \to \mathbb{P}(X)$ such that the filtering process (13) satisfies a recursive equation of the form

$$z_{t+1} = \Psi(z_t, a_t, y_{t+1}), \quad t \in \mathbb{N}, \tag{14}$$

with initial condition $z_1 = \nu \in \mathbb{P}(X)$ given in (12). Defining a new cost function $\tilde{c} : \mathbb{P}(X) \times A \to \mathbb{R}$ as

$$\tilde{c}(z, a) := \int_X c(x, a) z(\mathrm{d}x) \tag{15}$$

and following standard arguments on PO control systems, the performance index (10) can be written as

$$V(\pi, \nu) := E^\pi_\nu \sum_{t=1}^{\infty} \alpha^{t-1} \tilde{c}(z_t, a_t). \tag{16}$$

Hence, the CO control problem is to find a policy that minimize (16) subject to (12)–(14). Moreover, both CO and PO control problems are equivalent in the sense that an optimal policy for one is optimal for the other.

The solution of the CO control problem can be obtained by applying the following dynamic programming (DP) procedure. For a function $U : \mathbb{P}(X) \to \mathbb{R}$, we define the DP operator

$$TU(z) = \min_{a \in A} T_a U(z), \quad z \in \mathbb{P}(X), \tag{17}$$

where, for $z_1 = z$,

$$\begin{aligned} T_a U(z) &:= \tilde{c}(z, a) + \alpha E^\pi_z \left[U\left(z_2\right)\right] \\ &= \tilde{c}(z, a) + \alpha E^\pi_z \left[U\left(\Psi(z, a, y_2)\right)\right], \quad z \in \mathbb{P}(X), a \in A. \end{aligned} \tag{18}$$

In addition, we define the sequence of value iteration functions $\{v_t\}$ as

$$\begin{aligned} v_1 &= 0; \\ v_t(\cdot) &= T v_{t-1}(\cdot), \quad t > 1. \end{aligned} \tag{19}$$

Then the DP procedure consists in finding a solution of the optimality equation

$$TU(z) = U(z), \quad z \in \mathbb{P}(X), \tag{20}$$

and its respective minimizers. Specifically, it needs to be shown that the value function satisfies $TV^* = V^*$ and $v_t \to V^*$, as $t \to \infty$. For this purpose, it is necessary to impose continuity and compactness conditions as these introduced below.

Assumption 2.1. (a) A is a compact set.

(b) $\tilde{c} \in \mathcal{L}_+(\mathbb{P}(X) \times A)$.

(c) $T_a u \in \mathcal{L}_+(\mathbb{P}(X) \times A)$ for all $u \in \mathcal{L}_+(\mathbb{P}(X))$.

Following practically the same arguments as in [9, Th. 4.2], we can establish the following result:

Theorem 2.2. Suppose that Assumption 2.1 holds. Then:

(a) $v_t \nearrow V^*$, as $t \to \infty$.

(b) The value function $V^* : \mathbb{P}(X) \to \mathbb{R}$ is the minimal solution in $\mathcal{L}_+(\mathbb{P}(X))$ of the optimality equation, that is,

$$\begin{aligned} V^*(\nu) &= \min_{a\in A} \{\tilde{c}(\nu, a) + \alpha E_\nu^\pi [V^*(z_2)]\} \\ &= \min_{a\in A} \{\tilde{c}(\nu, a) + \alpha E_\nu^\pi [V^*(\Psi(\nu, a, y_2))]\}, \quad \nu \in \mathbb{P}(X). \end{aligned} \tag{21}$$

(c) There exists $f^* : \mathbb{P}(X) \to A$ such that

$$V^*(\nu) = \tilde{c}(\nu, f^*(\nu)) + \alpha E_\nu^\pi [V^*(\Psi(\nu, f^*(\nu), y_2))], \quad \nu \in \mathbb{P}(X).$$

Moreover, the policy $\pi^* = \{a_t^*\} \in \Pi$, determined by $a_t^* = f^*(z_t)$, $t \in \mathbb{N}$, is an optimal control policy.

Since the cost function $\tilde{c}$ is not necessarily bounded, observe that the uniqueness of the solution to the optimality equation (20) is not ensured under Assumption 2.1. There is another set of conditions implying uniqueness of V^* which fall within the weighted-norm approach, widely studied under several settings in the field of Markov decision processes (see, e. g., [12, 14] and reference therein). Such approach allows a possibly unbounded one-stage cost $\tilde{c}$ provided that it is upper bounded by some weight function W, which in turns must satisfy a growth condition. However, it is worth noting that for PO systems that condition could be very strong because the function W would be defined on the space $\mathbb{P}(X)$.

3. DESCRIPTION OF THE PO CONTROLLED QUEUEING SYSTEM

We consider a PO $GI/GI/1$ queueing system where the waiting times are observable only when they take zero value. The objective in this section is to define the corresponding PO optimal control problem in the sense of the previous section. To describe precisely the model, we define the following variables:

- x_t : waiting time of the tth customer, which represents the state of the system, taking values in $X = \mathbb{R}^+$, and observable only when $x_t = 0$;
- η_t : base service time of the tth customer;
- a_t : control variable taking values in $A = [a_*, a^*]$, $a^* > a_* > 0$, defined as the reciprocal of the service rate u_t for the tth customer i. e., $a_t = \dfrac{1}{u_t}$;
- ξ_t : interarrival time between the tth and $(t+1)-$th customers.

Considering these definitions, the waiting time process $\{x_t\}$ evolves in X according to the difference equation

$$x_{t+1} = (x_t + a_t\eta_t - \xi_t)^+, \quad t \in \mathbb{N}, \tag{22}$$

whose observation process is given by

$$y_t := I_{[x_t=0]}, \quad t \in \mathbb{N}. \tag{23}$$

We assume that the queueing system satisfies the following conditions:

- $\{\xi_t\}$ and $\{\eta_t\}$ are independent sequences of nonnegative i.i.d. random variables with distribution functions F_ξ and F_η and continuous densities f_ξ and f_η, respectively.

- At initial time $t = 1$, either $x_1 = 0$, therefore it is observed, or $x_1 > 0$ and has a density function $\kappa \in \mathbb{D}$.

