A MARKOV GAME APPROACH FOR OPTIMAL ROUTING INTO A QUEUEING NETWORK

Eitan ALTMAN
INRIA
2004 Route des Lucioles
BP93, 06902 Sophia-Antipolis Cedex, France
altman@epic.inria.fr, tel. 93 65 77 86

November 1994

Abstract

We study a dynamic optimal routing problem, where a controller has to decide to which of two queues should arriving customers (representing packets, messages, call etc...) be sent. The service rate in each queue may depend on the state of the system, may change in time and is unknown to the controller. The goal of the controller is to design a strategy that guarantees the best performance under the worst case service conditions. The payoff is composed of a holding cost, an admission cost, and a cost that depends on the quality of the service. We consider both the finite and infinite horizon discounted cost. The problem is studied in the framework of zero-sum Markov games where the server, called player 1, is assumed to play against the router, called player 2. Each player is assumed to have the information of all previous actions of both players as well as the current and past states of the system. We show that there exist pure optimal strategies for both players. A value iteration algorithm is used to establish properties of the value of the game, which are related to supermodularity and to convexity. This is then shown to imply the existence of optimal strategies described by monotone switching curves for both players.

Keywords: zero-sum stochastic games, value iteration, monotone switching curve strategies, control of queueing networks, routing control.

1 Introduction

We consider in this paper a min-max type optimal control of customers routing into two infinite capacity queues. Whenever a customer arrives, the controller has to decide to which queue it will be sent. The service rate in each queue is known to remain within some interval. However, the presence of customers arriving from other controlled sources as well as congestion phenomena is modeled by allowing the service rate in each queue to depend on the state of the system, and to change in time in a way that is unknown to the router. The goal of the router is to design a strategy that guarantees the best performance under the worst case service conditions. We formulate this problem as a zero-sum stochastic game, where the server, called player 1 (or "nature"), is assumed to play against the router, called player 2. Each player is assumed to have the information of all previous actions of both players as well as the current and past states of the system, namely, the length of the queues.

Our main result is to identify optimal strategies for both players which have a simple structure, which implies that the optimal min-max strategy for the router is easy to implement. We show that the router has an optimal strategy of monotone switching curve type (see [7]). This strategy has the following monotonicity property. If it is optimal to route a customer to queue 1 for a given length of the queues $s = (s_1, s_2)$, then it is also optimal to route the customer to queue 1 when the length of the queues is $t = (t_1, t_2)$ provided that $t_1 \leq s_1$ and $t_2 \geq s_2$. A similar monotonicity property holds for routing to queue 2. We then identify a worst-case service conditions, for which each server uses a bang-bang strategy, i.e., depending on the state of the system, either the largest or the smallest service rate will be chosen. Moreover, the decision rule for each server, between the highest or lowest service rate, is again characterized by a monotone switching curve strategy.

In order to establish the structure of the optimal strategies we use the following approach. We first identify properties that the value function may have which would imply the desired structure of the optimal strategies. We then use value iteration in order to show that the value function indeed possesses these properties. This approach was used in the past to obtain structural results in several other stochastic games arrising in queueing models. In [1] and [2] optimal threshold and optimal monotone strategies are shown to exist for min-max flow control problems into a single queue with unknown service rate; a similar structure is obtained when service rate is controlled [6]. In all cases, the property of the value function that induces the structure of the optimal strategies was convexity. A min-max routing problem was considered by Altman and Shimkin [5], where the router has to decide to which of N queues an arriving customer should be routed. A symmetric setting was considered, where the service rate in all queues are the same. In addition, an extra service capacity was assumed to be allocated to the queues in a way unknown to the router. Routing to the shortest queue was identified as an optimal strategy. The property of the value function that induces the structure of the optimal strategies was Schur-convexity. A related routing problem as well as a scheduling problem was considered by Altman and Koole [3]. In the present paper, the properties that will induce the structure of the optimal strategies are related to supermodularity and convexity.

Structural properties of optimal strategies in queueing systems were also obtained in non-zero sum games using different techniques, see e.g. Hsiao and Lazar [8], Altman and Shimkin [4], [5].

The structure of the paper is as follows: in Section 2 we describe the model. In Section 3 we solve the finite horizon problem, and in Section 4 we solve the infinite horizon problem. Generalizations are discussed finally in Section 5.

2 The model

Consider two infinite queues. Customers arrive to the system according to a Poisson process with rate λ'. Upon arrival of a customer, the router chooses an action 1 or 2, with the interpretation that action i corresponds to routing the customer to queue i. In each queue, customers are served according to the order of FCFS: first come first served. The service duration of a customer in queue i is exponentially distributed with a parameter $a_1(i)$ that lies in the interval $[\underline{\mu}', \overline{\mu}']$. This parameter, called the service rate, may change in time in a way unknown to the router.

