E. Everardo Martinez-Garcia; J. Adolfo Minjárez-Sosa; Oscar Vega-Amaya

Partially observable Markov decision processes with partially observable random discount factors

Kybernetika, Vol. 58 (2022), No. 6, 960–983

Persistent URL: http://dml.cz/dmlcz/151538

KYBERNETIKA — VOLUME 58 (2022), NUMBER 6, PAGES 960–983

PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES WITH PARTIALLY OBSERVABLE RANDOM DISCOUNT FACTORS

E. Everardo Martinez-Garcia, J. Adolfo Minjárez-Sosa and Oscar Vega-Amaya

This paper deals with a class of partially observable discounted Markov decision processes defined on Borel state and action spaces, under unbounded one-stage cost. The discount rate is a stochastic process evolving according to a difference equation, which is also assumed to be partially observable. Introducing a suitable control model and filtering processes, we prove the existence of optimal control policies. In addition, we illustrate our results in a class of GI/GI/1 queueing systems where we obtain explicitly the corresponding optimality equation and the filtering process.

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

Classification: 90C39, 90B22

1. INTRODUCTION

The paper concerns a study of a class of discounted partially observable (PO) Markov decision processes (MDPs) on Borel spaces and unbounded costs with random discount factors. Specifically we consider a controlled process whose evolution is given by the system equations:

$$x_{t+1} = F_1\left(x_t, a_t, w_t^{(1)}\right), \quad y_t = F_2\left(x_t, w_t^{(2)}\right), \tag{1}$$

$$\alpha_{t+1} = G_1\left(\alpha_t, \xi_t^{(1)}\right), \quad \beta_t = G_2\left(\alpha_t, \xi_t^{(2)}\right) \quad t \in \mathbb{N}_0, \tag{2}$$

where $\{x_t\}$ and $\{\alpha_t\}$, $\alpha_t > 0$, are the state and discount processes, respectively, which are partially observable through processes $\{y_t\}$ and $\{\beta_t\}$ called state observation process and discount observation process, respectively. In addition, F_i and G_i, $i = 1, 2$, are known functions, a_t represents the control at time t, and $\{w_t^{(i)}\}$ and $\{\xi_t^{(i)}\}$, $i = 1, 2$, are independent sequences of independent and identically distributed (i.i.d.) random vectors.

DOI: 10.14736/kyb-2022-6-0960

The discount process $\{\alpha_t\}$ defines a performance index with random discount factor in the following sense. Let c be the one-stage cost function, then the cost incurred at time $t \in \mathbb{N}_0$ takes the form $c(x_0, a_0)$ for $t = 0$, and

$$e^{-\sum_{k=0}^{t-1} \alpha_k} c(x_t, a_t), \quad t \in \mathbb{N}. \tag{3}$$

The expectation of the accumulation of the costs (3) during the evolution of the system determines the optimality criterion we are interested in studying. That is, our objective is to show the existence of an optimal policy minimizing such a performance index, subject to the equations (1) – (2). Furthermore, we dedicate an important part of the paper to the application of our results to a class of GI/GI/1 queueing systems with controlled service rates.

To achieve our goal we follow the usual approach (see, e. g., [2,14,24,25] and references therein), but adapted to the nonstandard system (1) – (2). This approach consists of transforming the PO control problem into an equivalent completely observable (CO) control problem with nonconstant discount factor, evolving on the space $\mathbb{P}(X) \times \mathbb{P}(\Gamma)$, where $\mathbb{P}(X)$ and $\mathbb{P}(\Gamma)$ are the families of probability measures on the state space X and on the discount factor space Γ, respectively. In this case, the CO state process and the CO discount process are sequence of measures $\{\nu_t\} \subset \mathbb{P}(X)$ and $\{\eta_t\} \subset \mathbb{P}(\Gamma)$ where ν_t and η_t are the conditional distributions of x_t and α_t, respectively, given the observed history.

Essentially, the transformation PO→CO is based in the application of suitable filtering techniques with which it is possible to prove the existence of functions Ψ and Φ such that

$$\nu_{t+1} = \Psi(\nu_t, a_t, y_{t+1}), \quad \eta_{t+1} = \Phi(\eta_t, \beta_{t+1}), \quad t \geq 0.$$

Hence, by appropriately analyzing the coupled process $\{(\nu_t, \eta_t)\} \subset \mathbb{P}(X) \times \mathbb{P}(\Gamma)$, we can apply the dynamic programming approach to solve the CO control problem.

It is worth emphasizing that this transformation procedure is merely theoretical and only guarantees the existence of the functions Ψ and Φ. Therefore, an important challenge, from the applications point of view, is to explicitly exhibit such functions in specific situations, which could be a non-trivial problem. Thus, as an additional result, in this paper we extensively illustrate the transformation procedure PO→CO in a G1/G1/1 queueing system with controlled service rate and PO waiting times. In this case, to obtain the functions Ψ and Φ, we assume that the conditional distributions ν_t and η_t have densities, which yields that our analysis is done on suitable spaces of density functions.

The discounted optimality criterion has been widely studied under several settings (see, e. g., [3,4,7–9,11–13,18,19,21,22,26] and references therein); in fact, within the field of applications, it is one of the most studied performance indices because the natural economic and financial interpretation of the discount factor as function of the interest rate. It is precisely by this interpretation that many recent paper have addressed the problem of assuming nonconstant and/or random discount factors (see [3, 7–9, 19, 20,

22, 26]). Hence, our contribution is framed in this aspect. However, to the best of our knowledge, partially observable MDPs with discount factors modelled by a partially observable stochastic process have not been previously studied.

The remainder of the paper is organized as follows. In Section 2 we define the PO system we are interested in. Next, in Section 3 we introduce the transformation procedure which yields the CO control problem whose solution is analyzed in Section 4. In addition, Section 5 contains the analysis of a PO queueing system with controlled service rate. Finally we conclude with some remarks.

1.1. Notation and terminology

Throughout the paper we will use the following notation and terminology

Symbols

- $\mathbb{N}$, set of positive integers.
- $\mathbb{N}_0$, set of nonnegative integers. So, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.
- $\Re$, set of real numbers.
- $\Re^+$, set of nonnegative real numbers.

Functions

- $(x)^+ := \max\{0, x\}$.
- $I_D(\cdot)$, indicator function of the set D.
- $\delta_d(x) = \begin{cases} 1, & x = d; \\ 0, & \text{otherwise.} \end{cases}$
  Observe that $\delta_0(\cdot) = \delta(\cdot)$ is the Dirac delta function.

Space of functions

A Borel space is a Borel subset of a complete and separable metric space. Let W be a Borel space.

- $\mathcal{B}(W)$, Borel $\sigma-$algebra in W. Further, "measurability" for sets and functions always means measurability with respect to $\mathcal{B}(W)$.
- $\mathbb{P}(W)$, set of all probability measures on W, which is also a Borel space with respect to the weak topology.
- $\mathcal{L}(W)$, space of lower semicontinuous (l.s.c.) and bounded below functions on W.
- $\mathcal{L}_+(W)$, space of nonnegative and l.s.c. functions. Hence $\mathcal{L}_+(W) \subset \mathcal{L}(W)$.
- Given the Borel space W', a stochastic kernel $\varphi(\cdot|\cdot)$ on W given W' is a function such that $\varphi(\cdot|w') \in \mathbb{P}(W)$ for each $w' \in W'$, and $\varphi(B|\cdot)$ is a measurable function on W' for each $B \in \mathcal{B}(W)$. We denote by $\mathbb{P}(W|W')$ the family of stochastic kernels on W given W'.

2. THE PARTIALLY OBSERVABLE SYSTEM

We consider a partially observable (PO) Markov decision process evolving according to the system equations

$$x_{t+1} = F_1\left(x_t, a_t, w_t^{(1)}\right), \tag{4}$$

$$y_t = F_2\left(x_t, w_t^{(2)}\right), \quad t \in \mathbb{N}_0, \tag{5}$$

x_0 given or with known distribution $\nu \in \mathbb{P}(X)$, where F_1 and F_2 are known functions, x_t, a_t, and y_t represent the state, control, and observation at time t, taking values in Borel spaces X, A, and Y, respectively. In addition $\left\{w_t^{(1)}\right\}$ and $\left\{w_t^{(2)}\right\}$ are independent sequences of independent and identically distributed (i.i.d.) random vectors with values in Borel spaces S_1 and S_2 with distributions $\theta_1 \in \mathbb{P}(S_1)$ and $\theta_2 \in \mathbb{P}(S_2)$ respectively. Both sequences are also assumed to be independent of the initial state $x_0 \in X$. Let $Q \in \mathbb{P}(X|X \times A)$ and $K \in \mathbb{P}(Y|X)$ be the state transition kernel and the observation kernel determined by F_1 and F_2. That is, for $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(Y)$, $x \in X$, and $a \in A$,

$$\begin{aligned} Q(B|x,a) &:= \Pr\left(x_{t+1} \in B | x_t = x, a_t = a\right) \\ &= \int_{S_1} I_B\left[F_1(x,a,s)\right] \theta_1(\mathrm{d}s), \end{aligned} \tag{6}$$

and

$$\begin{aligned} K(C|x) &:= \Pr\left(y_t \in C | x_t = x\right) \\ &= \int_{S_2} I_C\left[F_2(x,s)\right] \theta_2(\mathrm{d}s). \end{aligned} \tag{7}$$

We denote by $c : X \times A \to \Re$ the one-stage cost which is a nonnegative continuous function. In addition, we consider a stochastic process $\{\alpha_t\}$ representing the discount rate whose evolution is given by the equation

$$\alpha_{t+1} = G_1\left(\alpha_t, \xi_t^{(1)}\right), \quad t \in \mathbb{N}_0, \tag{8}$$