- For $t > 1$, when $x_t > 0$ (i. e., it is a non observed state) it has a conditional density κ_t given the observed history. That is

$$\kappa_t(\cdot): \text{ conditional density of } x_t \text{ given } \mathcal{Y}_{t-1} \text{ and } x_t > 0. \tag{24}$$

- The one-stage cost $c(x, a)$ is a continuous and nonnegative function. Moreover, the function

$$(a, \kappa) \to \int_0^\infty c(x, a)\kappa(x)\mathrm{d}x, \quad (a, \kappa) \in A \times \mathbb{D}, \tag{25}$$

  is continuous.

Observe that the events $[x_{t+1} = 0]$ and $[\xi_t > x_t + a_t\eta_t]$ are equivalent.

Under this settings, the initial condition of the PO process is given by a pair $(y_1, \kappa_1) := (y, \kappa) \in \{0, 1\} \times \mathbb{D}$ with the following meaning. If $y_1 = 1$ then $x_1 = 0$, and if $y_1 = 0$ then $x_1 > 0$ with density $\kappa \in \mathbb{D}$. Thus, the initial distribution $\nu \in \mathbb{P}(X)$ of the PO process takes the form

$$\nu(B) := y\delta_0(x_1) + (1 - y)\int_B \kappa(\omega)\,\mathrm{d}\omega, \quad B \in \mathcal{B}(X), \tag{26}$$

where δ_0 is the Dirac delta function. Observe that the distribution $\nu \in \mathbb{P}(X)$ is completely determined by the initial condition. Hence, we use the notation $P^\pi_{(y,\kappa)}$ and $E^\pi_{(y,\kappa)}$ instead of P^π_ν and E^π_ν, respectively. In this context, our objective is to show the existence of optimal policies for the PO optimal control problem, defined in general form in (10) – (11). To this end, we apply the standard approach described in previous section which we can summarize in the following steps:

- Once defined the PO control problem, transform it into an equivalent CO control problem by means of a filtered process $\{z_t\}$ (see (12)−(13)), which satisfies a recursive equation of the form (14). In our case we will see that the distribution z_t is determined by the pair $(y_t, \kappa_t) \in \{0,1\} \times \mathbb{D}$, leading that our goal is to find a recursive equation for the densities κ_t.
- Next, define the CO control problem in terms of the new process $\{\kappa_t\} \subset \mathbb{D}$ as well as the corresponding optimality equation.
- Finally, prove that the CO control problem satisfy the appropriate continuity condition in Assumption 2.1, to be able to apply Theorem 2.2.

The next sections are concerned with giving solutions to these points.

4. RECURSIVE EQUATION FOR DENSITIES

Let $\kappa_t(\cdot)$ be the conditional density of x_t given $\mathcal{Y}_t$ and $x_t > 0$ (see (24)). Then, for each policy $\pi \in \Pi$ and initial condition $(y, \kappa) \in \{0,1\} \times \mathbb{D}$, the filtered process $\{z_t\}$ takes the form ((see (12), (13), and (26)), for $B \in \mathcal{B}(X)$,

$$z_1(B) := P^{\pi}_{(y,\kappa)}(x_1 \in B) = y\delta_0(x_1) + (1-y)\int_B \kappa(\omega)\,\mathrm{d}\omega, \tag{27}$$

$$z_{t+1}(B) := P^{\pi}_{(y,\kappa)}(x_{t+1} \in B \mid \mathcal{Y}_t) = y_{t+1}\delta_0(x_{t+1}) + (1-y_{t+1})\int_B \kappa_{t+1}(\omega)\,\mathrm{d}\omega, \quad t \in \mathbb{N}. \tag{28}$$

Hence, the distribution z_t is completely determined by the pair $(y_t, \kappa_t) \in \{0,1\} \times \mathbb{D}$. Moreover, because the events $[x_{t+1} = 0]$ are observed, the densities k_{t+1} only take part when $x_{t+1} > 0$, i. e., if $y_t = 0$. In this sense, to study the evolution of the filtered process $\{z_t\}$ it is sufficient to obtain a recursive equation that generates the process $\{\kappa_t\} \subset \mathbb{D}$.

For each $\kappa \in \mathbb{D}$, $x \in X$, and $a \in A$, we define the function $\Psi(\kappa, x, a)(\cdot)$ as

$$\Psi(\kappa,x,a)(s) = \begin{cases} \dfrac{\int_{s/a_{t-1}}^{\infty} f_\xi\left((av-s)^+\right) f_\eta(v)\,\mathrm{d}v}{\int_0^\infty F_\xi(av) f_\eta(v)\,\mathrm{d}v} & \text{if } x = 0, \\[2ex] \dfrac{\int_0^\infty \int_{(s-av)^+}^{\infty} f_\xi\left((\omega+av-s)^+\right) f_\eta(v)\,\kappa(\omega)\,\mathrm{d}\omega\mathrm{d}v}{\int_0^\infty\int_0^\infty F_\xi(\omega+av) f_\eta(v)\,\kappa(\omega)\,\mathrm{d}\omega\mathrm{d}v} & \text{if } x > 0. \end{cases}$$

A straightforward calculation shows that $\Psi(\kappa, x, a) \in \mathbb{D}$, that is, for each $\kappa \in \mathbb{D}$, $x \in X$, and $a \in A$, we have $\Psi \geq 0$ and $\int_0^\infty \Psi(\kappa,x,a)(s)\,\mathrm{d}s = 1$.

From this fact, we can state our result as follows.

Theorem 4.1. The density process $\{\kappa_t\}$ evolves in $\mathbb{D}$ according to the system $\kappa_1 = \kappa \in \mathbb{D}$, $\kappa_t = \Psi(\kappa_{t-1}, x_{t-1}, a_{t-1})$, $t \in \mathbb{N}$. That is, for $t \in \mathbb{N}$,

$$\kappa_t(s) = I_{[x_{t-1}=0]} \left\{ \frac{\int\limits_{s/a_{t-1}}^{\infty} f_{\xi_{t-1}}\left((a_{t-1}v - s)^+\right) f_{\eta_{t-1}}(v)\,\mathrm{d}v}{\int\limits_0^{\infty} F_{\xi_{t-1}}(a_{t-1}v) f_{\eta_{t-1}}(v)\,\mathrm{d}v} \right\}$$
$$+ I_{[x_{t-1}>0]} \left\{ \frac{\int\limits_0^{\infty} \int\limits_{(s-a_{t-1}v)^+}^{\infty} f_{\xi_{t-1}}\left((\omega + a_{t-1}v - s)^+\right) f_{\eta_{t-1}}(v)\,\kappa_{t-1}(\omega)\,\mathrm{d}\omega\mathrm{d}v}{\int\limits_0^{\infty}\int\limits_0^{\infty} F_{\xi_{t-1}}(\omega + a_{t-1}v) f_{\eta_{t-1}}(v)\,\kappa_{t-1}(\omega)\,\mathrm{d}\omega\mathrm{d}v} \right\}. \tag{29}$$

The proof of Theorem 4.1 will be consequence of the Lemmas 4.2 and 4.3 below.

