Mathematical Methods of Operations Research (1997) 45:281–301

Optimal Service Control against Worst Case Admission Policies: A Multichained Stochastic Game

ARIE HORDIJK, OLAF PASSCHIER AND FLOSKE SPIEKSMA

Leiden University, Department of Mathematics and Computer Science, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Abstract: In this paper we will consider two-person zero-sum games and derive a general approach for solving them. We apply this approach to a queueing problem. In section 1 we will introduce the model and formulate the Key-theorem. In section 2 we develop the theory that we will use in section 3 to prove the Key-theorem. This includes a general and useful result in Lemma 2.1 on the sufficiency of stationary policies.

1 Introduction

Model: We consider the following network consisting of two single server queues each with an infinite buffer. The system manager controls the service rates in both queues, whilst basing his decision on the present and past queue sizes. He has to pay some costs per unit time. Upon his arrival to the system a customer has to decide which queue to join. How he chooses is unknown to the system manager. We want to guarantee a minimum service level and therefore we assume that the customers choose in the worst possible way for the system manager. We model this by introducing an arrival controller, called nature, playing against the system manager by assigning arriving customers to a queue in the worst possible way. Our objective is to find a strategy that minimises the expected average costs for this behaviour of the customers under weak stability conditions and separable immediate costs. Weak stability means that there exist controls under which the associated queue size process becomes transient, but there is a non-empty set of strategies of the system manager under which the process is always positive recurrent no matter how arriving customers select what queue to join. This model is a continuous time two-person zero-sum Markov game. Because we assume additive costs and transition probabilities we can restrict ourselves to state dependent strategies [9], apply a standard uniformisation procedure [8] and formulate the problem as an equivalent discrete time stochastic game.

We will show that there is a stationary optimal strategy for the system manager prescribing a non-decreasing service rate as a function of the queue size and that there is a stationary optimal strategy for nature of the following form: if

1432–2994/97/45:2/281–301 $2.50

arriving customers choose queue i in state $x = (x_1, x_2)$, they also choose queue i in state $y = (y_1, y_2)$ if $y_i \geq x_i$ and $y_{3-i} \leq x_{3-i}$ for $i = 1, 2$. These structural properties have already been derived in Altman [1] for the ξ-discounted cost criterion. Under strong stability conditions this follows from Altman & Hordijk [2], but the case of weak stability was left open.

Markov Games: We consider a countable state space X; at each state $x \in X$ the compact sets of actions $A(x)$, $B(x)$ are available to player 1 and 2 respectively. Let $A = \bigcup_x A(x)$, $B = \bigcup_x B(x)$. Let $K := \{(x, a, b): x \in X, a \in A(x), b \in B(x)\}$. $P = \{P_{xaby}\}$ are the transition probabilities, where P_{xaby} is the probability to go from state x to state y given that actions a and b are chosen by the players. $c: K \to R$ is an immediate cost paid by player 2 to player 1, assumed to be continuous in a and b. Let Π and R denote the set of strategies for player 1 and 2. A policy $\pi \in \Pi$ ($\rho \in R$) is a sequence $\pi = (\pi_1, \pi_2, \dots)$ ($\rho = (\rho_1, \rho_2, \dots)$), where $\pi_t(\rho_t)$ is the probability over $A(B)$ conditioned on the history of all actions of both players and on the state till time $t - 1$, as well as the state at time t. The actions of the two players at time t are chosen independently according to π_t and ρ_t. We denote by $\Pi(M)$, $R(M)$ the set of Markov strategies for player 1 and 2, and by $\Pi(S)$ and $R(S)$ the stationary randomised strategies. Let $\mathscr{P}_z^{\pi,\rho}$ and $E_z^{\pi,\rho}$ denote the (unique) probability measure induced by an initial state z and policies π, ρ, and the corresponding expectation. Let $\{X_t, A_t, B_t\}$ be the resulting stochastic process describing the evolution of states and actions.

We consider two cost criteria. First let the infinite horizon discounted cost be defined by: $V_{\pi,\rho}^{\xi}(z) = \mathsf{E}_z^{\pi,\rho} \sum_{s=1}^{\infty} \xi^{s-1} c(X_s, A_s, B_s)$ with $\xi < 1$. The infinite horizon expected average cost is defined as follows:

$$g_{\pi,\rho}(z) = \limsup_{t\to\infty} \mathsf{E}_z^{\pi,\rho} \frac{1}{t} \sum_{s=1}^{t} c(X_s, A_s, B_s) \ .$$

We define the following values:

$$\underline{V}^{\xi}(z) = \sup_{\pi} \inf_{\rho} V_{\pi,\rho}^{\xi}(z) \ , \qquad \overline{V}^{\xi}(z) = \inf_{\rho} \sup_{\pi} V_{\pi,\rho}^{\xi}(z) \ ,$$

$$\underline{g}(z) = \sup_{\pi} \inf_{\rho} g_{\pi,\rho}(z) \ , \qquad \overline{g}(z) = \inf_{\rho} \sup_{\pi} g_{\pi,\rho}(z) \ .$$

A policy π^* is called discounted optimal if $\inf_\rho V_{\pi^*,\rho}^{\xi}(z) = \underline{V}^{\xi}(z)$. Moreover, if $\inf_\rho V_{\pi^*,\rho}^{\xi}(z) = \overline{V}^{\xi}(z)$, π^* is called strongly optima. Similarly, ρ^* is (strongly) optimal if $\sup_\pi V_{\pi,\rho^*}^{\xi}(z) = \overline{V}^{\xi}(z)$ ($\sup_\pi V_{\pi,\rho^*}^{\xi}(z) = \underline{V}^{\xi}(z)$). For the average cost criteria, optimality of policies is defined similarly.

Mathematical Model: We assume that customers arrive according to a Poisson process with rate λ. The service duration of a customer in queue i is exponentially distributed with parameter $b_i \in B_i = [\underline{b}(i), \bar{b}(i)]$ for $i = 1, 2$. Since control is based on the queue sizes, we define the state by the queue lengths and hence the state space is $X = \mathbb{N}^2$. By $y = \mathscr{A}_i x$ we denote the state y with $y_i = x_i + 1$, $y_j = x_j$ for $i \neq j$, and by $y = \mathscr{D}_i x$ the state $y_i = \max\{0, x_i - 1\}$, $y_j = x_j$, $j \neq i$. We assume that the rates are normalised so that $\lambda + \bar{b}(1) + \bar{b}(2) \leq 1$. Then we have the following transition probabilities:

$$P_{xaby} = \begin{cases} \lambda \ , & a = i, y = \mathscr{A}_i x, i = 1, 2 \ ; \\ b_i \ , & y \neq x, y = \mathscr{D}_i x, i = 1, 2 \ ; \\ 1 - (\lambda - \sum_{i=1}^{2} b_i 1\{x_i > 0\}) \ , & y = x \ . \end{cases}$$

The immediate costs are of the form:

$$c(x, a, b) = h(x) + \sum_{i=1}^{2} \theta_i(b_i) + \sum_{i=1}^{2} \zeta_i 1\{a = i\} \ .$$

We see that the costs consist of three separable parts. Part one is the holding cost, which is nondecreasing in the queue lengths. The second part is the cost of providing service. The third and last part is the cost the system manager pays to a customer that is joining a certain queue. This cost can be negative, which means that the customer is paid for joining a queue.