α_0 given or with known distribution $\eta \in \mathbb{P}(\Gamma)$, where $\alpha_t \in \Gamma := (0, \infty)$, $\left\{\xi_t^{(1)}\right\}$ is a sequence of i.i.d. random vectors with values in a Borel space R_1 with common distribution $\rho_1 \in \mathbb{P}(R_1)$ and $G_1 : \Gamma \times R_1 \to \Gamma$ is a known function. The discount process plays the following role. Let $\epsilon(\alpha) := e^{-\alpha}$, $\alpha \in \Gamma$. Then the discounted cost incurred at time t is $c(x_0, a_0)$ for $t = 0$, and for $t \in \mathbb{N}$,

$$\epsilon(\alpha_0)\,\epsilon(\alpha_1) \cdots \epsilon(\alpha_{t-1})\, c(x_t, a_t). \tag{9}$$

In the setting of our problem, we assume that the discount process $\{\alpha_t\}$, given in (8), is partially observable with observation process defined as

$$\beta_t = G_2\left(\alpha_t, \xi_t^{(2)}\right) \quad t \in \mathbb{N}_0, \tag{10}$$

where $\beta_t \in \Sigma := (0, \infty)$, $G_2 : \Gamma \times R_2 \to \Sigma$ is a known function, $\left\{\xi_t^{(2)}\right\}$ is a sequence of i.i.d. random vectors, independent on $\left\{\xi_t^{(1)}\right\}$, with values in a Borel space R_2 and distribution $\rho_2 \in \mathbb{P}(R_2)$. Similarly as (6) – (7), let $Q' \in \mathbb{P}(\Gamma|\Gamma)$ and $K' \in \mathbb{P}(\Sigma|\Gamma)$ be the discount factor transition kernel and the corresponding observation kernel defined by G_1 and G_2 as

$$\begin{aligned} Q'(B'|\alpha) &:= \Pr(\alpha_{t+1} \in B'|\alpha_t = \alpha) \\ &= \int_{R_1} I_{B'}[G_1(\alpha, s)]\, \rho_1(\mathrm{d}s),\ B' \in \mathcal{B}(\Gamma), \alpha \in \Gamma, \end{aligned} \tag{11}$$

and

$$\begin{aligned} K'(C'|\alpha) &:= \Pr(\beta_t \in C'|\alpha_t = \alpha) \\ &= \int_{R_2} I_{C'}[G_2(\alpha, s)]\, \theta_2(\mathrm{d}s),\ C' \in \mathcal{B}(\Sigma), \alpha \in \Gamma. \end{aligned} \tag{12}$$

The previous elements define the following PO control model:

$$\mathcal{M}_{PO} := (X, Y, A, Q, K, \nu, \Gamma, \Sigma, Q', K', \eta, \epsilon, c). \tag{13}$$

The model $\mathcal{M}_{PO}$ has the following interpretation. At time $t = 0$, the initial state has a given distribution $\nu \in \mathbb{P}(X)$. A state observation y_0 is generated according to the observation kernel K and an action $a_0 \in A$ is selected. Next, the cost $c(x_0, a_0)$ is incurred and an initial discount factor $\alpha_0 \in \Gamma$ with initial distribution $\eta \in \mathbb{P}(\Gamma)$ comes in with observation β_0 generated by the kernel K'. Then, considering the kernel Q, the system moves to a state $x_1 \in X$. After that, a new state observation y_1 is obtained according to the kernel K and the action $a_1 \in A$ is selected. Then a discounted cost $\epsilon(\alpha_0)c(x_1, a_1)$ is incurred and a new discount factor $\alpha_1 \in \Gamma$ comes in with corresponding observation β_1 generated by the kernel K'. In general, at time $t \in \mathbb{N}$, when the system is in state $x_t \in X$, a state-observation $y_t \in Y$ is generated according to the kernel K. Next, a control $a_t \in A$ is selected, a discounted cost as (9) is incurred, and a new observation $\beta_t \in \Sigma$ is generated according to the kernel K' of the discount factor $\alpha_t \in \Gamma$. The system jumps to a new state and the process is repeated over and over again.

According to the previous description, the actions are selected taking into account the observed history. In this sense, let

$$\mathcal{Z}_0 := \sigma(y_0), \quad \mathcal{Z}_t := \sigma(z_0, z_1, ..., z_{t-1}, y_t),\ t \in \mathbb{N}, \text{ where } z_k = (y_k, \beta_k) \in Y \times \Sigma, \tag{14}$$

be the $\sigma-$algebra generated by the observations up to time t. Hence, a control policy is a sequence $\pi = \{a_t\}$ of $A-$valued random vectors such that a_t is $\mathcal{Z}_t-$measurable for each $t \in \mathbb{N}_0$. We denote by Π the set of all control policies.

According to (9), for each policy $\pi \in \Pi$ and initial distributions $\nu \in \mathbb{P}(X)$ and $\eta \in \mathbb{P}(\Gamma)$ of x_0 and α_0, respectively, we define the total expected discounted cost as

$$V(\pi, \nu, \eta) = E^{\pi}_{\nu,\eta} \sum_{t=0}^{\infty} \Lambda_t c(x_t, a_t), \tag{15}$$

where

$$\Lambda_t := \prod_{k=0}^{t-1} \epsilon(\alpha_k) \text{ for } t \in \mathbb{N}, \text{ and } \Lambda_0 := 1, \tag{16}$$

and $E^{\pi}_{\nu,\eta}$ is the expectation operator with respect to a probability measure $P^{\pi}_{\nu,\eta}$ induced by $\pi \in \Pi$ and $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$ (see, e. g., [5] for construction of $P^{\pi}_{\nu,\eta}$). Thus, if

$$V^*(\nu,\eta) := \inf_{\pi\in\Pi} V(\pi,\nu,\eta) \tag{17}$$

is the optimal cost, the PO optimal control problem is to find a policy $\pi^* \in \Pi$ satisfying

$$V^*(\nu,\eta) = V(\pi^*,\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma). \tag{18}$$

To avoid trivial situations, we assume that there is a policy $\pi \in \Pi$ such that $V(\pi,\nu,\eta) < \infty$, $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$.

3. THE COMPLETELY OBSERVABLE CONTROL PROBLEM

Following a standard procedure (see, e. g., [2,14,25]) the study of the PO optimal control problem (17) – (18) is based on its transformation into a completely observable (CO) optimal control problem by introducing suitable filtering processes $\{\nu_t\} \in \mathbb{P}(X)$ and $\{\eta_t\} \in \mathbb{P}(\Gamma)$, for both the state and the discount processes. Specifically, for each $\pi \in \Pi$, initial distributions $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$, and $B \in \mathcal{B}(X)$, let

$$\nu_0(B) := P^{\pi}_{\nu,\eta}(x_0 \in B) = \nu(B) \tag{19}$$

and

$$\nu_t(B) := P^{\pi}_{\nu,\eta}(x_t \in B | \mathcal{Z}_t), \quad t \in \mathbb{N}. \tag{20}$$

Then (see , e. g., [2,14,25]) there exists a measurable function $\Psi : \mathbb{P}(X) \times A \times Y \to \mathbb{P}(X)$ such that the process $\{\nu_t\} \in \mathbb{P}(X)$ satisfies

$$\nu_{t+1} = \Psi(\nu_t, a_t, y_{t+1}), \quad t \in \mathbb{N}_0, \tag{21}$$

with initial condition $\nu_0 = \nu$. Now let $k \in \mathbb{P}(Y|\mathbb{P}(X) \times A)$ be defined as

$$k(C|\nu,a) := \Pr[y_{t+1} \in C | \nu_t = \nu, a_t = a], \quad C \in \mathcal{B}(Y), (\nu,a) \in \mathbb{P}(X) \times A. \tag{22}$$

Then, from (6) and (7) we have

$$\begin{aligned} k(C|\nu,a) &= \int_X \int_X K(C|x') Q(\mathrm{d}x'|x,a)\nu(\mathrm{d}x) \\ &= \int_X \int_{S_1} \int_{S_2} I_C[F_2(F_1(x,a,s),s')] \theta_2(\mathrm{d}s')\theta_1(\mathrm{d}s)\nu(\mathrm{d}x). \end{aligned} \tag{23}$$

In addition, let $q \in \mathbb{P}(\mathbb{P}(X)|\mathbb{P}(X) \times A)$ be the transition kernel corresponding to the filtering process $\{\nu_t\}$, that is, for $D \in \mathcal{B}(\mathbb{P}(X)), (\nu,a) \in \mathbb{P}(X) \times A$,

$$q(D|\nu,a) := \Pr[\nu_{t+1} \in D | \nu_t = \nu, a_t = a]. \tag{24}$$

From (21) – (23) we have,

$$q\left(D|\nu,a\right)=\int_{Y}I_{D}\left[\Psi(\nu,a,y)\right]k\left(\mathrm{d}y|\nu,a\right),\ D\in\mathcal{B}(\mathbb{P}(X)),(\nu,a)\in\mathbb{P}(X)\times A. \tag{25}$$

For the filtering process $\{\eta_t\}\in\mathbb{P}(\Gamma)$ we proceed similar as (19) – (25). Indeed, for each $\pi\in\Pi$, initial distributions $(\nu,\eta)\in\mathbb{P}(X)\times\mathbb{P}(\Gamma)$, and $B'\in\mathcal{B}(\Gamma)$, we define

$$\eta_0(B'):=P_{\nu,\eta}^{\pi}\left(\alpha_0\in B'\right)=\eta(B') \tag{26}$$

and

$$\eta_t(B'):=P_{\nu,\eta}^{\pi}\left(\alpha_t\in B'|\mathcal{Z}_t,\beta_t\right),\quad t\in\mathbb{N}. \tag{27}$$

The process $\{\eta_t\}$ satisfies

$$\eta_{t+1}=\Phi(\eta_t,\beta_{t+1}),\quad t\in\mathbb{N}_0, \tag{28}$$

with initial condition $\eta_0=\eta$, for some measurable function $\Phi:\mathbb{P}(\Gamma)\times\Sigma\to\mathbb{P}(\Gamma)$. We define the kernels $k'\in\mathbb{P}\left(\Sigma|\mathbb{P}(\Gamma)\right)$ and $q'\in\mathbb{P}\left(\mathbb{P}(\Gamma)|\mathbb{P}(\Gamma)\right)$ as

$$k'\left(C'|\eta\right):=\Pr\left[\beta_{t+1}\in C'|\eta_t=\eta\right],\quad C'\in\mathcal{B}(\Sigma),\eta\in\mathbb{P}(\Gamma) \tag{29}$$

and

$$q'\left(D'|\eta\right):=\Pr\left[\eta_{t+1}\in D'|\eta_t=\eta\right],\quad D'\in\mathcal{B}(\mathbb{P}(\Gamma)),\eta\in\mathbb{P}(\Gamma). \tag{30}$$