At first sight, it seems that in order to model this process as an MDP (Markov Decision Process), the state space should include not only the queues' length, but also the identity of the last event that happens, since, the router can take decisions only at events of arrival of customers, whereas the servers can take decisions at any time. We construct, however, a simpler equivalent MDP model by allowing the router to take decisions at every event. If the router chooses action i at time t, it has the interpretation that if the following event is going to be an arrival, then the customer that will arrive will be routed to queue i. With this interpretation, it will suffice to consider the length of the queues as the state of the system. Let $\mathcal{A}_i s$, $i = 1,2$ denote the state obtained by an arrival of a customer to queue i when the state was s, and let $\mathcal{D}_i s$, $i = 1,2$ denote the state obtained by a departure of a customer from queue i when the state was s. If queue i is empty, then we understand $\mathcal{D}_i s = s$.

Consider a time interval Δ. The probability to be in state s' at time $t + \Delta$ if at time t the state is s and the actions are a_1, a_2 is given by:

$$q_\Delta(s' \mid s, a_1, a_2) := \begin{cases} \lambda'\Delta + o(\Delta), & \text{if } a_2 = i, s' = \mathcal{A}_i s, i = 1,2; \\ a_1(i)\Delta + o(\Delta), & \text{if } s' = \mathcal{D}_i s, s_i \neq 0, i = 1,2; \\ 1 - (\lambda' + \sum_{i=1}^{2} a_1(i) 1\{s_i > 0\})\Delta + o(\Delta), & \text{if } s' = s; \\ o(\Delta) & \text{otherwise.} \end{cases}$$

We shall use a time discretized model by fixing some small Δ. Define $\lambda = \lambda'\Delta$, $\underline{\mu}_i = \underline{\mu}'_i\Delta$, $\overline{\mu}_i = \overline{\mu}'_i\Delta$, $i = 1,2$. We assume that Δ is sufficiently small so that $\lambda + \overline{\mu}_1 + \overline{\mu}_2 \leq 1$. We obtain finally the following approximating discrete time MDP:

<u>The state space</u> is $\mathbf{S} = \mathbb{N}^2$ where $\mathbb{N}$ are the natural numbers, so that the state denotes the number of customers in the queues. An element $s \in \mathbf{S}$ is thus a two dimensional vector denoted by $s = (s_1, s_2)$, where s_i are the number of customers in queue i.

The action space of the servers (player 1) is the product of the compact intervals $A_1 = A_1(1) \times A_1(2) = [\underline{\mu}_1, \overline{\mu}_1] \times [\underline{\mu}_2, \overline{\mu}_2]$. (Note that although there are two servers, each of which is a controller, they can be considered together as a single player since they have a common objective). An action a in A_1 is thus a two dimensional vector: $a = \{a(1), a(2)\}$.

The action space of the router (player 2) is $A_2 = \{1, 2\}$.

The transition law:

$$q(s' \mid s, a_1, a_2) := \begin{cases} \lambda, & \text{if } a_2 = i, s' = \mathcal{A}_i s, i = 1, 2; \\ a_1(i), & \text{if } s' = \mathcal{D}_i s, i = 1, 2; \\ 1 - (\lambda + \sum_{i=1}^2 a_1(i) 1\{s_i > 0\}), & \text{if } s' = s. \end{cases}$$

We note that this law of motion is additive [10]: q can be expressed as $q(s' \mid s, a_1, a_2) = q_1(s' \mid s, a_1) + q(s' \mid s, a_2)$ where

$$q_1(s' \mid s, a_1) := \begin{cases} a_1(i), & \text{if } s' = \mathcal{D}_i s, i = 1, 2; \\ 1 - (\lambda + \sum_{i=1}^2 a_1(i) 1\{s_i > 0\}), & \text{if } s' = s, \end{cases}$$

and $q_2(s' \mid s, a_2) := \lambda$ if $a_2 = i, s' = \mathcal{A}_i s, i = 1, 2$. Moreover, q is continuous in the actions.

The immediate payoff: We assume that the payoff $r(s, a_1, a_2)$, which the router has to pay at each step if the state is s and the actions are a_1, a_2, is separable, and has the form:

$$r(s, a_1, a_2) = c(s) + \sum_{i=1}^{2} \theta_i(a_1) + \sum_{i=1}^{2} d_i 1\{a_2 = i\}.$$

It is composed of a holding cost c, a cost θ_i that depends on the quality of the service at queue i, and an admission cost d_i if a customer is to be admitted to queue i. We assume that θ_i (and thus r) are continuous in the actions.

The strategies We refer to [10] for the definition of strategies.