We use the following conditions.

Assumption 1.1:

i) $\dfrac{h(x)}{\delta^{|x|}}$ *is bounded for all* $\delta > 1$, $x \in X$.
ii) $\theta_i(b_i)$ *is bounded for* $b_i \in B_i$.
iii) $h(x)$ *is non-decreasing in both components and is a moment function, i.e. for every* $q \in \mathbb{R}$: $|\{y \in X: h(y) < q\}| < \infty$.
iv) θ_i *(and thus c) is a continuous function of the actions.*
v) *weak stability, i.e.* $\lambda < \min(\bar{b}(1), \bar{b}(2))$.

Because $c(x, a, b)$ is bounded from below we can assume that $c(x, a, b) \geq 0$ for all states x and actions a, b. To achieve this, we can add a constant to $h(x)$ and so the structure will not change.

Definition: We say that a decision rule is of *monotone switching curve type* (see Hajek [4]) if it has the following properties. There exists a curve in X with a non-negative slope separating X into two connected regions, X_1 and X_2 and there exist two actions, such that action a_i is optimal in region X_i.

Key-Theorem: Suppose that Assumption 1.1 *holds.*

i) *For nature there exists a optimal stationary strategy of the monotone switching curve type.*
ii) *There are positive constants y_1 and y_2 such that there exist an optimal policy ρ for the system manager with*

$$\rho_k(x_1, x_2) = \bar{b}(k) \qquad \text{for } x_k \geq y_k \quad \text{and} \quad k = 1, 2 \ ,$$

where $\rho_k(x_1, x_2)$ is the service rate at queue k in state $x = (x_1, x_2)$. Moreover, $\rho_k(x_1, x_2)$ is non-decreasing in x_k.
iii) *If θ_k is concave for $k = 1, 2$, then $\rho_k(x) \in \{\underline{b}(k), \bar{b}(k)\}$ for $x \in X$, and this is again a policy of the switching curve type.*

2 General Optimality Results

First we will show a general result on sufficiency of stationary policies in a Markov decision process (MDP), sufficiency is related to, but different, from the notion minimum pair in [6]. To this end we introduce some notation for MDP's. Let $g_\rho(i)$ be the expected average reward under policy ρ if we start in state i, similarly defined as in the game context.

Lemma 2.1: Consider a (possibly multichain) MDP with an immediate cost structure that is a non-negative moment function. Then the stationary recurrent policies are sufficient for the average cost criterion, i.e. for any policy $\rho \in R$ there exists a stationary policy $\beta(\rho) \in R(S)$ and a state $x \in X$ with

$$g_{\beta(\rho)}(x) = \inf_{i \in X} g_{\beta(\rho)}(i) \leq \inf_{i \in X} g_\rho(i) \ , \tag{1}$$

and x recurrent under $\beta(\rho)$.

Proof: First, some more notation for MDP's. We denote by $p(i, a, j)$ the probability to jump to state j if the present state is i and action a is chosen, while $c(i, a)$

is the cost associated with this state and action. Since we assume arbitrary chosen, but fixed, starting state i_1 and policy ρ, they are often omitted in the notation. Say g is the average expected cost under the fixed policy and starting state. Clearly there is nothing to prove if $|g| = \infty$ and so we assume that g is finite. More, let for $\tilde{X} \subset X$, $\tilde{A} \subset A$

$$f^t(\tilde{X}, \tilde{A}) = \mathbb{P}_\rho(X_t \in \tilde{X}, A_t \in \tilde{A} | X_1 = i_1) \ ,$$

$$f^t(\tilde{X}) = \mathbb{P}_\rho(X_t \in \tilde{X} | X_1 = i_1) \ ,$$

$$\bar{f}^t(\tilde{X}, \tilde{A}) = \frac{1}{t} \sum_{l=1}^{t} f^l(\tilde{X}, \tilde{A}) \ ,$$

$$\bar{f}^t(\tilde{X}) = \frac{1}{t} \sum_{l=1}^{t} f^l(\tilde{X}) \ .$$

Since $\liminf_{t\to\infty} \int_{j,a} c(j, a) d\bar{f}^t(j, a) \le \limsup_{t\to\infty} \int_{j,a} c(j, a) d\bar{f}^t(j, a) = g$, and c is a moment function, for every $\varepsilon > 0$ there exists a state $j(\varepsilon)$ such that

$$\int_{j \le j(\varepsilon)} \int_a d\bar{f}^t(j, a) \ge 1 - \varepsilon \ ,$$

and therefore there exists a sequence $\{t_k\}$ and $\bar{f}^\infty(X, Y)$, such that

$$\begin{aligned} &\lim_{k\to\infty} \bar{f}^{t_k}(\tilde{X}, \tilde{A}) = \bar{f}^\infty(\tilde{X}, \tilde{A}) \qquad \forall \tilde{X} \in X \ , \quad \tilde{A} \in A \ , \\ &\int_{j,a} d\bar{f}^\infty(j, a) = 1 \ . \end{aligned} \tag{2}$$

Using Fatou's lemma gives

$$\int_{j,a} c(j, a) d\bar{f}^\infty(j, a) \le g_\rho(i_1) \ .$$

Since

$$f^{t+1}(j) = \int_{i,a} p(i, a, j) df^t(i, a) \ ,$$

we have

$$\left| \bar{f}^{t+1}(j) - \int_{i,a} p(i,a,j) d\bar{f}^{t}(i,a) \right| \leq \frac{2}{t} \ .$$

Hence,

$$\bar{f}^{\infty}(j) = \int_{i,a} p(i,a,j) d\bar{f}^{\infty}(i,a)$$

$$= \sum_i \bar{f}^{\infty}(i) \int_a p(i,a,j) d\frac{\bar{f}^{\infty}(i,a)}{\bar{f}^{\infty}(i)} \ . \tag{3}$$

The first equality follows from Fatou's lemma and the fact that (2) holds. Define the stationary policy β by

$$\beta_i(\tilde{A}) = \begin{cases} \dfrac{\bar{f}^{\infty}(i,\tilde{A})}{\bar{f}^{\infty}(i)} & \text{if } \bar{f}^{\infty}(i) > 0 \ ; \\ \text{arbitrary} & \text{otherwise} \ . \end{cases}$$

Let $P(\beta)$ be the transition matrix for the stationary policy β, i.e.

$$[P(\beta)]_{ij} = P_{ij}(\beta) = \int_a p(i,a,j) d\beta_i(a) \ .$$

For π a stationary probability measure corresponding to $P(\beta)$ we have

$$\pi(j) = \sum_{i:\pi(i)>0} \pi(i) P_{ij}(\beta)$$

$$= \sum_{i:\pi(i)>0} \pi(i) \int_a p(i,a,j) d\beta_i(a)$$

$$= \sum_{i:\pi(i)>0} \pi(i) \int_a p(i,a,j) d\frac{\bar{f}^{\infty}(i,\tilde{A})}{\bar{f}^{\infty}(i)} \ .$$

So by virtue of (3), $\bar{f}^{\infty}$ is a stationary probability measure corresponding to $P(\beta)$.