Then, from (11) and (12) the kernels k' and q' can be written as

$$\begin{aligned}k'\left(C'|\eta\right) &= \int_{\Gamma}\int_{\Gamma}K'\left(C'|\alpha'\right)Q'(\mathrm{d}\alpha'|\alpha)\eta(\mathrm{d}\alpha)\\ &= \int_{\Gamma}\int_{R_1}\int_{R_2}I_{C'}\left[G_2\left(G_1\left(\alpha,s\right),s'\right)\right]\rho_2(\mathrm{d}s')\rho_1(\mathrm{d}s)\eta(\mathrm{d}\alpha)\end{aligned} \tag{31}$$

and

$$q'\left(D'|\eta\right)=\int_{\Sigma}I_{D'}\left[\Phi(\eta,\beta)\right]k'\left(\mathrm{d}\beta|\eta\right),\ D\in\mathcal{B}(\mathbb{P}(\Gamma)),\eta\in\mathbb{P}(\Gamma). \tag{32}$$

On the other hand, we define the one-stage cost $\tilde{c}:\mathbb{P}(X)\times A\to\Re$ as

$$\tilde{c}(\nu,a):=\int_{X}c(x,a)\nu(\mathrm{d}x) \tag{33}$$

and the function $\tilde{\epsilon}:\mathbb{P}(\Gamma)\to(0,\infty)$ as

$$\tilde{\epsilon}\left(\eta\right):=\int_{\Gamma}\epsilon(\alpha)\eta(\mathrm{d}\alpha). \tag{34}$$

From (33) and (34), for each policy $\pi\in\Pi$ and initial distributions $(\nu,\eta)\in\mathbb{P}(X)\times\mathbb{P}(\Gamma)$, we can write the performance index (15) as

$$V(\pi,\nu,\eta)=E_{\nu,\eta}^{\pi}\sum_{t=0}^{\infty}\tilde{\Lambda}_t\tilde{c}(\nu_t,a_t), \tag{35}$$

where

$$\tilde{\Lambda}_t := \prod_{k=0}^{t-1} \tilde{\epsilon}\,(\eta_k) \text{ for } t \in \mathbb{N}, \text{ and } \tilde{\Lambda}_0 := 1. \tag{36}$$

Thus, we can define the following CO optimal control problem which consists of finding a policy $\pi^* \in \Pi$ such that

$$V^*(\nu,\eta) := \inf_{\pi\in\Pi} V(\pi,\nu,\eta) = V(\pi^*,\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X)\times\mathbb{P}(\Gamma), \tag{37}$$

subject to (21) and (28). Furthermore, following standard procedures given, for instance, in [2, 14, 25], we have that a solution for the CO control problem (37) is also a solution of the PO control problem (18). In other words, both CO and PO control problems are equivalent.

4. SOLUTION OF THE COMPLETELY OBSERVABLE CONTROL PROBLEM

For a function $u : \mathbb{P}(X)\times\mathbb{P}(\Gamma) \to \Re$ we define the operator

$$Tu(\nu,\eta) = \min_{a\in A} T_a u(\nu,\eta),$$

where, for $(\nu_0,\eta_0) = (\nu,\eta)$

$$\begin{aligned} T_a u(\nu,\eta) &:= \tilde{c}(\nu,a) + \tilde{\epsilon}\,(\eta)\, E^{\pi}_{\nu,\eta}\left[u(\nu_1,\eta_1)\right] \\ &= \tilde{c}(\nu,a) + \tilde{\epsilon}\,(\eta)\, E^{\pi}_{\nu,\eta}\left[u(\Psi\,(\nu,a,y_1)\,,\Phi\,(\eta,\beta_1))\right] \\ &= \tilde{c}(\nu,a) + \tilde{\epsilon}\,(\eta)\int_{\mathbb{P}(\Gamma)}\int_{\mathbb{P}(X)} u(\nu',\eta')q\,(\mathrm{d}\nu'|\nu,a)\,q'\,(\mathrm{d}\eta'|\eta)\,. \end{aligned} \tag{38}$$

We also consider the following value iteration functions $\{v_t\}$ defined as

$$\begin{aligned} v_0 &= 0; \\ v_t(\nu,\eta) &:= Tv_{t-1}(\nu,\eta), \quad t \in \mathbb{N}. \end{aligned} \tag{39}$$

By applying dynamic programming arguments, it is easy to prove that v_t is the optimal cost for a $t-$stage optimal control problem. That is, defining

$$V_t(\pi,\nu,\eta) = E^{\pi}_{\nu,\eta}\sum_{k=0}^{t} \tilde{\Lambda}_k \tilde{c}(\nu_k,a_k), \tag{40}$$

we have

$$v_t(\nu,\eta) = \inf_{\pi\in\Pi} V_t(\pi,\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X)\times\mathbb{P}(\Gamma). \tag{41}$$

Assumption 4.1. (a) $\tilde{c}$ is inf-compact on $\mathbb{P}(X)\times A$, that is, the set $\left\{a \in A : \tilde{c}(\nu,a) \le r\right\}$ is compact for every $\nu \in \mathbb{P}(X)$ and $r \in \Re$;

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

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

Remark 4.2. (a) From [15, Lemma 3.2(f)] (see also [16, Lemma 2.7]), under Assumption 4.1 we have that for all $u \in \mathcal{L}_+\big(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\big)$ the function $u^*(\nu,\eta) := Tu(\nu,\eta)$ belongs to $\mathcal{L}_+\big(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\big)$. In addition, there exists a measurable function $f^* : \mathbb{P}(X)\times\mathbb{P}(\Gamma)\to A$ such that $u^*(\nu,\eta) = T_{f^*}u(\nu,\eta)$.

(b) It is clear that if the control set A is compact and $\tilde{c}$ is lower semicontinuous in $a \in A$, then $\tilde{c}$ is inf–compact.

Theorem 4.3. If Assumption 4.1 holds, then:

(a) The optimal cost V^* defined in (37) is the minimal solution in $\mathcal{L}_+\left(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\right)$ of the optimality equation, that is

$$V^*(\nu,\eta) = TV^*(\nu,\eta) = \min_{a\in A} T_a V^*(\nu,\eta), \quad (\nu,\eta)\in\mathbb{P}(X)\times\mathbb{P}(\Gamma), \tag{42}$$

and if $\tilde{V} \in \mathcal{L}_+\left(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\right)$ is another solution of the optimality equation then $V^* \le \tilde{V}$. Furthermore, $v_t \nearrow V^*$ as $t\to\infty$, where $\{v_t\}$ is the sequence of functions defined in (39).

(b) There exists a measurable function $f^* : \mathbb{P}(X)\times\mathbb{P}(\Gamma)\to A$ such that

$$V^*(\nu,\eta) = T_{f^*}V^*(\nu,\eta), \quad (\nu,\eta)\in\mathbb{P}(X)\times\mathbb{P}(\Gamma),$$

and the policy $\pi^* = \{a_t^*\}\in\Pi$ defined by $a_t^* = f^*(\nu_t,\eta_t)$, $t\in\mathbb{N}_0$, is optimal.

P r o o f. Observe that $\{v_t\}$ is a nondecreasing sequence in $\mathcal{L}_+\left(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\right)$. This implies that there exists a function $u\in\mathcal{L}_+\left(\mathbb{P}(X)\times\mathbb{P}(\Gamma)\right)$ such that $v_t \nearrow u$ as $t\to\infty$. Hence, for each $a\in A$ and $(\nu,\eta)\in\mathbb{P}(X)\times\mathbb{P}(\Gamma)$,

$$T_a v_{t-1}(\nu,\eta) \nearrow T_a u(\nu,\eta), \quad \text{as} \quad t\to\infty,$$

which implies, from [17, Lemma 4.2.4],

$$v_t(\nu,\eta) = \min_{a\in A} T_a v_{t-1}(\nu,\eta) \to \min_{a\in A} T_a u(\nu,\eta) = Tu(\nu,\eta).$$

Thus, since $v_t \nearrow u$ as $t\to\infty$, we get

$$u = Tu. \tag{43}$$

Now let $f : \mathbb{P}(X)\times\mathbb{P}(\Gamma)\to A$ be a measurable function such that (See Remark 4.2)

$$u(\nu,\eta) = \tilde{c}(\nu,f) + \tilde{\epsilon}\,(\eta)\int_{\mathbb{P}(\Gamma)}\int_{\mathbb{P}(X)} u(\nu',\eta')q\,(\mathrm{d}\nu'|\nu,f)\,q'\,(\mathrm{d}\eta'|\eta)\,.$$

Iteration of this inequality yields, for $\pi=\{a_t\}\in\Pi$, $a_t = f(\nu_t,\eta_t)$, $t\in\mathbb{N}_0$,

$$\begin{aligned} u(\nu,\eta) &= E^\pi_{\nu,\eta}\sum_{t=0}^{n-1}\tilde{\Lambda}_t\tilde{c}(\nu_t,a_t) + E^\pi_{\nu,\eta}\prod_{k=0}^{n-1}\tilde{\epsilon}\,(\eta_k)\,u(\nu_k,a_k) \\ &\ge E^\pi_{\nu,\eta}\sum_{t=0}^{n-1}\tilde{\Lambda}_t\tilde{c}(\nu_t,a_t), \end{aligned} \tag{44}$$

where the last inequality is because u is nonnegative. Letting $n \to \infty$, from (37) we obtain

$$u(\nu,\eta) \geq V(\pi,\nu,\eta) \geq V^*(\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma). \tag{45}$$

On the other hand, from (40) and (41), for all $\pi \in \Pi$, $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$, and $t \in \mathbb{N}_0$,

$$v_t(\nu,\eta) \leq V_t(\pi,\nu,\eta) \leq V(\pi,\nu,\eta).$$

Thus, letting $t \to \infty$, as $v_t \nearrow u$, we get

$$u(\nu,\eta) \leq V(\pi,\nu,\eta)$$

for all $\pi \in \Pi$ and $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$. This yields

$$u(\nu,\eta) \leq V^*(\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma). \tag{46}$$

Combining (45) and (46) we prove

$$u(\nu,\eta) = V^*(\nu,\eta), \quad (\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma). \tag{47}$$

Hence, from (43), V^* is a solution of the optimality equation and $v_t \nearrow V^*$ as $t \to \infty$. Now, if $\tilde{V} \in \mathcal{L}_+(\mathbb{P}(X) \times \mathbb{P}(\Gamma))$ is another solution of the optimality equation, from (45) with $\tilde{V}$ instead of u we obtain that V^* is the minimal solution in $\mathcal{L}_+(\mathbb{P}(X) \times \mathbb{P}(\Gamma))$ of the optimality equation. This proves the part (a).