The finite horizon and infinite horizon costs For given strategies f, g for the players and an initial state s, let $r_n(f, g)(s)$ be the corresponding expected payoff (that is paid by the router) at stage n. For a fixed discount factor $0 < \beta < 1$, we shall consider the finite horizon discounted cost $\phi_\beta(n; f, g)(s) = \sum_{m=1}^n \beta^{m-1} r_m(f, g)(s)$, and the infinite horizon discounted cost $\phi_\beta(f, g)(s) = \sum_{m=1}^\infty \beta^{m-1} r_m(f, g)(s)$. Let $v_\beta(n; s)$ and $v_\beta(s)$ denote the values of the finite and infinite horizon discounted games, respectively, i.e.,

$$v_\beta(n; s) = \sup_f \inf_g \phi_\beta(n; f, g)(s) = \inf_g \sup_f \phi_\beta(n; f, g)(s),$$

$$v_\beta(s) = \sup_f \inf_g \phi_\beta(f,g)(s) = \inf_g \sup_f \phi_\beta(f,g)(s).$$

A strategy f^* that satisfies $v_\beta(n;s) = \inf_g \phi_\beta(n; f^*, g)(s)$ is said to be optimal for player 1 for the finite horizon problem. Optimality for the infinite horizon problem and optimality for player 2 are defined similarly. Since for any fixed finite horizon n and initial state s, only a finite number of states can be reached, it follows that the finite horizon problem is equivalent to a stochastic game with a finite number of states, and the payoffs can thus be considered to be bounded. The existence of a value and of optimal Markov strategies for both players then follows from well known results, see e.g. [9] Theorem 4.1. We show in Section 4 that a value exists also for the infinite horizon problem within the stationary strategies.

Remark 2.1 *Combining the additivity of the transition probabilities and the separability of the payoffs, we conclude that the stochastic game has the AR-AT structure, see [10]. This implies that pure Markov optimal strategies exist for both player for the finite horizon case, and pure stationary strategies exist for both player for the infinite horizon case. This results from the fact that in the dynamic programming equation, the value operator decomposes to the sum of separate maximization and minimization operations.*

3 The finite horizon problem

3.1 The structure of optimal strategies and required properties of the value function

Consider a horizon of n steps. Define for all $s \in \mathbf{S}$ $v_\beta(0,s) = 0$, and denote

$$R_i^m(a,s) = a[v_\beta(m,\mathcal{D}_i s) - v_\beta(m,s)] + \theta_i(a) + \overline{\mu}_i v_\beta(m,s), \quad a \in A_1(i)$$

$$S^m(i,s) = v_\beta(m,\mathcal{A}_i s) + d_i, \quad i = 1,2.$$

for $m = 0,1,2,...,n$.

We shall make use of the following well known tool ([9] Theorem 4.1):

Lemma 3.1 *(i) The value function satisfies the following DP equation*

$$\begin{aligned} v_\beta(m+1,s) &= \mathbf{val}_{a_1,a_2}\left\{ r(s,a_1,a_2) + \beta \sum_{t\in\mathbf{S}} q(t|s,a_1,a_2) v_\beta(m,t) \right\} \qquad (1) \\ &= c(s) + \beta \sum_{i=1,2} \max_{a_1(i)\in A_1(i)} R_i^m(a_1(i),s) + \beta\lambda \min_{a_2=1,2} S^m(a_2,s) \\ &\quad + (1-\lambda-\overline{\mu}_1-\overline{\mu}_2) v_\beta(m,s), \end{aligned}$$

for $m = 0, ..., n-1$.

(ii) Any Markov strategies f^* *and* g^* *for the two players, that use at time* $k = n - m$ *any actions that achieve the max and min in (1) respectively, are optimal.*

We first establish a simple sufficient condition for the router to have an optimal strategy of a monotone switching curve type. We say that a strategy is of the monotone switching curve type (see [7]) if it has the following monotonicity property. It is described by a curve in $\mathbf{S}$ with a monotone slope, that separates $\mathbf{S}$ into two connected regions, $\mathbf{S}_1$ and $\mathbf{S}_2$; there exists two actions, $j = 1, 2$, such that in region $\mathbf{S}_j$ it is optimal to use action j, $j = 1, 2$.

We shall show, in particular, that the router has an optimal nondecreasing switching curve policy, i.e., if it is optimal to route a customer to queue 1 for a given length of the queues $s = (s_1, s_2)$, then it is also optimal to route the customer to queue 1 when the length of the queues is $t = (t_1, t_2)$ provided that $t_1 \le s_1$ and $t_2 \ge s_2$. (This implies indeed that the curve that separates $\mathbf{S}1$ and $\mathbf{S}2$ is nondecreasing). A similar monotonicity property should hold for routing to queue 2.

Consider the following property that a function $z : \mathbf{S} \to \mathbb{R}$ may possess:

$\mathbf{\Pi_1}$: $\quad z(\mathcal{A}_i^2 s) - z(\mathcal{A}_i \mathcal{A}_j s) \ge z(\mathcal{A}_i s) - z(\mathcal{A}_j s), \qquad i, j = 1, 2,\ i \ne j.$

It is easy to see that if $v_\beta(m, \cdot)$ satisfies $\mathbf{\Pi_1}$, then there exists an action for the minizer in (1) which has the above monotone switching curve structure.