Define $E_{\bar{f}^{\infty}} = \{i \in X : \bar{f}^{\infty}(i) > 0\}$. $E_{\bar{f}^{\infty}}$ is a set of recurrent states under $\bar{\beta}$, which we can write as the union of minimal closed subsets E_k, and so $E_{\bar{f}^{\infty}} = \bigcup_k E_k$. The expected average reward for initial states in E_k are all equal, $g^{(k)}$ say. Let $q^{(k)}(i,a)$ be the stationary state-action frequences under β for initial states from E_k. Then there exist $\lambda_k \geq 0$ with $\sum_k \lambda_k = 1$, such that

$$\bar{f}^{\infty}(i, a) = \sum_k \lambda_k q^{(k)}(i, a) \ .$$

Using this, we see that

$$\begin{aligned} g_\rho(i_1) &\geq \int_{j,a} c(j, a) d\bar{f}^{\infty}(j, a) \\ &= \int_{j,a} c(j, a) d\left(\sum_t \lambda_t q^{(t)}(j, a)\right) \\ &= \sum_k \lambda_k \int_{j,a} c(j, a) dq^{(k)}(j, a) \\ &= \sum_k \lambda_k g^{(k)} \ , \end{aligned}$$

where we used Fubini's theorem in the second equality. As a consequence there exists a set E_k with $g^{(k)} \leq g_\rho(i_1)$. This shows the assertion. □

In Proposition 4.7 of [5] for Borel state and action spaces and in Theorem 1 of [7] for countable state and finite action spaces it is shown that the assumption of strong duality of the linear programming formulation implies the existence of a minimum pair, i.e. a state x and a policy β such that (1) holds for all policies ρ. In Theorem 2 of [7] it is proved that the moment condition together with some technical assumptions implies the existence of a minimum pair in the class of stationary policies. Combining this result with our Lemma 2.1 gives that the pair is minimal in the class of all policies.

The remaining theory in this section is an extension of results that can be found in Altman, Hordijk & Spieksma [3], Altman & Hordijk [2] and in Sennott [11].

Assumption 2.1: There exists a positive function μ: $X \to [1, \infty)$ such that

i) *the immediate cost is positive and μ-bounded, i.e.*

$$c^* := \sup_{x \in X} \sup_{a \in A(x)} \sup_{b \in B(x)} \frac{c(x, a, b)}{\mu(x)} < \infty \ ;$$

ii) $\sum_y P_{xaby} \mu_y < \infty \ \forall x \in X,\ a \in A(x),\ b \in B(x)$;

iii) *the transition probabilities are μ-continuous, i.e. for all $x \in X$ and any sequences $a(n) \to n$, $b(n) \to b$,*

$$\lim_{n \to \infty} \sum_{y \in X} |P_{xa(n)b(n)y} - P_{xaby}| \mu_y = 0 \ .$$

Assumption 2.2(ξ): There exists a state x_0, a constant $\zeta < 1$ and for every $x \neq x_0$ an action $b(x) \in B(x)$, such that for any $x \neq x_0$, $a \in A(x)$,

$$\xi \sum_{y \in X} {}_0P_{xab(x)y}\mu_y < \zeta\mu_x \ , \tag{4}$$

for μ as in Assumption 2.1, where ${}_0P$ is the matrix of taboo probabilities, i.e. ${}_{x_0}P_{xaby}$ is equal to P_{xaby} and to 0, if $y \neq x_0$ and $y = x_0$ respectively.

Lemma 2.2: Under Assumptions 2.1 and 2.2(ξ), with $\xi < 1$, the following holds.

(i) *There exists a μ-bounded solution of the optimality equation,*

$$v^\xi(x) = \operatorname*{val}_{a,b}\left[c(x, a, b) + \xi \sum_y P_{xaby}v^\xi(y)\right] . \tag{5}$$

This solution is unique in the class of μ-bounded functions. The stochastic game has a value V^ξ and $V^\xi = v^\xi$.

(ii) *Let r, u be any decision rules that are optimal in the dummy game $[c(a, b) + \xi P_{ab}v^\xi]$. Then the stationary policies α, β such that $\alpha_t = r$, $\beta_t = u$ (for all t) are strongly optimal for both players.*

Proof: [3] proves this lemma for the strongly stable case. Our weakly stable case follows easily. □

Lemma 2.3: Assumptions 2.1 and 2.2(1) imply that $(1 - \xi)|v^\xi(0)|$ is bounded for $\xi < 1$.

Proof: First note that with r and u two deterministic and stationary policies, with $u(x) = b(x)$ for all x, it follows from Assumption 2.2(1) that

$$P_{ru}\mu \leq {}_0P_{ru}\mu + \mathbf{e}\mu_0 \leq \zeta\mu + \mu_0\mathbf{e} \ .$$

Iterating this gives

$$P_{ru}^n\mu \leq \zeta^n\mu + \mu_0\mathbf{e}\sum_{k=0}^{n-1}\zeta^k \ .$$

Combining this with the definition of the expected discounted costs we have that if r^ξ and u^ξ are stationary ξ-discounted optimal policies and u is the policy for player 2 as defined above:

$$
\begin{aligned}
v^\xi &= V^\xi_{r^\xi u^\xi} \\
&\leq V^\xi_{r^\xi u} \\
&\leq \sum_{n=0}^{\infty} \xi^n P^n_{r^\xi u} c(r^\xi, u) \\
&\leq c^* \sum_{n=0}^{\infty} \xi^n P^n_{r^\xi u} \mu \\
&\leq c^* \sum_{n=0}^{\infty} \xi^n \left\{ \zeta^n \mu + \mu_0 \mathbf{e} \sum_{k=0}^{n-1} \zeta^k \right\} \\
&\leq c^* \left\{ \sum_{n=0}^{\infty} \zeta^n \mu + \mu_0 \mathbf{e} \sum_{n=0}^{\infty} \xi^n \sum_{k=0}^{n-1} \zeta^k \right\} \\
&= c^* \left\{ \sum_{n=0}^{\infty} \zeta^n \mu + \mu_0 \mathbf{e} \sum_{k=0}^{\infty} \zeta^k \sum_{n=k+1}^{\infty} \xi^n \right\} \\
&< c^* \{ \mu (1-\zeta)^{-1} + \mu_0 \mathbf{e} (1-\xi)^{-1} (1-\zeta)^{-1} \}
\end{aligned}
$$