Finally, the existence of f^* follows from Remark 4.2. Furthermore, using the fact $u = V^*$, from (44) we have, for $\pi^* = \{a_t^*\}$ defined by $a_t^* = f^*(\nu_t, \eta_t)$, $t \in \mathbb{N}_0$,

$$V^*(\nu,\eta) \geq E_{\nu,\eta}^{\pi^*} \sum_{t=0}^{n-1} \tilde{\Lambda}_t \tilde{c}(\nu_t, a_t),$$

which, letting $n \to \infty$, yields $V^*(\nu,\eta) \geq V(\pi^*,\nu,\eta)$ for all $(\nu,\eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$. Therefore, from (37), π^* is optimal. □

Remark 4.4. There are situations where is convenient to express the filtering processes (19) – (20) and (26) – (27) in terms of densities. That is, assume that the measures ν_t and η_t are absolutely continuous with respect to the Lebesgue measure in $\Re$. This particular case is important from the practical point of view because it allows to obtain, in an easy way, the corresponding recursive equation (21) and (28) for specific application problems. Of course, the definition and solution of the corresponding CO optimal control problem depend on the observation process and must be done under the right assumptions. First of all, the state space is a subset of $\Re$. Thus it is necessary to reformulate Assumption 4.1. For instance, for the lower semicontinuous conditions we need to set down the notion of convergence in the set of densities, as will be specified in the study of the PO $GI/GI/1$ queueing system in next section.

5. EXAMPLE: A PO CONTROLLED QUEUEING SYSTEM

We denote by $\mathbb{D}_d$ the set of all density functions g on (d, ∞). In particular, $\mathbb{D}_0 := \mathbb{D}$ is the set of all densities on $\Re^+$. We consider the L_1-norm

$$\|g\|_{(d)} := \int_{\mathrm{d}}^{\infty} |g(s)| \, \mathrm{d}s.$$

Observe that $\mathbb{D}_d$ is a closed subset of $L_1(d, \infty)$, i. e., if $\{g_t\}$ is a sequence in $\mathbb{D}_d$ such that

$$\|g_t - g\|_{(d)} \to 0, \quad \text{as} \quad t \to \infty, \tag{48}$$

then $g \in \mathbb{D}_d$. We denote $g_t \to g$ when (48) holds.

We consider the PO $GI/GI/1$ queueing system introduced in [10], which evolves as

$$x_{t+1} = \left(x_t + a_t \bar{w}_t - w_t^{(1)}\right)^+ := \max\left\{x_t + a_t \bar{w}_t - w_t^{(1)}, 0\right\}, \quad t \in \mathbb{N}, \tag{49}$$

with observation process

$$y_t := I_{[x_t = 0]}, \quad t \in \mathbb{N}_0. \tag{50}$$

Here, x_t and $w_t^{(1)}$ represent the waiting time of the $t-$th customer and the interarrival time between the $t-$th and $(t+1)-$th customers, taking values in $X = S_1 = [0, \infty)$, respectively. In addition, $\bar{w}_t$ is a random "base" service time of the $t-$th customer and a_t, the control, is the reciprocal of a service rate u_t, i. e., $a_t = 1/u_t$, taking values in $A = [a_*, a^*]$, $a^* > a_* > 0$. We assume that the waiting time x_t only is observed when $x_t = 0$, that is when $y_t = 1$ (see (50)), which means that the controller only can register when the customer arrives directly to the server.

In setting of our results, we suppose that the discount process $\{\alpha_t\}$ evolves as an autoregressive process of the form

$$\alpha_{t+1} = \max\left\{\gamma \alpha_t + \xi_t^{(1)}, d\right\}, \quad t \in \mathbb{N}_0, \tag{51}$$

with observation process

$$\beta_t = I_{[\alpha_t = d]}, \tag{52}$$

where $\gamma > 0$, $\left\{\xi_t^{(1)}\right\}$ is a sequence of i.i.d. random variables taking values in $R_1 = [0, \infty)$, and $d > 0$ is a fixed constant. We assume that the discount factor only is observed while $\alpha_t = d$. Therefore, the events $[\beta_t = 1]$, $[\alpha_t = d]$, and $\left[\gamma \alpha_{t-1} + \xi_{t-1}^{(1)} \leq d\right]$ are equivalent.

Specifically, the controller only observes the events $[x_t = 0]$ and $[\alpha_t = d]$, defining a PO control system with PO discount factor. Such a control system will be analyzed assuming existence of densities for the involved random variables (see Remark 4.4), for which we impose the following assumptions.

Assumption 5.1. (a) $\{\bar{w}_t\}$, $\{w_t^{(1)}\}$, and $\{\xi_t^{(1)}\}$ are independent sequences of nonnegative i.i.d. random variables with distribution functions $F_{\bar{w}}$, F_w, and F_ξ, and continuous density functions $f_{\bar{w}}$, f_w, and f_ξ in $\mathbb{D}$.

(b) At initial time, if $x_0 = 0$ the initial state is observed, and if $x_0 > 0$, x_0 has a density function $g^1 \in \mathbb{D}$. Similarly, at initial time, either $\alpha_0 \leq d$, and therefore it is observed $\alpha_0 = d$, or $\alpha_0 > d$ and it has a density $g^2 \in \mathbb{D}_d$.

(c) For $t \in \mathbb{N}$, when $x_t > 0$ and $\alpha_t > d$ both are non observed, and they have conditional densities $g_t^1 \in \mathbb{D}$ and $g_t^2 \in \mathbb{D}_d$ given the observed history. That is (see (14))

$$\int_0^x g_t^1(\omega)\,\mathrm{d}\omega = P\left[x_t \leq x | x_t > 0, \mathcal{Z}_{t-1}\right]$$

and

$$\int_{\mathrm{d}}^{\alpha} g_t^2(s)\,\mathrm{d}s = P\left[\alpha_t \leq \alpha | \alpha_t > \mathrm{d}, \mathcal{Z}_t\right]. \tag{53}$$

(d) The cost function $c(\cdot,\cdot)$ is continuous and nonnegative. In addition, the function

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

is continuous.

It is worth noting that under Assumption 5.1 the initial condition of the PO queueing system is a vector $\left(y_0, g_0^1, \beta_0, g_0^2\right) = \left(y, g^1, \beta, g^2\right) \in \{0,1\} \times \mathbb{D} \times \{0,1\} \times \mathbb{D}_d$ because:

- if $y_0 = 1$ and $\beta_0 = 1$, then $x_0 = 0$ and $\alpha_0 = d$ are observed;
- if $y_0 = 1$ and $\beta_0 = 0$, then $x_0 = 0$ is observed and $\alpha_0 > d$ with density $g^2 \in \mathbb{D}_d$;
- if $y_0 = 0$ and $\beta_0 = 1$, then $x_0 > 0$ with density $g^1 \in \mathbb{D}$ and $\alpha_0 = d$ is observed;
- if $y_0 = 0$ and $\beta_0 = 0$, then $x_0 > 0$ with density $g^1 \in \mathbb{D}$ and $\alpha_0 > d$ with density $g^2 \in \mathbb{D}_d$.

Therefore, the initial distribution $(\nu, \eta) \in \mathbb{P}(X) \times \mathbb{P}(\Gamma)$ for $(x_0, \alpha_0) \in X \times \Gamma$ takes the form (see (19), (26))

$$\nu(B) := y\delta_0(x_0) + (1-y)\int_B g^1(\omega)\,\mathrm{d}\omega, \quad B \in \mathcal{B}(X)$$

and

$$\eta(B') := \beta\delta_{\mathrm{d}}(\alpha_0) + (1-\beta)\int_{B'} g^2(\omega)\,\mathrm{d}\omega, \quad B' \in \mathcal{B}(\Gamma).$$

Furthermore, the filtering processes (19) – (20) and (26) – (27) takes the form

$$\begin{aligned}
\nu_0(B) &:= P_{\nu,\eta}^{\pi}\left(x_0 \in B\right) = \nu(B), \\
\nu_t(B) &:= P_{\nu,\eta}^{\pi}\left(x_t \in B | \mathcal{Z}_t\right) = y_t\delta_0(x_t) + (1-y_t)\int_B g_t^1(\omega)\,\mathrm{d}\omega, \quad t \in \mathbb{N}, B \in \mathcal{B}(X);
\end{aligned}$$

and

$$\begin{aligned}\eta_0(B') &:= P^{\pi}_{\nu,\eta}(\alpha_0 \in B') = \eta(B')\\ \eta_t(B') &:= P^{\pi}_{\nu,\eta}(\alpha_t \in B' | \mathcal{Z}_t, \beta_t) = \beta_t \delta_d(\alpha_t) + (1-\beta_t)\int_{B'} g_t^2(\omega)\,\mathrm{d}\omega, \quad t \in \mathbb{N}, B' \in \mathcal{B}(\Gamma).\end{aligned}$$