Next we identify sufficient conditions for a similar structure to exist for optimal service strategies. Consider the following property that a function $z : \mathbf{S} \to \mathbb{R}$ may possess:

$\mathbf{\Pi_2}$: $\quad z(\mathcal{A}_i \mathcal{A}_j s) - z(\mathcal{A}_j s) \ge z(\mathcal{A}_i s) - z(s), \qquad i, j = 1, 2.$

It is easy to see that if $v_\beta(m, \cdot)$ satisfies $\mathbf{\Pi_2}$, then there exists an action for both maximizers in (1) which are monotone nonincreasing in s. Moreover, if θ_i is convex for some $i = 1, 2$, then server i has an action achieving the maximum in (1) which is among $\{\underline{\mu}_i, \overline{\mu}_i\}$ (strategies having this property are called bang-bang strategies). In that case, if $v_\beta(m, \cdot)$ satisfies $\mathbf{\Pi_2}$, server i has a nonincreasing monotone switching curve structure (this means in particular that if at some state s server i uses $\underline{\mu}_i$, then it also uses $\underline{\mu}_i$ for all states $t \ge s$, componentwise).

Summerizing the above and using Lemma 3.1, we get the following:

Lemma 3.2 *If* $v_\beta(m, \cdot)$, $m = 1, ..., n$ *satisfy* $\mathbf{\Pi_1}$*, then the router has an optimal pure Markov policy* $g = (g_1, ..., g_n)$*, such that* g_m *has a monotone nondecreasing switching curve structure,* $m = 1, ..., n$*. If* $v_\beta(m, \cdot)$ *satisfy* $\mathbf{\Pi_2}$*, then both servers have optimal pure Markov policies* $f = (f_1^i, ..., f_n^i)$, $i = 1, 2$*, such that all* f_m^i *are monotone nonincreasing in* s*. If, moreover,* θ_i *is convex for some* $i = 1, 2$*, then for all* $m = 1, ..., n$*, server* i *has an optimal pure bang-bang Markov policy with a monotone nonincreasing switching curve structure.*

We note that property $\mathbf{\Pi_2}$is equivalent to the properties of (integer-)convexity of the value function in s_i and to supermodularity.

Next we use the value iteration to show that $v_\beta(m,\cdot),\ m=1,...,n$ indeed possesses properties $\mathbf{\Pi_1}$and $\mathbf{\Pi_2}$.

3.2 The properties of the value function

Lemma 3.3 *Assume that c are nondecreasing, and satisfy* $\mathbf{\Pi_1}$ *and* $\mathbf{\Pi_2}$*. Then* $v_\beta(m,\cdot),\ m=1,...,n$ *are nondecreasing and satisfy* $\mathbf{\Pi_1}$ *and* $\mathbf{\Pi_2}$*.*

Proof. Clearly, $v_\beta(m,\cdot)$ are nondecreasing and satisfy $\mathbf{\Pi_1}$and $\mathbf{\Pi_2}$for $m=0$. Assume that $v_\beta(m,\cdot)$ are nondecreasing and satisfy $\mathbf{\Pi_1}$and $\mathbf{\Pi_2}$for some $m=0,...,n-1$. We show that this implies that $v_\beta(m+1,\cdot)$ also satisfies these properties. It suffices to show that both $\max_{a\in A_1(i)} R_i^m(a,s)$ and $\min_{a_2=1,2} S^m(a_2,s)$ have these properties.

We begin with $\mathbf{\Pi_1}$. Let $k\in\{1,2\}$ be an action achieving $argmin_a S^m(a,\mathcal{A}_i^2 s)$ and let $l\in\{1,2\}$ be an action achieving $argmin_a S^m(a,\mathcal{A}_j s)$. Then

$$\begin{aligned}
&\min_{a\in A_2} S^m(a,\mathcal{A}_i^2 s)-\min_{a\in A_2} S^m(a,\mathcal{A}_i\mathcal{A}_j s)-\min_{a\in A_2} S^m(a,\mathcal{A}_i s)+\min_{a\in A_2} S^m(a,\mathcal{A}_j s)\\
&\geq\ S^m(k,\mathcal{A}_i^2 s)-S^m(k,\mathcal{A}_i\mathcal{A}_j s)-S^m(l,\mathcal{A}_i s)+S^m(l,\mathcal{A}_j s)\\
&=\ v_\beta(m,\mathcal{A}_k\mathcal{A}_i^2 s)-v_\beta(m,\mathcal{A}_k\mathcal{A}_i\mathcal{A}_j s)-v_\beta(m,\mathcal{A}_l\mathcal{A}_i s)+v_\beta(m,\mathcal{A}_l\mathcal{A}_j s)\geq 0. \qquad (2)
\end{aligned}$$