and it follows that for $\tilde{c} = c^* \mu_0 (1-\zeta)^{-1}$

$$
v^\xi(0) \leq \tilde{c}\{1 + (1-\xi)^{-1}\}
$$

and so

$$
0 \leq (1-\xi) v^\xi(0) \leq 2\tilde{c} \ . \qquad \square
$$

Let (α^ξ, β^ξ) be a stationary discounted optimal policy pair for discount factor ξ. Now let $\xi(k)$ be a sequence along which the following limits exist:

$$
\lim_{k \to \infty} \xi(k) = 1 \ ,
$$

$$
\lim_{k \to \infty} \alpha^{\xi(k)} = \alpha^* \ ,
$$

$$
\lim_{k \to \infty} \beta^{\xi(k)} = \beta^* \ .
$$

Since $(1-\xi)v^{\xi}(0)$ is bounded for $\xi \in [0, 1)$ it follows that for $\{\xi_n\}_{n=0}^{\infty}$ a sequence as mentioned before, there exists a subsequence $\{\xi_{n_k}\}_{k=0}^{\infty}$ with $n_{k+1} > n_k$ such that $\lim_{k\to\infty}(1-\xi_{n_k})v^{\xi_{n_k}}(0) = g^*$ for some g^*.

Assumption 2.3: Let $w^{\xi(k)}(x) = v^{\xi(k)}(x) - v^{\xi(k)}(0)$. *Assume that there exists a function* $M(x)$ *and a positive number* N *such that*

(i) $-N \le w^{\xi(k)}(x) \le M(x) < \infty$;
(ii) $\sum_y P_{xaby} M(y) < \infty$.

Let $\{\xi_{n_m}\}_{m=0}^{\infty}$ be a subsequence of the previous sequences along which the following limit exists

$$\lim_{m\to\infty} w^{\xi_{n_m}} = w^* \ .$$

The existence of this limit is ensured by the bounds $-N$ and $M(x)$ using a diagonalisation procedure. It follows that

$$\lim_{m\to\infty}(1-\xi_{n_m})v^{\xi_{n_m}}(x) = \lim_{m\to\infty}(1-\xi_{n_m})(w^{\xi_{n_m}}(x) + v^{\xi_{n_m}}(0)) = g^* \ .$$

Assumption 2.4: $g_{\alpha^*,\rho} \ge g^*, \forall \rho \in R$.

Theorem 2.1: Suppose that the Assumptions 2.1, 2.2(1), 2.3 *and* 2.4 *hold. Then the pair* (g^*, w^*) *is a solution pair to the average optimality equation,*

$$w^*(x) = \operatorname*{val}_{a,b}\left[c(x, a, b) - g^*e + \sum_y P_{xaby} w^*(y)\right] , \tag{6}$$

the stochastic game has a value, given by g^*, *and* (α^*, β^*) *is an optimal policy pair.*

Proof: If we subtract $v^{\xi}(0)$ from both sides in equation (5), we obtain for any $\xi \in [0, 1)$

$$w^{\xi}(x) = \operatorname*{val}_{a,b}\left[c(x, a, b) - (1-\xi)v^{\xi}(0) + \xi \sum_y P_{xaby} w^{\xi}(y)\right] .$$

By taking the limit $m \to \infty$ along the sequence $\{\xi_{n_m}\}$ in both sides, (6) follows from Assumption 2.3 and dominated convergence. We see that (α^*, β^*) solves (6) for the solution (g^*, w^*). For an arbitraty strategy $\pi = (\pi_1, \pi_2, \dots)$ for player 1 we have that

$$\begin{aligned} w^* &\geq c(\pi_1, \beta^*) + P(\pi_1, \beta^*)w^* - g^*e \\ &\geq c(\pi_1, \beta^*) + P(\pi_1, \beta^*)c(\pi_2, \beta^*) + P(\pi_1, \beta^*)P(\pi_2, \beta^*)w^* - 2g^*e \\ &\geq \sum_{k=1}^{K} P^{(k-1)}(\pi, \beta^*)c(\pi_k, \beta^*) + P^{(K)}(\pi, \beta^*)w^* - Kg^*e \ . \end{aligned}$$

This implies for g^* that

$$g^* \geq \frac{1}{K}\sum_{k=1}^{K} P^{(k-1)}(\pi, \beta^*)c(\pi_k, \beta^*) + \frac{1}{K}P^{(K)}(\pi, \beta^*)w^* - \frac{1}{K}w^* \ ,$$

so that

$$g^* \geq \limsup_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} P^{(k-1)}(\pi, \beta^*)c(\pi_k, \beta^*) = g_{\pi,\beta^*} \ . \tag{7}$$

Next, let ρ be an arbitrary strategy for player 2. From Assumption 2.4 it follows that $g_{\alpha^*,\rho} \geq g^*$ and therefore we have

$$g_{\pi,\beta^*} \leq g^* = g_{\alpha^*,\beta^*} \leq g_{\alpha^*,\rho}$$

and the game has a value given by g^*. □

Assumption 2.4 is generally hard to check. Therefore we introduce Assumption 2.5, which will be shown to imply Assumption 2.4 in Lemma 2.4. This lemma and proof are generalisations of Assumption D and Lemma 4.7 in [2]. First we define a *minimal closed subset* (MCS) under a stationary policy pair (α, β). This is a set of states $C \in X$, such that $P_{x\alpha(x)\beta(x)y} = 0$ for $x \in C$, $y \notin C$. Moreover, C does not contain a set of states with the same property.

Assumption 2.5:

(i) *Assume Assumptions* 2.1 *and* 2.2 *and fix* $A \subset X$. *There exists a class of stationary policies for player* 2, *denoted by* $\tilde{R}(S)$, *which the property that for* *any policy $\beta \in \tilde{R}(S)$, the average cost $g_{\alpha^*,\beta}(x)$ is constant on A, for α^* defined above; moreover, A is a closed set under (α^*, β) and for all $\rho \in R$, there exists $\beta \in \tilde{R}(S)$ with*

$$g_{\alpha^*,\rho}(x) \geq g_{\alpha^*,\beta}(x)$$

for $x \in A$.