Let $\pi \in \Pi$ and $(y, \kappa) \in \{0,1\} \times \mathbb{D}$ fixed and arbitrary. In the remainder, to ease notation, we write P and E instead of $P^{\pi}_{(y,\kappa)}$ and $E^{\pi}_{(y,\kappa)}$, respectively.

Lemma 4.2. For every bounded function $\varphi : X \to \mathbb{R}$ and $t \in \mathbb{N}$,

$$E[\varphi(x_t) \mid \mathcal{Y}_t] = I_{[x_t=0]}\varphi(0) + I_{[x_t>0]} \frac{E[\varphi(x_t) I_{[x_t>0]} \mid \mathcal{Y}_{t-1}]}{P(x_t > 0 \mid \mathcal{Y}_{t-1})} \tag{30}$$
$$= I_{[x_t=0]}\varphi(0) + I_{[x_t>0]} E[\varphi(x_t) \mid \mathcal{Y}_{t-1}, x_t > 0]. \tag{31}$$

Proof. Observe that for each $t \in \mathbb{N}$,

$$\begin{aligned} E[\varphi(x_t) \mid \mathcal{Y}_t] &= E[\varphi(x_t)\left(I_{[x_t=0]} + I_{[x_t>0]}\right) \mid \mathcal{Y}_t] \\ &= E[\varphi(x_t)(I_{[x_t=0]} \mid \mathcal{Y}_t] + E[\varphi(x_t) I_{[x_t>0]} \mid \mathcal{Y}_t] \\ &= \varphi(0) I_{[x_t=0]} + E[\varphi(x_t) I_{[x_t>0]} \mid \mathcal{Y}_t] \end{aligned} \tag{32}$$

By definition of conditional expectation, there exists a measurable function Φ such that $E[\varphi(x_t) \mid \mathcal{Y}_t] = \Phi(y_1, \ldots, , y_{t-1}, y_t)$. Therefore, since the event $[x_t > 0]$ is $\mathcal{Y}_t-$measurable, the last term in (32) can be rewrite as

$$\begin{aligned} E[\varphi(x_t) I_{[x_t>0]} \mid \mathcal{Y}_t] &= I_{[x_t>0]} E[\varphi(x_t) \mid \mathcal{Y}_t] \\ &= I_{[x_t>0]} \Phi(y_1, \ldots, , y_{t-1}, y_t) \\ &= I_{[x_t>0]} \Phi(y_1, \ldots, y_{t-1}, 0), \end{aligned} \tag{33}$$

where the last equality comes from the equivalence of the event $[x_t > 0]$ and $[y_t = 0]$. Applying conditional expectation respect to $\mathcal{Y}_{t-1}$ in (33), and taking into account that $\mathcal{Y}_{t-1} \subseteq \mathcal{Y}_t$ and $\Phi(y_1, \ldots, , y_{t-1}, 0)$ is $\mathcal{Y}_{t-1}-$measurable, we obtain

$$E[\varphi(x_t) I_{[x_t>0]} \mid \mathcal{Y}_{t-1}] = \Phi(y_1, \ldots, , y_{t-1}, 0,) P(x_t > 0 \mid \mathcal{Y}_{t-1}). \tag{34}$$

Combining (32) – (34) we get equality (30), that is

$$E[\varphi(x_t) \mid \mathcal{Y}_t] = I_{[x_t=0]}\varphi(0) + I_{[x_t>0]}\frac{E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}]}{P(x_t > 0 \mid \mathcal{Y}_{t-1})}. \tag{35}$$

To prove (31), by the Conditional Bayes Theorem (see [5]), we have that

$$\begin{aligned}\int_0^\infty \varphi(s)\kappa_t(s)\,\mathrm{d}s &= E[\varphi(x_t) \mid \mathcal{Y}_{t-1}, x_t > 0]\\ &= \frac{E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}]}{E[I_{[x_t>0]} \mid \mathcal{Y}_{t-1})}\\ &= \frac{E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}]}{P(x_t > 0 \mid \mathcal{Y}_{t-1})}.\end{aligned} \tag{36}$$

Hence, using (36) in (35) we get (31). □

Lemma 4.3. For every bounded function $\varphi : X \to \mathbb{R}$ and $t \in \mathbb{N}$,

$$\begin{aligned}&E[\varphi(x_t) \mid \mathcal{Y}_t]I_{[x_t>0]}\\ &= I_{[x_{t-1}=0]}\left\{\frac{\int_0^\infty \int_{s/a_{t-1}}^\infty \varphi(s)f_{\xi_{t-1}}\left((a_{t-1}v - s)^+\right)f_{\eta_{t-1}}(v)\,\mathrm{d}v\mathrm{d}s}{\int_0^\infty F_{\xi_{t-1}}(a_{t-1}v)\,f_{\eta_{t-1}}(v)\,\mathrm{d}v}\right\} + I_{[x_{t-1}>0]}\\ &\quad\cdot\left\{\frac{\int_0^\infty\int_0^\infty \varphi(s)\int_{(s-a_{t-1}v)^+}^\infty f_{\xi_{t-1}}\left((\omega + a_{t-1}v - s)^+\right)f_{\eta_{t-1}}(v)\,\kappa_{t-1}(\omega)\,\mathrm{d}\omega\,\mathrm{d}v\mathrm{d}s}{\int_0^\infty\int_0^\infty F_{\xi_{t-1}}(\omega + a_{t-1}v)\,f_{\eta_{t-1}}(v)\,\kappa_{t-1}(\omega)\,\mathrm{d}\omega\,\mathrm{d}v}\right\}\end{aligned} \tag{37}$$