The last inequality follows for $k=l$ since $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_1}$(with $s'=\mathcal{A}_l s$). It holds also for the other two cases $(l=i,k=j)$ and $(k=i,l=j)$ since

$$\begin{aligned}
v_\beta(m,\mathcal{A}_i^3 s)-v_\beta(m,\mathcal{A}_j\mathcal{A}_i^2 s) &\geq v_\beta(m,\mathcal{A}_i^2 s)-v_\beta(m,\mathcal{A}_j\mathcal{A}_i s) &(3)\\
&\geq v_\beta(m,\mathcal{A}_j\mathcal{A}_i^2 s)-v_\beta(m,\mathcal{A}_j^2\mathcal{A}_i s) &(4)\\
&\geq v_\beta(m,\mathcal{A}_j\mathcal{A}_i s)-v_\beta(m,\mathcal{A}_j^2 s) &(5)
\end{aligned}$$

where both (3) and (4) follow from the fact that $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_1}$, with $s'=\mathcal{A}_i s$, and (5) follows from the fact that $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_1}$with $s'=\mathcal{A}_j s$. Hence $\min_{a_2=1,2} S^m(a_2,s)$ satisfies $\mathbf{\Pi_1}$.

Fix $i\in\{1,2\}$. Let $\alpha\in A_1(i)$ be an action achieving $argmax_a R_i^m(a,\mathcal{A}_i\mathcal{A}_j s)$ and let $\gamma\in A_1(i)$ be an action achieving $argmax_a R_i^m(a,\mathcal{A}_i s)$. Then

$$\max_{a\in A_1(i)} R_i^m(a,\mathcal{A}_i^2 s)-\max_{a\in A_1(i)} R_i^m(a,\mathcal{A}_i\mathcal{A}_j s)-\max_{a\in A_1(i)} R_i^m(a,\mathcal{A}_i s)+\max_{a\in A_1(i)} R_i^m(a,\mathcal{A}_j s)$$

$$
\begin{aligned}
&\geq R_i^m(\alpha, \mathcal{A}_i^2 s) - R_i^m(\alpha, \mathcal{A}_i\mathcal{A}_j s) - R_i^m(\gamma, \mathcal{A}_i s) + R_i^m(\gamma, \mathcal{A}_j s) \\
&= \alpha \left[v_\beta(m, \mathcal{A}_i s) - v_\beta(m, \mathcal{A}_i^2 s) - v_\beta(m, \mathcal{A}_j s) + v_\beta(m, \mathcal{A}_i\mathcal{A}_j s)\right] \\
&\quad -\gamma \left[v_\beta(m, \mathcal{D}_i\mathcal{A}_i s) - v_\beta(m, \mathcal{A}_i s) - v_\beta(m, \mathcal{A}_j\mathcal{D}_i s) + v_\beta(m, \mathcal{A}_j s)\right] \\
&\quad +\overline{\mu}_i \left[v_\beta(m, \mathcal{A}_i^2 s) - v_\beta(m, \mathcal{A}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_i s) + v_\beta(m, \mathcal{A}_j s)\right] \\
&= (\overline{\mu}_i - \alpha) \left[v_\beta(m, \mathcal{A}_i^2 s) - v_\beta(m, \mathcal{A}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_i s) + v_\beta(m, \mathcal{A}_j s)\right] \qquad (6) \\
&\quad +\gamma \left[v_\beta(m, \mathcal{A}_i s) - v_\beta(m, s) - v_\beta(m, \mathcal{A}_j s) + v_\beta(m, \mathcal{A}_j\mathcal{D}_i s)\right] \qquad (7) \\
&\geq 0.
\end{aligned}
$$

Indeed, (6) itself is nonnegative since $\overline{\mu}_i - \alpha$ are nonnegative, and since the term in square brackets is nonnegative as $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_1}$. To see that (7) is nonnegative we distinguish between two cases. If s_i is zero (queue i is empty) then $\mathcal{D}_i s = s$, and (7) reduces to $\gamma\left[v_\beta(m, \mathcal{A}_i s) - v_\beta(m, s)\right]$. This is nonnegative by the monotonicity of $v_\beta(m,\cdot)$. To see that (7) is nonnegative if s_i is nonzero, we use $\mathbf{\Pi_1}$for $v_\beta(m,\cdot)$ evaluated in $s' = \mathcal{D}_i s$. This establishes that $\max_{a\in A_1(i)} R_i^m(a,s)$ satisfies $\mathbf{\Pi_1}$.