(ii) *There is some (partial) order on A; for all $\beta \in \tilde{R}(S)$, there exists a sequence pair $\{\alpha^k, \beta^k\}$ of stationary policies with $\lim_{k\to\infty} \alpha^k = \alpha^*$ and $\lim_{k\to\infty} \beta^k = \beta$, such that $P^T_{x\alpha^k\beta^k\cdot} \leq P^T_{x\alpha^*\beta\cdot}$ for all $x \in A$ and $T \in \mathbb{N}$ in the stochastic order corresponding to the partial order on the states. Moreover, A is a closed set under (α^k, β^k).*

(iii) *The immediate cost is separable, i.e. $c(x, a, b) = c_1(a, b) + h(x)$; h is a monotone nondecreasing function in x and satisfies the moment condition, i.e. for any constant $q \in \mathbb{R}$, the set $\{y \in X: h(y) < q\}$ is finite. Moreover, c_1 is bounded.*

(iv) $g_{\alpha^k,\beta^k}(x) \geq g^*$, $\forall k \in \mathbb{N}$, $x \in A$.

Lemma 2.4: Assumption 2.5 implies $g_{\alpha^*,\rho}(x) \geq g^*$ for $x \in A$.

Proof: For checking $g_{\alpha^*,\beta}(x) \geq g^*$ we may restrict to $\beta \in \tilde{R}(S)$. Fix x and β. First assume that the Markov chain induced by α^*, β and starting state x is not positive recurrent. The moment condition implies that $g_{\alpha^*,\beta}(x) = \infty$, thus proving thhe assertion for x and β. Next assume that this Markov chain is positive recurrent. Let

$$P^\infty_{x\cdot}(\alpha^*, \beta) = \lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} P^n_{x\cdot}(\alpha^*, \beta) \ .$$

Assume that for some $x \in A$, $P^\infty_{x\cdot}(\alpha^*, \beta)h = \infty$. Itfollows from the fact that c_1 is bounded together with Fatou's lemma that $g_{\alpha^*,\beta}(x) = \infty > g^*$. As a consequence it suffices to fix some $x \in A$ for which $P^\infty_{x\cdot}(\alpha^*, \beta)h < \infty$. Hence,

$$\sum_{y\in A} P^\infty_{xy}(\alpha^*, \beta) = 1 \ ,$$

by virtue of the above and the fact that A is closed under all considered policies. Next we show that this implies that the sequence $\{P^\infty_{x\cdot}(\alpha^k, \beta^k)\}_{k=1}^\infty$ is tight. Fix some $\varepsilon > 0$, and let $q(\varepsilon)$ be such that for all T,

$$\sum_{y\leq q(\varepsilon)} [P^T]_{xy}(\alpha^*, \beta) \geq 1 - \varepsilon \ .$$

Then by Assumption 2.5(ii) it follows that

$$\sum_{y \le q(\varepsilon)} [P^T]_{xy}(\alpha^k, \beta^k) \ge \sum_{y \le q(\varepsilon)} [P^T]_{xy}(\alpha^*, \beta) \ge 1 - \varepsilon \ ,$$

for any k, T. Hence $\{P^{\infty}_{x\cdot}(\alpha^k, \beta^k)\}_{k=1}^{\infty}$ is tight. Suppose $\hat{P}^{\infty}$ is a limit measure of this sequence, obtained along some subsequence $k(n)$. Clearly it is a proability measure on A. Since

$$\sum_{y \in A} P^{\infty}_{xy}(\alpha^{k(n)}, \beta^{k(n)}) P_{yz}(\alpha^{k(n)}, \beta^{k(n)}) = P^{\infty}_{xz}(\alpha^{k(n)}, \beta^{k(n)}) \ ,$$

it follows from the bounded convergence theorem that

$$\hat{P}^{\infty} P(\alpha^*, \beta) = \hat{P}^{\infty} \ . \tag{8}$$

Hence $\hat{P}^{\infty}$ is an invariant measure of $P(\alpha^*, \beta)$. Since

$$P^{\infty}_{x\cdot}(\alpha^{k(n)}, \beta^{k(n)}) \le P^{\infty}_{x\cdot}(\alpha^*, \beta)$$

and

$$P^{\infty}_{x\cdot}(\alpha^*, \beta) h < \infty \ ,$$

it follows since h non-decreasing

$$\sum_{y \ge q} P^{\infty}_{xy}(\alpha^k, \beta^k) h(y) \le \sum_{y \ge q} P^{\infty}_{xy}(\alpha^*, \beta) h(y) \ .$$

Using that

$$\lim_{q \to \infty} \sum_{y \ge q} P^{\infty}_{xy}(\alpha^*, \beta) h(y) = 0$$

we find that h is uniformly integrable with respect $\{P^{\infty}_{x\cdot}(\alpha^k, \beta^k)\}_{k=1}^{\infty}$ and hence

$$\lim_{n \to \infty} P^{\infty}_{x\cdot}(\alpha^{k(n)}, \beta^{k(n)}) h = \hat{P}^{\infty} h \ . \tag{9}$$

Since c_1 is bounded, we have

$$\lim_{n\to\infty} P_{x\cdot}^{\infty}(\alpha^{k(n)}, \beta^{k(n)})c_1 = \hat{P}^{\infty}c_1 \ . \tag{10}$$

Combining (8), (9) and (10) gives for $x \in A$

$$\begin{aligned}\lim_{n\to\infty} P_{x\cdot}^{\infty}(\alpha^{k(n)}, \beta^{k(n)})\{h + c_1\} &= \hat{P}^{\infty}P^{\infty}(\alpha^*, \beta)\{h + c_1\} \\ &= \sum_{y\in A} \hat{P}^{\infty}(y) g_{\alpha^*,\beta}(y) \\ &= g_{\alpha^*,\beta}(x).\end{aligned}$$

The last equality follows because $g_{\alpha^*,\beta}(y)$ is constant on A. Hence

$$g_{\alpha^*,\beta}(x) = \lim_{n\to\infty} P_{x\cdot}^{\infty}(\alpha^{k(n)}, \beta^{k(n)})\{h + c_1\} = \lim_{n\to\infty} g_{\alpha^{k(n)},\beta^{k(n)}}(x) \geq g^* \ . \qquad \square$$

3 Proof of the Key-Theorem

Definition: We define the properties $\mathbf{\Pi_1}$ and $\mathbf{\Pi_2}$ for a function $z: X \to \mathsf{R}$:

$$\mathbf{\Pi_1}: z(\mathscr{A}_i^2 x) - z(\mathscr{A}_i\mathscr{A}_j x) \geq z(\mathscr{A}_i x) - z(\mathscr{A}_j x) \ , \qquad i, j = 1, 2 \ , \quad i \neq j \ .$$

$$\mathbf{\Pi_2}: z(\mathscr{A}_i\mathscr{A}_j x) - z(\mathscr{A}_j x) \geq z(\mathscr{A}_i x) - z(x) \ , \qquad i, j = 1, 2 \ .$$

Lemma 3.1: Let $g^* \in \mathsf{R}$ and $w: X \to \mathsf{R}^{\infty}$ satisfy the following optimality equation.

$$\begin{aligned} g^* = h(x) &+ \max_a \{\zeta_a + \lambda[w(\mathscr{A}_a x) - w(x)]\} \\ &+ \sum_{i=1}^{2} \min_{b(i)} \{\theta(b(i)) + b(i)[w(\mathscr{D}_i x) - w(x)]\} \ . \end{aligned} \tag{11}$$