Proof. For each $t \in \mathbb{N}$, using independence of the sequences $\{\xi_t\}$ and $\{\eta_t\}$, and that the density of the random vector $\rho_{t-1} = (\xi_{t-1}, \eta_{t-1})$ is $f_{\rho_{t-1}}(u, v) = f_{\xi_{t-1}}(u)f_{\eta_{t-1}}(v)$ we have

$$\begin{aligned}&E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}]\\ &= E[\varphi\left((x_{t-1} + a_{t-1}\eta_{t-1} - \xi_{t-1})^+\right)I_{[x_{t-1}+a_{t-1}\eta_{t-1}-\xi_{t-1}>0]} \mid \mathcal{Y}_{t-1}]\\ &= E\left[E\left[\varphi\left((x_{t-1} + a_{t-1}\eta_{t-1} - \xi_{t-1})^+\right)I_{[x_{t-1}+a_{t-1}\eta_{t-1}-\xi_{t-1}>0]} \mid \mathcal{Y}_{t-1}, x_{t-1}\right] \mid \mathcal{Y}_{t-1}\right]\\ &= E\left[\iint_{\{(v,u):x_{t-1}+a_{t-1}v-u>0,\,u>0,v>0\}}\varphi\left((x_{t-1}+a_{t-1}v-u)^+\right)f_{\xi_{t-1}}(u)f_{\eta_{t-1}}(v)\,\mathrm{d}u\mathrm{d}v \mid \mathcal{Y}_{t-1}\right]\\ &= E\left[\int_0^\infty\int_0^{x_{t-1}+a_{t-1}v}\varphi\left((x_{t-1} + a_{t-1}v - u)^+\right)f_{\xi_{t-1}}(u)f_{\eta_{t-1}}(v)\,\mathrm{d}u\mathrm{d}v \mid \mathcal{Y}_{t-1}\right]\end{aligned}$$

By applying the change of variable $s := x_{t-1} + a_{t-1}v - u$ and standard integration properties, it follows

$$
\begin{aligned}
&E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}] \\
&= E\left[\int_0^\infty \left[\int_0^{x_{t-1}+a_{t-1}v} \varphi(s) f_{\xi_{t-1}}\left((x_{t-1}+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\,\mathrm{d}s\right]\mathrm{d}v \mid \mathcal{Y}_{t-1}\right] \\
&= E\left[\int_0^\infty \int_0^\infty \varphi(s) f_{\xi_{t-1}}\left((x_{t-1}+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[x_{t-1}+a_{t-1}v-s>0]}\,\mathrm{d}s\mathrm{d}v \mid \mathcal{Y}_{t-1}\right] \\
&= \int_0^\infty \int_0^\infty \varphi(s) E\left[f_{\xi_{t-1}}\left((x_{t-1}+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[x_{t-1}+a_{t-1}v-s>0]} | \mathcal{Y}_{t-1}\right]\mathrm{d}s\mathrm{d}v.
\end{aligned}
\tag{38}
$$

On the other hand, note that from (31)

$$
E[\varphi(x_{t-1}) \mid \mathcal{Y}_{t-1}] = I_{[x_{t-1}=0]}\varphi(0) + I_{[x_{t-1}>0]} \int_0^\infty \varphi(s)\kappa_{t-1}(s)\,\mathrm{d}s, \tag{39}
$$

for any real bounded function φ. Then we can take

$$
\varphi(x_{t-1}) = f_{\xi_{t-1}}\left((x_{t-1}+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[x_{t-1}+a_{t-1}v-s>0]}
$$

in (39) to obtain, for each $s, v \in \mathbb{R}^+$,

$$
\begin{aligned}
&E\left[f_{\xi_{t-1}}\left((x_{t-1}+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[x_{t-1}+a_{t-1}v-s>0]} \mid \mathcal{Y}_{t-1}\right] \\
&= I_{[x_{t-1}=0]} f_{\xi_{t-1}}\left((a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[a_{t-1}v-s>0]} \\
&+ I_{[x_{t-1}>0]} \int_0^\infty f_{\xi_{t-1}}\left((\omega+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[\omega+a_{t-1}v-s>0]}\kappa_{t-1}(\omega)\,\mathrm{d}\omega.
\end{aligned}
\tag{40}
$$

Hence, combination of (38) and (40) yields

$$
\begin{aligned}
&E[\varphi(x_t)I_{[x_t>0]} \mid \mathcal{Y}_{t-1}] \\
&= I_{[x_{t-1}=0]} \int_0^\infty \int_0^\infty \varphi(s) f_{\xi_{t-1}}\left((a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\, I_{[a_{t-1}v-s>0]}\,\mathrm{d}s\mathrm{d}v \\
&+ I_{[x_{t-1}>0]} \int_0^\infty \int_0^\infty \varphi(s) \int_{(s-a_{t-1}v)^+}^\infty f_{\xi_{t-1}}\left((\omega+a_{t-1}v-s)^+\right) f_{\eta_{t-1}}(v)\,\kappa_{t-1}(\omega)\,\mathrm{d}\omega\mathrm{d}s\mathrm{d}v,
\end{aligned}
\tag{41}
$$

which is the numerator in (37). Also, the denominator in (37) can be expressed as

$$\begin{aligned} P[x_t > 0 \mid \mathcal{Y}_{t-1}] &= E\left[I_{[x_{t-1}+a_{t-1}\eta_{t-1}-\xi_{t-1}>0]} \mid \mathcal{Y}_{t-1}\right] \\ &= E\left[E[I_{[x_{t-1}+a_{t-1}\eta_{t-1}-\xi_{t-1}>0]} \mid \mathcal{Y}_{t-1}, x_{t-1}] \mid \mathcal{Y}_{t-1}\right] \\ &= E\left[E[I_{[\xi_{t-1}<x_{t-1}+a_{t-1}\eta_{t-1}]} \mid \mathcal{Y}_{t-1}, x_{t-1}] \mid \mathcal{Y}_{t-1}\right] \\ &= E\left[P\left(I_{[\xi_{t-1}<x_{t-1}+a_{t-1}\eta_{t-1}]} \mid \mathcal{Y}_{t-1}, x_{t-1}\right) \mid \mathcal{Y}_{t-1}\right] \\ &= E\left[\int_0^\infty F_{\xi_{t-1}}\left(x_{t-1}+a_{t-1}v\right) f_{\eta_{t-1}}\left(v\right) \mathrm{d}v \mid \mathcal{Y}_{t-1}\right] \\ &= I_{[x_{t-1}=0]} \int_0^\infty F_{\xi_{t-1}}\left(a_{t-1}v\right) f_{\eta_{t-1}}\left(v\right) \mathrm{d}v \\ &+ I_{[x_{t-1}>0]} \int_0^\infty \int_0^\infty F_{\xi_{t-1}}\left(\omega+a_{t-1}v\right) f_{\eta_{t-1}}\left(v\right) \kappa_{t-1}(\omega) \,\mathrm{d}\omega \mathrm{d}v, \end{aligned} \tag{42}$$

where the last equality follows by applying (39) with

$$\varphi(x_{t-1}) = \int_0^\infty F_{\xi_{t-1}}\left(x_{t-1}+a_{t-1}v\right) f_{\eta_{t-1}}\left(v\right) \mathrm{d}v.$$