Hence, both $\{\nu_t\}$ and $\{\eta_t\}$ are completely determined by the pairs $\left(y_t, g_t^1\right)$ and $\left(\beta_t, g_t^2\right)$ respectively. Under this context, and abusing the notation, we can consider the cost function $\tilde{c}$ and the discount factor function $\tilde{\epsilon}$, defined in (33) and (34) as

$$\begin{aligned}\tilde{c}(\nu_t, a_t) &= \int_0^{\infty} c(x, a_t)\nu_t(\mathrm{d}x) := \tilde{c}\left(y_t, g_t^1, a_t\right)\\ &= y_t c(0, a_t) + (1-y_t)\int_0^{\infty} c(x, a_t) g_t^1(x)\,\mathrm{d}x \qquad (54)\end{aligned}$$

and

$$\begin{aligned}\tilde{\epsilon}(\eta_t) &= \int_0^{\infty} \epsilon(\alpha)\eta_t(\mathrm{d}\alpha) := \tilde{\epsilon}\left(\beta_t, g_t^2\right)\\ &= \beta_t \epsilon(d) + (1-\beta_t)\int_{\mathrm{d}}^{\infty} \epsilon(\alpha) g_t^2(\alpha)\,\mathrm{d}\alpha. \qquad (55)\end{aligned}$$

Taking into account these facts, if $\left(y, g^1, \beta, g^2\right) \in \{0,1\} \times \mathbb{D} \times \{0,1\} \times \mathbb{D}_d$ is the initial condition, we express the performance index (35) as

$$V(\pi, y, g^1, \beta, g^2) = E\sum_{t=0}^{\infty} \tilde{\Lambda}_t \tilde{c}(y_t, g_t^1, a_t),$$

where

$$\tilde{\Lambda}_t := \prod_{k=0}^{t-1} \tilde{\epsilon}\left(\beta_k, g_k^2\right) \text{ for } t \in \mathbb{N}, \text{ and } \tilde{\Lambda}_0 := 1,$$

and E is the corresponding expectation operator $E^{\pi}_{y,g^1,\beta,g^2}$. Furthermore, a policy $\pi^* \in \Pi$ is optimal for the CO queueing control problem if (see (37)) for all $\left(y, g^1, \beta, g^2\right) \in \{0,1\} \times \mathbb{D} \times \{0,1\} \times \mathbb{D}_d$

$$V^*(y, g^1, \beta, g^2) := \inf_{\pi \in \Pi} V(\pi, y, g^1, \beta, g^2) = V(\pi^*, y, g^1, \beta, g^2). \qquad (56)$$

Now, in order to show the existence of an optimal policy via an optimality equation, we need to obtain recursive equations (see (21) and (28)) for the sequences of densities $\left\{g_t^1\right\} \in \mathbb{D}$ and $\left\{g_t^2\right\} \in \mathbb{D}_d$.

5.1. Recursive evolution of densities

We will borrow the recursive equation for $\left\{g_t^1\right\}$ from [10], by changing what has to be changing as follows.

For measurable functions $\varphi_1, \varphi_2 : \Re^+ \to \Re^+$ and $a \in A$, we denote

$$< \varphi_1, \varphi_2 >:= \int_0^\infty \varphi_1(s)\varphi_2(s)\,\mathrm{d}s, \tag{57}$$

$$\rho_{(a,\varphi_1)}(\omega) := \int_0^\infty \int_{(\omega - a\bar{\omega})^+}^\infty f_w\left((s + a\bar{\omega} - \omega)^+\right) f_{\bar{w}}\left(\bar{\omega}\right) \varphi_1(s)\,\mathrm{d}s\mathrm{d}\bar{\omega} \tag{58}$$

and

$$\theta(\omega; a, \varphi_1) := \frac{\rho_{(a,\varphi_1)}(\omega)}{< \rho_{(a,\varphi_1)}, 1 >}. \tag{59}$$

If $\varphi_1 = \delta$, we have

$$\rho_{(a,\delta)}(\omega) = \int_0^\infty f_w\left((a\bar{\omega} - \omega)^+\right) f_{\bar{w}}\left(\bar{\omega}\right) I_{[\omega < a\bar{\omega}]}\mathrm{d}\bar{\omega}. \tag{60}$$

Then (see [10, Th. 4.1, eq. (47)]) $\left\{g_t^1\right\}$ evolves in $\mathbb{D}$ according to the equation

$$\begin{aligned} g_0^1 &= g^1; \\ g_t^1(s) &= y_{t-1}\theta(s; a_{t-1}, \delta) + (1 - y_{t-1})\theta(s; a_{t-1}, g_{t-1}^1), \quad t \in \mathbb{N}. \end{aligned} \tag{61}$$

On the other hand, the evolution of the densities $\left\{g_t^2\right\} \in \mathbb{D}_d$ is given in next result which is proved in Appendix.

Theorem 5.2. The density process $\left\{g_t^2\right\} \in \mathbb{D}_d$ satisfies

$$\begin{aligned} g_0^2 &= g^2; \\ g_t^2(s) &= \beta_{t-1}\left\{\frac{f_\xi(s - \gamma d) I_{[s-\gamma d > 0]}}{\bar{F}_\xi\left(d - \gamma d\right)}\right\} \\ &\quad + (1 - \beta_{t-1})\left\{\frac{\int_{\mathrm{d}}^\infty f_\xi(s - \gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega}{\int_{\mathrm{d}}^\infty \bar{F}_\xi(d - \gamma\omega) g_{t-1}^2(\omega)\,\mathrm{d}\omega}\right\} \end{aligned} \tag{62}$$

for $t \in \mathbb{N}$, where $\bar{F}_\xi(\cdot) = 1 - F_\xi(\cdot)$.

In order to simplify the equation (62), we introduce similar notation as (57) – (60). For measurable functions $\varphi_1, \varphi_2 : (d, \infty) \to \Re^+$ we define

$$< \varphi_1, \varphi_2 >_d := \int_{\mathrm{d}}^\infty \varphi_1(s)\varphi_2(s)\,\mathrm{d}s, \tag{63}$$

$$\rho_{\varphi_1}(s) := \int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} \varphi_1(\omega)\,\mathrm{d}\omega, \tag{64}$$

and

$$\theta_{\varphi_1}(s) := \frac{\rho_{\varphi_1}(s)}{< \rho_{\varphi_1}, 1 >_d}. \tag{65}$$

Thus,

$$\begin{aligned} \rho_{\delta_d}(s) &= \int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} \delta_{\mathrm{d}}(\omega)\,\mathrm{d}\omega \\ &= f_\xi(s-\gamma d) I_{[s-\gamma \mathrm{d}>0]}. \end{aligned} \tag{66}$$

Taking into account (63) – (66) we have

$$\begin{aligned} \theta_{g_{t-1}^2}(s) &= \frac{\rho_{g_{t-1}^2}(s)}{< \rho_{g_{t-1}^2}, 1 >_d} \\ &= \frac{\int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega}{\int_d^{\infty}\int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega \mathrm{d}s}. \end{aligned}$$

Observe that from Fubini Theorem and letting $u = s - \gamma\omega$ we have

$$\int_d^{\infty}\int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega \mathrm{d}s$$

$$= \int_{\mathrm{d}}^{\infty} \left[\int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]}\,\mathrm{d}s\right] g_{t-1}^2(\omega)\,\mathrm{d}\omega = \int_{\mathrm{d}}^{\infty} \left[\int_{\mathrm{d}-\gamma\omega}^{\infty} f_\xi(u) I_{[u>0]}\,\mathrm{d}u\right] g_{t-1}^2(\omega)\,\mathrm{d}\omega$$

$$= \int_{\mathrm{d}}^{\infty} \bar{F}_\xi\left(\mathrm{d}-\gamma\omega\right) g_{t-1}^2(\omega)\,\mathrm{d}\omega.$$

Therefore

$$\theta_{g_{t-1}^2}(s) = \frac{\int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega}{\int_d^{\infty} \bar{F}_\xi\left(\mathrm{d}-\gamma\omega\right) g_{t-1}^2(\omega)\,\mathrm{d}\omega}. \tag{67}$$

In addition

$$\begin{aligned} \theta_{\delta_d}(s) &= \frac{\rho_{\delta_d}(s)}{< \rho_{\delta_d}, 1 >_d} = \frac{f_\xi(s-\gamma d) I_{[s-\gamma d>0]}}{\int_{\mathrm{d}}^{\infty} f_\xi\left(s-\gamma \mathrm{d}\right) I_{[s-\gamma \mathrm{d}>0]} \mathrm{d}s} \\ &= \frac{f_\xi(s-\gamma d) I_{[s-\gamma d>0]}}{\bar{F}_\xi\left(d-\gamma d\right)}. \end{aligned} \tag{68}$$

Hence, using (67) and (68) we obtain the following expression of (62):

$$\begin{aligned} g_0^2 &= g^2 \in \mathbb{D}_d; \\ g_t^2(s) &= \beta_{t-1}\theta_{\delta_d}(s) + (1-\beta_{t-1})\theta_{g_{t-1}^2}(s), \quad t \in \mathbb{N}. \end{aligned} \tag{69}$$

Observe that the cost function and the discount factor function (54) and (55) can be written as

$$\tilde{c}\left(y_t, g_t^1, a_t\right) = y_t c(0, a_t) + (1 - y_t)\left\langle c(\cdot, a_t), g_t^1(\cdot)\right\rangle \tag{70}$$

and

$$\tilde{\epsilon}\left(\beta_t, g_t^2\right) = \beta_t \epsilon(d) + (1 - \beta_t)\left\langle \epsilon(\cdot), g_t^2(\cdot)\right\rangle_d . \tag{71}$$

5.2. The optimality equation

It is worth noting that we are dealing with a CO control process that evolves according to the equations (61) – (69). In this sense, the corresponding optimality equation is obtained by applying similar dynamic programming techniques to the case when the processes $\{x_t\}$ and $\{\alpha_t\}$ are completely observable. We explain such a procedure in order to make it clearer to obtain the optimality equation for processes (61) – (69).