If w satisfies Π_1 and Π_2, then for the maximiser in (11) *there exists an optimal action of the switching curve type and for both minimisers one that is monotone* *non-decreasing. Moreover, if θ_i is concave for $i = 1, 2$, then the optimal service rate for server i is one of $\{\underline{b}(i), \bar{b}(i)\}$, this again is a switching curve type structure.*

Proof: The maximiser in (11) chooses action 1 if

$$\zeta_1 + \lambda[w(\mathscr{A}_1 x) - w(x)] \geq \zeta_2 + \lambda[w(\mathscr{A}_2 x) - w(x)] ,$$

and hence if

$$w(\mathscr{A}_1 x) - w(\mathscr{A}_2 x) \geq \frac{\zeta_2 - \zeta_1}{\lambda} .$$

The left-hand side is, according to Π_1 non-decreasing in x_1 and non-increasing in x_2, so if in state x action 1 is optimal for nature the same holds for y with $y_1 \geq x_1$ and $y_2 \leq x_2$. Identical results can be shown for action 2.

If w satisfies Π_2 then $w(\mathscr{D}_i x) - w(x)$ is non-increasing in x_1 and x_2 and so there exist optimal minimisers that are non-decreasing in x_1 and x_2. If $\theta_i(\rho_i)$ is concave in ρ_i then $\theta_i(\rho_i) + \rho_i[w(\mathscr{D}_i x) - w(x)]$ is also concave in ρ_i. Since the minimum of a concave function on a bounded interval is always achieved at one of the end-points, it is one of $\{\underline{b}(i), \bar{b}(i)\}$. □

Proof of the Key-Theorem: First we will show that the assumptions for Theorem 2.1 from the previous section are satisfied.

Ass. 2.1: Let $\mu_x = (1 + c)^{x_1 + x_2}$. Assumption 1.1 implies Assumption 2.1(i) for every $c > 0$. Assumption 2.1 ii) and iii) follow from the fact the jump distributions from each state have a finite support that is uniformly bounded in the action pairs.

Ass. 2.2(1): Because $\lambda < \bar{b}(i)$ there exist $\varepsilon > 0$ and $c > 0$ such that $\left(\lambda + \frac{\varepsilon}{c}\right)(1 + c) < \bar{b}(i)$. Substituting this in (4) with $b(x) = (\bar{b}(1), \bar{b}(2))$ for all $x \in X$ we have three different cases.

(1) $x_1 > 0, x_2 > 0$.

$$\sum_y P_{xaby}\mu_y = (1+c)^{|x|}\left\{1 + \lambda[(1+c)-1] + (\bar{b}(1)+\bar{b}(2))\left[\frac{1}{c+1} - 1\right]\right\}$$

$$\leq \mu_x\left\{1 + c\left[\left(\lambda - \frac{\bar{b}(1)}{c+1}\right) + \left(\lambda - \frac{\bar{b}(2)}{c+1}\right)\right]\right\}$$

$$< \mu_x\{1 - 2\varepsilon\} \ ;$$

(2) $x_1 > 0, x_2 = 0.$

$$\sum_y P_{xaby}\mu_y = (1+c)^{x_1}\left\{1 + \lambda[(1+c)-1] + \bar{b}(1)\left[\frac{1}{c+1} - 1\right]\right\}$$

$$\leq \mu_x\left\{1 + c\left[\left(\lambda - \frac{\bar{b}(1)}{c+1}\right)\right]\right\}$$

$$< \mu_x\{1 - \varepsilon\} \ ;$$

(3) $x_1 = 0, x_2 > 0.$

$$\sum_y P_{xaby}\mu_y = (1+c)^{x_2}\left\{1 + \lambda[(1+c)-1] + \bar{b}(2)\left[\frac{1}{c+1} - 1\right]\right\}$$

$$\leq \mu_x\left\{1 + c\left[\left(\lambda - \frac{\bar{b}(2)}{c+1}\right)\right]\right\}$$

$$< \mu_x\{1 - \varepsilon\} \ .$$

Ass. 2.3: Let ρ^* be the stationary policy for player 2 that serves at the highest possible rate in all states, i.e. $\rho^*(i) = (\bar{b}(1), \bar{b}(2))$. We define

$$M(x) = \sup_{\xi} \sum_{0}^{\infty} {}_0P^n(\pi^{\xi}, \rho^*)c(\pi^{\xi}, \rho^*) \ .$$

Since $\lambda < \min\{\bar{b}(1), \bar{b}(2)\}$ and player 2 uses strategy ρ^*, it follows that the process will always be μ-geometric recurrent for any α and so $M(x) < \infty$. We obtain that

$$
\begin{aligned}
w^{\xi} &\leq c(\pi^{\xi}, \rho^*) - (1-\xi)v^{\xi}(0)\mathbf{e} + \xi_0 P(\pi^{\xi}, \rho^*)w^{\xi} \\
&< c(\pi^{\xi}, \rho^*) + {}_0P(\pi^{\xi}, \rho^*)w^{\xi} \\
&< \sum_{n=1}^{\infty} {}_0P^n(\pi^{\xi}, \rho^*)c(\pi^{\xi}, \rho^*) \\
&\leq M(x) \ .
\end{aligned}
$$

We can take $N = 0$, since $v^{\xi}(x)$ is increasing in x_1 and x_2. The second part of Assumption 2.3 is easy because the number of states that can be reached in one step is finite.

Ass. 2.4: Figure 1 illustrates some aspects of this proof. It shows a part of the state space and the thick line is the switching curve corresponding to the policy α^*. Recall that α^* is a limit policy of $\alpha^{\xi(k)}$ and that $\alpha^{\xi(k)}$ is of the monotone switching curve type (see [1]), and so α^* has this structure. Customers are assigned to the first queue in the region containing the x_1-axis and to the second queue in the other region. Assume that $\alpha^*(0, 0) = 1$ (the case $\alpha^*(0, 0) = 2$ is similar), so if the system is empty, an arriving customer is routed to the first queue. We will show that there exists $\tilde{\beta} \in R(S)$ such that $g_{\alpha^*,\tilde{\beta}}(x) =$