Finally, interchanging the order of integration in (41), substituting (41) and (42) in (30), multiplying both sides of the resulting equality by $I_{[x_t>0]}$, and using properties of indicator functions, we get (37). □

P r o o f of Theorem 4.1. Observe that the conditional expectation in Lemma 4.2 can be expressed in terms of the density κ_t as follows

$$E[\varphi(x_t) \mid \mathcal{Y}_t] = I_{[x_t=0]}\varphi(0) + I_{[x_t>0]} \int_0^\infty \varphi(s)\kappa_t(s)\,\mathrm{d}s. \tag{43}$$

Then we obtain the desired result by comparing equations (37) and (43). □

For functions $\varphi_1, \varphi_2 : \mathbb{R}^+ \to \mathbb{R}^+$ and $a \in A$, we define

$$< \varphi_1, \varphi_2 >:= \int_0^\infty \varphi_1(s)\varphi_2(s)\,\mathrm{d}s \tag{44}$$

and

$$\rho_{(a,\varphi_1)}(s) := \int_0^\infty \int_{(s-av)^+}^\infty f_\xi\left((\omega+av-s)^+\right) f_\eta\left(v\right) \varphi_1(\omega)\,\mathrm{d}\omega\mathrm{d}v. \tag{45}$$

In the particular case when φ_1 is the Dirac delta function we have

$$\begin{aligned} \rho_{(a,\delta)}(s) &= \int_0^\infty \int_{(s-av)^+}^\infty f_\xi((\omega+av-s)^+) f_\eta\left(v\right) \delta(\omega)\, \omega\mathrm{d}v \\ &= \int_0^\infty f_\xi((av-s)^+) f_\eta\left(v\right) I_{[s<av]}\,\mathrm{d}v. \end{aligned}$$

Additionally we define

$$\theta(s;a,\varphi_1) := \frac{\rho_{(a,\varphi_1)}(s)}{< \rho_{(a,\varphi_1)}, 1 >}. \tag{46}$$

Then, considering these definitions and the observation process (23), and following straightforward calculations we can get a simplified version of the equation (29) of the form

$$\begin{aligned} \kappa_1 &= \kappa, \\ \kappa_t(s) &= y_{t-1}\theta(s;a_{t-1},\delta) + (1-y_{t-1})\theta(s;a_{t-1},\kappa_{t-1}), \quad t>1. \end{aligned} \tag{47}$$

5. THE CO CONTROL PROBLEM

Following the outlined scheme, in this section we will define the CO control problem in terms of the dynamic of densities given in (47), and obtain the corresponding optimality equation.

We define the one-stage cost (see (15) and (28)) as

$$\begin{aligned} \overline{c}(z_t,a_t) &:= \tilde{c}(y_t,\kappa_t,a_t) \\ &= \int_0^\infty c(x,a_t) z_t(\mathrm{d}x) \\ &= y_t c(0,a_t) + (1-y_t)\int_0^\infty c(x,a_t)\kappa_t(x)\,\mathrm{d}x \\ &= y_t c(0,a_t) + (1-y_t)\left\langle c(\cdot,a_t),\kappa_t(\cdot)\right\rangle. \end{aligned} \tag{48}$$

Thus, if $(y,\kappa) \in \{0,1\} \times \mathbb{D}$ is the initial condition defining the initial distribution $z_1 = \nu \in \mathbb{P}(X)$ (see (27)), then for each $\pi \in \Pi$ we define the total discounted cost for the CO control problem as

$$V(\pi,y,\kappa) = \sum_{t=1}^\infty \alpha^{t-1} E^\pi_{(y,k)}[\tilde{c}(y_t,\kappa_t,a_t)], \tag{49}$$

with optimal value function

$$V^*(y,\kappa) = \inf_{\pi\in\Pi} V(\pi,y,\kappa), \quad (y,\kappa)\in\{0,1\}\times\mathbb{D}. \tag{50}$$

Hence, the control problem consists of to find a policy $\pi^* \in \Pi$ such that

$$V^*(y,\kappa) = V(\pi^*,y,\kappa), \quad \text{for all } (y,\kappa)\in\{0,1\}\times\mathbb{D}. \tag{51}$$

5.1. Optimality equation

Similarly as (20), a function $U : \{0,1\}\times\mathbb{D} \to \mathbb{R}$ satisfies the optimality equation if

$$U(y,\kappa) = \min_{a\in A}\left\{ yc(0,a) + (1-y)\left\langle c(\cdot,a),\kappa(\cdot)\right\rangle + \alpha E^\pi_{(y,k)}[U(y_2,\kappa_2)]\right\}, \tag{52}$$

where, from (47),

$$\begin{aligned} E^{\pi}_{(y,k)}\left[U\left(y_2\,,\,\kappa_2\right)\right] &= E^{\pi}_{(y,k)}\left[U\left(y_2, y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right)\right] \\ &= U\left(1, y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right) P^{\pi}_{(y,k)}[x_2=0] \\ &+ U\left(0, y\theta(a,\delta)(x)+(1-y)\theta(a,\kappa)(x)\right) P^{\pi}_{(y,k)}[x_2>0]. \end{aligned} \tag{53}$$

By properties of conditional expectation

$$\begin{aligned} P^{\pi}_{(y,k)}[x_2=0] &= P^{\pi}_{(y,k)}\left[\xi \geq x + a\eta\right] \\ &= 1 - P^{\pi}_{(y,k)}\left[\xi \leq x + a\eta\right] \\ &= 1 - E^{\pi}_{(y,k)}\left[E^{\pi}_{(y,k)}\left[I_{[\xi<x+a\eta]}\right] \mid x\right] \\ &= 1 - \int_0^\infty P^{\pi}_{(y,k)}(\xi < \omega + a\eta)\kappa(\omega)\,\mathrm{d}\omega \\ &= 1 - \int_0^\infty \int_0^\infty F_\xi(\omega + av) f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega \end{aligned} \tag{54}$$

and

$$\begin{aligned} P^{\pi}_{(y,k)}[x_2 &> 0] = P^{\pi}_{(y,k)}\left[\xi < x + a\eta\right] \\ &= \int_0^\infty \int_0^\infty F_\xi(\omega + av) f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega. \end{aligned} \tag{55}$$