Suppose that the processes $\{x_t\}$ and $\{\alpha_t\}$ are completely observable. We can write the system (49) and (51) as

$$\kappa_{t+1} = H\left(\kappa_t, a_t, \chi_t\right), \quad t \in \mathbb{N}_0,$$

where $\kappa_t = (x_t, \alpha_t)$, $\chi_t = (w_t^1, \xi_t^1)$, and $H : X \times \Gamma \times A \times [0, \infty) \times [0, \infty) \to X \times \Gamma$ is a function defined as

$$H\left(\kappa_t, a_t, \chi_t\right) = \left(\left(x_t + a_t \bar{w}_t - w_t^{(1)}\right)^+, \max\left\{\gamma\alpha_t + \xi_t^{(1)}, d\right\}\right).$$

Following standard dynamic programming arguments, for a function $\hat{U} : X \times \Gamma \to \Re$, $\kappa_0 = \kappa \in X \times \Gamma$, and $a \in A$, we define the operator

$$\begin{aligned} \hat{T}_a \hat{U}(\kappa) &= c(x, a) + \epsilon(\alpha) E\left[\hat{U}(\kappa_1)\right] \\ &= c(x, a) + \epsilon(\alpha) E\left[\hat{U}(H\left(\kappa, a, \chi_0\right))\right], \quad \kappa = (x, \alpha) \in X \times \Gamma, \end{aligned} \tag{72}$$

where $\epsilon(\alpha) = e^{-\alpha}$. Hence, the optimality equation takes the form

$$\hat{U}(\kappa) = \min_{a \in A} \hat{T}_a \hat{U}(\kappa) = \min_{a \in A}\left\{c(x, a) + \epsilon(\alpha) E\left[\hat{U}(\kappa_1)\right]\right\}, \quad \kappa = (x, \alpha) \in X \times \Gamma.$$

Now, to obtain the optimality equation corresponding to the process $\left\{\left(y_t, g_t^1, \beta_t, g_t^2\right)\right\}$ whose evolution is determined by the relations (50), (61), (52), and (69), we consider the cost function $\tilde{c}$ (see (70)) and the discount factor function $\tilde{\epsilon}$ (see (71)). Then, for a function $U : \{0, 1\} \times \mathbb{D} \times \{0, 1\} \times \mathbb{D}_d \to \Re$, $\left(y_0, g_0^1, \beta_0, g_0^2\right) = \left(y, g^1, \beta, g^2\right) \in \{0, 1\} \times \mathbb{D} \times \{0, 1\} \times \mathbb{D}_d$, and $a \in A$ we define the operator (see (38), (70),(71), (72))

$$T_a U\left(y, g^1, \beta, g^2\right) := y c(0, a) + (1 - y)\left\langle c(\cdot, a), g^1(\cdot)\right\rangle + \tilde{\epsilon}\left(\beta, g^2\right) E\left[U\left(y_1, g_1^1, \beta_1, g_1^2\right)\right]. \tag{73}$$

We then proceed to calculate $E\left[U\left(y_1, g_1^1, \beta_1, g_1^2\right)\right]$. From (61) and (69) we have

$$\begin{aligned} & E\left[U\left(y_1, g_1^1, \beta_1, g_1^2\right)\right] \\ = \; & E\left[U\left(y_1, y\theta(s; a, \delta) + (1 - y)\theta(s; a, g^1), \beta_1, \beta\theta_{\delta_d}(\alpha) + (1 - \beta)\theta_{g_{t-1}^2}(\alpha)\right)\right] \end{aligned}$$

$$
\begin{aligned}
&= U\left(1, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 1, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2_{t-1}}(\alpha)\right) P\left[x_1 = 0\right] P\left[\alpha_1 = d\right] \\
&+U\left(1, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 0, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2_{t-1}}(\alpha)\right) P\left[x_1 = 0\right] P\left[\alpha_1 > d\right] \\
&+U\left(0, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 1, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2_{t-1}}(\alpha)\right) P\left[x_1 > 0\right] P\left[\alpha_1 = d\right] \\
&+U\left(0, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 0, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2_{t-1}}(\alpha)\right) P\left[x_1 > 0\right] P\left[\alpha_1 > d\right].
\end{aligned} \tag{74}
$$

Now observe that

$$
\begin{aligned}
P\left[x_1 = 0\right] &= P\left[w^{(1)} > x + a\bar{w}\right] = 1 - P\left[w^{(1)} \leq x + a\bar{w}\right] \\
&= 1 - E\left[E\left[I_{\left[w^{(1)} \leq x + a\bar{w}\right]} | x\right]\right] \\
&= 1 - \int_0^\infty P\left[w^{(1)} \leq \omega + a\bar{w}\right] g^1(\omega)\, \mathrm{d}\omega \\
&= 1 - \int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{w}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\, \mathrm{d}\omega
\end{aligned} \tag{75}
$$

Thus

$$
P\left[x_1 > 0\right] = 1 - P\left[x_1 = 0\right] = \int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{w}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\, \mathrm{d}\omega. \tag{76}
$$

Similarly we get

$$
P\left[\alpha_1 = d\right] = \int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha \tag{77}
$$

and

$$
P\left[\alpha_1 > d\right] = 1 - \int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha. \tag{78}
$$

Then, combining (73) – (78) we obtain

$$
\begin{aligned}
T_a U\left(y, g^1, \beta, g^2\right) &:= yc(0,a) + (1-y)\left\langle c(\cdot, a), g^1(\cdot)\right\rangle \\
&+\tilde{\epsilon}\left(\beta, g^2\right) U\left(1, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 1, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2}(\alpha)\right) \\
&\cdot \left(1 - \int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{\omega}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\mathrm{d}\omega\right)\left(\int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha\right) \\
&+\tilde{\epsilon}\left(\beta, g^2\right) U\left(1, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 0, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2}(\alpha)\right) \\
&\cdot \left(1 - \int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{\omega}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\mathrm{d}\omega\right)\left(1 - \int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha\right) \\
&+\tilde{\epsilon}\left(\beta, g^2\right) U\left(0, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 0, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2}(\alpha)\right) \\
&\cdot \left(\int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{\omega}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\, \mathrm{d}\omega\right)\left(\int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha\right)
\end{aligned}
$$

$$+\tilde{\epsilon}\left(\beta, g^2\right) U\left(0, y\theta(s;a,\delta) + (1-y)\theta(s;a,g^1), 1, \beta\theta_{\delta_d}(\alpha) + (1-\beta)\theta_{g^2}(\alpha)\right)$$
$$\cdot \left(\int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{\omega}\right) f_{\bar{w}}(\bar{\omega}) g^1(\omega)\, \mathrm{d}\bar{\omega}\mathrm{d}\omega\right)\left(1 - \int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) g^2(\alpha)\, \mathrm{d}\alpha\right).$$

Therefore the dynamic programming operator T is defined as

$$TU\left(y, g^1, \beta, g^2\right) = \min_{a\in A} T_a U\left(y, g^1, \beta, g^2\right), \quad \left(y, g^1, \beta, g^2\right) \in \{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d. \tag{79}$$

Observe that from Assumption 5.1 (d) and because the control set A is compact (see Remark 4.2 (b)), the function $\tilde{c}$, defined in (54), is inf-compact on $\{0,1\}\times\mathbb{D}\times A$. Thus, in the setting of Theorem 4.3, to prove that the optimal value function V^* defined in (56) is the minimal solution in $\mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\right)$ and to ensure the existence of an optimal policy, it is sufficient to verify that (see Assumption 4.1)

a) $\tilde{c} \in \mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\right)$;

b) $T_a U \in \mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\times A\right)$ for all $U \in \mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\right)$.

Condition a) easily follows from (54) (see (70)) and Assumption 5.1 (d), while Condition b) is obtained from the continuity of the following functions whose proofs are given in Appendix,

$$(a,\varphi_1) \to \int_0^\infty \int_0^\infty f_w\left((s + a\bar{\omega} - \omega)^+\right) f_{\bar{w}}\left(\bar{\omega}\right) \varphi_1(s)\, \mathrm{d}s\mathrm{d}\bar{\omega}, \ (a,\varphi_1) \in A\times\mathbb{D}, \tag{80}$$

$$\varphi_2 \to \int_{\mathrm{d}}^\infty f_\xi(s - \gamma\omega) I_{[s-\gamma\omega>0]}\varphi_2(\omega)\, \mathrm{d}\omega, \quad \varphi_2 \in \mathbb{D}_{\mathrm{d}}, \tag{81}$$

$$(a,\varphi_1) \to \int_0^\infty \int_0^\infty F_w\left(\omega + a\bar{\omega}\right) f_{\bar{w}}(\bar{\omega})\varphi_1(\omega)\, \mathrm{d}\bar{\omega}\mathrm{d}\omega, \ (a,\varphi_1) \in A\times\mathbb{D}, \tag{82}$$

and

$$\varphi_2 \to \int_{\mathrm{d}}^\infty F_\xi\left(\mathrm{d} - \gamma\alpha\right) \varphi_2(\alpha)\, \mathrm{d}\alpha, \quad \varphi_2 \in \mathbb{D}_{\mathrm{d}}. \tag{83}$$

Indeed, observe that (80) implies the continuity of the functions

$$(a,\varphi_1) \to \rho_{(a,\varphi_1)}(\cdot) \text{ and } (a,\varphi_1) \to \theta(\cdot;a,\varphi_1), \quad (a,\varphi_1) \in A\times\mathbb{D}, \tag{84}$$

and (81) yields the continuity of the functions

$$\varphi_2 \to \rho_{\varphi_2}(\cdot) \ \text{ and } \ \varphi_2 \to \theta_{\varphi_2}(\cdot), \quad \varphi_2 \in \mathbb{D}_d. \tag{85}$$

Therefore, for all $U \in \mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\right)$, form (80) – (85) we can conclude that $T_a U \in \mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\times A\right)$, that is Condition b) holds.

In conclusion, from Theorem 4.3 we have that V^* is the minimal solution of the optimality equation (79) in the space of functions $\mathcal{L}_+\left(\{0,1\}\times\mathbb{D}\times\{0,1\}\times\mathbb{D}_d\right)$, and there exists an optimal policy for the PO queueing system (49) – (50) with PO discount factor (51) – (52).

6. APPENDIX: PROOFS

Lemma 6.1. The functions defined in (80), (81), (82), and (83) are continuous.