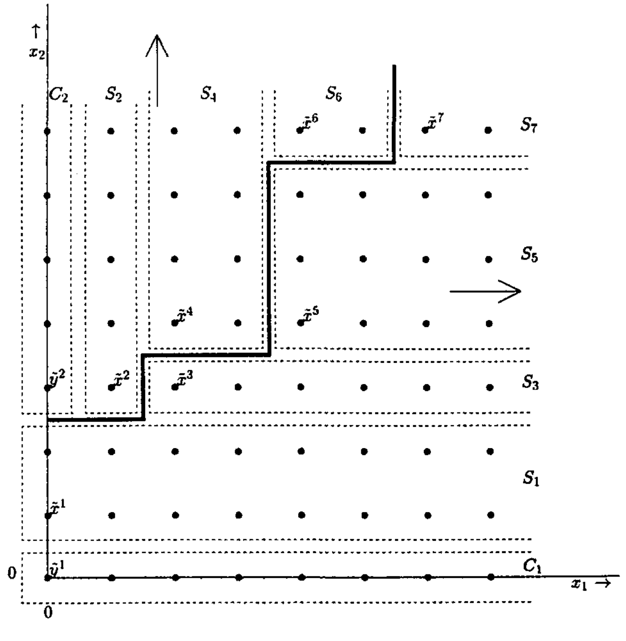

Fig. 1. The state space with some structure

$\min_{\rho \in R} g_{\alpha^*,\rho}(x)$ for all $x \in X$. To this end we define the sets

$$C_1 = \{(x_1, 0): x_1 \in \mathbb{N}\} \ ,$$

$$C_2 = \{(0, x_2): x_2 \in \mathbb{N}, \alpha^*(x) = 2\} \ .$$

Let $\tilde{y}^1 = (0, 0)$ and denote by $\tilde{y}^2 \in C_2$ as the state in C_2 with minimum second coordinate. For $k = 1, 2, \ldots$ let

$$D_k = X \Big\backslash \Big(C_1 \cup C_2 \bigcup_{s<k} S_s \Big) \ ,$$

$\tilde{x}^k \in D_k$ is the state in D_k with minimum coordinates and

$$S_k = \begin{cases} \{(x_1, x_2) \in D_k\colon \alpha^*(x_1, \tilde{x}_2^k)\ 2\} & \text{if } k \text{ even } ; \\ \{(x_1, x_2) \in D_k\colon \alpha^*(\tilde{x}_1^k, x_2) = 1\} & \text{if } k \text{ odd } . \end{cases}$$

C_1 is closed under all policies, so apply Proposition 5 in [10] to C_1 gives the existence of a stationary policy $\tilde{\beta}_1$ for player 2, such that $\tilde{\beta}_1$ is optimal against α^* and $g_{\alpha^*,\tilde{\beta}_1}(x) = g_1$ for all $x \in C_1$. Next, consider C_2. The only state where player 2 can force the process to leave C_2 is in $\tilde{y}^2$. If we restrict player 2 to policies such that the process cannot leave C_2, it is clear (again using [10]) that there exists a stationary policy $\tilde{\beta}_2$ that is optimal on C_2 in this class and $g_{\alpha^*,\tilde{\beta}_2}(x) = g_2$ for all $x \in C_2$. We can distinguish two cases.

(i) $g_1 > g_2$. We will show that the minimum average cost is equal to g_1 for the process starting in $S_1 \cup C_1$, and for the process starting outside these sets to g_2. To analyse the states in S_1, we have to introduce dummy transitions. If the process is in a state with no jobs in queue 2, player 2 can still decide to serve this queue. The duration of such a service has the same distribution as a normal service. Upon its completion, there is no transition, but a policy might depend on the number of completed dummy services in the past. Suppose, that there exists a state y and a policy $\rho \in R$ with expected average cost $\tilde{g} := g_{\alpha^*,\rho}(y) < g_1$. We define the state $\tilde{y}$ and the policy $\tilde{\rho} \in R$ on C_1 as $\tilde{y} = (y_1, 0)$ and $\tilde{\rho}_t(x_1, x_2, N_2(t)) = \rho_t(x_1, y_2 - N_2(t))$ with $N_2(t)$ the number of previous completed dummy services at the second queue. We see that

$$\mathbb{P}_{\alpha^*,\rho}(X_t = (x_1, x_2) | X_0 = y) = \mathbb{P}_{\alpha^*,\tilde{\rho}}(X_t = (x_1, 0), N_2(t) = y_2 - x_2 | X_0 = \tilde{y}) \ .$$

We can now couple the processes starting in y and in $\tilde{y}$ respectively, the first under the policy pair (α^*, ρ) and the second under $(\alpha^*, \tilde{\rho})$. Then

$$g_1 > \tilde{g} = g_{\alpha^*,\rho}(y) \geq g_{\alpha^*,\rho}(\tilde{y}) \geq g_{\alpha^*,\tilde{\beta}_1}(\tilde{y}) = g_1 \ ,$$

since $\tilde{\beta}_1$ is optimal, $h(x)$ is non-decreasing and the number of dummy transitions is finite and therefore does not influence the average cost. Next we show that the minimal cost in the states in $X \backslash (S_1 \cup C_1)$ is equal to g_2. First, consider the states in C_2. If the process leaves C_2 it will be absorbed in $C_1 \cup S_1$ with expected average cost equal to $g_1 > g_2$, and so this action will not be chosen. To show that the minimum average cost for the states in D_2 equals g_2, we define for $k \geq 1$, $\tilde{g}_k := \inf_{x \in S_k} \inf_\rho g_{\alpha^*,\rho}(x)$. If there exists a strategy ρ and a state $x \in D_2$ such that $g_{\alpha^*,\rho}(x) < g_2$, we can define S_k as the set with the smallest index containing a state x with $\inf_\rho g_{\alpha^*,\rho}(x) < g_2$. It follows that $\tilde{g}_k < g_2$. Consider the MDP with state space $C_1 \cup C_2 \bigcup_{l \leq k} S_l$. From Lemma 2.1 we know that there exists a stationary policy $\bar{\beta}$ and a recurrent set E' corresponding to $(\alpha^*, \bar{\beta})$, such that $g_{\alpha^*,\bar{\beta}}(x) \leq \inf_{x \in X} g_{\alpha^*,\rho}(x) < g_2$ for some state $x \in E'$. Clearly, $E' \cap (C_1 \cup C_1 \bigcup_{l<k} S_l) = \emptyset$, and so $E' \subset S_k$. Without loss of generality we assume that k is odd. From the structure of α it follows that E' must be a subset of one row. So let $a, b, z \in \mathsf{N}$ (b might be infinite) be such that $E' = \{x \in S_k \colon a \leq x_1 \leq b, x_2 = z\}$. We define the stationary policy β' on $D := \{x \in C_1 \colon a \leq x_1 \leq b\}$ as follows: $\beta'(x_1, 0) = \bar{\beta}(x_1, z)$. D cannot be left if we start there, and so using the non-decreasingness of h and the optimality of $\tilde{\beta}_1$, we can derive

$$g_2 > g_{\alpha^*,\bar{\beta}}(x_1, z) \geq g_{\alpha^*,\beta'}(x_1, 0) \geq g_{\alpha^*,\tilde{\beta}_1}(x_1, 0) = g_1 > g_2 \ ,$$

yielding a contradiction. This implies that g_2 is the optimal average cost when we start in $X \backslash (C_1 \cup S_1)$.