Thus, substituting (55) and (54) in (53), and then in (52), we obtain the optimality equation

$$\begin{aligned} U\left(y,\kappa\right) &= \min_{a\in A}\Big\{ yc(0,a) + (1-y)\left\langle c(\cdot,a),\kappa(\cdot)\right\rangle \\ &+ \alpha U\left(1, y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right)\left(1 - \int_0^\infty\int_0^\infty F_\xi(\omega+av)f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega\right) \\ &+ \alpha U\left(0\,,\,y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right)\left(\int_0^\infty\int_0^\infty F_\xi(\omega+av)f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega\right)\Big\}. \end{aligned} \tag{56}$$

Hence, we define the operators (see (17), (18))

$$\begin{aligned} T_aU(y,\kappa) &= yc(0,a) + (1-y)\left\langle c(\cdot,a),\kappa(\cdot)\right\rangle \\ &+ \alpha U\left(1, y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right)\left(1 - \int_0^\infty\int_0^\infty F_\xi(\omega+av)f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega\right) \\ &+ \alpha U\left(0\,,\,y\theta(x;a,\delta)+(1-y)\theta(x;a,\kappa)\right)\left(\int_0^\infty\int_0^\infty F_\xi(\omega+av)f_\eta\left(v\right)\kappa(\omega)\,\mathrm{d}v\mathrm{d}\omega\right), \end{aligned} \tag{57}$$

and

$$TU(y,\kappa) = \min_{a\in A} T_aU(y,\kappa). \tag{58}$$

6. EXISTENCE OF OPTIMAL POLICIES

According to the approach given in Theorem 2.2, to prove the existence of optimal policies for the PO queueing system, it is sufficient to show that Assumption 2.1 is satisfied within the scenario stated in previous section. In this case, we consider the convergence under the L_1-norm (5) in $\mathbb{D}$ for the lower semicontinuity of the costs and the corresponding dynamic programming operator. That is, we need to prove the following properties:

P1 The one-stage cost $\tilde{c}$ belongs to $\mathcal{L}_+(\{0,1\} \times \mathbb{D} \times A)$ (see (48)).

P2 $T_a U \in \mathcal{L}_+(\{0,1\} \times \mathbb{D} \times A)$, for all function $U \in \mathcal{L}_+(\{0,1\} \times \mathbb{D})$.

Specifically our main results can be stated as follows.

Theorem 6.1. If Properties P1 and P2 hold, then:

(a) $U_t(y,\kappa) \nearrow V^*(y,\kappa)$, as $t \to \infty$, $(y,\kappa) \in \{0,1\} \times \mathbb{D}$, where $\{U_t\}$ is the sequence of value iteration functions defined as $U_0 := 0$ and

$$U_t(y,\kappa) = TU_{t-1}(y,\kappa), \quad t > 1, (y,\kappa) \in \{0,1\} \times \mathbb{D},$$

and T is the operator (58).

(b) The function $V^* : \{0,1\} \times \mathbb{D} \to \mathbb{R}$ is the minimal solution in $\mathcal{L}_+(\{0,1\} \times \mathbb{D})$ of the optimality equation, i. e., $TV^* = V^*$.

(c) There exists $f^* : \{0,1\} \times \mathbb{D} \to A$ such that (see (57))

$$V^*(y,\kappa) = T_{f^*}V^*(y,\kappa), \quad (y,\kappa) \in \{0,1\} \times \mathbb{D},$$

and $\pi^* = \{f^*\} \in \Pi$ is a stationary optimal policy for the queueing system.

We then proceed to prove that properties P1 and P2 hold for the queueing system. Property P1 follows from the continuity of the function c and the condition (25), whereas P2 will be consequence of the following facts.

Lemma 6.2. The function

$$(a,\varphi) \mapsto \int_0^\infty \int_0^\infty f_\xi((\omega + av - s)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v, \quad (a,\varphi) \in A \times \mathbb{D} \tag{59}$$

is continuous.

P r o o f. Let $\{(a_t, \varphi_t)\}$ be a sequence in $A \times \mathbb{D}$ converging to $(a, \varphi) \in A \times \mathbb{D}$ (see (5)). By adding and subtracting the term

$$\int_0^\infty \int_0^\infty f_\xi((\omega + a_t v - s)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v$$

we get

$$\begin{aligned}
&\left| \int_0^\infty \int_0^\infty f_\xi((\omega + a_t v - s)^+) f_\eta(v)\, \varphi_t(\omega)\, \mathrm{d}\omega \mathrm{d}v - \int_0^\infty \int_0^\infty f_\xi((\omega + a v - s)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v \right| \\
\leq\ & \int_0^\infty \int_0^\infty f_\xi((\omega + a_t v - s)^+) f_\eta(v)\, |\varphi_t(\omega) - \varphi(\omega)|\, \mathrm{d}\omega \mathrm{d}v \\
&+ \int_0^\infty \int_0^\infty \left| f_\xi((\omega + a_t v - s)^+) f_\eta(v) - f_\xi((\omega + a v - s)^+) f_\eta(v) \right| \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v.
\end{aligned}$$

Since f_ξ and f_η are bounded functions, there exists a constant $M > 0$ such that $f_\xi(\cdot) f_\eta(\cdot) \leq M$. Then, letting $t \to \infty$, from the continuity of f_ξ and f_η and the Dominated Convergence Theorem, we obtain

$$\begin{aligned}
&\left| \int_0^\infty \int_0^\infty f_\xi((\omega + a_t v - s)^+) f_\eta(v)\, \varphi_t(\omega)\, \mathrm{d}\omega \mathrm{d}v - \int_0^\infty \int_0^\infty f_\xi((\omega + a v - s)^+) f_\eta(v))\varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v \right| \\
\leq\ & M \lim_{t \to \infty} \int_0^\infty |\varphi_t(\omega) - \varphi(\omega)|\, \mathrm{d}\omega \\
&+ \lim_{t \to \infty} \int_0^\infty \int_0^\infty \left| f_\xi((\omega + a_t v - s)^+) f_\eta(v) - f_\xi((\omega + a v - s)^+) f_\eta(v)) \right| \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v = 0,
\end{aligned}$$

which proves the continuity of function (59). □

Applying similar arguments as the proof of Lemma 6.2, it is easy to prove the continuity of the following functions:

$$(a, \varphi) \mapsto 1 - \int_0^\infty \int_0^\infty F_\xi(\omega + a v) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v,$$

$$(a, \varphi) \mapsto \int_0^\infty \int_0^\infty F_\xi(\omega + a v) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v,$$

$$(a, \varphi) \mapsto \int_0^\infty \int_{(s - a v)^+}^\infty f_\xi((\omega + a v - s)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v$$

and

$$(a, \varphi) \mapsto \int_0^\infty \int_0^\infty \int_0^\infty f_\xi((\omega + a v - s)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v \mathrm{d}s.$$