P r o o f . The continuity of the function (80) is proved following similar arguments as in [10, Lemma 6.2]. Now, let $\{\varphi_n\}$ be a sequence in $\mathbb{D}_d$ such that $\varphi_n \to \varphi \in \mathbb{D}_d$ (see (48)). Then, since f_ξ is bounded, from Convergence Dominated Theorem we have that

$$\lim_{n\to\infty} \left| \int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} \varphi_n(\omega)\,\mathrm{d}\omega - \int_{\mathrm{d}}^{\infty} f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} \varphi(\omega)\,\mathrm{d}\omega \right| = 0,$$

which implies that the function (81) is continuous. Similarly is proved the continuity of functions (82) and (83). □

6.1. Proof of Theorem 5.2

Lemma 6.2. For any real valued and bounded function $\varphi : \Gamma \to \Re$ and $t \in \mathbb{N}$,

$$\begin{aligned} \text{(a)}\quad E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] &= I_{[\alpha_t=d]}\varphi(d) + I_{[\alpha_t>d]} \frac{E[\varphi(\alpha_t) I_{[\alpha_t>d]} \mid \mathcal{Z}_t]}{P(\alpha_t > d \mid \mathcal{Z}_t)} \\ &= I_{[\alpha_t=d]}\varphi(d) + I_{[\alpha_t>d]} E[\varphi(\alpha_t) \mid \alpha_t > d, \mathcal{Z}_t]. \end{aligned} \tag{86}$$

$$\begin{aligned} \text{(b)}\ E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] I_{[\alpha_t>d]} &= I_{[\alpha_{t-1}=d]} \frac{\int_{\mathrm{d}}^{\infty} \varphi(s) f_\xi\left(s-\gamma d\right) I_{[s-\gamma d>0]}\,\mathrm{d}s}{\bar{F}_\xi\left(\mathrm{d}-\gamma \mathrm{d}\right)} \\ &+ I_{[\alpha_{t-1}>\mathrm{d}]} \frac{\int_{\mathrm{d}}^{\infty} \varphi(s) \int_{\mathrm{d}}^{\infty} f_\xi\left(s-\gamma\omega\right) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega\mathrm{d}s}{\int_{\mathrm{d}}^{\infty} \bar{F}_\xi\left(\mathrm{d}-\gamma \mathrm{d}\right) g_{t-1}^2(\omega)\,\mathrm{d}\omega} \end{aligned} \tag{87}$$

P r o o f . (a) Let $\varphi : \Gamma \to \Re$ be any real valued and bounded function. For each $t \in \mathbb{N}$,

$$\begin{aligned} E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] &= E[\varphi(\alpha_t) I_{[\alpha_t=d]} \mid \mathcal{Z}_t, \beta_t] + E[\varphi(\alpha_t) I_{[\alpha_t>d]} \mid \mathcal{Z}_t, \beta_t] \\ &= I_{[\alpha_t=d]}\varphi(d) + E[\varphi(\alpha_t) I_{[\alpha_t>d]} \mid \mathcal{Z}_t, \beta_t] \end{aligned} \tag{88}$$

On the other hand, from definition of conditional expectation, there exists a measurable function G such that

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] = G(z_0, \ldots, , z_{t-1}, z_t, \beta_t).$$

Hence,

$$\begin{aligned} E[\varphi(\alpha_t) I_{[\alpha_t>d]} \mid \mathcal{Z}_t, \beta_t] &= I_{[\alpha_t>d]} E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] \\ &= I_{[\alpha_t>d]} G(z_0, \ldots, z_{t-1}, z_t, \beta_t) \\ &= I_{[\alpha_t>d]} G(z_0, \ldots, z_{t-1}, z_t, 0). \end{aligned} \tag{89}$$

Taking expectation given $\mathcal{Z}_t$ in (89), using the fact $\mathcal{Z}_t \subseteq (\mathcal{Z}_t, \beta_t)$, and because $G(z_0, \ldots$ $\ldots, z_{t-1}, z_t, 0)$ is $\mathcal{Z}_t-$measurable, we obtain

$$E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t] = G(z_0, \ldots, , z_{t-1}, z_t, 0) P(\alpha_t > d \mid \mathcal{Z}_t).$$

Hence

$$G(z_0, \ldots, , z_{t-1}, z_t, 0) = \frac{E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t]}{P(\alpha_t > d \mid \mathcal{Z}_t)}. \tag{90}$$

Combination of (88) – (90) yields

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] = I_{[\alpha_t = d]} \varphi(d) + I_{[\alpha_t > d]} \frac{E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t]}{P(\alpha_t > d \mid \mathcal{Z}_t)}. \tag{91}$$

This proves the first equality in (86).

On the other hand, from Conditional Bayes Theorem (see [6]) and definition of g_t^2 (see (53)) we have

$$\begin{aligned} \int_{\mathrm{d}}^{\infty} \varphi(s) g_t^2(s) \, \mathrm{d}s = E[\varphi(\alpha_t) \mid \alpha_t > \mathrm{d}, \mathcal{Z}_t] &= \frac{E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t]}{E[I_{[\alpha_t > d]} \mid \mathcal{Z}_t)} \\ &= \frac{E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t]}{P(\alpha_t > d \mid \mathcal{Z}_t)}. \end{aligned}$$

Thus, from (91)

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] = I_{[\alpha_t = d]} \varphi(d) + I_{[\alpha_t > d]} E[\varphi(\alpha_t) \mid \alpha_t > d, \mathcal{Z}_t],$$

which proves the second equality in (86).

(b) Observe that from (53), (86), and the independence of the state and discount factor observation processes we have

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] = I_{[\alpha_t = d]} \varphi(d) + I_{[\alpha_t > d]} \int_{\mathrm{d}}^{\infty} \varphi(\omega) g_t^2(\omega) \, \mathrm{d}\omega \tag{92}$$

and

$$E[\varphi(\alpha_{t-1}) \mid \mathcal{Z}_t] = I_{[\alpha_{t-1} = d]} \varphi(\mathrm{d}) + I_{[\alpha_{t-1} > \mathrm{d}]} \int_{\mathrm{d}}^{\infty} \varphi(\omega) g_{t-1}^2(\omega) \, \mathrm{d}\omega. \tag{93}$$

Now, since $\mathcal{Z}_t = \sigma(z_0, \ldots, z_t) \subseteq \sigma(z_0, \ldots, z_t, \alpha_{t-1})$, we have

$$\begin{aligned} &E[\varphi(\alpha_t) I_{[\alpha_t > d]} \mid \mathcal{Z}_t] \\ &\qquad = E\left[\varphi\left(\gamma \alpha_{t-1} + \xi_{t-1}^{(1)}\right) I_{[\gamma \alpha_{t-1} + \xi_{t-1}^{(1)} > d]} \mid \mathcal{Z}_t\right] \\ &\qquad = E\left[E\left[\varphi\left(\gamma \alpha_{t-1} + \xi_{t-1}^{(1)}\right) I_{[\gamma \alpha_{t-1} + \xi_{t-1}^{(1)} > d]} \mid \mathcal{Z}_t, \alpha_{t-1}\right] \mid \mathcal{Z}_t\right] \\ &\qquad = E\left[\int_0^{\infty} \varphi\left(\gamma \alpha_{t-1} + \omega\right) I_{[\gamma \alpha_{t-1} + \omega > \mathrm{d}]} f_{\xi}(\omega) \, \mathrm{d}\omega \mid \mathcal{Z}_t\right] \\ &\qquad = E\left[\int_{\mathrm{d} - \gamma \alpha_{t-1}}^{\infty} \varphi\left(\gamma \alpha_{t-1} + \omega\right) f_{\xi}(\omega) \, \mathrm{d}\omega \mid \mathcal{Z}_t\right]. \end{aligned}$$

Letting $s = \gamma\alpha_{t-1} + \omega$ we get

$$E[\varphi(\alpha_t)I_{[\alpha_t>d]} \mid \mathcal{Z}_t] = E\left[\int_{\mathrm{d}}^{\infty} \varphi(s) f_\xi\left(s-\gamma\alpha_{t-1}\right) I_{[s-\gamma\alpha_{t-1}>0]}\mathrm{d}s \mid \mathcal{Z}_t\right]$$
$$= \int_{\mathrm{d}}^{\infty} \varphi(s) E\left[f_\xi\left(s-\gamma\alpha_{t-1}\right) I_{[s-\gamma\alpha_{t-1}>0]}|\mathcal{Z}_t\right]\mathrm{d}s. \tag{94}$$

From (93) with $\varphi(\alpha_{t-1}) = f_\xi\left(s-\gamma\alpha_{t-1}\right) I_{[s-\gamma\alpha_{t-1}>0]}$ we obtain

$$E\left[f_\xi\left(s-\gamma\alpha_{t-1}\right) I_{[s-\gamma\alpha_{t-1}>0]}|\mathcal{Z}_t\right] = I_{[\alpha_{t-1}=d]} f_\xi\left(s-\gamma d\right) I_{[s-\gamma d>0]}$$
$$+ I_{[\alpha_{t-1}>d]} \int_{\mathrm{d}}^{\infty} f_\xi\left(s-\gamma\omega\right) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega. \tag{95}$$

Thus, (94) takes the form

$$E[\varphi(\alpha_t)I_{[\alpha_t>d]} \mid \mathcal{Z}_t]$$
$$= \int_{\mathrm{d}}^{\infty} \varphi(s)\left[I_{[\alpha_{t-1}=\mathrm{d}]} f_\xi\left(s-\gamma\mathrm{d}\right) I_{[s-\gamma\mathrm{d}>0]} + I_{[\alpha_{t-1}>\mathrm{d}]} \int_{\mathrm{d}}^{\infty} f_\xi\left(s-\gamma\omega\right) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega) d\omega\right]\mathrm{d}s$$
$$= I_{[\alpha_{t-1}=\mathrm{d}]} \int_{\mathrm{d}}^{\infty} \varphi(s) f_\xi\left(s-\gamma\mathrm{d}\right) I_{[s-\gamma\mathrm{d}>0]}\mathrm{d}s$$
$$+ I_{[\alpha_{t-1}>d]} \int_{d}^{\infty} \varphi(s) \int_{d}^{\infty} f_\xi\left(s-\gamma\omega\right) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\omega\mathrm{d}s. \tag{96}$$