Next we establish $\mathbf{\Pi_2}$. Let $k \in \{1,2\}$ be an action achieving $argmin_a S^m(a, \mathcal{A}_i\mathcal{A}_j s)$ and let $l \in \{1,2\}$ be an action achieving $argmin_a S^m(a, s)$. If $l = k$ then

$$
\begin{aligned}
&\min_{a\in A_2} S^m(a, \mathcal{A}_i\mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) + \min_{a\in A_2} S^m(a, s) \\
&\quad \geq S^m(k, \mathcal{A}_i\mathcal{A}_j s) - S^m(k, \mathcal{A}_i s) - S^m(l, \mathcal{A}_j s) + S^m(l, s) \\
&\quad = v_\beta(m, \mathcal{A}_i\mathcal{A}_j(\mathcal{A}_k s)) - v_\beta(m, \mathcal{A}_j(\mathcal{A}_k s)) - v_\beta(m, \mathcal{A}_i(\mathcal{A}_k s)) + v_\beta(m, \mathcal{A}_j(\mathcal{A}_k s)) \\
&\quad \geq 0,
\end{aligned}
$$

since $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_2}$. If $l = j$ then

$$
\begin{aligned}
&\min_{a\in A_2} S^m(a, \mathcal{A}_i\mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) + \min_{a\in A_2} S^m(a, s) \\
&\quad \geq S^m(k, \mathcal{A}_i\mathcal{A}_j s) - S^m(k, \mathcal{A}_i s) - S^m(l, \mathcal{A}_j s) + S^m(l, s) \\
&\quad = v_\beta(m, \mathcal{A}_i\mathcal{A}_k(\mathcal{A}_j s)) - v_\beta(m, \mathcal{A}_k(\mathcal{A}_j s)) - v_\beta(m, (\mathcal{A}_j s)) + v_\beta(m, (\mathcal{A}_j s)) \\
&\quad \geq 0,
\end{aligned}
$$

since $v_\beta(m,\cdot)$ satisfies $\mathbf{\Pi_2}$. A similar argument holds for $l = i$. It remains to check the case $l \neq i, l \neq j$. We have

$$
\begin{aligned}
&\min_{a\in A_2} S^m(a, \mathcal{A}_i\mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) - \min_{a\in A_2} S^m(a, \mathcal{A}_j s) + \min_{a\in A_2} S^m(a, s) \\
&\quad \geq S^m(k, \mathcal{A}_i\mathcal{A}_i s) - S^m(k, \mathcal{A}_i s) - S^m(l, \mathcal{A}_i s) + S^m(l, s) \\
&\quad = v_\beta(m, \mathcal{A}_i^2\mathcal{A}_k s) - v_\beta(m, \mathcal{A}_i\mathcal{A}_k s) - v_\beta(m, \mathcal{A}_i\mathcal{A}_l s) + v_\beta(m, \mathcal{A}_l s)
\end{aligned}
$$

$$
\begin{aligned}
&= v_\beta(m, \mathcal{A}_i\mathcal{A}_k(\mathcal{A}_i s)) - v_\beta(m, \mathcal{A}_k(\mathcal{A}_i s)) - v_\beta(m, \mathcal{A}_i(\mathcal{A}_i s)) + v_\beta(m, (\mathcal{A}_i s)) && (8)\\
&\quad + v_\beta(m, \mathcal{A}_i^2 s) - v_\beta(m, \mathcal{A}_i\mathcal{A}_l s) - v_\beta(m, \mathcal{A}_i s) + v_\beta(m, \mathcal{A}_l s) && (9)\\
&\geq 0\ .
\end{aligned}
$$

(8) is nonnegative due to $\mathbf{\Pi_2}$, and (9) is nonnegative due to $\mathbf{\Pi_1}$. Hence, $\min_{a_2=1,2} S^m(a_2, s)$ satisfies $\mathbf{\Pi_2}$.

Next we show that $\max_{a\in A_1(i)} R_i^m(a, s)$ satisfies $\mathbf{\Pi_2}$. Fix $i \in \{1, 2\}$. Let $\alpha \in A_1(i)$ be an action achieving $argmax_a R_i^m(a, \mathcal{A}_i\mathcal{A}_j s)$ and let $\gamma \in A_1(i)$ be an action achieving $argmax_a R_i^m(a, \mathcal{A}_i s)$. Then

$$
\begin{aligned}
&\max_{a\in A_1(i)} R_i^m(a, \mathcal{A}_i\mathcal{A}_j s) - \max_{a\in A_1(i)} R_i^m(a, \mathcal{A}_i s) - \max_{a\in A_1(i)} R_i^m(a, \mathcal{A}_j s) + \max_{a\in A_1(i)} R_i^m(a, \mathcal{A}s)\\
&\geq R_i^m(\gamma, \mathcal{A}_i\mathcal{A}_j s) - R_i^m(\gamma, \mathcal{A}_i s) - R_i^m(\alpha, \mathcal{A}_j s) + R_i^m(\alpha, s)\\
&= \gamma\,[v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j\mathcal{A}_i s) - v_\beta(m, s) + v_\beta(m, \mathcal{A}_i s)]\\
&\quad -\alpha\,[v_\beta(m, \mathcal{D}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{D}_i s) + v_\beta(m, s)]\\
&\quad +\overline{\mu}_i\,[v_\beta(m, \mathcal{A}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{A}_i s) + v_\beta(m, s)]\\
&= (\overline{\mu}_i - \gamma)\,[v_\beta(m, \mathcal{A}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{A}_i s) + v_\beta(m, s)] && (10)\\
&\quad +\alpha\,[v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j\mathcal{D}_i s) - v_\beta(m, s) + v_\beta(m, \mathcal{D}_i s)] && (11)\\
&\geq 0\ .
\end{aligned}
$$

Indeed, (10) is itself nonnegative due to $\mathbf{\Pi_2}$, and (11) is zero if $s_i = 0$, and is otherwise nonnegative due to $\mathbf{\Pi_2}$.