Hence, by observing that

$$\begin{aligned}\theta(x;a,\varphi) \quad : &= \frac{\rho_{(a,\varphi)}(x)}{\langle \rho_{(a,\varphi)}, 1 \rangle} \\ &= \frac{\int_0^\infty \int_{(x-av)^+}^\infty f_\xi((\omega+av-x)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v}{\int_0^\infty \int_0^\infty \int_{(x-av)^+}^\infty f_\xi((\omega+av-x)^+) f_\eta(v)\, \varphi(\omega)\, \mathrm{d}\omega \mathrm{d}v \mathrm{d}x},\end{aligned}$$

we have that, for each $x \in [0,\infty)$, the function

$$(a,\varphi) \mapsto \theta(x;a,\varphi_1)$$

is continuous. Therefore, standard arguments yield that $T_a U \in \mathcal{L}_+ (\{0,1\} \times \mathbb{D} \times A)$ for all function $U \in \mathcal{L}_+ (\{0,1\} \times \mathbb{D})$, that is property P2 holds.

7. CONCLUDING REMARKS

In this paper we have analyzed a $GI/GI/1$ queueing system with controlled service rate where the waiting times x_t are partially observed; that is the controller only observe when $x_t = 0$. We have followed the standard approach which consists of to transform the partially observed control problem into a completely observable control problem where the state's process evolves on a space of probability measures. In general, the key point in this standard procedure is precisely to obtain the function that defines the dynamical system. However, commonly, many works in the literature on partially observable control problems assume the existence of such a function, also asking that it fulfill the necessary hypotheses. In this paper, assuming that the measures have density, we have obtained an explicit form of the dynamical process, which in turn defines the filtering process, satisfying the appropriate properties to prove the existence of optimal policies. In order to illustrate, in a certain way, the behavior of the filtering process, we consider the following particular case.

We assume that the base service time for the tth customer, η_t, has a uniform distribution on $[0,1]$ and the interarrival time between the tth and $(t+1)$th customers, ξ_t, follows an exponential distribution with parameter $\gamma > 0$. Moreover, we assume that the controller can select only two possible service rates, namely, a normal service (ns) and an express service (es). Hence, the corresponding controls are $a_{ns} := \frac{1}{ns}$ and $a_{es} := \frac{1}{es}$ (recall $a_t = \frac{1}{u_t}$ where u_t is the service rate for the tth customer). Under this scenario we have $X = [0,\infty)$ and $A = \{a_{es}, a_{ns}\}$, and assuming $x_1 \backsim \mathrm{Exp}(\lambda)$, $\lambda > 0$, the evolution of the conditional densities (29) defining the filtering process is given by

$$\kappa_1(s) = \lambda e^{-\lambda s}, \quad s \in X;$$

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

$$\begin{aligned}\kappa_{t+1}(s) &= y_t \frac{\int\limits_{s/a_t}^{1} \gamma e^{-\gamma(a_t v - s)^+}\, \mathrm{d}v}{\int\limits_0^1 (1 - e^{-\gamma a_t v})\, \mathrm{d}v} + (1-y_t) \frac{\int\limits_0^1 \int\limits_{(s-a_t v)^+}^{\infty} \gamma e^{-\gamma(\omega + a_t v - s)^+} \kappa_t(\omega)\, \mathrm{d}\omega \mathrm{d}v}{\int\limits_0^1 \int\limits_0^{\infty} \left(1 - e^{-\gamma(\omega + a_t v)}\right) \kappa_t(\omega)\, \mathrm{d}\omega \mathrm{d}v} \\ &= y_t \frac{\frac{1}{a_t}\left(1 - e^{-\gamma(a_t - s)^+}\right)}{1 - \frac{1}{\gamma a_t}\left(1 - e^{-\gamma a_t}\right)} + (1-y_t) \frac{\int\limits_0^1 \int\limits_{(s-a_t v)^+}^{\infty} \gamma e^{-\gamma(\omega + a_t v - s)^+} \kappa_t(\omega)\, \mathrm{d}\omega \mathrm{d}v}{1 - \int\limits_0^1 \int\limits_0^{\infty} e^{-\gamma(\omega + a_t v)} \kappa_t(\omega)\, \mathrm{d}\omega \mathrm{d}v}. \qquad (60)\end{aligned}$$

For the specific values $a_{ns} = 0.9$, $a_{es} = 0.8$, $\lambda = 0.4$, and $\gamma = 0.3$, Figures 1 and 2 show the behavior of the distribution functions $F_{\kappa_t}(s) = \int_0^s \kappa_t(\omega \mid y_t, a_t)\, \mathrm{d}\omega$, $s \geq 0$, for two different trajectories.

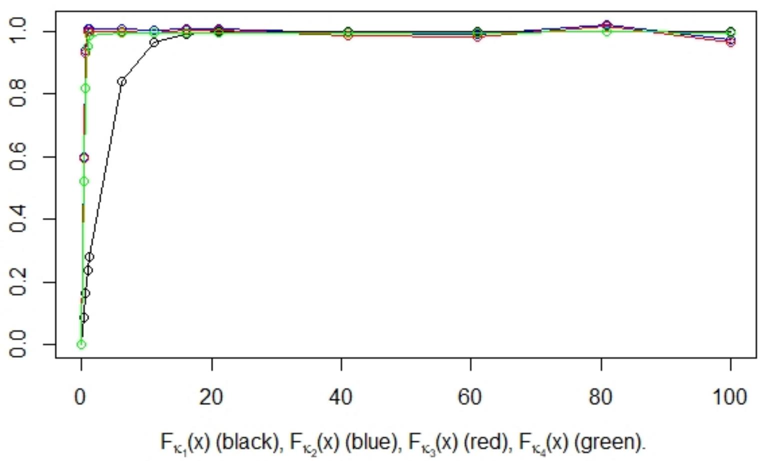

Fig. 1. Graph approximations of F_{κ_1}, F_{κ_2}, F_{κ_3} and F_{κ_4} conditioned to $y_1 = 1, a_1 = a_{es}$, $y_2 = 1, a_2 = a_{es}$, $y_3 = 0, a_2 = a_{ns}$.