As a consequence the following stationary policy is optimal.

$$\tilde{\beta}(x) = \begin{cases} \tilde{\beta}_1(x) & \text{if } x \in C_1 \ ; \\ \tilde{\beta}_2(x) & \text{if } x \in C_2 \ ; \\ (\bar{b}(1), \bar{b}(2)) & \text{if } x \in D_1 \text{ and } x_2 \neq \tilde{y}_2^2 \ ; \\ (\bar{b}(1), 0)) & \text{if } x \in D_1 \text{ and } x_2 = \tilde{y}_2^2 \ . \end{cases}$$

In order to verify Assumption 2.4 it remains to be shown that $g_2 \geq g^*$. There are two different ways of doing this. The first is to use Assumption 2.5 and Lemma 2.4. Let $A = C_2$, then $P_{x\alpha^*\tilde{\beta}y} = 0$ if $x \in A$, $y \notin A$. We take $\tilde{R}(S) = \{\tilde{\beta}\}$, and the order on A defined by $x < y \Leftrightarrow x_2 < y_2$. The sequences $\{\alpha^k\}$ and $\{\beta^k\}$ are defined as follows: $\alpha^k = \alpha^*$ and

$$\beta^k(0, x_2) = \begin{cases} \tilde{\beta}(0, x_2) \ , & \text{if } x_2 \leq k \ ; \\ (0, \bar{b}(2)) \ , & \text{otherwise} \ . \end{cases}$$

It is easily checked that the process is μ-geometric recurrent under policy pair (α^k, β^k) and thus (see Lemma 2.1 in [3]) $\lim_{n\to\infty} n^{-1} P^{(n)}(\alpha^k, \beta^k) w^* = 0$. Using this we can show samilarly to (7) that from

$$w^* \leq c(\alpha^*, \beta^k) + P(\alpha^*, \beta^k) w^* - g^* e$$

it follows that

$$g^* \leq g_{\alpha^*, \beta^k} = g_{\alpha^k, \beta^k} \ .$$

Applying Lemma 2.4 yields $g_2 = g_{\alpha^*, \tilde{\beta}}(x) \geq g^*$.

The second way of proving $g_2 \geq g^*$ is straightforward. Since for $\tilde{y}^2$ there exists a positive number m_0 such that $\alpha^{\xi_m}(\tilde{y}^2) = \alpha^*(\tilde{y}^2)$ for $m \geq m_0$, it follows by the monotonicity of α^* that $\alpha^{\xi_m}(x) = \alpha^*(x) = 2$ for all $x \in C_2$, $m \geq m_0$. Therefore, for $x \in C_2$ and $m \geq m_0$:

$$\begin{aligned} g_2 &= g_{\alpha^*, \tilde{\beta}}(x) \\ &= g_{\alpha^{\xi_m}, \tilde{\beta}}(x) \\ &= \lim_{n\to\infty} (1 - \xi_n) V^{\xi_n}_{\alpha^{\xi_m}, \tilde{\beta}}(x) \\ &= \lim_{n\to\infty} (1 - \xi_n) V^{\xi_n}_{\alpha^{\xi_n}, \tilde{\beta}}(x) \\ &\geq \lim_{n\to\infty} (1 - \xi_n) V^{\xi_n}_{\alpha^{\xi_n}, \beta^{\xi}_n}(x) \\ &= g^* \ . \end{aligned}$$

(ii) $g_1 \leq g_2$. This case is similar, but we now choose $\tilde{\beta}$ as follows:

$$\tilde{\beta}(x) = \begin{cases} \tilde{\beta}_1(x) & \text{if } x \in C_1 \ ; \\ (\bar{b}(1), \bar{b}(2)) & \text{if } x \in X \backslash C_1 \ ; \end{cases}$$

and $g_{\alpha^*, \tilde{\beta}}(x) = g_1$ for all $x \in X$. Again we can prove with the previous methods that $g_{\alpha^*, \tilde{\beta}}(x) = g_1 \geq g^*$.

Finally we derive some structural properties. Denote

$$F^{\xi}(a, x) = \zeta_a + \xi\lambda[V^{\xi}(\mathscr{A}_a x) - V^{\xi}(x)] \ ,$$

$$G^{\xi}_i(b, x) = \theta_i(b) + \xi b[V^{\xi}(\mathscr{D}_i x) - V^{\xi}(x)] \ .$$

The dynamic programming equation has the form:

$$V^{\xi}(x) = h(x) + \xi V^{\xi}(x) + \max_{a} F^{\xi}(a, x) + \sum_{i=1}^{2} \min_{b(i) \in B_i} G_i^{\xi}(b(i), x) \ . \tag{12}$$

We know that for every $\xi < 1$ this equation has a unique solution for V^{ξ}. If we take a sequence $\xi(k)$, $k = 1, 2, \dots$ with $\lim_{k\to\infty} \xi(k) = 1$ then the previous theory implies the existence of a subsequence $\xi(k_i)$ such that the limits $g^* = \lim_{i\to\infty}(1 - \xi(k_i))V^{\xi(k_i)}$ and $w = \lim_{i\to\infty} w^{\xi(k_i)}$ exist and satisfy (11).

Altman [1] shows that $V^{\xi(k)}$ satisfies Π_1 and Π_2. Therefore w does and so the Key-theorem follows from Lemma 3.1. □

References

[1] Altman E (1994) A Markov game approach for optimal routing into a queueing network. Technical Report 2178, INRIA, Submitted

[2] Altman E, Hordijk A (1994) Zero-sum Markov games and worst-case optimal control of queueing systems. To appear in Questa

[3] Altman E, Hordijk A, Spieksma F (1994) Contraction conditions for average and α-discount optimality in countable state markov games with unbounded rewards. Technical Report TW-93-16, Leiden University, Submitted to MOR

[4] Hajek B (1984) Optimal control of two interacting service stations. IEEE Transactions on Automatic Control 29(6):491–499

[5] Hernández-Lerma O, Lasserre JB (1994) Lineaire programming and Markov control processes. SIAM Journal on Control and Optimization 32(2):480–500

[6] Kurano M (1989) The existence of a minimum pair of state and policy for Markov decision processes under the hypothesis of Doeblin. SIAM Journal on Control and Optimisation 27:296–307

[7] Lasserre JB (1994) Average optimal stationary policies and linear programming in countable space Markov decision processes. Journal of mathematical analysis and applications 183: 233–249.

[8] Lippman SA (1975) Appling a new device in the optimization of exponential queueing systems. Operations Research 23:687–710

[9] Raghavan TES, Tijs SH, Vrieze OJ (1985) On stochastic games with additive reward and transition structure. Journal of Optimization Theory and Applications 47(4):451–464

[10] Sennott LI (1989) Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. Operations Research 37(4):626–633

[11] Sennott LI (1994) Zero-sum stochastic games with unbounded costs: discounted and average cost cases. ZOR 40:145–162

Received: April 1996