It remains to establish the monotonicity of $\max_{a\in A_1(i)} R_i^m(a, s)$ and $\min_{a_2=1,2} S^m(a_2, s)$. Choose some $i \in \{1, 2\}$, and let $j \in \{1, 2\}$ be an action achieving $argmin_a S^m(a, \mathcal{A}_i s)$. Then

$$
\begin{aligned}
&\min_{a\in A_2} S^m(a, \mathcal{A}_i s) - \min_{a\in A_2} S^m(a, s)\\
&\geq S^m(j, \mathcal{A}_i s) - S^m(j, s)\\
&= v_\beta(m, \mathcal{A}_i(\mathcal{A}_j s)) - v_\beta(m, (\mathcal{A}_j s))\\
&\geq 0,
\end{aligned}
$$

since $v_\beta(m, \cdot)$ is monotone nondecreasing. Next we show that $\max_{a\in A_1(i)} R_i^m(a, s)$ is monotone nondecreasing. Fix $i \in \{1, 2\}$. Let $\alpha \in A_1(i)$ be an action achieving $argmax_a R_i^m(a, s)$. Then

$$
\begin{aligned}
&\max_{a\in A_1(i)} R_i^m(a, \mathcal{A}_j s) - \max_{a\in A_1(i)} R_i^m(a, s)\\
&\geq R_i^m(\alpha, \mathcal{A}_j s) - R_i^m(\alpha, s)\\
&= \alpha\,[v_\beta(m, \mathcal{D}_i\mathcal{A}_j s) - v_\beta(m, \mathcal{A}_j s) - v_\beta(m, \mathcal{D}_i s) + v_\beta(m, s)] + \overline{\mu}_i\,[v_\beta(m, \mathcal{A}_j s) - v_\beta(m, s)]
\end{aligned}
$$

$$= \quad (\overline{\mu}_i - \alpha)\,[v_\beta(m, \mathcal{A}_j s) - v_\beta(m, s)] + \alpha\,[v_\beta(m, \mathcal{D}_i \mathcal{A}_j s) - v_\beta(m, \mathcal{D}_i s)]$$

$$\geq \quad 0\,, \tag{12}$$

again by the monotonicity of $v_\beta(m, \cdot)$. This establishes the monotonicity of $\max_{a \in A_1(i)} R_i^m(a, s)$ and $\min_{a_2=1,2} S^m(a_2, s)$, and hence the proof. ∎

Combining Lemma 3.2 with Lemma 3.3 we obtain the main result of the section:

Proposition 3.1 *Consider the finite horizon discounted problem. Assume that c is nondecreasing, and satisfy* $\mathbf{\Pi_1}$ *and* $\mathbf{\Pi_2}$*. Then the router has an optimal pure Markov policy which has at each step a monotone nondecreasing switching curve structure. Moreover, both servers have optimal pure Markov policies which are monotone nonincreasing in s at each time. If, moreover,* θ_i *is convex for some* $i = 1, 2$*, then server i has an optimal pure bang-bang Markov policy at each step, with a monotone nonincreasing switching curve structure.*

4 The infinite horizon problem

Under some mild growth conditions on the immediate cost, optimal stationary exist for both players with the same structure as described in Proposition 3.1. This is summerized in the following:

Proposition 4.1 *Consider the infinite horizon discounted problem with discount factor* $\beta < 1$*. Assume that c is nondecreasing, satisfy* $\mathbf{\Pi_1}$ *and* $\mathbf{\Pi_2}$*. Assume moreover that for some* $1 < \gamma < \beta^{-1}$,

$$\sup_{s \in \mathbf{S}} \frac{|c(s)|}{\gamma^{(s_1+s_2)}} < \infty. \tag{13}$$

Then the router has an optimal pure stationary policy with a monotone nondecreasing switching curve structure. Moreover, both servers have optimal pure stationary policies which are monotone nondecreasing in s. If, moreover, θ_i *is convex for some* $i = 1, 2$*, then server i has an optimal pure bang-bang stationary policy with a monotone nonincreasing switching curve structure.*

Proof. Let f_m, g_m be maximizing and minimizing actions in the DP (1), $m = 1, 2, \ldots$. It follows from Proposition 3.1 that f_m, g_m can be chosen to be pure and monotone. Choose any pure stationary policies f, g obtained as some limit points of f_m, g_m. It is easily seen that f, g inherits the structural properties of f_m, g_m. Moreover, by Theorem 3.3 in [6], f and g are optimal, provided that some conditions are satisfied, which we check below. This establishes the proof.