A challenging problem, which in itself is interesting, is the numerical implementation of the value iteration algorithm established in Theorem 6.1. Indeed, as it is well known, for its implementation the obstacle represented by the so-called "*curse of dimensionality*" has to be overcome. Such an obstacle intensifies in partially observable problems, as the one studied in this paper, due to the fact that we are dealing with spaces of

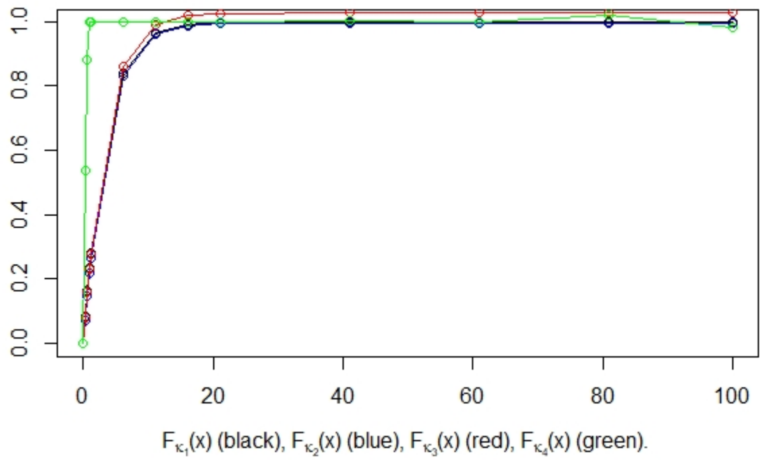

Fig. 2. Graph approximations of F_{κ_1}, F_{κ_2}, F_{κ_3} and F_{κ_4} conditioned to $y_1 = 0, a_1 = a_{es}$, $y_2 = 0, a_2 = a_{ns}$, $y_3 = 1, a_2 = a_{ns}$.

infinite dimension, e. g., the set of densities $\mathbb{D}$. The authors are currently working on obtaining general approximation algorithms for both the optimal value function and the optimal policy in spaces of infinite dimension, within the context of approximate dynamic programming.

ACKNOWLEDGEMENT

The authors acknowledge support from grant DGIP 01/FYM/RPR/022/19, FCFM-UNACH and Consejo Nacional de Ciencia y Tecnología (CONACYT-México) grant Ciencia Frontera 2019-87787.

(Received July 31, 2020)

REFERENCES

[1] A. Bensoussan, M. Cakanyildirim, and S. P. Sethi: Partially observed inventory systems: the case of zero-balance walk. SIAM J. Control Optim. *46* (2007), 176–209. DOI:10.1137/040620321

[2] N. Bauerle and U. Rieder: Markov Decision Processes with Applications to Finance. Springer, Berlin 2011.

[3] D- P. Bertsekas and S. E. Shreve: Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York 1978.

[4] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer-Verlag, New York 1979.

[5] R. J. Elliott, L. Aggoun, and J. B. Moore: Hidden Markov Models: Estimation and Control. Springer-Verlag, New York 1994.

[6] E. Gordienko and O. Hernandez-Lerma: Average cost Markov control processes with weighted norms: value iteration. Appl. Math. *23* (1995), 219—237. DOI:10.4064/am-23-2-219-237

[7] E. Gordienko and J. A. Minjarez-Sosa: Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. Kybernetika *34* (1998), 217—234.

[8] O. Hernandez-Lerma: Adaptive Markov Control Processes. Springer-Verlag, New York 1989.

[9] O. Hernandez-Lerma and M. Munoz-de-Ozak: Discrete-time Markov control processes with discounted unbounded costs: optimality criteria. Kybernetika *28* (1992), 191–221.

[10] M. Y. Kitaev and V. V. Rykov: Controlled Queueing Systems. CRC Press, Boca Raton 1995.

[11] D. V. Lindley: The theory of queues with a single server. Proc. Cambridge Philos Soc. *48* (1952), 277–289. DOI:10.1017/S0305004100027638

[12] N. Hilgert and J. A. Minjarez-Sosa: Adaptive policies for time-varying stochastic systems under discounted criterion. Math. Methods Oper. Res. *54* (2001), 491–505. DOI:10.1007/s001860100170

[13] J. A. Minjarez-Sosa: Approximation and estimation in Markov control processes under discounted criterion. Kybernetika *40* (2004), 681–690.

[14] J. A. Minjarez-Sosa: Markov control models with unknown random state-action-dependent discount factors. TOP *23* (2015), 743–772. DOI:10.1007/s11750-015-0360-5

[15] W. J. Runggaldier and L. Stettner: Approximations of Discrete Time Partially Observed Control Problems. Appl. Math. Monographs CNR 6, Giardini, Pisa 1994.

[16] L. I. Sennott: Stochastic Dynamic Programming and the Control of Queueing Systems. Wiley, New York 1999.

[17] C. Striebel: Optimal Control of Discrete Time Stochastic Systems. Lecture Notes Econ. Math. Syst. 110, Springer-Verlag, Berlin 1975.

[18] A. A. Yushkevich: Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces. Theory Probab. Appl.*21* (1976), 153–158. DOI:10.1137/1121014

Yofre H. García, ORCID:0000-0003-0243-1392, Universidad Autónoma de Chiapas, Ciudad Universitaria, Facultad en Física y Matemáticas, Carretera Emiliano Zapata km. 8, Rancho San Francisco, Col. Terán, 29050 Tuxtla Gutiérrez, Chiapas. México.
e-mail: yofregarcia@gmail.com

Saul Diaz-Infante, ORCID:0000-0001-9559-1293, CONACYT-Universidad de Sonora, Departamento de Matemáticas, Universidad de Sonora. Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora. México.
e-mail: saul.diazinfante@unison.mx

J. Adolfo Minjárez-Sosa, Corresponding author. Departamento de Matemáticas, Universidad de Sonora. Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora. México.
e-mail: aminjare@gauss.mat.uson.mx