On the other hand

$$P[\alpha_t > d \mid \mathcal{Z}_t] = E\left[I_{[\gamma\alpha_{t-1}+\xi_{t-1}>d]} \mid \mathcal{Z}_t\right]$$
$$= E\left[E[I_{[\gamma\alpha_{t-1}+\xi_{t-1}>d]} \mid \mathcal{Z}_t, \alpha_{t-1}] \mid \mathcal{Z}_t\right]$$
$$= E\left[P\left[\xi_{t-1} > d-\gamma\alpha_{t-1}\right] \mid \mathcal{Z}_t, \alpha_{t-1} \mid \mathcal{Z}_t\right]$$
$$= E\left[\bar{F}_\xi\left(d-\gamma\alpha_{t-1}\right) \mid \mathcal{Z}_t\right]$$

$$= I_{[\alpha_{t-1}=d]} \bar{F}_\xi\left(d-\gamma d\right) + I_{[\alpha_{t-1}>d]} \int_{d}^{\infty} \bar{F}_\xi\left(d-\gamma\omega\right) g_{t-1}^2(\omega)\,\mathrm{d}\omega, \tag{97}$$

where the last equality comes from (93). Combining (86), (96), and (97) we get

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] = I_{[\alpha_t=d]}\varphi(d)$$
$$+ \frac{I_{[\alpha_t>d]}\left[I_{[\alpha_{t-1}=d]} \int_{d}^{\infty} \varphi(s) f_\xi\left(s-\gamma d\right) I_{[s-\gamma d>0]}\,\mathrm{d}s + I_{[\alpha_{t-1}>d]} \int_{d}^{\infty} \varphi(s) \int_{d}^{\infty} f_\xi\left(s-\gamma\omega\right) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega\mathrm{d}s\right]}{I_{[\alpha_{t-1}=d]} \bar{F}_\xi\left(d-\gamma d\right) + I_{[\alpha_{t-1}>d]} \int_{d}^{\infty} \bar{F}_\xi\left(d-\gamma\omega\right) g_{t-1}^2(\omega)\,\mathrm{d}\omega}. \tag{98}$$

Hence, multiplying by $I_{[\alpha_t>d]}$ in (98) and using indicator functions properties we get

$$E[\varphi(\alpha_t) \quad | \quad \mathcal{Z}_t, \beta_t] I_{[\alpha_t>d]} = I_{[\alpha_{t-1}=d]} \frac{\int_d^\infty \varphi(s) f_\xi\,(s-\gamma d)\, I_{[s-\gamma d>0]}\,\mathrm{d}s}{\bar{F}_\xi\,(d-\gamma d)}$$
$$+ I_{[\alpha_{t-1}>d]} \frac{\int_d^\infty \varphi(s) \int_d^\infty f_\xi\,(s-\gamma\omega)\, I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega\mathrm{d}s}{\int_d^\infty \bar{F}_\xi\,(d-\gamma d)\, g_{t-1}^2(\omega)\,\mathrm{d}\omega},$$

that is, (87) holds. □

P r o o f . of Theorem 5.2. From (92) we have

$$E[\varphi(\alpha_t) \mid \mathcal{Z}_t, \beta_t] I_{[\alpha_t>d]} = \int_d^\infty \varphi(\omega) g_t^2(\omega)\,\mathrm{d}\omega. \tag{99}$$

Thus, comparing (87) and (99), and using the equivalence of the events $[\beta_{t-1}=1]$, $[\alpha_{t-1}\le d]$, and $[\alpha_{t-1}=d]$ (see (52)), we conclude, for $t\in\mathbb{N}$,

$$g_t^2(s) = \beta_{t-1}\left\{\frac{f_\xi(s-\gamma d) I_{[s-\gamma d>0]}}{\bar{F}_\xi\,(d-\gamma d)}\right\} + (1-\beta_{t-1})\left\{\frac{\int_d^\infty f_\xi(s-\gamma\omega) I_{[s-\gamma\omega>0]} g_{t-1}^2(\omega)\,\mathrm{d}\omega}{\int_d^\infty \bar{F}_\xi(d-\gamma\omega) g_{t-1}^2(\omega)\,\mathrm{d}\omega}\right\},$$

which is the relation (62). □

7. CONCLUDING REMARKS

In this paper we have studied a class of partially observable MDPs under a discounted criterion where the discount factor is modelled by a stochastic process which is also partially observable. The importance in considering random discount factors lies in its applications in economic and financial models. Indeed, as it is well known, usually the discount factor is a function of the interest rate, which, from a realistic point of view, should be considered random and even more, partially observable, unlike the ordinary case that assumes it to be constant. The analysis that we have proposed was based on the application of a standard procedure, that is, transforming the partially observable problem into a completely observable problem through a filtering process.

On the other hand, although the transformation procedure is merely theoretical, we have extensively illustrated it with a queueing system with controlled service rate and partially observable waiting times, which itself is interesting in the field of operations research. In fact, the queueing system (49) can be considered as a generalization of the classical inventory system by letting $\bar{w}_t = 1$. In this case, x_t and a_t represent the stock and the ordered quantity at the beginning of stage t, while $w_t^{(1)}$ is the random demand during the stage t. Under a partially observation scenario, we can assume that the only possible observation is when the stock is zero. This case is known in the literature of partially inventory systems as "zero balance walk" (see, e. g., [1]). Hence, according to our results, we can analyze the zero balance walk case for inventory systems considering partially observable discount factors.

An interesting work that the authors are currently working on is the computational implementation of the results. It is clear that this constitutes a non-trivial challenge due to the fact of dealing with processes in infinite-dimensional spaces, which implies proposing efficient discretization methods.

ACKNOWLEDGEMENT

This work was partially supported by Consejo Nacional de Ciencia y Tecnología (CONACYT-México) grant Ciencia Frontera 2019-87787.

(Received April 25, 2022)

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] D. P. Bertsekas and S. E. Shreve: Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York 1978. DOI:10.1137/1022042

[3] Y. Carmon and A. Shwartz: Markov decision processes with exponentially representable discounting. Oper. Res. Lett. *37* (2009), 51–55. DOI:10.1016/j.orl.2008.10.005

[4] H. Cruz-Suárez and R. Montes-de-Oca: Discounted Markov control processes induced by deterministic systems. Kybernetika *42* (2006), 647–664.

[5] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer-Verlag, New York 1979. DOI:10.1137/1023056

[6] R. J. Elliott, L. Aggoun, and J. B. Moore: Hidden Markov Models: Estimation and Control. Springer-Verlag, New York 1994. DOI10.1007/978-0-387-84854-9:

[7] E. A. Feinberg and A. Shwartz: Constrained dynamic programming with two discount factors: applications and an algorithm. IEEE Trans. Automat. Control *44* (1999), 628–631. DOI:10.1109/9.751365

[8] J. González-Hernández, R R. López-Martínez, and J. A. Minjárez-Sosa: Approximation, estimation and control of stochastic systems under a randomized discounted cost criterion. Kybernetika *45* (2009), 737–754. DOI:10.1017/S0022226709990132

[9] J. González-Hernández, R. R. López-Martínez, J. A. Minjárez-Sosa, and J. R.Gabriel-Arguelles: Constrained Markov control processes with randomized discounted rate: infinite linear programming approach. Optim. Control Appl. Meth. *35* (2014), 575–591. DOI:10.1002/oca.2089

[10] Y. H. García, S. Diaz-Infante, and J. A. Minjarez-Sosa: Partially observable queueing systems with controlled service rates under a discounted optimality criterion. Kybernetika *57* (2021), 493–512. DOI:10.14736/kyb-2021-3-0493

[11] E- I- Gordienko and F. S. Salem: Robustness inequality for Markov control processes with unbounded costs. Syst. Control Lett. *33* (1998), 125–130. DOI:10.1016/S0167-6911(97)00077-7

[12] E. Gordienko, E. Lemus-Rodríguez, and R. Montes-de-Oca: Discounted cost optimality problem: stability with respect to weak metrics. Math. Methods Oper. Res. *68* (2008), 77–96. DOI:10.1007/s00186-007-0171-z

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

[14] O. Hernandez-Lerma: Adaptive Markov Control Processes. Springer-Verlag, New York 1989. DOI:10.1137/1033169

[15] O. Hernandez-Lerma and W. Runggaldier: Monotone approximations for convex stochastic control problems. J. Math. Syst. Estim. Control *4* (1994), 99–140.

[16] O. Hernandez-Lerma and M. Munoz-de-Ozak: Discrete-time Markov control processes with discounted unbounded costs: optimality criteria. Kybernetika *28* (1992), 191–221. DOI:10.1016/S0010-9452(13)80050-0

[17] O. Hernández-Lerma and J. B. Lasserre: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer-Verlag, New York 1996.

[18] 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

[19] K. Hinderer: Foundations of Non-stationary Dynamic Programming with Discrete Time parameter. In: Lecture Notes Oper. Res. *33*, Springer, New York 1979.

[20] H. Jasso-Fuentes, J. L. Menaldi, and T. Prieto-Rumeau: Discrete-time control with non-constant discount factor. Math. Methods Oper. Res. *92* (2020), 377–399. DOI:10.1007/s00186-020-00716-8

[21] J. A. Minjarez-Sosa: Approximation and estimation in Markov control processes under discounted criterion. Kybernetika *40* (2004), 681–690. DOI:10.1016/j.jvs.2004.07.005

[22] 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

[23] U. Rieder: Measurable selection theorems for optimization problems. Manuscripta Math. *24* (1978), 115–131. DOI:10.1007/BF01168566

[24] W. J. Runggaldier and L. Stettner: Approximations of Discrete Time Partially Observed Control Problems. Applied Mathematics Monographs CNR *6,* Giardini, Pisa 1994. DOI:10.1007/BFb0006563

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

[26] Q. Wei and X. Guo: Markov decision processes with state-dependent discount factors and unbounded rewards/costs. Oper. Res. Lett. *39* (2011), 368–274. DOI:10.1016/j.orl.2011.06.014

E. Everardo Martinez-Garcia, Departamento de Matemáticas, Universidad de Sonora. Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora. México.
e-mail: emartine@live.com

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

Oscar Vega-Amaya, Departamento de Matemáticas, Universidad de Sonora. Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora. México.
e-mail: ovega@mat.uson.mx