It remains thus to check the following conditions of Theorem 3.3 in [6]. There exists some function $\mu : \mathbf{S} \to [1, \infty)$ such that

(i) $\beta \times \sup_{s\in\mathbf{S},a_1\in A_1,a_2\in A_2} \left\{ \mu_s^{-1} \sum_{t\in\mathbf{S}} q(t|s,a_1,a_2)\mu_t \right\} < 1,$

(ii) $\sup_{s\in\mathbf{S},a_1\in A_1,a_2\in A_2} \left\{ \mu_s^{-1} |r(s,a_1,a_2|) \right\} < \infty.$

We choose $\mu_s = \gamma^{(s_1+s_2)}$. Then (i) is satisfied since

$$\beta\frac{\sum_{t\in\mathbf{S}} q(t|s,a_1,a_2)\mu_t}{\mu_s} = \beta\frac{\sum_{t\in\mathbf{S}} q(t|s,a_1,a_2)\gamma^{t_1+t_2}}{\gamma^{s_1+s_2}} \le \beta\frac{\gamma^{s_1+s_2+1}}{\gamma^{s_1+s_2}} = \beta\gamma < 1.$$

(ii) holds by (13). ∎

Remark 4.1 *Using [6] it esaily follows that the results in Proposition 4.1 hold also for the expected average cost, provided that the following strong stability condition holds instead of (13):* $\lambda < \underline{\mu}_i$, $i = 1, 2$.

5 Concluding remarks

In this paper we established conditions for obtaining monotonicity properties of strategies, in general, and monotone switching curve type strategies in particular in the context of stochastic games, where we focused on a routing problem. This extends to the game setting methods already known in the context of control, see e.g. [7]. Our results can easily be extended to handle more involved control problems. Indeed, the main results remain valid in the presence of additional uncontrolled Poissonian flows of customers to both queues. Moreover, one may consider a situation where, with some probability, customers have to rerouted to get extra service in one of the queues, as in the model studied by Hajek [7]. Extra controllers have then to be designed to decide to which queue should such customers be rerouted. Assuming that the input router controller, which we considered in our paper, and the new routers have the same common objective, it can again be shown, as in the control case studied by Hajek [7], that all routers have monotone switching curves strategies.

An interesting observation is that a flow control problem similar to the discrete time flow control models studied in [1, 2] can be considered as a special case of the routing model which we solved. Indeed, if the holding cost, the admission cost, and the service cost at queue 2 are zero, then the routing problem transforms into a flow control problem into the first queue. (This becomes even more apparent when the service rate in the second queue is extremely high, so that it is almost always empty).

References

[1] E. Altman, "Flow control using the theory of zero-sum Markov games", To appear in *IEEE Trans. Automatic Control*, 1994.

[2] E. Altman, "Monotonicity of optimal policies in a zero sum game: a flow control model", to appear in *Advances of dynamic games and applications*, 1993.

[3] E. Altman and G. Koole, "Stochastic Scheduling Games with Markov Decision Arrival Processes", *Journal Computers and Mathematics with Appl.*, 3rd special issue on Differential Games, pp. 141-148, 1993.

[4] E. Altman and N. Shimkin, "Individually Optimal Dynamic Routing in a Processor Sharing System: Stochastic Game Analysis", EE Pub No. 849, August 1992. Submitted to *Operations Research.*

[5] E. Altman and N. Shimkin, "Worst-case and Nash routing policies in parallel queues with uncertain service allocations", IMA Preprint Series No. 1120, Institute for Mathematics and Applications, University of Minnesota, Minneapolis, USA, 1993, submitted to *Operations Research.*

[6] E. Altman, A. Hordijk and F. M. Spieksma, "Contraction conditions for average and α-discounted optimality in countable state Markov games with unbounded rewards", submitted to *MOR*, 1994.

[7] B. Hajek, "Optimal control of two interacting service stations", *IEEE Trans. Automatic Control*, **29**. No. 6, pp. 491-499, 1984.

[8] M. T. Hsiao and A. A. Lazar, "A game theoretic approach to decentralized flow control of Markovian queueing Networks", *Performance '87*, Courtois & Latouche (eds.), pp. 55-73, 1988.

[9] A. S. Nowak, "On zero-sum stochastic games with general state space I". Prob and Math Statistics, Vol. IV, Fasc. 1, pp. 13-32, 1984.

[10] T.E.S. Raghavan and J.A. Filar, "Algorithms for Stochastic Games - A survey", *Zeitschrift für OR*, vol 35, pp. 437-472, 1991.